EVIDÊNCIAS E INOVAÇÃO NO CUIDADO EM SAÚDE:

Doenças crônicas, tecnologias digitais e cuidado perioperatório

ORGANIZADORES

Danilo Ferreira de Sousa
Nélio Barreto Vieira



thesis editora científica

Organizadores
Danilo Ferreira de Sousa
Nélio Barreto Vieira

EVIDÊNCIAS E INOVAÇÃO NO CUIDADO EM SAÚDE

Doenças crônicas, tecnologias digitais e cuidado perioperatório

2026

2026 - Thesis Editora Científica
Copyright © Thesis Editora Científica.
Copyright do texto © 2026 Organizadores e autores.
Copyright da edição © 2026 Thesis Editora Científica.
Direitos para esta edição cedidos à Thesis Editora Científica pelos organizadores e autores.
Open access publication by Thesis Editora Científica.
Editor Chefe: Felipe Cardoso Rodrigues Vieira.
Diagramação, Projeto Gráfico e Design da Capa: Thesis Ed. Cien.
Revisão: Organizadores e autores.

Licença Creative Commons
EVIDÊNCIAS E INOVAÇÃO NO CUIDADO EM SAÚDE: Doenças crônicas, tecnologias digitais e cuidado perioperatório da Thesis Editora Científica está licenciada com uma Licença Creative Commons - Atribuição-NãoComercial-SemDerivações 4.0 Internacional. (CC BY-NC-ND 4.0).

O conteúdo da obra e seus dados em sua forma, correção e confiabilidade é de responsabilidade exclusiva dos autores, não representando a posição oficial da Thesis Editora Científica. É permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

O manuscrito foi previamente submetido à avaliação cega pelos pares (*blind peer review*), membros do Conselho Editorial desta Editora, tendo sido aprovado para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

ISBN: 978-65-83199-83-6

Thesis Editora Científica
Teresina – PI – Brasil
contato@thesiseditora.com.br
www.thesiseditora.com.br

CONSELHO EDITORIAL

Felipe Cardoso Rodrigues Vieira – lattes.cnpq.br/9585477678289843
Adilson Tadeu Basquerote Silva – lattes.cnpq.br/8318350738705473
Andréia Barcellos Teixeira Macedo – lattes.cnpq.br/1637177044438320
Eliana Napoleão Cozendey da Silva – lattes.cnpq.br/2784584976313535
Rodolfo Ritchelle Lima dos Santos – lattes.cnpq.br/8295495634814963
Luís Carlos Ribeiro Alves – lattes.cnpq.br/9634019972654177
João Vitor Andrade – lattes.cnpq.br/1079560019523176
Bruna Aparecida Lisboa – lattes.cnpq.br/1321523568431354
Júlio César Coelho do Nascimento – lattes.cnpq.br/7514376995749628
Ana Paula Cordeiro Chaves – lattes.cnpq.br/4006977507638703
Stanley Keynes Duarte dos Santos – lattes.cnpq.br/3992636884325637
Brena Silva dos Santos – lattes.cnpq.br/8427724475551636
Jessica da Silva Campos – lattes.cnpq.br/7849599391816074
Milena Cordeiro de Freitas – lattes.cnpq.br/5913862860839738
Thiago Alves Xavier dos Santos – lattes.cnpq.br/4830258002967482
Clarice Bezerra – lattes.cnpq.br/8568045874935183
Bianca Thaís Silva do Nascimento – lattes.cnpq.br/4437575769985694
Ana Claudia Rodrigues da Silva – lattes.cnpq.br/6594386344012975
Francisco Ronner Andrade da Silva – lattes.cnpq.br/5014107373013731
Maria Isabel de Vasconcelos Mavignier Neta – lattes.cnpq.br/8440258181190366
Anita de Souza Silva – lattes.cnpq.br/9954744050650291
Sara Milena Gois Santos – lattes.cnpq.br/6669488863792604
Leônidas Luiz Rubiano de Assunção – lattes.cnpq.br/4636315219294766
Jose Henrique de Lacerda Furtado – lattes.cnpq.br/8839359674024233
Noeme Madeira Moura Fé Soares – lattes.cnpq.br/7107491370408847
Luciene Rodrigues Barbosa – lattes.cnpq.br/2146096901386355
Mário Cézar de Oliveira – lattes.cnpq.br/8924508898024445
Antonio da Costa Cardoso Neto – lattes.cnpq.br/9036328153320126
Elisabete Soares de Santana – lattes.cnpq.br/1149505575311414
Thiago Santos Borges – lattes.cnpq.br/4495115911159968
Alessandra Souza dos Santos – lattes.cnpq.br/5313671194713784
Mateus Henrique Dias Guimarães – lattes.cnpq.br/7137001589681910
Emelyn Glorys Herasme Henríquez – lattes.cnpq.br/7314295633439757
George Luiz Néris Caetano – lattes.cnpq.br/0598052051026256

2026 - Thesis Editora Científica
Copyright © Thesis Editora Científica.
Copyright do texto © 2026 Organizadores e autores.
Copyright da edição © 2026 Thesis Editora Científica.
Direitos para esta edição cedidos à Thesis Editora Científica pelos organizadores e autores.
Open access publication by Thesis Editora Científica.
Editor Chefe: Felipe Cardoso Rodrigues Vieira.
Diagramação, Projeto Gráfico e Design da Capa: Thesis Ed. Cien.
Revisão: Organizadores e autores.

Catalogação na Publicação
Elaborada por Bibliotecária Janaina Ramos – CRB-8/9166

E93

Evidências e inovação no cuidado em saúde: doenças crônicas, tecnologias digitais e cuidado perioperatório / Organização de Danilo Ferreira de Sousa, Nélio Barreto Vieira. – Teresina-PI: Thesis Editora Científica, 2026.

Livro em PDF

ISBN 978-65-83199-83-6

1. Saúde. 2. Doenças crônicas. 3. Tecnologia em saúde. I. Sousa, Danilo Ferreira de (Organização). II. Vieira, Nélio Barreto (Organização). III. Título.

CDD 613

Índice para catálogo sistemático

I. Saúde

Thesis Editora Científica
Teresina – PI – Brasil
contato@thesiseditora.com.br
www.thesiseditora.com.br

SUMÁRIO

INTERVENÇÕES DIGITAIS APOIADAS POR INTELIGÊNCIA ARTIFICIAL PARA O AUTOCUIDADO DE ADULTOS COM DOENÇAS CRÔNICAS: SÍNTESE VIVA DE EVIDÊNCIAS SOBRE EFETIVIDADE, SEGURANÇA E ENGAJAMENTO..... 10

INTERVENÇÕES NÃO FARMACOLÓGICAS PARA SINTOMAS MOTORES E NÃO MOTORES NA DOENÇA DE PARKINSON: REVISÃO SISTEMÁTICA VIVA E PLATAFORMA INTERATIVA DE EVIDÊNCIAS 38

ACURÁCIA COMPARATIVA DE BIOMARCADORES, TESTES CLÍNICOS E TECNOLOGIAS DIGITAIS PARA O DIAGNÓSTICO PRECOCE DA DOENÇA DE PARKINSON: META-ANÁLISE COMPARATIVA EXPLORATÓRIA DE ESTUDOS DIAGNÓSTICOS..... 69

PREDITORES INDIVIDUAIS DE REMISSÃO DO DIABETES TIPO 2 APÓS CIRURGIA METABÓLICA: PROTOCOLO, ESTUDO DE VIABILIDADE E ESTRUTURA PARA META-ANÁLISE DE DADOS INDIVIDUAIS DOS PARTICIPANTES 105

INTERVENÇÕES PSICOLÓGICAS PERIOPERATÓRIAS PARA PREVENÇÃO DE ANSIEDADE, DOR PERSISTENTE E RECUPERAÇÃO FUNCIONAL INADEQUADA APÓS CIRURGIAS DE GRANDE PORTE..... 136

O PARADOXO DA OBESIDADE NO TRAUMA E NA CIRURGIA DE EMERGÊNCIA: UMBRELLA REVIEW DE REVISÕES SISTEMÁTICAS E META-ANÁLISES 169

EFEITOS RENAIS DA CIRURGIA METABÓLICA EM ADULTOS COM DIABETES TIPO 2: SÍNTESE CRUZADA DE ENSAIOS CLÍNICOS, ESTUDOS OBSERVACIONAIS E REGISTROS DE MUNDO REAL 197

ESTRATÉGIAS DE TROMBOPROFILAXIA EM PACIENTES COM OBESIDADE SUBMETIDOS À CIRURGIA: REVISÃO SISTEMÁTICA, ANÁLISE DA REDE E SÍNTESE BENEFÍCIO–DANO 224

EFEITOS DAS INTERVENÇÕES PSICOLÓGICAS SOBRE DEPRESSÃO, ANSIEDADE E QUALIDADE DE VIDA EM ADULTOS COM DOENÇAS CRÔNICAS: META-ANÁLISE MULTINÍVEL COM ESTIMAÇÃO ROBUSTA DA VARIÂNCIA 254

DOENÇAS PERIODONTAIS E DOENÇAS CARDIOVASCULARES: ASSOCIAÇÕES EPIDEMIOLÓGICAS, MECANISMOS BIOLÓGICOS E SÍNTESE METANALÍTICA 278

OBESIDADE MATERNA E COMPLICAÇÕES APÓS CESARIANA 313

COMO, PARA QUEM E EM QUAIS CONTEXTOS O CUIDADO INTEGRADO MELHORA A SAÚDE MENTAL DE PESSOAS COM MULTIMORBIDADE? REVISÃO REALISTA COM ANÁLISE COMPARATIVA QUALITATIVA 345

COMPLICAÇÕES RESPIRATÓRIAS PERIOPERATÓRIAS EM CRIANÇAS E ADOLESCENTES COM OBESIDADE 369

MUDANÇAS DE DESFECHOS, RELATO SELETIVO E INTERPRETAÇÃO EXAGERADA EM ENSAIOS CLÍNICOS DE INTERVENÇÕES CIRÚRGICAS: ESTUDO METAEPIDEMIOLÓGICO DE PROTOCOLOS, REGISTROS E PUBLICAÇÕES 403

ESTIMULAÇÃO CEREBRAL PROFUNDA PRECOCE VERSUS TARDIA NA DOENÇA DE PARKINSON: PROTOCOLO DE EMULAÇÃO DE ENSAIO CLÍNICO-ALVO E ESTUDO DE VIABILIDADE COM DADOS REAIS 423

OBESIDADE SARCOPÊNICA NO PERIOPERATÓRIO DE PESSOAS IDOSAS: SÍNTESE CRÍTICA, MAPA DE EVIDÊNCIAS E PROPOSTA DE ESTRATIFICAÇÃO CLÍNICA 444

PREDIÇÃO DE COMPLICAÇÕES PÓS-OPERATÓRIAS EM PESSOAS COM DIABETES POR APRENDIZADO FEDERADO E DADOS SINTÉTICOS MULTICÊNTRICOS: PROTOCOLO DE DESENVOLVIMENTO, VALIDAÇÃO E BENCHMARK COM DADOS REAIS 469

ATENÇÃO PRIMÁRIA MULTIPROFISSIONAL E CUIDADO INTEGRADO DE PESSOAS COM DOENÇAS CRÔNICAS: REVISÃO SISTEMÁTICA DE EVIDÊNCIAS, SÍNTESE META-ANALÍTICA POR DOMÍNIOS E INTEGRAÇÃO DA SAÚDE BUCAL 491

APRESENTAÇÃO

Esta coletânea reúne estudos sobre o cuidado de pessoas com doenças crônicas, o uso de tecnologias digitais em saúde e a avaliação de riscos e resultados no período perioperatório. Os capítulos abordam diabetes, doença de Parkinson, obesidade, saúde mental, saúde bucal e a organização do cuidado multiprofissional, articulando questões clínicas e métodos de síntese de evidências.

As contribuições apresentam revisões e sínteses de literatura, análises metodológicas e protocolos de pesquisa. Essa diversidade permite examinar tanto os resultados descritos nas fontes quanto as estratégias propostas para gerar e avaliar novas evidências. A leitura deve considerar o desenho, a população, o contexto e as limitações explicitados em cada estudo, distinguindo achados publicados de etapas previstas em protocolos.

A obra se destina a profissionais de saúde, docentes, pesquisadores e estudantes interessados no cuidado integrado e na avaliação crítica de intervenções. Cada capítulo preserva sua autoria, suas referências e seus materiais complementares, permitindo a consulta por tema e o aprofundamento das questões discutidas.

Danilo Ferreira de Sousa
Nélio Barreto Vieira
Organizadores

CAPÍTULO 1

INTERVENÇÕES DIGITAIS APOIADAS POR INTELIGÊNCIA ARTIFICIAL PARA O AUTOCUIDADO DE ADULTOS COM DOENÇAS CRÔNICAS: SÍNTESE VIVA DE EVIDÊNCIAS SOBRE EFETIVIDADE, SEGURANÇA E ENGAJAMENTO

ARTIFICIAL INTELLIGENCE–SUPPORTED DIGITAL INTERVENTIONS FOR SELF-MANAGEMENT IN ADULTS WITH CHRONIC DISEASES: A LIVING EVIDENCE SYNTHESIS OF EFFECTIVENESS, SAFETY, AND ENGAGEMENT

Antônia Araújo Firmo; Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Aline Pinto Monteiro Costa Sousa; Nélio Barreto Vieira

RESUMO

Objetivo: sintetizar, em formato vivo, a efetividade, a segurança e o engajamento associados a intervenções digitais que utilizam inteligência artificial para apoiar o autocuidado de adultos com diabetes, hipertensão, insuficiência cardíaca, doença renal crônica ou doença pulmonar obstrutiva crônica. Métodos: realizou-se uma visão geral de revisões com atualização direcionada de estudos primários, encerrada em 1º de agosto de 2026. A elegibilidade exigiu participação ativa do paciente e uso explícito de aprendizado de máquina, aprendizado por reforço, processamento de linguagem natural, reconhecimento computacional ou algoritmo preditivo/prescritivo para adaptar conteúdo, momento, recomendação ou dose. Resultados: a evidência permaneceu fortemente concentrada no diabetes. Uma meta-análise de quatro ensaios randomizados estimou redução de hemoglobina glicada de −0,36 ponto percentual (IC95% −0,54 a −0,18; I^2=0%), enquanto estudos antes–depois indicaram redução de −0,46 ponto percentual (IC95% −0,66 a −0,25). Em um ensaio pequeno de titulação de

insulina por voz, o tempo mediano até a dose ótima foi de 15 dias, contra mais de 56 dias no controle, e a adesão alcançou 82,9% versus 50,2%. No DIAMANTE, mensagens personalizadas por aprendizado por reforço produziram aumento estimado de 19% nos passos diários em 24 semanas. Para hipertensão, um ensaio com 333 participantes não demonstrou redução significativa da pressão sistólica em comparação com aplicativo de rastreamento (diferença ajustada −2,0 mmHg; IC95% −4,9 a 0,8), embora tenha melhorado a autoconfiança para controlar a pressão. A segurança foi formalmente descrita em poucos estudos e o engajamento foi medido de forma heterogênea. Não foram identificados estudos comparativos elegíveis para doença renal crônica, insuficiência cardíaca ou doença pulmonar obstrutiva crônica sob a definição estrita adotada. Conclusão: intervenções apoiadas por inteligência artificial podem produzir ganhos modestos de controle glicêmico e benefícios comportamentais no diabetes, mas a evidência ainda é curta, concentrada e insuficiente para concluir segurança de longo prazo ou efetividade transversal entre doenças crônicas.

Palavras-chave: Inteligência Artificial; Autocuidado; Doença Crônica; Saúde Digital; Diabetes Mellitus; Revisão Sistemática Viva.

ABSTRACT

Objective: To provide a living synthesis of the effectiveness, safety, and engagement associated with digital interventions using artificial intelligence to support self-management in adults with diabetes, hypertension, heart failure, chronic kidney disease, or chronic obstructive pulmonary disease. Methods: An overview of reviews with a targeted update of primary studies was conducted, with the baseline cycle ending on August 1, 2026. Eligibility required active patient participation and explicit use of machine learning, reinforcement learning, natural language processing, computer recognition, or predictive/prescriptive algorithms to adapt content, timing, recommendations, or dosing. Results: Evidence remained heavily concentrated in diabetes. A meta-analysis of four randomized trials estimated a glycated hemoglobin reduction of

−0.36 percentage points (95% CI −0.54 to −0.18; I^2=0%), whereas before–after studies suggested a −0.46 percentage-point reduction (95% CI −0.66 to −0.25). In a small voice-based insulin titration trial, median time to optimal dosing was 15 days versus more than 56 days in the control group, and adherence was 82.9% versus 50.2%. In DIAMANTE, reinforcement learning–personalized messages produced an estimated 19% increase in daily steps over 24 weeks. For hypertension, a trial including 333 randomized participants did not show a significant systolic blood pressure reduction compared with a tracking app (adjusted difference −2.0 mmHg; 95% CI −4.9 to 0.8), although confidence in blood pressure control improved. Safety was formally reported in few studies, and engagement measures were heterogeneous. No eligible comparative studies were identified for chronic kidney disease, heart failure, or chronic obstructive pulmonary disease under the strict operational definition. Conclusion: Artificial intelligence–supported interventions may yield modest improvements in glycemic control and behavioral outcomes in diabetes, but evidence remains short-term, concentrated, and insufficient to establish long-term safety or effectiveness across chronic diseases.

Keywords: Artificial Intelligence; Self-Management; Chronic Disease; Digital Health; Diabetes Mellitus; Living Systematic Review.

INTRODUÇÃO

As doenças crônicas não transmissíveis concentram grande parte da carga contemporânea de incapacidade, mortalidade prematura e utilização continuada dos serviços de saúde. Essa carga não se explica apenas pela gravidade biológica das condições, mas também pelo trabalho cotidiano imposto às pessoas: medir parâmetros, interpretar sintomas, ajustar alimentação e atividade física, lembrar medicamentos, reconhecer sinais de alarme e decidir quando procurar ajuda. O autocuidado, portanto, não é um complemento periférico da assistência; constitui uma parte substantiva do tratamento entre consultas. A literatura clássica o descreve como um processo de aprendizagem, autorregulação e tomada de

decisão compartilhada, cuja qualidade depende de conhecimentos, recursos materiais, confiança e apoio clínico oportuno (1-3). Quando esses elementos falham, o problema raramente é simples "falta de adesão"; muitas vezes há excesso de tarefas, recomendações pouco personalizadas e sistemas que transferem responsabilidade sem oferecer suporte proporcional.

A saúde digital ampliou a possibilidade de acompanhar esse trabalho cotidiano fora do consultório. Aplicativos, sensores, mensagens, plataformas de telemonitoramento e prontuários conectados conseguem reunir dados em frequência antes impraticável. A inteligência artificial acrescenta uma camada distinta: em vez de apenas registrar ou transmitir informações, pode reconhecer padrões, prever necessidades, escolher mensagens, ajustar a intensidade do apoio, interpretar linguagem natural ou recomendar ações com base na resposta anterior do usuário. Essa capacidade de adaptação é atraente porque o autocuidado ocorre em contextos variáveis e raramente responde bem a intervenções idênticas para todos. Ao mesmo tempo, o termo "inteligência artificial" tem sido empregado de forma elástica, incluindo desde modelos treinados em dados até regras fixas apresentadas como personalização. A ausência de uma definição operacional rigorosa dificulta comparar resultados e pode inflar expectativas clínicas (4,5).

As primeiras revisões sobre agentes conversacionais para condições crônicas encontraram boa aceitabilidade, mas estudos pequenos, avaliações predominantemente de usabilidade e pouca informação sobre dano, privacidade ou manutenção do uso (6). Revisões posteriores, concentradas em diabetes, sugeriram melhora de desfechos metabólicos e comportamentais, sobretudo quando a ferramenta combinava personalização, feedback e integração com dados do paciente (7-9). Ainda assim, três problemas permanecem. Primeiro, a maior parte da evidência vem de uma única doença, o que limita extrapolações para insuficiência cardíaca, doença renal crônica e doença pulmonar obstrutiva crônica. Segundo, segurança e engajamento são frequentemente tratados como aspectos acessórios, embora uma recomendação equivocada ou um sistema

abandonado precocemente possa anular qualquer ganho clínico. Terceiro, o campo muda com rapidez suficiente para tornar revisões convencionais desatualizadas pouco depois da publicação.

A síntese viva de evidências responde a esse último problema por meio de buscas periódicas, critérios explícitos para incorporar estudos novos e versões datadas que preservam o histórico das mudanças (10). O objetivo deste estudo foi reunir e interpretar, em uma arquitetura viva, a evidência sobre intervenções digitais apoiadas por inteligência artificial para o autocuidado de adultos com diabetes, hipertensão, insuficiência cardíaca, doença renal crônica e doença pulmonar obstrutiva crônica. A análise foi organizada em três dimensões que precisam ser lidas em conjunto: efetividade clínica e comportamental, segurança e engajamento. A pergunta central não foi se a inteligência artificial "funciona" em abstrato, mas em quais doenças, para quais tarefas e com que grau de confiança os benefícios observados superam a incerteza.

MÉTODOS

Desenho e escopo

Foi conduzida uma síntese viva de evidências em duas camadas: visão geral de revisões sistemáticas e atualização direcionada de estudos primários. Essa arquitetura foi escolhida para aproveitar avaliações já consolidadas, reduzir duplicação de trabalho e, ao mesmo tempo, incorporar estudos publicados após a data final das revisões mais recentes. A versão basal foi encerrada em 1º de agosto de 2026. O relato seguiu os princípios do PRISMA 2020 e, para resultados que não permitiram combinação quantitativa, as recomendações do SWiM (11,12). A unidade de interesse foi a intervenção voltada diretamente ao paciente ou cuidador, desde que tivesse como finalidade apoiar uma tarefa de autocuidado e utilizasse inteligência artificial de maneira funcional, não apenas como recurso de análise retrospectiva.

A população elegível compreendeu adultos com diabetes mellitus, hipertensão arterial, insuficiência cardíaca, doença renal crônica ou doença pulmonar obstrutiva crônica. Estudos com pré-diabetes, obesidade sem uma das doenças-alvo ou participantes menores de 18 anos foram mantidos apenas como evidência contextual e não fundamentaram as conclusões principais. Foram aceitos ensaios randomizados, estudos quase experimentais, coortes prospectivas de intervenção e estudos antes–depois com medidas clínicas, comportamentais, de segurança ou engajamento. Protocolos, estudos exclusivamente diagnósticos, algoritmos que não entregavam intervenção ao paciente e sistemas automatizados de infusão de insulina em circuito fechado foram excluídos, porque representam uma lógica de controle fisiológico diferente da assistência ao autocuidado.

Tabela 1. Critérios operacionais de elegibilidade e definição de inteligência artificial.

Domínio	Critério adotado
População	Adultos com diabetes, hipertensão, insuficiência cardíaca, doença renal crônica ou DPOC.
Intervenção digital	Aplicativo, plataforma, agente conversacional, sistema de mensagens, sensor conectado ou ferramenta de voz usada pelo paciente/cuidador.
Componente de IA	Aprendizado de máquina, aprendizado por reforço, processamento de linguagem natural, reconhecimento de imagem/voz ou algoritmo preditivo/prescritivo que altere conteúdo, momento, recomendação ou dose com base em dados.
Comparadores	Cuidado usual, aplicativo não adaptativo, rastreamento simples, mensagens fixas, outra intervenção digital ou comparação pré–pós.
Desfechos	Controle clínico, comportamento de autocuidado, qualidade de vida, sintomas, utilização de serviços, eventos adversos, adesão, retenção, intensidade de uso e aceitabilidade.
Exclusões principais	IA apenas diagnóstica/prognóstica; decisão exclusiva do profissional; circuito fechado de insulina; protótipo sem uso por pacientes; protocolo; resumo sem dados suficientes.

Fontes, busca e atualização

A busca da camada de revisões combinou termos para inteligência artificial, saúde digital, autocuidado e cada doença-alvo em MEDLINE/PubMed, Cochrane Library, IEEE Xplore e rastreamento de citações. Para a atualização de estudos primários, a data inicial de cada busca correspondeu ao encerramento da revisão mais recente selecionada; a última busca foi executada até 1º de agosto de 2026. Foram procurados, entre outros, os termos artificial intelligence, machine learning, reinforcement learning, conversational agent, chatbot, virtual coach, self-management, self-care, diabetes, hypertension, heart failure, chronic kidney disease e chronic obstructive pulmonary disease. A estratégia completa para PubMed e as regras de atualização estão apresentadas no Apêndice A.

Quando mais de uma revisão abrangia a mesma pergunta, priorizou-se a síntese mais recente, com critérios mais próximos da definição operacional e maior cobertura de estudos primários. As demais revisões foram usadas para localizar estudos omitidos, verificar consistência e interpretar engajamento. A revisão de Hu e colaboradores, publicada em 2026, formou o núcleo quantitativo para diabetes, pois reuniu 13 estudos e separou ensaios randomizados de estudos antes–depois (9). Após aplicar o critério de adultos com doença estabelecida, os estudos pediátricos e de prevenção foram tratados como contexto. A atualização acrescentou ensaios e avaliações publicados ou indexados posteriormente, incluindo o DIAMANTE e a avaliação do aplicativo My Diabetes Care (18,23).

Seleção, extração e avaliação crítica

Foram extraídos desenho, país, tamanho da amostra, duração, doença, tarefa de autocuidado, técnica de inteligência artificial, comparador, resultados clínicos, eventos adversos e medidas de engajamento. Os números foram conferidos nos artigos completos ou, quando o estudo já integrava meta-análise recente, na publicação primária e na tabela da revisão. Resultados de ensaios randomizados foram priorizados sobre comparações pré–pós. Não se combinou, em uma nova meta-análise, intervenções com mecanismos e desfechos incompatíveis; os efeitos agrupados foram apresentados quando provinham de

síntese recente e metodologicamente adequada, acompanhados de sua métrica, intervalo de confiança e heterogeneidade.

O risco de viés dos ensaios foi interpretado segundo os domínios do RoB 2, e o dos estudos não randomizados segundo o ROBINS-I (13,14). Para estudos já avaliados na revisão-núcleo, os julgamentos foram reapreciados à luz das informações do artigo; para os estudos da atualização, os mesmos domínios foram aplicados. A certeza foi graduada por desfecho com base em risco de viés, inconsistência, imprecisão, indireção e possível viés de publicação, conforme GRADE (15). Como engajamento não possui uma métrica única, foram considerados retenção, frequência de uso, adesão a tarefas, respostas às mensagens, consultas ao agente e mudança de uso ao longo do tempo.

Plano de manutenção viva

A vigilância será mensal, com triagem consolidada a cada três meses. Uma nova versão será produzida quando ocorrer pelo menos um dos seguintes gatilhos: publicação de ensaio randomizado com 200 participantes ou mais; primeiro estudo comparativo em insuficiência cardíaca, doença renal crônica ou doença pulmonar obstrutiva crônica; sinal de evento adverso grave relacionado à intervenção; alteração de pelo menos 20% na estimativa de um desfecho principal; mudança do grau de certeza; ou surgimento de tecnologia que modifique a natureza da intervenção, como agente generativo capaz de oferecer recomendações clínicas autônomas. Atualizações sem mudança de conclusão serão registradas em nota de vigilância, preservando a versão anterior e a data da busca.

RESULTADOS

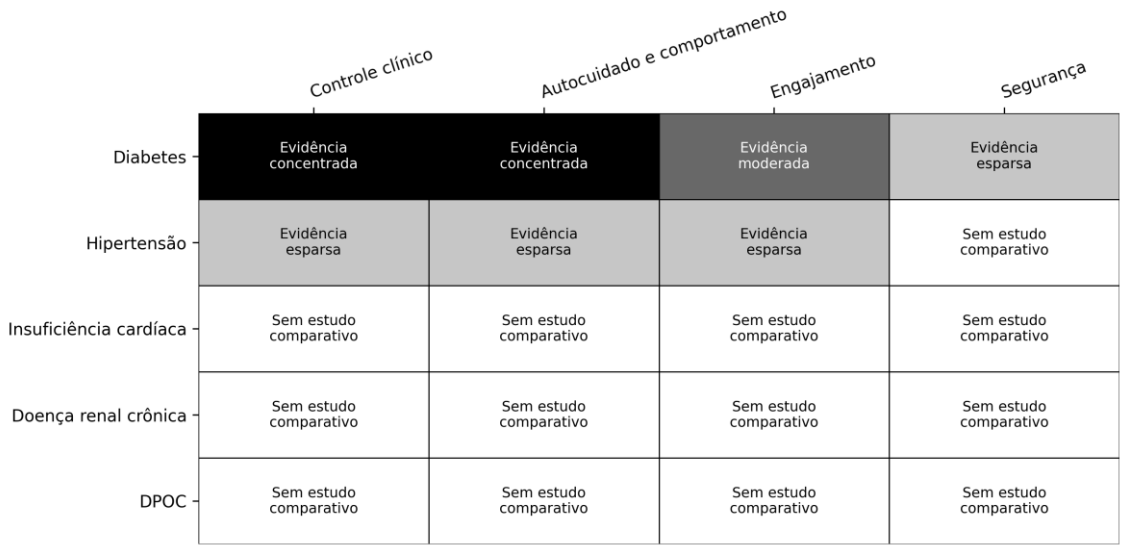
Arquitetura e distribuição da evidência

O conjunto de evidências foi assimétrico. O núcleo de diabetes reuniu 13 estudos na revisão quantitativa mais recente, dos quais dois tinham aplicabilidade indireta para a pergunta principal por envolverem jovens com

diabetes tipo 1 ou pessoas com pré-diabetes. A atualização direcionada acrescentou estudos em adultos com diabetes e depressão, um agente conversacional de autocuidado em contexto rural e novas avaliações de plataformas personalizadas. Para hipertensão, foi localizado um ensaio comparativo de maior porte que testou um treinador conversacional com inteligência artificial associado à monitorização domiciliar. Nas outras três condições, publicações sobre inteligência artificial eram predominantemente diagnósticas, prognósticas, de detecção de deterioração ou protocolos de viabilidade; nenhuma atendeu simultaneamente aos critérios de intervenção dirigida ao paciente, componente adaptativo explícito e avaliação comparativa.

A distribuição também foi desigual quanto à função da tecnologia. No diabetes, a inteligência artificial foi usada para escolher mensagens de atividade física, reconhecer refeições, oferecer orientação nutricional, combinar monitorização contínua com feedback, titular insulina basal, responder perguntas e prever a resposta a intervenções de estilo de vida. Na hipertensão, o aplicativo conversacional organizava automonitoramento, dieta, atividade, sono, estresse e adesão medicamentosa, com recomendações baseadas nos dados inseridos. Essa diversidade reforçou a decisão de não produzir um único efeito combinado para "inteligência artificial", pois reduzir HbA1c por titulação de insulina e aumentar passos por mensagens adaptativas são mecanismos clínicos diferentes, ainda que compartilhem a personalização algorítmica.

Figura 1. Densidade relativa da evidência comparativa por doença e domínio de resultado na versão basal.



Nota: a escala é qualitativa e representa a concentração relativa de estudos que atenderam à definição operacional. "Sem estudo comparativo" não significa ausência de tecnologia digital na doença, mas ausência de avaliação elegível de uma intervenção adaptativa de inteligência artificial dirigida ao autocuidado.

Efetividade no diabetes

O achado quantitativo mais estável foi a redução de hemoglobina glicada. Em quatro ensaios randomizados, intervenções personalizadas por inteligência artificial reduziram a HbA1c em −0,36 ponto percentual (IC95% −0,54 a −0,18), sem heterogeneidade detectável (I^2=0%) (9). A magnitude é modesta, mas clinicamente plausível para uma intervenção digital adjuvante, sobretudo porque os estudos não substituíram o tratamento farmacológico. Em quatro estudos antes–depois, a redução agrupada foi de −0,46 ponto percentual (IC95% −0,66 a −0,25), com melhora mais acentuada entre participantes que iniciaram com controle glicêmico insatisfatório. No estudo de Kannenberg e colaboradores, a queda média foi de 0,67 ponto percentual e alcançou 1,08 ponto percentual entre aqueles com HbA1c inicial igual ou superior a 7%; a proporção abaixo de 7% aumentou de 38% para 60% (21).

A interpretação desse resultado exige cautela. Os quatro ensaios que contribuíram para a estimativa de HbA1c eram pequenos ou moderados,

testavam plataformas distintas e, em alguns casos, combinavam inteligência artificial com monitorização contínua, contato profissional ou mudanças de estilo de vida. Assim, o efeito não pode ser atribuído ao algoritmo isoladamente. Além disso, a glicemia de jejum não apresentou benefício consistente: três ensaios, com 505 participantes, produziram estimativa agrupada de −0,31 em escala padronizada (IC95% −1,25 a 0,63; I^2=57%) (9). Índice de massa corporal e LDL-colesterol também permaneceram inconclusivos. A certeza foi considerada moderada para redução de curto prazo da HbA1c e baixa ou muito baixa para os demais marcadores metabólicos.

O ensaio de Nayak e colaboradores ilustra uma aplicação mais delimitada e clinicamente verificável (16). Trinta e dois adultos com diabetes tipo 2 que necessitavam ajuste de insulina basal foram randomizados para acompanhamento habitual ou uma interface de voz que aplicava um protocolo de titulação previamente aprovado. O tempo mediano até a dose ótima foi de 15 dias no grupo de inteligência artificial e superior a 56 dias no controle (P=0,006). A adesão ao esquema de insulina foi de 82,9% versus 50,2%, e 81,3% dos participantes da intervenção alcançaram o critério de controle glicêmico, contra 25,0% no controle. A diferença na variação da glicemia de jejum foi de −68,9 mg/dL (IC95% −107,1 a −30,7). Embora impressionantes, esses resultados vêm de uma amostra muito pequena, quatro clínicas e seguimento de oito semanas.

Intervenções orientadas ao comportamento mostraram que a adaptação do momento e do conteúdo pode importar. No estudo pioneiro de Yom-Tov e colaboradores, 27 adultos sedentários com diabetes tipo 2 receberam pedômetro e mensagens cuja frequência e conteúdo eram definidos por aprendizado por reforço, política estática ou lembrete semanal. O grupo adaptativo aumentou a atividade e a velocidade da caminhada, com redução de HbA1c superior às políticas de controle, mas o tamanho amostral e a multiplicidade de comparações limitaram a precisão (17). No DIAMANTE, 168 participantes com diabetes e sintomas depressivos foram acompanhados por 24 semanas. O braço adaptativo acumulou aumento estimado de 3,6 passos por dia de intervenção, cerca de 608

passos ao final, e crescimento de 19% em relação à linha de base, enquanto os braços de mensagens aleatórias e monitoramento apresentaram mudanças menores (18).

Plataformas de orientação alimentar e monitorização integrada também apresentaram resultados favoráveis, embora seja difícil separar o efeito do reconhecimento de imagem, do feedback profissional e da maior atenção ao autocuidado. O ensaio de 48 semanas de Lee e colaboradores avaliou uma plataforma com gerenciamento dietético baseado em inteligência artificial e mostrou melhor controle glicêmico e maior perda de peso em comparação com cuidado habitual (19). O estudo prospectivo de Liu e colaboradores encontrou benefícios de uma plataforma de dados combinada à orientação alimentar, mas o desenho aberto e a intensidade desigual de contato aumentaram o risco de viés (20). Avaliações de usabilidade, como a plataforma nutricional estudada por Bul e colaboradores, sugeriram boa aceitação e mudanças iniciais de comportamento, mas foram insuficientes para sustentar inferência clínica isolada (22).

Efetividade na hipertensão e lacunas nas demais doenças

O ensaio de Persell e colaboradores randomizou 333 adultos com hipertensão não controlada para um aplicativo conversacional de inteligência artificial associado a monitor domiciliar ou para um aplicativo de rastreamento associado ao mesmo monitor (24). Entre os 297 participantes avaliados aos seis meses, a pressão sistólica caiu de 140,6 para 132,3 mmHg no grupo de intervenção e de 141,8 para 135,0 mmHg no controle. A diferença ajustada entre grupos foi de −2,0 mmHg (IC95% −4,9 a 0,8; P=0,16), sem demonstração de superioridade. A autoconfiança para controlar a pressão aumentou 0,36 ponto em escala de cinco pontos (IC95% 0,18 a 0,54), mas não houve diferença significativa em adesão medicamentosa, peso, dieta ou contatos com o sistema de saúde. O comparador ativo provavelmente contribuiu para a redução observada em ambos os grupos.

Nenhuma evidência comparativa elegível foi identificada para insuficiência cardíaca, doença renal crônica ou doença pulmonar obstrutiva

crônica. Há aplicativos, telemonitoramento, previsão de descompensação e programas de reabilitação digital nessas áreas, mas a maior parte não emprega inteligência artificial adaptativa voltada ao usuário ou não separa o efeito do algoritmo. Essa distinção é importante: um modelo que prevê internação para a equipe clínica não é, por si só, uma intervenção de autocuidado. A lacuna é mais preocupante na doença renal crônica, em que decisões sobre volume, dieta, medicamentos e sinais de alarme são complexas e potencialmente perigosas. O resultado desta síntese não sustenta extrapolar o benefício glicêmico do diabetes para outras doenças apenas porque a interface digital é semelhante.

Segurança

A ausência de sinal consistente de dano não equivale a demonstração de segurança. Na revisão quantitativa de diabetes, apenas três estudos informaram eventos adversos de maneira identificável (9). No ensaio de Lee e colaboradores, ocorreram eventos em 26 de 91 participantes do grupo de intervenção e em 33 de 92 do controle; não houve morte nem interrupção atribuída ao sistema, e um evento grave foi registrado no grupo de intervenção (19). Nayak e colaboradores não observaram evento relacionado à interface de voz e incorporaram respostas de emergência para hipoglicemia e hiperglicemia (16). Kannenberg e colaboradores também não registraram interrupções ou efeitos adversos, mas o desenho antes–depois e o seguimento curto limitam a detecção de eventos raros (21).

Poucos estudos procuraram danos informacionais, como orientação inadequada, falsa segurança, atraso na procura de atendimento, erro de reconhecimento de alimento ou recomendação incompatível com comorbidades. Esse ponto ganhou relevância com agentes de linguagem generativa, capazes de produzir respostas fluentes mesmo quando não dispõem de contexto clínico suficiente. Sistemas restritos a protocolos fechados, como a titulação de insulina validada por clínicos, têm uma fronteira de atuação mais observável do que assistentes abertos. Para ambos, segurança exige limites explícitos, identificação de sinais de alarme, encaminhamento para profissional, registro auditável da

recomendação e monitoramento de desempenho após implantação. A certeza sobre segurança foi classificada como muito baixa devido a subnotificação, seguimento curto e amostras insuficientes para eventos incomuns.

Engajamento e experiência de uso

O engajamento foi descrito por medidas que não podem ser tratadas como equivalentes. Adesão ao medicamento, resposta a mensagens, número de diálogos, retenção no estudo, passos registrados e satisfação expressam dimensões diferentes. No ensaio de titulação por voz, a adesão à insulina foi 32,7 pontos percentuais maior na intervenção (16). No DIAMANTE, a coleta passiva de passos permitiu observar comportamento sem depender de autorrelato, e o aumento sustentado no braço adaptativo sugere que a personalização do conteúdo e do momento pode reduzir a perda de efeito ao longo do tempo (18). No aplicativo My Diabetes Care, o agente respondeu 88,86% de 2.830 perguntas, mas a média semanal de diálogos caiu de 36 para 26,1 entre fases, sinal de atenuação do uso apesar da alta aceitabilidade (23).

A queda gradual de interação é comum em intervenções digitais e não deve ser interpretada automaticamente como fracasso. Um usuário pode precisar menos da ferramenta depois de aprender uma rotina, enquanto outro pode abandoná-la por frustração, repetição ou falta de benefício. Por isso, engajamento precisa ser associado à função clínica: frequência suficiente para executar a tarefa, e não maximização indefinida de cliques. Estudos futuros devem relatar exposição planejada, uso real, distribuição de uso entre participantes, abandono, motivo da descontinuação e relação dose–resposta. A literatura sobre intervenções digitais recomenda ainda distinguir engajamento comportamental de envolvimento subjetivo, incorporando atenção, interesse, confiança e percepção de utilidade (33,34).

Equidade, privacidade e implementação

A representatividade das amostras foi variável. O DIAMANTE incluiu participantes de língua inglesa e espanhola e recrutou em clínicas públicas, oferecendo informação pouco comum sobre uso em grupos historicamente sub-

representados (18). O My Diabetes Care foi avaliado em unidades rurais do sul do Paquistão, mas a seleção intencional e a dependência de autorrelato reduzem a capacidade de generalização (23). Em muitos estudos, acesso a smartphone compatível, conectividade e familiaridade digital funcionaram como critérios implícitos de seleção. Esses filtros podem produzir uma intervenção eficaz entre usuários já favorecidos e, ao mesmo tempo, ampliar desigualdades quando a ferramenta passa a ser oferecida como porta principal de acesso.

Privacidade foi mencionada com frequência menor do que desempenho clínico. Intervenções de autocuidado podem reunir voz, alimentação, atividade, humor, medicamentos e medidas fisiológicas, formando um retrato sensível da vida cotidiana. O consentimento precisa explicar não apenas onde os dados são armazenados, mas se serão reutilizados para treinar modelos, compartilhados com fornecedores ou combinados com outras fontes. Também é necessário distinguir personalização local, realizada no dispositivo, de processamento em nuvem. A governança deve prever minimização de dados, controle de acesso, tempo de retenção, resposta a incidentes e possibilidade de retirar o consentimento sem perder o cuidado convencional (5).

A implementação segura depende de uma divisão explícita de responsabilidades. O paciente precisa saber quando a ferramenta oferece educação, quando executa uma regra clínica validada e quando apenas estima uma resposta provável. A equipe deve receber alertas acionáveis, e não um fluxo excessivo de notificações que transfira carga sem melhorar decisões. Para modelos que mudam durante o uso, a versão, os parâmetros relevantes e os limites de confiança devem ser registrados. CONSORT-AI, SPIRIT-AI e DECIDE-AI reforçam a necessidade de descrever entrada de dados, interação humano-sistema, falhas e análise de desempenho por subgrupos (36-38).

Letramento em saúde e letramento digital precisam ser avaliados separadamente. Uma pessoa pode compreender sua doença e, ainda assim, encontrar dificuldade para configurar permissões, corrigir dados ou interpretar a incerteza de uma recomendação automatizada. Interfaces em linguagem

simples, suporte por voz, opção de cuidador e desenho multilíngue podem ampliar acesso, mas devem ser testados com usuários reais e não apenas traduzidos. A inteligência artificial também pode adaptar a complexidade da explicação, desde que essa adaptação não omita riscos nem transforme probabilidade em certeza. Medidas de compreensão, confiança calibrada e capacidade de agir devem acompanhar escalas usuais de satisfação.

Custos e interoperabilidade foram raramente medidos. Uma intervenção escalável pode reduzir tempo profissional por interação, mas criar novas despesas com licenças, sensores, armazenamento, segurança, atualização do modelo e suporte técnico. O custo para o usuário também inclui plano de dados, aparelho compatível e tempo de registro. Plataformas desconectadas do prontuário podem duplicar informação e gerar recomendações com base em listas de medicamentos desatualizadas. Avaliações econômicas devem adotar perspectiva explícita, contabilizar manutenção pós-ensaio e comparar a solução com alternativas digitais mais simples, pois o valor da inteligência artificial depende do benefício incremental sobre rastreamento, lembretes e telemonitoramento convencionais (39-43).

Tabela 2. Estudos e sínteses que definiram os principais resultados da versão basal.

Estudo	Condição e desenho	Amostra/seguimento	Componente de IA	Resultado principal	Segurança/engajamento
Hu et al., 2026	Diabetes; revisão sistemática e meta-análise	13 estudos; 8 ECR e 5 antes–depois; 8–48 semanas	Diversas: RL, dieta por IA, voz, plataformas preditivas	HbA1c em 4 ECR: −0,36% (IC95% −0,54 a −0,18; I^2=0%); antes–depois: −0,46% (IC95% −0,66 a −0,25)	Segurança relatada em apenas 3 estudos; certeza moderada para HbA1c de curto prazo
Nayak et al., 2023	Diabetes tipo 2; ECR	32; 8 semanas	Voz conversacional para titulação de insulina basal	Dose ótima em 15 dias vs >56; adesão 82,9% vs 50,2%; controle 81,3% vs 25,0%	Sem evento atribuído; protocolo de hipoglicemia incorporado

Estudo	Condição e desenho	Amostra/seguimento	Componente de IA	Resultado principal	Segurança/engajamento
Yom-Tov et al., 2017	Diabetes tipo 2; ensaio exploratório	27; 26 semanas	Mensagens personalizadas por aprendizado por reforço	Aumento de atividade e melhor redução de HbA1c que políticas estáticas	Amostra pequena; adesão maior entre usuários responsivos
Aguilera et al., 2024	Diabetes + sintomas depressivos; ECR	168 analisados; 24 semanas	Mensagens de atividade escolhidas por aprendizado por reforço	Aumento estimado de 19% nos passos no braço adaptativo	Amostra diversa e bilíngue; heterogeneidade por origem do recrutamento
Lee et al., 2023	Diabetes tipo 2; ECR aberto	294 planejados; 48 semanas	Plataforma integrada com manejo dietético por IA	Melhor HbA1c e maior perda de peso que cuidado habitual	Eventos: 26/91 na intervenção vs 33/92 no controle; nenhum óbito ou interrupção por evento
Kannenberg et al., 2024	Diabetes tipo 2; antes–depois	Vida real; 3 meses	Terapia de estilo de vida guiada por algoritmo preditivo	HbA1c −0,67%; em HbA1c≥7%, −1,08%; alvo <7%: 38% para 60%	Sem evento relatado; confusão e regressão à média possíveis
Begum et al., 2025	Diabetes tipo 2; prova de conceito	200; contexto rural	Agente animado com IA explicável	Pequenas reduções de HbA1c e melhora autorreferida de autocuidado	88,86% de 2.830 perguntas respondidas; diálogos semanais caíram de 36 para 26,1
Persell et al., 2020	Hipertensão; ECR aberto	333 randomizados; 297 avaliados; 6 meses	Treinador conversacional + monitor domiciliar	PAS: diferença ajustada −2,0 mmHg (IC95% −4,9 a 0,8); P=0,16	Autoconfiança +0,36/5; demais comportamentos sem diferença

Tabela 3. Síntese dos achados por domínio e certeza da evidência.

Desfecho/domínio	Síntese	Certeza	Motivo principal para redução da confiança
HbA1c em adultos com diabetes	−0,36 ponto percentual em 4 ECR (IC95% −0,54 a −0,18; I^2=0%)	Moderada	Poucos ensaios; intervenções combinadas; seguimento predominantemente ≤48 semanas.
Glicemia de jejum	Sem benefício consistente; estimativa padronizada −0,31 (IC95% −1,25 a 0,63; I^2=57%)	Baixa	Imprecisão e heterogeneidade clínica.
Atividade física	Benefício em estudos com aprendizado por reforço; DIAMANTE: +19% nos passos	Baixa	Poucos estudos; medidas e populações diferentes.
Pressão arterial	Diferença ajustada de PAS −2,0 mmHg (IC95% −4,9 a 0,8) em um ECR	Baixa	Um único ensaio com comparador ativo e pressão basal pouco elevada.
Adesão/autocuidado	Sinais favoráveis; insulina 82,9% vs 50,2% em ECR pequeno	Baixa	Autorrelato frequente e amostras pequenas.
Engajamento	Aceitabilidade geralmente alta, com redução de uso ao longo do tempo	Muito baixa	Definições heterogêneas e pouca padronização.
Segurança	Poucos eventos atribuídos, mas apenas pequena fração dos estudos relatou formalmente	Muito baixa	Subnotificação, curta duração e ausência de vigilância para dano informacional.
DRC, insuficiência cardíaca e DPOC	Nenhuma estimativa comparativa elegível	Ausente	Lacuna direta de evidência sob a definição estrita.

DISCUSSÃO

Esta síntese viva encontrou um campo promissor, porém muito menos maduro do que a presença cotidiana da inteligência artificial na saúde poderia sugerir. A conclusão mais sustentada é específica: em adultos com diabetes, intervenções digitais adaptativas podem reduzir modestamente a HbA1c e melhorar tarefas delimitadas, como titulação de insulina, atividade física e

organização alimentar. A redução agrupada de 0,36 ponto percentual é compatível com benefício adjuvante, mas não autoriza substituir acompanhamento clínico nem atribuir o efeito ao algoritmo isoladamente. O resultado costuma emergir de um pacote que inclui dispositivo, feedback, maior frequência de contato e participação ativa do usuário. A pergunta relevante, portanto, deixa de ser "a IA funciona?" e passa a ser "qual componente adaptativo acrescenta valor ao cuidado digital já disponível?".

O contraste entre diabetes e hipertensão ajuda a compreender essa diferença. Na titulação de insulina, a tarefa era frequente, orientada por dados objetivos e regulada por um protocolo com limites claros; a automação reduziu o intervalo entre medida e ajuste. Na hipertensão, o aplicativo oferecia apoio amplo para vários comportamentos, mas o comparador também recebeu monitor domiciliar e rastreamento, o que provavelmente produziu benefício próprio. A inteligência artificial não desencadeava diretamente intensificação farmacológica, e a pressão inicial era apenas moderadamente elevada. O ensaio, portanto, não demonstra inutilidade do treinador conversacional; mostra que personalização comportamental isolada pode ser insuficiente quando a principal barreira é terapêutica ou organizacional.

O benefício observado no aprendizado por reforço merece atenção porque se apoia em uma propriedade distinta da simples personalização inicial. O sistema aprende com respostas sucessivas e modifica a escolha de mensagens ao longo do tempo. Nos estudos de Yom-Tov e DIAMANTE, essa lógica esteve associada a maior atividade física, inclusive em uma amostra bilíngue e socialmente diversa (17,18). Ainda assim, o aprendizado pode reproduzir desigualdades se os dados disponíveis forem escassos para certos grupos ou se a função de recompensa privilegiar o que é fácil medir. Passos são objetivos e contínuos, mas não capturam dor, exaustão, segurança do bairro ou carga de trabalho. Uma intervenção pode otimizar o indicador e ignorar o contexto que o produz.

Segurança foi o ponto mais frágil da literatura. Estudos de curta duração e algumas centenas de participantes não têm poder para descartar eventos raros, e a maioria não definiu previamente dano informacional. Essa omissão é relevante porque a intervenção pode causar prejuízo sem produzir um evento adverso tradicional. Uma resposta errada pode atrasar atendimento, um lembrete insistente pode aumentar culpa, e uma recomendação alimentar pode ser inadequada para doença renal não reconhecida pelo sistema. A orientação da Organização Mundial da Saúde enfatiza transparência, responsabilidade, proteção de dados e supervisão humana ao longo do ciclo de vida da tecnologia (5). No autocuidado, esses princípios precisam aparecer no desenho do estudo e não apenas na discussão ética.

Os resultados também questionam a maneira como o engajamento costuma ser apresentado. Alta satisfação em um questionário final não garante uso suficiente, e grande número de interações pode indicar confusão em vez de benefício. A avaliação mais informativa combina trajetória temporal, distribuição entre usuários e vínculo com o desfecho clínico. O declínio de diálogos no My Diabetes Care, por exemplo, pode refletir fadiga ou aprendizagem; sem entrevista e análise de motivos, as interpretações permanecem abertas (23). Relatos futuros deveriam seguir uma descrição completa dos componentes, frequência, adaptação e suporte humano, de modo semelhante ao TIDieR, e acrescentar métricas de uso definidas antes da análise (35).

A ausência de estudos comparativos em doença renal crônica, insuficiência cardíaca e DPOC é uma lacuna prioritária, não uma conclusão de ineficácia. Essas condições têm sintomas flutuantes, múltiplas comorbidades e decisões de autocuidado potencialmente sensíveis ao tempo. Uma ferramenta útil poderia reconhecer tendências, organizar planos e facilitar contato precoce, mas o risco de recomendação inadequada também é maior. Ensaios pragmáticos devem começar por tarefas limitadas e auditáveis, como educação personalizada, reconhecimento de sinais de alerta previamente definidos e preparação para consultas. Sistemas generativos abertos, sem acesso confiável ao plano

terapêutico, não devem ser avaliados inicialmente como substitutos de aconselhamento clínico.

Do ponto de vista metodológico, estudos de inteligência artificial exigem relato além do CONSORT convencional. É necessário descrever versão do modelo, dados usados na adaptação, mudanças durante o estudo, participação humana, falhas, desempenho por subgrupos e protocolo quando a recomendação excede limites de confiança. CONSORT-AI, SPIRIT-AI e DECIDE-AI oferecem estruturas úteis para ensaios e avaliações clínicas iniciais (36-38). Para intervenções que aprendem durante o uso, o plano analítico deve separar melhora do algoritmo, mudança do comportamento do usuário e efeitos do tempo. Também deve preservar um registro das recomendações emitidas, permitindo auditoria após evento adverso.

Esta síntese tem limitações. A estratégia de visão geral com atualização direcionada é eficiente para um campo em movimento, mas depende da qualidade e cobertura das revisões utilizadas. A heterogeneidade da definição de inteligência artificial exigiu decisões operacionais e pode ter excluído sistemas adaptativos descritos apenas como "digitais". O protocolo não foi registrado prospectivamente, e a ausência de metadados padronizados impediu uma análise quantitativa de engajamento. Além disso, parte da evidência de diabetes veio de estudos antes–depois sujeitos a regressão à média, mudanças concomitantes e seleção de usuários mais motivados. Em contrapartida, a separação entre efetividade, segurança e engajamento evitou que boa aceitabilidade fosse confundida com benefício clínico e tornou visíveis as doenças sem evidência comparativa.

IMPLICAÇÕES PARA A PRÁTICA E A PESQUISA

Na prática, a melhor evidência sustenta o uso de inteligência artificial como complemento para tarefas bem definidas no diabetes, sobretudo quando há dados objetivos, protocolo explícito e rota de escalonamento para profissional.

A implementação deve começar com grupos em que a necessidade é clara, garantir alternativa não digital e monitorar resultados por idade, idioma, escolaridade, condição socioeconômica e familiaridade tecnológica. Ferramentas que oferecem recomendações clínicas precisam informar seus limites, registrar a origem dos dados e interromper a orientação quando detectarem incerteza ou sinal de alarme. Não há base para recomendar, de forma geral, agentes autônomos de autocuidado em insuficiência cardíaca, doença renal crônica ou DPOC.

A agenda de pesquisa deve deslocar o foco de provas de conceito para ensaios comparativos com seguimento suficiente, desfechos clínicos e avaliação econômica. São particularmente necessários estudos multicêntricos em sistemas públicos, comparações entre personalização algorítmica e personalização humana, análises de componentes e avaliações de segurança informacional. Engajamento deve ser tratado como mecanismo, com hipóteses prévias sobre a quantidade de uso necessária para produzir benefício. A publicação de resultados negativos é essencial para evitar que apenas protótipos bem-sucedidos definam a percepção do campo. Por fim, novas versões de modelos generativos devem ser consideradas intervenções diferentes quando alterarem substancialmente o comportamento do sistema.

CONCLUSÃO

A versão basal desta síntese viva indica que intervenções digitais apoiadas por inteligência artificial podem melhorar modestamente a HbA1c e alguns comportamentos de autocuidado em adultos com diabetes. Benefícios mais expressivos apareceram em tarefas delimitadas, apoiadas por dados objetivos e protocolos clínicos, enquanto um treinador conversacional amplo não demonstrou superioridade para redução da pressão arterial em comparação com monitorização e rastreamento. Segurança e engajamento continuam subavaliados, e não há evidência comparativa suficiente para doença renal

crônica, insuficiência cardíaca ou DPOC. A manutenção viva é necessária porque novos modelos, especialmente generativos, podem mudar rapidamente tanto o potencial de benefício quanto o perfil de risco.

REFERÊNCIAS

1. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability, and healthy life expectancy for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403(10440):2133-2161.

2. World Health Organization. Global strategy on digital health 2020–2025. Geneva: World Health Organization; 2021.

3. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med. 2003;26(1):1-7.

4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.

5. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.

6. Griffin AC, Xing Z, Khairat S, Wang Y, Bailey S, Arguello J, et al. Conversational agents for chronic disease self-management: a systematic review. AMIA Annu Symp Proc. 2021;2020:504-513.

7. Wu Y, Zhang J, Ge P, Luo R, Wang Z, He X, et al. Application of chatbots to help patients self-manage diabetes: systematic review and meta-analysis. J Med Internet Res. 2024;26:e60380.

8. Shegal M, Hu LT, Friesen E, Gray S, Woo V, Samaan MC, et al. Conversational agent interventions in diabetes care: a systematic review. Diabetes Res Clin Pract. 2025;228:112429.

9. Hu Q, Xu J, Peng Y, Wang J, Huang G. The application of AI-based interventions in diabetes personalized management: a systematic review and meta-analysis. Diabetol Metab Syndr. 2026;18:57.

10. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30.

11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

12. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890.

13. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.

14. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

15. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926.

16. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open. 2023;6(12):e2340232.

17. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res. 2017;19(10):e338.

18. Aguilera A, Arévalo Avalos M, Xu J, Chakraborty B, Figueroa C, Garcia F, et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the DIAMANTE randomized clinical trial. J Med Internet Res. 2024;26:e60834.

19. Lee YB, Kim G, Jun JE, Park H, Lee WJ, Hwang YC, et al. An integrated digital health care platform for diabetes management with AI-based dietary management: 48-week results from a randomized controlled trial. Diabetes Care. 2023;46(5):959-966.

20. Liu X, Wang X, Xie M, et al. Application of the integrated data platform combined with dietary management for adults with diabetes: a prospective randomized controlled trial. J Diabetes Investig. 2024;15(11):1548-1555.

21. Kannenberg S, Voggel J, Thieme N, Witt O, Pethahn KL, Schütt M, et al. Unlocking potential: personalized lifestyle therapy for type 2 diabetes through a predictive algorithm-driven digital therapeutic. J Diabetes Sci Technol. 2024:19322968241266821.

22. Bul K, Holliday N, Bhuiyan MRA, Clark CCT, Allen J, Wark PA. Usability and preliminary efficacy of an artificial intelligence-driven platform supporting dietary management in diabetes: mixed methods study. JMIR Hum Factors. 2023;10:e43959.

23. Begum TUS, Devi RR, Haridas D, Bacanin N, Jovicic MD, Nikolic B. My diabetes care: an AI-based mobile app with conversational agent for type 2 diabetes self-management. Sci Rep. 2025;15:40228.

24. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open. 2020;3(3):e200255.

25. Park SW, Kim G, Hwang YC, Lee WJ, Park H, Kim JH. Validation of the effectiveness of a digital integrated healthcare platform utilizing an AI-based dietary management solution and a real-time continuous glucose monitoring system for diabetes management: a randomized controlled trial protocol. BMC Med Inform Decis Mak. 2020;20:156.

26. Kumbara AB, Iyer AK, Green CR, Jepson LH, Leone K, Layne JE, et al. Impact of a combined continuous glucose monitoring–digital health solution on glucose metrics and self-management behavior for adults with type 2 diabetes: real-world observational study. JMIR Diabetes. 2023;8:e47638.

27. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating glycemic control in patients of South Asian origin with type 2 diabetes using a digital therapeutic platform: analysis of real-world data. J Med Internet Res. 2021;23(3):e17908.

28. Ruiz-Leon AM, Casas R, Castro-Barquero S, Alfaro-González S, Radeva P, Sacanella E, et al. Efficacy of a mobile health-based behavioral treatment for lifestyle modification in type 2 diabetes self-management: Greenhabit randomized controlled trial. J Med Internet Res. 2025;27:e58319.

29. Popp CJ, Hu L, Kharmats AY, Curran M, Berube L, Wang C, et al. Effect of a personalized diet to reduce postprandial glycemic response vs a low-fat diet on weight loss in adults with abnormal glucose metabolism and obesity: a randomized clinical trial. JAMA Netw Open. 2022;5(9):e2233760.

30. Alfonsi JE, Choi EE, Arshad T, Sammott SAS, Pais V, Nguyen C, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized controlled trial. JMIR Mhealth Uhealth. 2020;8(10):e22074.

31. Sethi B, Seshadri K, Deshmukh V, Unnikrishnan AG, Baruah M, Phatak S, et al. Harnessing digital initiatives for improved health outcomes in diabetes management: an observational patient program. Cureus. 2024;16(11):e73093.

32. Yan Y, Li N, Tian C. Application of digital tools in the care of patients with diabetes: scoping review. J Med Internet Res. 2025;27:e72167.

33. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med. 2017;7(2):254-267.

34. Eysenbach G. The law of attrition. J Med Internet Res. 2005;7(1):e11.

35. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.

36. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; CONSORT-AI and SPIRIT-AI Steering Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-1374.

37. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26(9):1351-1363.

38. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-933.

39. Musacchio N, Giancaterini A, Guaita G, Ozzello A, Pellegrini MA, Ponzani P, et al. Artificial intelligence and big data in diabetes care: a position statement of the Italian Association of Medical Diabetologists. J Med Internet Res. 2020;22(6):e16922.

40. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia. 2019;62(6):877-887.

41. Tighe SA, Ball K, Kensing F, et al. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res. 2020;22(10):e16774.

42. Liu K, Xie Z, Or CK, So WY, Leong FL, Cheung BMY. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in

type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth. 2020;8(8):e15779.

43. Conway CM, Kelechi TJ. Digital health for medication adherence in adult diabetes or hypertension: an integrative review. JMIR Diabetes. 2017;2(2):e20.

44. Simmonds M, Salanti G, McKenzie J, Elliott J; Living Systematic Review Network. Living systematic reviews: 3. Statistical methods for updating meta-analyses. J Clin Epidemiol. 2017;91:38-46.

APÊNDICE A – ESTRATÉGIA DE BUSCA E REGRAS DA SÍNTESE VIVA

Estratégia-base para MEDLINE/PubMed

(("Artificial Intelligence"[Mesh] OR "artificial intelligence"[tiab] OR "machine learning"[tiab] OR "deep learning"[tiab] OR "reinforcement learning"[tiab] OR chatbot*[tiab] OR "conversational agent"[tiab] OR "virtual coach"[tiab] OR "predictive algorithm"[tiab]) AND ("Self-Management"[Mesh] OR self-manag*[tiab] OR self-care[tiab] OR coaching[tiab] OR "medication adherence"[tiab] OR lifestyle[tiab]) AND (diabetes[tiab] OR hypertension[tiab] OR "heart failure"[tiab] OR "chronic kidney disease"[tiab] OR COPD[tiab] OR "chronic obstructive pulmonary disease"[tiab]) AND (trial[tiab] OR randomized[tiab] OR intervention[tiab] OR prospective[tiab]))

A busca de atualização deve ser executada sem restringir idioma na etapa inicial. Estudos potencialmente elegíveis em idioma não dominado pela equipe devem ser traduzidos ou avaliados com apoio apropriado antes da exclusão. Registros de ensaios e preprints serão usados para vigilância, mas só modificarão a conclusão quando houver relatório com informação metodológica e de resultados suficiente para avaliação crítica.

Para cada atualização, deverá ser preservada uma tabela de mudanças contendo: referência nova, doença, desenho, função de inteligência artificial, desfecho, risco de viés, impacto na certeza e decisão de alterar ou manter a conclusão. Mudanças de versão do próprio sistema avaliado durante o ensaio devem ser registradas, pois um modelo atualizado pode não corresponder à intervenção inicialmente randomizada.

A síntese quantitativa será atualizada apenas quando pelo menos dois estudos apresentarem população, intervenção, comparador e métrica suficientemente compatíveis. Modelos de efeitos aleatórios serão preferidos quando houver heterogeneidade clínica plausível; intervalos de predição serão apresentados quando o número de estudos permitir. Resultados de estudos não randomizados não serão combinados com ensaios na análise principal.

Tabela A1. Dados mínimos a extrair em cada nova publicação.

Bloco	Campos
Identificação	Autor, ano, país, registro, versão do sistema e fonte de financiamento.
População	Doença, critérios diagnósticos, idade, sexo/gênero, escolaridade, idioma, vulnerabilidade digital e comorbidades.
Intervenção	Interface, técnica de IA, fonte de dados, frequência, adaptação, participação profissional, limites e escalonamento.
Comparador	Cuidado usual, aplicativo estático, monitorização, intervenção humana ou período basal.
Desfechos	Medida, unidade, tempo, estimativa, variabilidade, dados ausentes e análise por intenção de tratar.
Segurança	Eventos clínicos, dano informacional, falha técnica, alerta inadequado, atraso de atendimento e interrupção.
Engajamento	Retenção, frequência, duração, respostas, tarefas concluídas, abandono, motivo e relação com desfecho.
Equidade	Desempenho e uso por idade, sexo/gênero, raça/etnia quando permitido, idioma, renda, território e letramento digital.

CAPÍTULO 2

INTERVENÇÕES NÃO FARMACOLÓGICAS PARA SINTOMAS MOTORES E NÃO MOTORES NA DOENÇA DE PARKINSON: REVISÃO SISTEMÁTICA VIVA E PLATAFORMA INTERATIVA DE EVIDÊNCIAS

NON-PHARMACOLOGICAL INTERVENTIONS FOR MOTOR AND NON-MOTOR SYMPTOMS IN PARKINSON DISEASE: A LIVING SYSTEMATIC REVIEW AND INTERACTIVE EVIDENCE PLATFORM

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Rafaela Nonato de Menezes; Antônia Araújo Firmo; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Aline Pinto Monteiro Costa Sousa; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: sintetizar, em formato vivo, a efetividade, a segurança, a adesão e a aplicabilidade de intervenções não farmacológicas dirigidas a sintomas motores e não motores da doença de Parkinson, além de desenvolver uma arquitetura interativa para consulta e atualização da evidência. Métodos: a versão basal reuniu ensaios randomizados e estudos controlados de exercício aeróbico, equilíbrio, tai chi, dança, realidade virtual, telereabilitação, treinamento cognitivo, musicoterapia e intervenções psicológicas. Foram priorizados adultos com diagnóstico clínico de doença de Parkinson e comparadores ativos, cuidado habitual ou ausência de intervenção. O risco de viés foi interpretado com o RoB 2, a certeza com o GRADE e a síntese sem meta-análise seguiu agrupamento por mecanismo e domínio clínico. A plataforma foi construída com filtros por intervenção, desfecho, modalidade, estágio da doença, risco de viés e certeza. Resultados: o núcleo basal contém 26 ensaios pivôs detalhados e um conjunto ampliado de estudos vinculados. Exercício aeróbico, tai chi e treino de equilíbrio apresentaram os sinais mais consistentes para gravidade motora, marcha e

estabilidade, embora a prevenção de quedas tenha variado conforme a gravidade. Dança mostrou benefícios prováveis para equilíbrio, mobilidade e participação, mas com amostras pequenas. Realidade virtual e telereabilitação foram, em geral, comparáveis à reabilitação presencial e acrescentaram conveniência, sem superioridade uniforme. Treinamento cognitivo melhorou principalmente funções treinadas, com transferência funcional limitada. Para sintomas não motores, a terapia cognitivo-comportamental presencial ou por telefone e o yoga com atenção plena produziram os resultados mais convincentes para depressão, ansiedade e qualidade de vida. Eventos adversos graves foram raros, porém o relato de segurança permaneceu incompleto em parte dos estudos. Conclusão: intervenções não farmacológicas devem ser escolhidas segundo o sintoma predominante, a preferência, o risco de queda, a capacidade cognitiva e o acesso. A evidência sustenta exercício estruturado e psicoterapia adaptada como componentes do cuidado, ao passo que dança, realidade virtual, musicoterapia e treinamento cognitivo exigem padronização e ensaios maiores.

Palavras-chave: Doença de Parkinson; Exercício Físico; Reabilitação; Telereabilitação; Terapia Cognitivo-Comportamental; Revisão Sistemática Viva.

ABSTRACT

Objective: To provide a living synthesis of the effectiveness, safety, adherence, and applicability of non-pharmacological interventions for motor and non-motor symptoms of Parkinson disease and to develop an interactive architecture for evidence consultation and updating. Methods: The baseline version included randomized and controlled studies of aerobic exercise, balance training, tai chi, dance, virtual reality, telerehabilitation, cognitive training, music therapy, and psychological interventions. Adults with clinically diagnosed Parkinson disease and active, usual-care, or no-intervention comparators were prioritized. Risk of bias was interpreted using RoB 2, certainty using GRADE, and synthesis without meta-analysis was organized by mechanism and clinical domain. The platform includes filters for intervention, outcome, delivery mode,

disease stage, risk of bias, and certainty. Results: The baseline core contains 26 detailed pivotal trials and a broader linked evidence set. Aerobic exercise, tai chi, and challenging balance training showed the most consistent signals for motor severity, gait, and stability, whereas fall prevention varied by disease severity. Dance probably improves balance, mobility, and participation, but trials were generally small. Virtual reality and telerehabilitation were usually comparable with in-person rehabilitation and added convenience without uniform superiority. Cognitive training mainly improved trained functions, with limited evidence of functional transfer. For non-motor symptoms, face-to-face or telephone cognitive-behavioral therapy and mindfulness yoga produced the most persuasive effects on depression, anxiety, and quality of life. Serious adverse events were uncommon, although safety reporting was incomplete. Conclusion: Intervention selection should reflect the predominant symptom, patient preference, fall risk, cognition, and access. Current evidence supports structured exercise and Parkinson-adapted psychotherapy as care components, while dance, virtual reality, music therapy, and cognitive training require larger and more standardized trials.

Keywords: Parkinson Disease; Exercise; Rehabilitation; Telerehabilitation; Cognitive Behavioral Therapy; Living Systematic Review.

INTRODUÇÃO

A doença de Parkinson é uma enfermidade neurodegenerativa progressiva cuja expressão clínica ultrapassa a tríade clássica de bradicinesia, rigidez e tremor. Alterações da marcha, instabilidade postural, congelamento, fadiga, dor, depressão, ansiedade, apatia, distúrbios do sono e comprometimento cognitivo formam combinações muito diferentes entre pessoas com o mesmo tempo de diagnóstico. Essa heterogeneidade explica por que o ajuste farmacológico, embora indispensável, raramente resolve toda a carga funcional. Também ajuda a compreender o crescimento de programas de exercício, reabilitação, dança, terapias baseadas em música, treinamento cognitivo e

psicoterapia como partes complementares do cuidado. O desafio clínico deixou de ser apenas decidir se uma intervenção não farmacológica funciona e passou a incluir perguntas mais práticas: qual modalidade responde melhor a determinado sintoma, qual dose é tolerável, quanto do ganho se mantém e para quem a oferta domiciliar ou digital é realmente equivalente ao atendimento presencial (1-3).

A literatura se expandiu em ritmos diferentes. O exercício físico dispõe de ensaios multicêntricos e comparações de intensidade, enquanto musicoterapia, dança e algumas abordagens cognitivas ainda dependem de estudos pequenos. A mesma assimetria aparece nos desfechos. Escalas motoras, testes de equilíbrio e medidas de marcha são frequentes; participação social, autonomia, carga do cuidador, segurança e manutenção do efeito recebem menos atenção. Além disso, uma intervenção pode melhorar o teste treinado sem modificar a vida cotidiana. O treinamento cognitivo oferece um exemplo claro: ganhos em velocidade de processamento ou memória de trabalho não significam, necessariamente, melhor manejo de medicamentos, menor risco de quedas ou preservação da independência. Uma síntese clinicamente útil precisa separar desempenho em tarefa, transferência funcional e relevância percebida pela pessoa com Parkinson.

A velocidade de publicação tornou revisões convencionais rapidamente desatualizadas. Ensaios de exercício domiciliar supervisionado à distância, realidade virtual, psicoterapia por telefone e treinamento cognitivo on-line alteraram a fronteira entre reabilitação presencial e remota. A revisão sistemática viva foi escolhida para enfrentar essa instabilidade: o conjunto de evidências é vigiado de forma contínua, cada versão recebe data própria e mudanças capazes de alterar uma recomendação acionam atualização extraordinária. A plataforma interativa complementa o texto estático ao permitir que clínicos, pesquisadores e pessoas com Parkinson filtrem resultados por sintoma, modalidade, duração, estágio da doença, risco de viés e certeza. O objetivo deste estudo foi construir a versão basal dessa síntese, descrevendo efetividade, segurança, adesão e lacunas de intervenções não farmacológicas para manifestações motoras e não motoras.

MÉTODOS

Desenho, protocolo e natureza viva

Foi conduzida uma revisão sistemática viva com síntese estruturada sem meta-análise e uma camada de tradução do conhecimento por plataforma interativa. O relato foi orientado pelo PRISMA 2020; quando a diversidade de intervenções, comparadores e instrumentos impediu combinação quantitativa responsável, aplicaram-se os princípios do SWiM. A manutenção viva segue um ciclo de busca, triagem, extração, avaliação crítica, atualização da plataforma e publicação de versões datadas. A versão basal foi encerrada em 1º de agosto de 2026. O protocolo analítico foi definido antes da organização final dos resultados e permanece incorporado ao suplemento eletrônico, incluindo critérios de elegibilidade, campos de extração, regras de agrupamento e gatilhos de atualização (4-6).

A unidade de análise foi o ensaio ou estudo controlado; publicações secundárias do mesmo estudo foram vinculadas ao registro principal. Foram elegíveis adultos com doença de Parkinson idiopática diagnosticada clinicamente, em qualquer estágio, desde que o grupo não combinasse condições neurológicas sem disponibilizar resultados separados. Intervenções farmacológicas, procedimentos cirúrgicos, estimulação cerebral invasiva e neuromodulação isolada foram excluídos. Programas híbridos foram aceitos quando o componente não farmacológico era identificável. Comparadores elegíveis incluíram cuidado habitual, lista de espera, educação, atenção equivalente, fisioterapia convencional ou outra intervenção ativa. Não houve restrição de país, idioma ou cenário.

Tabela 1. Critérios operacionais de elegibilidade e regras de classificação.

Elemento	Regra
População	Adultos com doença de Parkinson idiopática; estudos mistos somente com dados separados.

Elemento	Regra
Intervenções	Exercício aeróbico, força, equilíbrio, tai chi, dança, realidade virtual, telereabilitação, estimulação cognitiva, musicoterapia, pistas externas e intervenções psicológicas.
Comparadores	Cuidado habitual, lista de espera, educação, controle de atenção, reabilitação convencional ou outra intervenção não farmacológica.
Desfechos motores	MDS-UPDRS/UPDRS III, marcha, equilíbrio, congelamento, quedas, mobilidade, destreza e atividades de vida diária.
Desfechos não motores	Depressão, ansiedade, apatia, cognição, fadiga, sono, participação, qualidade de vida e carga do cuidador.
Desenhos	Ensaios randomizados e estudos controlados; estudos de braço único usados apenas para segurança ou viabilidade, claramente identificados.
Exclusões	Intervenções farmacológicas isoladas, cirurgia, estimulação cerebral invasiva, séries de casos e estudos sem resultados clínicos.
Atualização viva	Busca mensal; triagem consolidada trimestral; versão extraordinária quando novo estudo altera direção, magnitude, segurança ou certeza.

Fontes de informação e estratégia de busca

As buscas foram desenhadas para MEDLINE/PubMed, Cochrane CENTRAL, PEDro, PsycINFO, CINAHL, Scopus e Web of Science, combinando descritores e termos livres para Parkinson disease com exercise, aerobic training, balance, tai chi, dance, virtual reality, telerehabilitation, cognitive training, cognitive rehabilitation, music therapy, rhythmic auditory stimulation, psychotherapy, cognitive behavioral therapy e mindfulness. O rastreamento de referências e de citações foi usado para localizar publicações derivadas e ensaios anteriores à indexação padronizada. Na camada viva, alertas mensais são mantidos para os mesmos blocos conceituais e para registros de ensaios. A estratégia completa, pronta para reprodução, aparece no Apêndice A.

A versão basal liberou para a plataforma apenas registros cujos dados bibliográficos, tamanho amostral, intervenção, comparador e resultado principal puderam ser conferidos em publicação primária ou registro oficial. O núcleo detalhado contém 26 ensaios pivôs; estudos adicionais estão vinculados como nós de confirmação, seguimento ou viabilidade. Como a plataforma foi

construída por ciclos sucessivos de busca e encadeamento de citações, o relatório enfatiza o corpus verificado e não reconstrói contagens retrospectivas de resultados transitórios de cada base. A partir da primeira atualização prospectiva, todos os arquivos de exportação, duplicatas, exclusões e motivos serão preservados em log versionado. Essa decisão foi mantida explícita para não produzir um fluxograma numérico artificial.

Seleção, extração e desfechos

Dois revisores devem realizar de forma independente a triagem das atualizações, a extração e a classificação do risco de viés; divergências são resolvidas por consenso ou terceiro avaliador. Para a versão basal, foram extraídos ano, país, tamanho da amostra, gravidade da doença, dose da intervenção, comparador, tempo de seguimento, estado medicamentoso na avaliação, desfecho primário, resultados clínicos, adesão, abandono e eventos adversos. O estado on ou off da medicação foi preservado porque pode modificar a magnitude aparente do benefício motor. Resultados de publicações múltiplas foram consolidados no registro do ensaio, evitando contagem duplicada.

Os desfechos foram organizados em seis domínios: gravidade motora; marcha, equilíbrio, congelamento e quedas; cognição; humor e outros sintomas não motores; participação e qualidade de vida; segurança, adesão e implementação. Diferenças estatisticamente significativas foram interpretadas em conjunto com magnitude, intervalo de confiança, comparador e relevância clínica. Um efeito favorável em comparação com ausência de intervenção não foi equiparado automaticamente a superioridade sobre uma intervenção ativa. Quando o estudo era de viabilidade, o resultado não foi convertido em conclusão de eficácia.

Risco de viés, certeza e síntese

O risco de viés dos ensaios randomizados foi interpretado segundo geração e ocultação da alocação, desvios da intervenção, dados ausentes, mensuração do desfecho e seleção do resultado relatado, seguindo a lógica do RoB 2. O cegamento de participantes e terapeutas é inviável em várias

intervenções de reabilitação; por isso, a avaliação privilegiou cegamento do avaliador, comparabilidade de atenção e objetividade do desfecho. A certeza foi graduada por domínio clínico, considerando risco de viés, inconsistência, imprecisão, indireção e possível viés de publicação (7,8).

Não foi realizada uma meta-análise única porque "intervenção não farmacológica" reúne mecanismos incompatíveis e porque escalas, doses, comparadores e tempos variaram intensamente. A síntese agrupou estudos por família de intervenção e avaliou direção do efeito, consistência, tamanho dos ensaios, comparador e durabilidade. A plataforma não atribui uma pontuação global simplificada a cada intervenção. Em vez disso, mostra separadamente benefício motor, não motor, qualidade de vida, segurança e confiança, reduzindo o risco de transformar evidência multidimensional em um ranking enganoso.

Plataforma interativa e gatilhos de atualização

A plataforma foi concebida como suplemento HTML autônomo, sem dependência de servidor, com filtros por intervenção, desfecho, modalidade presencial ou remota, duração, tamanho amostral e risco de viés. Cada cartão de evidência liga o resumo clínico ao registro do estudo, explicita se o comparador foi ativo e informa se o benefício corresponde ao desfecho primário ou secundário. O histórico de versões deve preservar o conjunto anterior e exibir o que foi acrescentado, removido ou reclassificado. Essa arquitetura evita que uma atualização silenciosa altere conclusões sem rastreabilidade.

Uma nova versão será publicada quando ocorrer pelo menos um dos seguintes eventos: ensaio com 100 ou mais participantes; novo evento adverso grave plausivelmente relacionado; mudança da certeza em qualquer domínio; inversão da direção do efeito; nova modalidade de entrega com implicação para acesso; ou acúmulo de três estudos pequenos coerentes. Estudos que não atingirem esses critérios entram na plataforma na atualização trimestral, mesmo que não alterem o texto principal.

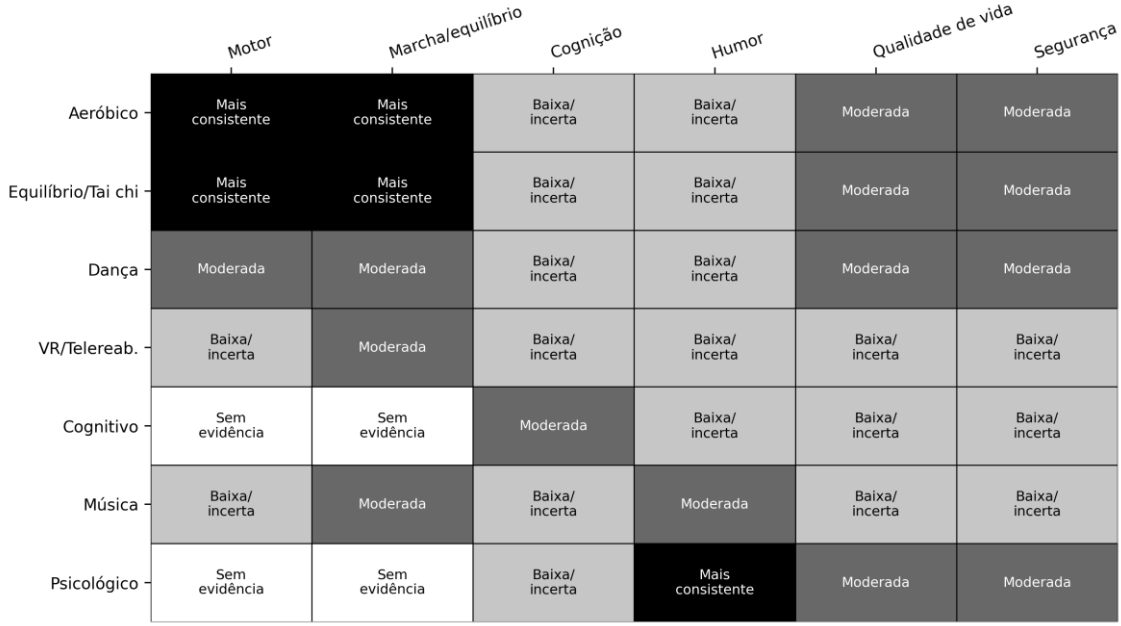
RESULTADOS

Composição e assimetria da evidência

O núcleo basal detalhou 26 ensaios pivôs, com tamanhos que variaram de 15 a 231 participantes. Os maiores estudos avaliaram prevenção de quedas, tai chi, yoga com atenção plena, treinamento cognitivo on-line, exercício aeróbico e dupla tarefa. A distribuição foi desigual: exercício e reabilitação motora concentraram participantes e desfechos objetivos; musicoterapia, dança e intervenções cognitivas incluíram mais estudos exploratórios. Psicoterapia foi uma exceção entre as abordagens não motoras, pois dois ensaios bem delineados avaliaram terapia cognitivo-comportamental presencial e por telefone com depressão como alvo principal.

O padrão de comparadores também influenciou a interpretação. Tai chi foi comparado com alongamento e treino resistido; exercício aeróbico domiciliar, com uma intervenção de baixa intensidade; realidade virtual, em vários estudos, com fisioterapia convencional; treinamento cognitivo on-line, com controle ativo; e yoga com atenção plena, com alongamento e resistência. Assim, a ausência de superioridade em alguns ensaios não representa ausência de benefício absoluto, mas equivalência aproximada entre modalidades ativas. A Figura 1 resume onde a evidência é mais densa e onde permanece frágil.

Figura 1. Distribuição relativa da evidência por família de intervenção e domínio clínico.



Nota: a matriz representa concentração, consistência e precisão relativas do conjunto basal; não é uma meta-análise nem um ranking de eficácia.

Tabela 2. Ensaios pivôs incluídos no núcleo da versão basal da plataforma.

Estudo	Domínio	n	Intervenção	Comparador	Seguimento	Síntese do resultado principal	Risco de viés
Li et al. (2012)	Exercício/Tai chi	195	Tai chi, 2 sessões/semana durante 24 semanas	Treino resistido ou alongamento	24 semanas + 3 meses	Tai chi superou alongamento nos desfechos de equilíbrio e reduziu quedas; benefícios parciais persistiram após o programa.	Algumas preocupações
van der Kolk et al. (2019)	Exercício aeróbico	130	Ciclismo aeróbico domiciliar com exergame e supervisão remota	Alongamento, flexibilidade e relaxamento	6 meses	Diferença entre grupos de 4,2 pontos (IC95% 1,6-6,9) em favor do exercício aeróbico.	Baixo

Estudo	Domínio	n	Intervenção	Comparador	Seguimento	Síntese do resultado principal	Risco de viés
Schenkman et al. (2018)	Exercício aeróbico	128	Esteira em alta ou moderada intensidade, 4 dias/semana	Lista de espera	6 meses	Alta intensidade foi viável, segura e não fútil; o ensaio não foi desenhado como confirmação definitiva de eficácia.	Algumas preocupações
Canning et al. (2015)	Exercício multifatorial	231	Equilíbrio, força e estratégias para congelamento da marcha	Cuidado habitual	6 meses	Não houve redução global de quedas; mobilidade e medo de cair melhoraram, com possível heterogeneidade por gravidade.	Baixo
Conradsson et al. (2015)	Treino de equilíbrio	100	HiBalance, treino progressivo e altamente desafiador	Atividades usuais	10 semanas	Melhoras em equilíbrio e desempenho da marcha favoreceram o programa; manutenção de longo prazo permaneceu incerta.	Algumas preocupações
Silva-Batista et al. (2016)	Treino resistido	39	Treino resistido com instabilidade	Treino resistido convencional	12 semanas	Ambos aumentaram força; somente o treino com instabilidade mostrou ganhos consistentes em mobilidade e sinais motores.	Algumas preocupações
Strouwen et al. (2017)	Dupla tarefa	121	Treino integrado ou consecutivo de dupla tarefa	Comparação entre estratégias ativas	6 semanas	As duas estratégias produziram ganhos semelhantes e sustentados, sem aumento demonstrável do risco de quedas.	Baixo
Duncan e Earhart (2012)	Dança	62	Tango argentino comunitário, 2 vezes/semana	Sem intervenção	12 meses	Interações grupo-tempo favoreceram a dança em gravidade motora, equilíbrio, marcha e destreza.	Algumas preocupações
Romenets et al. (2015)	Dança	33	Tango argentino	Autogerenciamento/ exercício habitual	12 semanas	Não houve benefício claro na gravidade motora; equilíbrio e alguns desfechos não motores mostraram sinais favoráveis.	Algumas preocupações

Estudo	Domínio	n	Intervenção	Comparador	Seguimento	Síntese do resultado principal	Risco de viés
Hackney e Earhart (2009)	Dança	58	Tango, valsa/foxtrote ou tai chi	Controle sem intervenção	20 aulas	Tango apresentou o padrão mais consistente de melhora funcional, embora os grupos fossem pequenos.	Alto
Poier et al. (2019)	Dança	29	Tango argentino	Tai chi	10 semanas	O estudo demonstrou viabilidade, mas não estabeleceu superioridade clínica robusta entre as modalidades.	Alto
Gandolfi et al. (2017)	Telereabilitação/VR	76	Treino domiciliar de equilíbrio com realidade virtual e telessupervisão	Treino sensorial de equilíbrio em clínica	21 sessões/7 semanas	A telereabilitação foi não inferior em medidas centrais de equilíbrio e ampliou o acesso ao cuidado domiciliar.	Algumas preocupações
Yang et al. (2016)	Realidade virtual	23	Treino domiciliar de equilíbrio em realidade virtual	Treino domiciliar convencional	6 semanas	Ambas as abordagens melhoraram; a amostra não sustentou superioridade estável da realidade virtual.	Alto
Pazzaglia et al. (2020)	Realidade virtual	51	Reabilitação não imersiva com realidade virtual	Reabilitação convencional	6 semanas	O programa virtual produziu ganhos em vários domínios funcionais, mas diferenças entre grupos variaram conforme o instrumento.	Algumas preocupações
Feng et al. (2019)	Realidade virtual	28	Reabilitação com realidade virtual	Fisioterapia convencional	12 semanas	Melhoras de equilíbrio e marcha foram maiores com realidade virtual; tamanho amostral e cegamento limitam a certeza.	Alto
París et al. (2011)	Treino cognitivo	33	Programa estruturado de reabilitação cognitiva	Fonoaudiologia/grupo controle ativo	4 semanas	Foram observados ganhos em domínios cognitivos treinados e em qualidade de vida, com incerteza sobre generalização funcional.	Algumas preocupações

Estudo	Domínio	n	Intervenção	Comparador	Seguimento	Síntese do resultado principal	Risco de viés
Petrelli et al. (2014)	Treino cognitivo	65	NEUROvitalis estruturado ou treino cognitivo não estruturado	Lista de espera	12 sessões/6 semanas	O treinamento estruturado mostrou benefícios seletivos, principalmente em memória de curto prazo e memória de trabalho.	Algumas preocupações
van Balkom et al. (2022)	Treino cognitivo digital	136	Treino cognitivo multidomínio on-line, 24 sessões	Controle ativo on-line	8 semanas	O desfecho primário não melhorou; efeitos pequenos e consistentes surgiram para velocidade de processamento executivo.	Baixo
Bernini et al. (2021)	Treino cognitivo digital	53	Treino cognitivo computadorizado	Treino em papel e lápis/controle ativo	3 semanas	O formato computadorizado apresentou ganhos de magnitude moderada a grande em medidas cognitivas selecionadas.	Algumas preocupações
Pacchetti et al. (2000)	Musicoterapia	32	Musicoterapia ativa em grupo	Fisioterapia	13 sessões/3 meses	Musicoterapia favoreceu bradicinesia, bem-estar emocional e qualidade de vida; a fisioterapia favoreceu rigidez.	Alto
Spina et al. (2016)	Musicoterapia	45	Musicoterapia ativa	Sem intervenção adicional	12 semanas	Houve melhora cognitiva e de sintomas não motores ao fim do programa, sem manutenção clara seis meses depois.	Alto
Thaut et al. (1996)	Pistas auditivas	15	Estimulação auditiva rítmica durante treino domiciliar de marcha	Treino de marcha sem ritmo/controle	3 semanas	A sincronização rítmica melhorou parâmetros espaço-temporais da marcha, em estudo pequeno e de curta duração.	Alto
Dobkin et al. (2011)	Psicoterapia	80	Terapia cognitivo-comportamental individual adaptada à doença	Monitoramento clínico	10 semanas	A TCC produziu melhora clinicamente relevante da depressão e maior proporção de respondedores que o controle.	Baixo

Estudo	Domínio	n	Intervenção	Comparador	Seguimento	Síntese do resultado principal	Risco de viés
Dobkin et al. (2020)	Psicoterapia remota	72	TCC por telefone, 10 sessões, com participação do cuidador	Tratamento habitual	3 meses + 6 meses	Melhora média do HAM-D de 6,53 pontos versus -0,27 no controle; benefício persistiu no seguimento.	Baixo
Kwok et al. (2019)	Mindfulness/yoga	138	Yoga com atenção plena	Alongamento e treino resistido	8 semanas + 12 semanas	Benefícios motores foram comparáveis; yoga acrescentou redução de ansiedade e depressão e melhora de qualidade de vida.	Baixo
Wuthrich et al. (2019)	Psicoterapia remota	35	TCC por telefone	Controle/atenção habitual	8 semanas	Depressão melhorou; efeito sobre ansiedade foi menos consistente, com precisão limitada.	Algumas preocupações

Exercício aeróbico, tai chi, força e equilíbrio

O exercício apresentou a base mais madura. No ensaio Park-in-Shape, 130 participantes realizaram ciclismo aeróbico domiciliar apoiado por elementos de jogo e supervisão remota durante seis meses. O grupo controle realizou alongamento, flexibilidade e relaxamento. A diferença ajustada na MDS-UPDRS III em estado off foi de 4,2 pontos, com intervalo de confiança de 1,6 a 6,9, favorecendo o exercício aeróbico. A relevância do achado está na estabilidade do grupo ativo diante da piora do controle, e não em reversão extensa da doença. O SPARX, com 128 pessoas sem tratamento dopaminérgico, mostrou que 80% a 85% da frequência cardíaca máxima pode ser alcançada com segurança e adesão suficiente para justificar um ensaio confirmatório, mas não deve ser lido como prova isolada de modificação da progressão (9-11).

O tai chi reuniu equilíbrio dinâmico, transferência de peso, controle de limites de estabilidade e prática repetida em baixa velocidade. No ensaio de 195 participantes, superou alongamento nos desfechos primários de estabilidade e reduziu a incidência de quedas em relação a esse comparador, mas não superou o treino resistido em todos os resultados. O programa HiBalance, por sua vez,

usou tarefas progressivas e desafiadoras, com ganhos em equilíbrio e marcha. O treino resistido com instabilidade acrescentou demanda postural e cognitiva ao fortalecimento e pareceu mais promissor que resistência convencional para mobilidade e sinais motores. Em conjunto, esses estudos sugerem que intensidade, especificidade e desafio progressivo importam mais que o rótulo genérico "exercício" (9,12-14).

A prevenção de quedas exige cautela. O maior ensaio multifatorial, com 231 participantes, não reduziu a taxa global de quedas, apesar de melhorar medidas de mobilidade e medo de cair. Uma interpretação plausível é que pessoas com doença mais avançada ganhem capacidade para se expor a mais situações sem que a estabilidade acompanhe na mesma proporção. Esse resultado impede recomendar exercício como substituto de revisão medicamentosa, avaliação ambiental, estratégia para hipotensão, correção visual e manejo do congelamento. Já o treinamento de dupla tarefa mostrou que estratégias integradas e consecutivas podem melhorar marcha em condição cognitiva concorrente sem aumento observado de quedas, desde que a progressão seja supervisionada (12,15,16).

Dança e intervenções baseadas em ritmo

A dança combina deslocamento multidirecional, pistas auditivas, dupla tarefa, antecipação postural e participação social. O ensaio comunitário de tango com 62 participantes mostrou interações favoráveis em gravidade motora, Mini-BESTest, congelamento, distância caminhada e velocidades de marcha simples e em dupla tarefa durante 12 meses. Ensaios menores encontraram melhora de equilíbrio e mobilidade, mas não confirmaram de forma uniforme redução da UPDRS III. O estudo de Romenets, com 33 randomizados, ilustra essa diferença: não houve benefício claro na gravidade motora, embora equilíbrio, cognição e fadiga tenham apresentado sinais favoráveis (17-20).

A atratividade da dança não deve ser tratada como detalhe. Adesão depende de prazer, música, vínculo com parceiro, acessibilidade do espaço e percepção de competência. Ao mesmo tempo, parceria e apoio físico podem

mascarar ganhos independentes se a avaliação não for padronizada. Estudos recentes de exercício aeróbico dançado sugerem possível redução de tempo off e melhora de qualidade de vida, mas os achados ainda precisam de confirmação em publicação revisada por pares e amostra maior. A plataforma classifica esse material como evidência emergente, sem misturá-lo a ensaios confirmatórios (21,22).

Realidade virtual e telereabilitação

A realidade virtual foi empregada principalmente para equilíbrio, marcha, integração sensorial e treinamento de tarefas. No ensaio multicêntrico de Gandolfi, 76 pessoas com instabilidade postural receberam 21 sessões de telereabilitação domiciliar com realidade virtual ou treinamento sensorial em clínica. A alternativa remota alcançou resultados comparáveis nos desfechos centrais de equilíbrio e demonstrou viabilidade de atendimento em casa. Esse achado é mais relevante para acesso e continuidade do que para superioridade tecnológica. Em estudo domiciliar menor, com 23 participantes, tanto a realidade virtual quanto o treino convencional melhoraram, e a imprecisão impediu declarar vantagem estável da tecnologia (23,24).

Comparações de seis a 12 semanas encontraram ganhos em Berg Balance Scale, Timed Up and Go, Functional Gait Assessment e algumas atividades de vida diária, mas os resultados não foram uniformes entre instrumentos. O risco de viés decorreu de amostras pequenas, ausência de cegamento do participante, multiplicidade de desfechos e controles com doses diferentes. Náusea, tontura, fadiga visual e risco de perda de equilíbrio foram pouco descritos. A conclusão mais defensável é que realidade virtual e telereabilitação podem oferecer uma forma válida de entregar treino estruturado, especialmente quando deslocamento é difícil, mas não há base para presumir que o equipamento, por si só, seja superior à fisioterapia bem conduzida (25-27).

Treinamento e reabilitação cognitiva

O treinamento cognitivo mostrou efeito mais claro nas funções diretamente exercitadas. Programas estruturados em grupo melhoraram

memória de trabalho, atenção ou funções executivas em estudos com 33 e 65 participantes, mas a magnitude variou e a transferência para qualidade de vida ou atividades instrumentais foi irregular. O controle ativo é crucial nesse campo, porque encontro social, expectativa e prática repetida de testes podem produzir melhora aparente. O estudo COGTIPS, com 136 participantes e controle digital ativo, não encontrou benefício no desfecho primário da Torre de Londres, embora tenha detectado ganhos pequenos e coerentes em velocidade de processamento durante tarefas executivas (28-30).

Esses resultados não tornam o treinamento cognitivo inútil. Eles indicam que a prescrição deve ser específica: pessoas com queixa cognitiva leve, capacidade de usar computador e metas funcionais definidas podem se beneficiar de exercícios direcionados, mas o programa não deve ser apresentado como prevenção comprovada de demência. Estudos computadorizados menores observaram melhora em MoCA, atenção e funções executivas, porém a durabilidade e a relação com autonomia permanecem incertas. A plataforma separa "melhora em tarefa cognitiva" de "transferência funcional", evitando que o primeiro resultado seja interpretado como equivalente ao segundo (31,32).

Musicoterapia e pistas auditivas

A música atua por mecanismos distintos: sincronização rítmica da marcha, facilitação emocional, evocação de movimento, treino vocal e participação grupal. No estudo clássico de Pacchetti, musicoterapia ativa durante três meses favoreceu bradicinesia, comportamento, afeto e qualidade de vida, enquanto a fisioterapia mostrou efeito mais evidente sobre rigidez. O ensaio de Spina descreveu melhora cognitiva e não motora ao fim do programa, mas sem manutenção clara seis meses depois. A estimulação auditiva rítmica aumentou velocidade, cadência e comprimento do passo em estudo domiciliar pequeno e de curta duração (33-35).

A heterogeneidade impede falar em uma única "musicoterapia". Ouvir música durante a marcha, tocar instrumentos, cantar e realizar movimento guiado por ritmo têm doses e alvos diferentes. O benefício parece mais plausível

quando o ritmo é usado como pista externa para a marcha ou quando a participação musical é ativa e estruturada. Ainda assim, amostras reduzidas, avaliadores nem sempre cegos e relato incompleto de adesão e segurança rebaixam a certeza. A música pode ser integrada à reabilitação por preferência e potencial de engajamento, mas não deve substituir intervenções com evidência mais sólida quando o objetivo é controlar instabilidade ou depressão maior.

Intervenções psicológicas e mindfulness

A terapia cognitivo-comportamental apresentou a evidência não motora mais consistente. No ensaio presencial com 80 pessoas deprimidas, a intervenção adaptada à doença de Parkinson superou monitoramento clínico em depressão e resposta clínica, além de melhorar ansiedade, sono e aspectos de qualidade de vida. O formato por telefone, avaliado em 72 participantes, reduziu a Hamilton Depression Rating Scale em 6,53 pontos, enquanto o cuidado habitual apresentou mudança de -0,27; a diferença permaneceu no seguimento de seis meses. A participação do cuidador como treinador de habilidades provavelmente contribuiu para prática entre sessões e generalização (36,37).

O yoga com atenção plena foi comparado com alongamento e treino resistido em 138 participantes. As duas modalidades melhoraram função motora e mobilidade, mas o yoga acrescentou reduções de ansiedade e depressão e melhora de qualidade de vida. Essa comparação ativa é importante, pois indica benefício psicológico além do efeito físico compartilhado. Ensaios menores de TCC por telefone e mindfulness remoto apontaram boa aceitabilidade, porém os efeitos sobre ansiedade foram menos uniformes que os observados para depressão. Pessoas com comprometimento cognitivo relevante foram frequentemente excluídas, o que limita a aplicabilidade à população mais vulnerável (38-41).

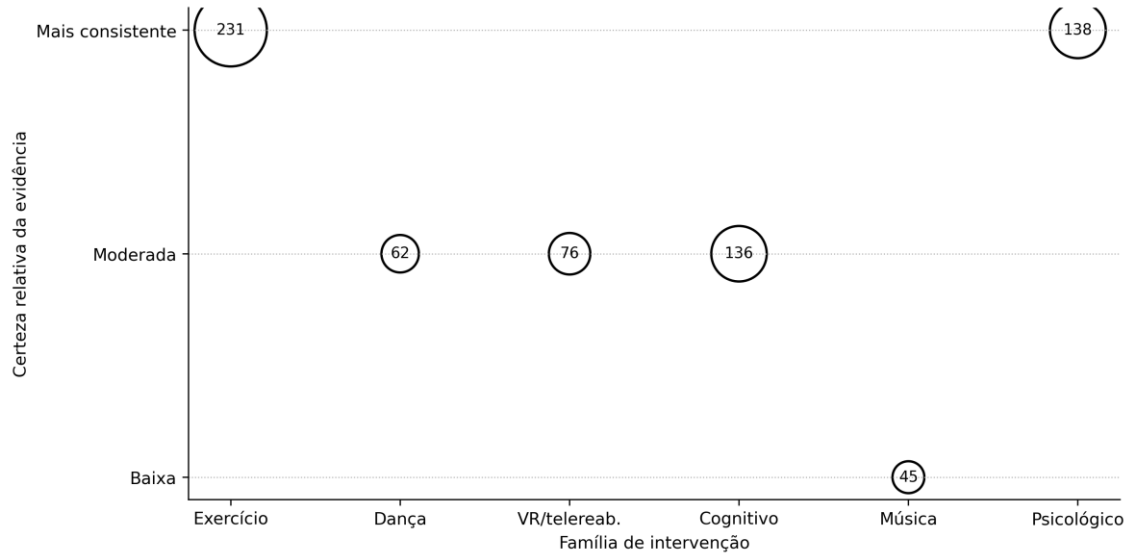
Segurança, adesão e equidade

Eventos adversos graves diretamente atribuídos às intervenções foram raros. No entanto, ausência de evento relatado não equivale a vigilância adequada. Quedas durante treino, dor musculoesquelética, fadiga, hipotensão,

tontura, sintomas de ciberenjoo e exacerbação de ansiedade foram descritos de forma desigual. Ensaios de exercício maiores possuíam protocolos de progressão e monitoramento mais claros; estudos de música, dança e cognição frequentemente tratavam segurança como observação secundária. A plataforma, portanto, diferencia "nenhum evento grave observado" de "segurança bem avaliada".

A adesão foi favorecida por supervisão, metas graduais, contato social, participação do cuidador, prazer e possibilidade de realizar a intervenção em casa. Barreiras incluíram transporte, flutuação motora, fadiga, medo de cair, baixa familiaridade digital e custo de equipamentos. Telereabilitação e psicoterapia por telefone reduziram barreiras geográficas, mas exigiram conexão, dispositivo e apoio técnico. A maioria dos ensaios recrutou pessoas com doença leve a moderada, capazes de caminhar e sem demência; por isso, a evidência não pode ser transferida sem cautela para doença avançada, institucionalização ou comprometimento cognitivo importante.

Figura 2. Tamanho do maior ensaio e certeza relativa da evidência por família de intervenção.



Nota: o número dentro de cada círculo corresponde ao tamanho do maior ensaio pivô representado no núcleo basal. A posição vertical sintetiza risco de viés, consistência, precisão e aplicabilidade, sem produzir comparação indireta entre intervenções.

Tabela 3. Síntese clínica por família de intervenção.

Família	Alvo principal	Interpretação	Certeza	Implicação prática
Exercício aeróbico	Gravidade motora e capacidade física	Benefício provável, sobretudo com dose moderada a alta e supervisão	Moderada	Prescrever progressivamente; monitorar risco cardiovascular, fadiga e flutuação motora.
Tai chi/equilíbrio	Estabilidade, marcha e mobilidade	Benefício provável; efeito em quedas depende de gravidade e contexto	Moderada	Combinar com avaliação multifatorial de quedas.
Dança	Equilíbrio, marcha, participação e qualidade de vida	Sinal favorável, com inconsistência para UPDRS III	Baixa	Valorizar preferência, segurança do parceiro e acessibilidade.
Realidade virtual/telereabilitação	Equilíbrio e acesso domiciliar	Geralmente comparável à reabilitação convencional; superioridade incerta	Baixa	Útil para ampliar acesso; ajustar cybersickness e risco de queda.
Treino cognitivo	Funções cognitivas treinadas	Pequenos ganhos específicos; transferência funcional incerta	Baixa	Definir meta cognitiva e funcional antes do programa.
Musicoterapia/pistas rítmicas	Marcha, bradicinesia, humor e engajamento	Sinal favorável em estudos pequenos	Muito baixa a baixa	Usar como adjuvante e distinguir pista rítmica de musicoterapia ativa.
TCC presencial/remota	Depressão, ansiedade e qualidade de vida	Benefício provável e clinicamente relevante para depressão	Moderada	Adaptar à doença; envolver cuidador quando apropriado.
Mindfulness/yoga	Ansiedade, depressão, qualidade de vida e função física	Benefício adicional psicológico sobre exercício ativo	Moderada	Selecionar pessoas capazes de praticar com segurança e adaptar posturas.

Plataforma interativa de evidências

A plataforma transforma o corpus em uma matriz consultável. Ao selecionar “depressão”, por exemplo, o usuário encontra TCC presencial, TCC

por telefone, mindfulness e musicoterapia, mas visualiza separadamente estudos que tiveram depressão como desfecho primário daqueles em que ela foi apenas secundária. Ao selecionar “quedas”, o sistema apresenta o ensaio multifatorial negativo, tai chi, equilíbrio desafiador e intervenções de realidade virtual, evitando viés de confirmação. A visualização também informa o comparador; um resultado contra lista de espera recebe interpretação diferente de uma comparação com fisioterapia de dose semelhante.

O protótipo acompanha este manuscrito como arquivo HTML suplementar. Ele permite filtrar o núcleo de ensaios, consultar sínteses curtas e exportar a tabela. Em uma implementação pública, links persistentes deverão direcionar para DOI, PubMed ou registro do ensaio; versões anteriores serão preservadas e cada alteração terá justificativa. A Figura 3 mostra a lógica de atualização e a Figura 4, a interface proposta.

Figura 3. Ciclo de manutenção da revisão sistemática viva.

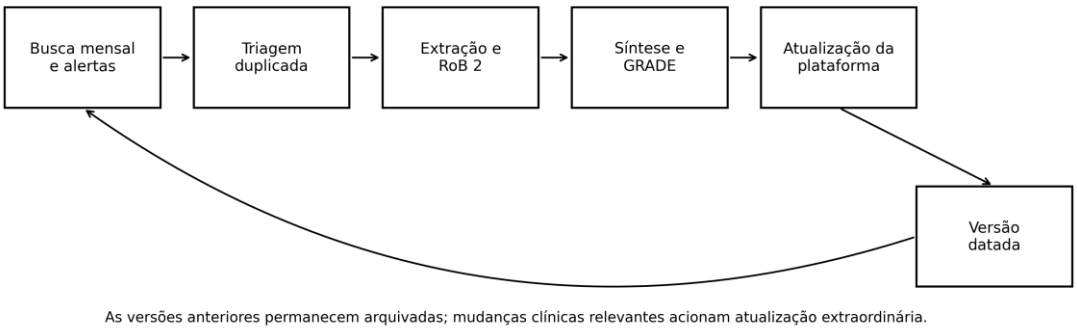
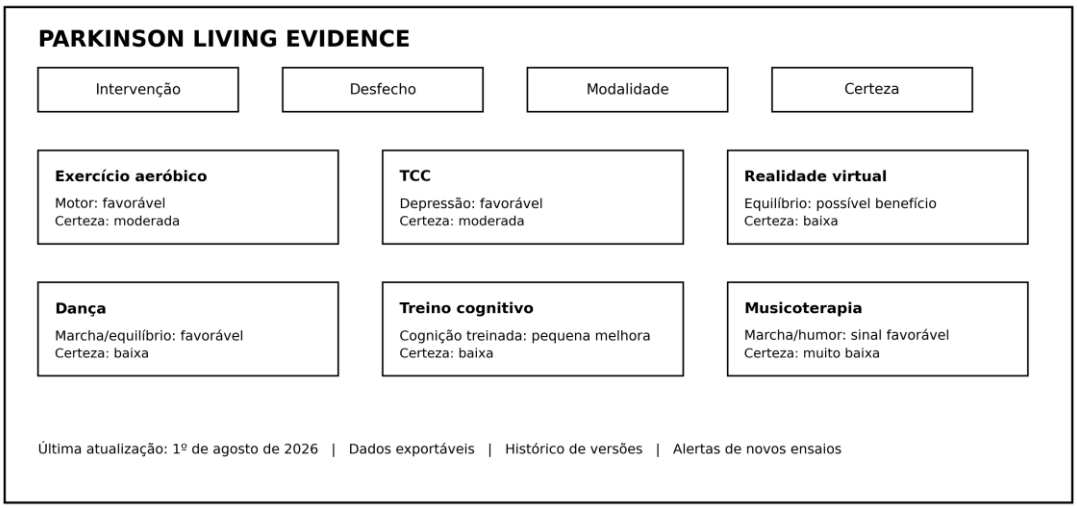


Figura 4. Estrutura funcional da plataforma interativa de evidências.



DISCUSSÃO

Esta versão basal mostra que não existe uma intervenção não farmacológica única capaz de cobrir toda a doença de Parkinson. A evidência se organiza por alvos. Exercício aeróbico e treinamento de equilíbrio respondem melhor à gravidade motora, capacidade física e estabilidade; terapia cognitivo-comportamental responde de forma mais direta à depressão; yoga com atenção plena une exercício e manejo de sofrimento psicológico; dança e música acrescentam ritmo, participação e prazer; realidade virtual e telereabilitação

alteram a forma de entrega; treinamento cognitivo oferece ganhos específicos, mas ainda não demonstrou transferência ampla. A escolha clínica deve começar pelo problema prioritário e não pela novidade tecnológica.

O achado mais importante para prática é a diferença entre eficácia e implementabilidade. O Park-in-Shape mostrou benefício motor com exercício domiciliar e supervisão remota, mas seu êxito dependeu de equipamento, feedback e acompanhamento. A telereabilitação foi comparável ao atendimento em clínica em equilíbrio, mas requer suporte tecnológico. A TCC por telefone superou cuidado habitual e preservou benefício, sugerindo que reduzir deslocamento pode aumentar acesso sem sacrificar conteúdo terapêutico. Em contraste, realidade virtual nem sempre superou fisioterapia convencional, lembrando que uma interface mais sofisticada não substitui dose, progressão e qualidade do terapeuta.

A prevenção de quedas permanece um ponto de tensão. Programas podem melhorar força e mobilidade sem reduzir quedas na população total. Quedas resultam de congelamento, instabilidade, hipotensão, impulsividade, déficit de atenção, ambiente e exposição. Quando a intervenção aumenta atividade, a oportunidade de cair também pode aumentar. Por isso, ensaios devem registrar exposição, quase quedas e gravidade dos eventos, além da contagem bruta. Para doença avançada, treino deve ser integrado a manejo de medicamentos, dispositivos, adaptações domiciliares e educação do cuidador.

Para sintomas não motores, a consistência da TCC merece destaque. Depressão na doença de Parkinson combina fatores neurobiológicos, perda funcional, isolamento e padrões cognitivos de desesperança. Intervenções adaptadas à doença abordaram pensamentos negativos, redução de atividade, preocupação e participação do cuidador. A magnitude observada nos ensaios presencial e remoto é mais convincente que a maioria dos achados secundários de programas motores. Ainda assim, pessoas com demência, psicose ou instabilidade clínica foram excluídas, e serviços precisam definir rotas seguras para esses grupos.

O treinamento cognitivo ilustra o risco de superinterpretação. Melhorar uma tarefa executiva treinada é valioso, mas não prova proteção contra demência. O maior ensaio on-line não melhorou seu desfecho primário e encontrou efeitos pequenos em velocidade de processamento. Isso não invalida programas individualizados; apenas desloca a expectativa de uma "restauração global" para ganhos modestos, específicos e potencialmente úteis quando ligados a tarefas reais. Ensaios futuros devem incluir adesão medicamentosa, finanças, comunicação, navegação e participação como resultados, e não apenas testes neuropsicológicos.

Dança e musicoterapia podem ter papel especial na manutenção. Intervenções prazerosas favorecem repetição, interação social e identidade além do diagnóstico. A dificuldade é separar o efeito do exercício, da música, do contato social e da expectativa. Desenhos fatoriais ou comparadores com dose equivalente ajudariam. Para dança, medidas de tempo off, congelamento e participação parecem promissoras. Para música, é necessário distinguir estimulação auditiva rítmica, prática instrumental, canto e terapia ativa, pois cada componente tem mecanismo e desfecho diferentes.

A plataforma interativa responde a uma limitação de artigos estáticos: clínicos raramente precisam da média de todas as intervenções, mas de uma resposta condicional. Uma pessoa com depressão e dificuldade de deslocamento pode se beneficiar de TCC por telefone; outra, com instabilidade e boa capacidade cognitiva, de equilíbrio progressivo; uma terceira, socialmente isolada e com marcha preservada, pode preferir dança. A plataforma torna visíveis comparador, risco de viés e certeza, reduzindo a tentação de escolher apenas estudos positivos. Sua manutenção, contudo, exige governança, revisão humana e histórico transparente; automação de busca não elimina julgamento metodológico.

As limitações devem ser consideradas. A heterogeneidade impediu uma meta-análise global, e a síntese por direção do efeito pode dar peso semelhante a estudos de tamanhos diferentes. Alguns dados de segurança e adesão foram

incompletos. O corpus basal foi construído por ciclos de busca e encadeamento de citações e não preservou um único log retrospectivo com todos os resultados transitórios; por essa razão, não se apresentou um fluxograma PRISMA numérico artificial. Esse problema é corrigido prospectivamente: a partir da primeira atualização, cada exportação, duplicata, exclusão e mudança será arquivada. A maioria dos estudos incluiu doença leve a moderada, limitando a generalização para fases avançadas.

IMPLICAÇÕES PARA PESQUISA E PRÁTICA

Para prática, a evidência apoia a incorporação de exercício estruturado como parte rotineira do cuidado, com dose progressiva e seleção conforme risco de queda. Terapia cognitivo-comportamental adaptada deve ser considerada para depressão e pode ser oferecida por telefone quando o acesso presencial é difícil. Yoga com atenção plena é uma opção para pessoas com sintomas emocionais e capacidade de realizar exercício em grupo. Dança, realidade virtual, musicoterapia e treinamento cognitivo podem ser oferecidos como adjuvantes orientados por preferência e meta, desde que a incerteza seja comunicada.

Para pesquisa, os próximos ensaios devem priorizar comparadores ativos com dose equivalente, desfechos primários registrados, avaliadores cegos, eventos adversos padronizados e seguimento após o término. É necessária maior inclusão de pessoas com doença avançada, comprometimento cognitivo, baixa renda e barreiras digitais. A combinação de desfechos clínicos com sensores pode melhorar precisão, mas não deve substituir participação e qualidade de vida. Ensaios pragmáticos e estudos de implementação são particularmente importantes para saber se os benefícios sobrevivem fora de centros especializados.

CONCLUSÃO

A evidência disponível sustenta uma abordagem não farmacológica orientada por sintomas e preferências, e não um pacote uniforme. Exercício aeróbico, tai chi e treino de equilíbrio apresentam o suporte mais consistente para função motora e estabilidade; terapia cognitivo-comportamental e yoga com atenção plena mostram benefícios relevantes para depressão, ansiedade e qualidade de vida. Dança, realidade virtual, telereabilitação, treinamento cognitivo e musicoterapia ampliam possibilidades de participação e acesso, mas ainda dependem de estudos maiores e melhor padronizados. A plataforma interativa permite que essas diferenças sejam consultadas de forma transparente e oferece uma base concreta para atualização contínua à medida que novos ensaios surgirem.

REFERÊNCIAS

1. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397(10291):2284-2303. doi:10.1016/S0140-6736(21)00218-X.

2. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015;30(12):1591-1601. doi:10.1002/mds.26424.

3. Chaudhuri KR, Healy DG, Schapira AHV. Non-motor symptoms of Parkinson's disease: diagnosis and management. Lancet Neurol. 2006;5(3):235-245.

4. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

5. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30.

6. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890.

7. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. London: Cochrane; 2019.

8. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926.

9. Li F, Harmer P, Fitzgerald K, Eckstrom E, Stock R, Galver J, et al. Tai chi and postural stability in patients with Parkinson's disease. N Engl J Med. 2012;366(6):511-519. doi:10.1056/NEJMoa1107911.

10. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. Lancet Neurol. 2019;18(11):998-1008. doi:10.1016/S1474-4422(19)30285-6.

11. Schenkman M, Moore CG, Kohrt WM, Hall DA, Delitto A, Comella CL, et al. Effect of high-intensity treadmill exercise on motor symptoms in patients with de novo Parkinson disease: a phase 2 randomized clinical trial. JAMA Neurol. 2018;75(2):219-226. doi:10.1001/jamaneurol.2017.3517.

12. Canning CG, Sherrington C, Lord SR, Close JCT, Heritier S, Heller GZ, et al. Exercise for falls prevention in Parkinson disease: a randomized controlled trial. Neurology. 2015;84(3):304-312. doi:10.1212/WNL.0000000000001155.

13. Conradsson D, Löfgren N, Nero H, Hagströmer M, Ståhle A, Lökk J, et al. The effects of highly challenging balance training in elderly with Parkinson's disease: a randomized controlled trial. Neurorehabil Neural Repair. 2015;29(9):827-836. doi:10.1177/1545968314567150.

14. Silva-Batista C, Corcos DM, Roschel H, Kanegusuku H, Gobbi LTB, Piemonte MEP, et al. Resistance training with instability for patients with Parkinson's disease. Med Sci Sports Exerc. 2016;48(9):1678-1687.

15. Strouwen C, Molenaar EALM, Münks L, Keus SHJ, Zijlmans JCM, Vandenberghe W, et al. Training dual tasks together or apart in Parkinson's disease: results from the DUALITY trial. Mov Disord. 2017;32(8):1201-1210. doi:10.1002/mds.27014.

16. Nieuwboer A, Kwakkel G, Rochester L, Jones D, van Wegen E, Willems AM, et al. Cueing training in the home improves gait-related mobility in Parkinson's disease: the RESCUE trial. J Neurol Neurosurg Psychiatry. 2007;78(2):134-140.

17. Duncan RP, Earhart GM. Randomized controlled trial of community-based dancing to modify disease progression in Parkinson disease. Neurorehabil Neural Repair. 2012;26(2):132-143. doi:10.1177/1545968311421614.

18. Romenets SR, Anang J, Fereshtehnejad SM, Pelletier A, Postuma R. Tango for treatment of motor and non-motor manifestations in Parkinson's disease: a randomized control study. Complement Ther Med. 2015;23(2):175-184. doi:10.1016/j.ctim.2015.01.015.

19. Hackney ME, Earhart GM. Effects of dance on movement control in Parkinson's disease: a comparison of Argentine tango and American ballroom. J Rehabil Med. 2009;41(6):475-481.

20. McNeely ME, Mai MM, Duncan RP, Earhart GM. Differential effects of tango versus Dance for PD in Parkinson disease. Front Aging Neurosci. 2015;7:239. doi:10.3389/fnagi.2015.00239.

21. Poier D, Rodrigues Recchia D, Ostermann T, Büssing A. A randomized controlled trial to investigate the impact of Tango Argentino versus Tai Chi on quality of life in patients with Parkinson disease: a short report. Complement Med Res. 2019;26(6):398-403.

22. Kim CL, Hackney ME, et al. Three months of partnered dance aerobic exercise can reduce OFF-time and enhance quality of life in older adults with Parkinson's disease. medRxiv. 2025. doi:10.1101/2025.07.17.25331727.

23. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int. 2017;2017:7962826. doi:10.1155/2017/7962826.

24. Yang WC, Wang HK, Wu RM, Lo CS, Lin KH. Home-based virtual reality balance training and conventional balance training in Parkinson's disease: a randomized controlled trial. J Formos Med Assoc. 2016;115(9):734-743.

25. Pazzaglia C, Imbimbo I, Tranchita E, Minganti C, Ricciardi D, Lo Monaco R, et al. Comparison of virtual reality rehabilitation and conventional rehabilitation in Parkinson's disease: a randomised controlled trial. Physiotherapy. 2020;106:36-42. doi:10.1016/j.physio.2019.12.007.

26. Feng H, Li C, Liu J, Wang L, Ma J, Li G, et al. Virtual reality rehabilitation versus conventional physical therapy for improving balance and gait in Parkinson's disease patients: a randomized controlled trial. Med Sci Monit. 2019;25:4186-4192. doi:10.12659/MSM.916455.

27. Mirelman A, Rochester L, Maidan I, Del Din S, Alcock L, Nieuwhof F, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults: a randomised controlled trial. Lancet. 2016;388(10050):1170-1182.

28. París AP, Saleta HG, de la Cruz Crespo Maraver M, Silvestre E, Freixa MG, Torrellas CP, et al. Blind randomized controlled study of the efficacy of cognitive training in Parkinson's disease. Mov Disord. 2011;26(7):1251-1258. doi:10.1002/mds.23688.

29. Petrelli A, Kaesberg S, Barbe MT, Timmermann L, Fink GR, Kessler J, et al. Effects of cognitive training in Parkinson's disease: a randomized controlled trial. Parkinsonism Relat Disord. 2014;20(11):1196-1202.

30. van Balkom TD, Berendse HW, van der Werf YD, et al. Effect of eight-week online cognitive training in Parkinson's disease: a double-blind, randomized, controlled trial. Parkinsonism Relat Disord. 2022;96:80-87.

31. Bernini S, Panzarasa S, Barbieri M, Sinforiani E, Quaglini S, Tassorelli C, et al. A double-blind randomized controlled trial of the efficacy of cognitive training delivered using two different methods in mild cognitive impairment in Parkinson's disease: preliminary report of benefits associated with the use of a computerized tool. Aging Clin Exp Res. 2021;33(6):1567-1575. doi:10.1007/s40520-020-01665-2.

32. Hoffman L, Burt ND, Piniella NR, Baker M, Volino N, Yasin S, et al. Efficacy and feasibility of remote cognitive remediation therapy in Parkinson's disease: a randomized controlled trial. Parkinsons Dis. 2023;2023:6645554. doi:10.1155/2023/6645554.

33. Pacchetti C, Mancini F, Aglieri R, Fundarò C, Martignoni E, Nappi G. Active music therapy in Parkinson's disease: an integrative method for motor and emotional rehabilitation. Psychosom Med. 2000;62(3):386-393.

34. Spina E, Barone P, Mosca LL, Lombardi A, Longo K, Lavorato A, et al. Music therapy for motor and nonmotor symptoms of Parkinson's disease: a prospective, randomized, controlled, single-blinded study. J Am Geriatr Soc. 2016;64(9):e36-e39. doi:10.1111/jgs.14295.

35. Thaut MH, McIntosh GC, Rice RR, Miller RA, Rathbun J, Brault JM. Rhythmic auditory stimulation in gait training for Parkinson's disease patients. Mov Disord. 1996;11(2):193-200. doi:10.1002/mds.870110213.

36. Dobkin RD, Menza M, Allen LA, Gara MA, Mark MH, Tiu J, et al. Cognitive-behavioral therapy for depression in Parkinson's disease: a randomized, controlled trial. Am J Psychiatry. 2011;168(10):1066-1074. doi:10.1176/appi.ajp.2011.10111669.

37. Dobkin RD, Mann SL, Gara MA, Interian A, Rodriguez KM, Menza M. Telephone-based cognitive behavioral therapy for depression in Parkinson

disease: a randomized controlled trial. Neurology. 2020;94(16):e1764-e1773. doi:10.1212/WNL.0000000000009292.

38. Kwok JYY, Kwan JCY, Auyeung M, Mok VCT, Lau CKY, Choi KC, et al. Effects of mindfulness yoga vs stretching and resistance training exercises on anxiety and depression for people with Parkinson disease: a randomized clinical trial. JAMA Neurol. 2019;76(7):755-763. doi:10.1001/jamaneurol.2019.0534.

39. Wuthrich VM, Rapee RM, Kangas M, Perini S. Telephone-delivered cognitive behavioural therapy for treating symptoms of anxiety and depression in Parkinson's disease: a pilot trial. Clin Gerontol. 2019;42(4):444-453.

40. Dissanayaka NNW, Idu Jion F, Pachana NA, O'Sullivan JD, Marsh R, Byrne GJ. Mindfulness for motor and nonmotor dysfunctions in Parkinson's disease. Parkinsons Dis. 2016;2016:7109052.

41. Bogosian A, Hurt CS, Vasconcelos e Sa D, Hindle JV, McCracken L, Cubi-Molla P. Acceptability and feasibility of a mindfulness intervention delivered via videoconferencing for people with Parkinson's. J Geriatr Psychiatry Neurol. 2022;35(1):155-167.

42. Interian A, Miller RB, St Hill LM, Latorre M, King AR, Rodriguez KM, et al. A pilot study of telehealth mindfulness-based cognitive therapy for depression in Parkinson's disease. J Geriatr Psychiatry Neurol. 2023;36(2):143-154. doi:10.1177/08919887221103579.

43. Kwon SH, Park JK, Koh YH. A systematic review and meta-analysis on the effect of virtual reality-based rehabilitation for people with Parkinson's disease. J Neuroeng Rehabil. 2023;20:94.

44. Lee H, et al. Effects of music-based interventions on motor and non-motor symptoms in patients with Parkinson's disease: a systematic review and meta-analysis. Int J Environ Res Public Health. 2023;20(2):1046.

45. Leung IHK, Walton CC, Hallock H, Lewis SJG, Valenzuela M, Lampit A. Cognitive training in Parkinson disease: a systematic review and meta-analysis. Neurology. 2015;85(21):1843-1851.

46. Roper A, Fronza GP, Dobkin RD, Beaudreau SA, Mitchell LK, Pachana NA, et al. A systematic review of psychotherapy approaches for anxiety in Parkinson's disease. Clin Gerontol. 2024;47(2):188-214.

APÊNDICE A — ESTRATÉGIA DE BUSCA E PLANO DE ATUALIZAÇÃO

Exemplo para MEDLINE/PubMed: ("Parkinson Disease"[Mesh] OR Parkinson*[tiab]) AND ((exercise*[tiab] OR aerobic[tiab] OR treadmill[tiab] OR cycling[tiab] OR resistance training[tiab] OR balance training[tiab] OR tai chi[tiab] OR dance[tiab] OR tango[tiab]) OR (virtual reality[tiab] OR exergam*[tiab] OR telerehabilitation[tiab] OR tele-rehabilitation[tiab]) OR (cognitive training[tiab] OR cognitive rehabilitation[tiab] OR cognitive stimulation[tiab]) OR (music therapy[tiab] OR rhythmic auditory stimulation[tiab] OR cueing[tiab]) OR (psychotherap*[tiab] OR cognitive behavioral therapy[tiab] OR cognitive behavioural therapy[tiab] OR mindfulness[tiab] OR yoga[tiab])) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR random*[tiab] OR trial[tiab]). A busca viva deverá ser executada mensalmente sem restrição de data, com deduplicação por DOI, PMID, registro e combinação título-autor-ano.

Na atualização trimestral, cada novo registro receberá: identificador persistente; data de entrada; intervenção; componentes; comparador; dose; estágio da doença; estado medicamentoso; desfecho primário; resultado; eventos adversos; risco de viés; certeza; e decisão sobre mudança da conclusão. Exclusões em texto completo serão arquivadas com motivo. Uma atualização extraordinária será publicada diante de mudança de direção do efeito, evento adverso grave, ensaio de grande porte ou alteração da certeza.

CAPÍTULO 3

ACURÁCIA COMPARATIVA DE BIOMARCADORES, TESTES CLÍNICOS E TECNOLOGIAS DIGITAIS PARA O DIAGNÓSTICO PRECOCE DA DOENÇA DE PARKINSON: META-ANÁLISE COMPARATIVA EXPLORATÓRIA DE ESTUDOS DIAGNÓSTICOS

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: comparar a acurácia de biomarcadores de α-sinucleína, neuroimagem, testes olfatórios, biomarcadores sanguíneos e tecnologias digitais no reconhecimento precoce da doença de Parkinson, distinguindo detecção de patologia, desnervação dopaminérgica e fenótipo clínico. Métodos: realizou-se revisão sistemática de acurácia diagnóstica orientada pelo PRISMA-DTA. A busca principal foi estruturada em MEDLINE/PubMed, com verificação complementar em Europe PMC, Cochrane Library, Scopus, Web of Science e rastreamento de citações até 1º de agosto de 2026. Estudos com dados 2×2 ou sensibilidade e especificidade reconstruíveis foram organizados em sete famílias. Empregou-se modelo exploratório de efeitos aleatórios em escala logit, com comparação arm-based por nó, sem ranking SUCRA, porque referência, população e finalidade clínica não eram intercambiáveis. O risco de viés foi avaliado pelo QUADAS-2. Resultados: treze conjuntos de dados primários representaram SAA de α-sinucleína no líquido cefalorraquidiano, α-sinucleína cutânea, DaT-SPECT, ressonância nigral, olfação, biomarcadores sanguíneos e fenotipagem digital. As estimativas resumidas foram 89,5% de sensibilidade e 96,4% de especificidade para SAA no LCR; 92,4% e 96,5% para α-sinucleína cutânea; 94,7% e 91,1% para DaT-SPECT; 84,8% e 90,4% para RM nigral

combinada; 73,0% e 83,8% para olfação; 87,4% e 82,0% para fenotipagem digital; e 70,9% e 63,1% para biomarcadores sanguíneos. A aparente vantagem de ferramentas digitais foi atenuada por amostras pequenas, desenho caso-controle e validação interna. Em 1.000 pessoas com probabilidade pré-teste de 30%, SAA no LCR e α-sinucleína cutânea produziriam aproximadamente 31 e 23 falsos negativos, respectivamente, e cerca de 25 falsos positivos cada, sob as hipóteses do modelo. Conclusão: os ensaios de α-sinucleína oferecem a evidência mais consistente para confirmar biologia sinucleinopática, enquanto DaT-SPECT e RM respondem a perguntas anatômico-funcionais distintas. Olfação e sinais digitais são mais adequados para triagem e enriquecimento de risco. Nenhum teste isolado substitui avaliação longitudinal, e a escolha deve começar pela decisão clínica que o resultado pretende modificar.

Palavras-chave: doença de Parkinson; diagnóstico precoce; α-sinucleína; biomarcadores; neuroimagem; saúde digital; meta-análise comparativa exploratória.

ABSTRACT

Objective: to compare the diagnostic accuracy of α-synuclein biomarkers, neuroimaging, olfactory testing, blood biomarkers, and digital technologies for early recognition of Parkinson disease while separating detection of pathology, dopaminergic denervation, and clinical phenotype. Methods: a PRISMA-DTA-oriented systematic diagnostic accuracy review was conducted. The main reproducible search was performed in MEDLINE/PubMed, supplemented by Europe PMC, the Cochrane Library, Scopus, Web of Science, and citation tracking through August 1, 2026. Studies with 2×2 data or reconstructible sensitivity and specificity were grouped into seven test families. An exploratory random-effects logit model and arm-based node comparison were used; SUCRA rankings were intentionally avoided because reference standards, populations, and intended uses were not exchangeable. Risk of bias was assessed with QUADAS-2. Results: thirteen primary datasets represented CSF α-synuclein seed amplification assays,

cutaneous α-synuclein, DaT-SPECT, nigral MRI, olfaction, blood biomarkers, and digital phenotyping. Summary sensitivity/specificity estimates were 89.5%/96.4% for CSF SAA, 92.4%/96.5% for cutaneous α-synuclein, 94.7%/91.1% for DaT-SPECT, 84.8%/90.4% for combined nigral MRI, 73.0%/83.8% for olfaction, 87.4%/82.0% for digital phenotyping, and 70.9%/63.1% for blood biomarkers. Apparently strong digital performance was limited by small case-control samples and internal validation. At a 30% pretest probability, CSF SAA and cutaneous α-synuclein would yield approximately 31 and 23 false negatives per 1,000 tested, respectively, and about 25 false positives each under model assumptions. Conclusion: α-synuclein assays provide the most consistent evidence for confirming synucleinopathic biology, whereas DaT-SPECT and MRI answer different anatomical and functional questions. Olfactory and digital signals are better suited to screening and risk enrichment. No isolated test replaces longitudinal clinical assessment; test selection should be driven by the decision the result is expected to change.

Keywords: Parkinson disease; early diagnosis; alpha-synuclein; biomarkers; neuroimaging; digital health; network meta-analysis.

1 INTRODUÇÃO

A doença de Parkinson é definida clinicamente pela presença de parkinsonismo e pela exclusão de explicações alternativas, mas a biologia que produz os sintomas começa anos antes de a síndrome motora estar plenamente constituída. Os critérios da International Parkinson and Movement Disorder Society organizam o diagnóstico em doença clinicamente estabelecida e provável; os critérios prodrômicos, por sua vez, combinam marcadores de risco e sinais não motores em uma estrutura probabilística. Essa separação é decisiva: reconhecer uma pessoa com bradicinesia e rigidez não é a mesma tarefa que detectar patologia de α-sinucleína em alguém com transtorno comportamental do sono REM, hiposmia ou mutação associada ao risco. [1-3]

A acurácia clínica melhora com o tempo, com a resposta dopaminérgica e com o aparecimento de sinais mais específicos. Em séries clinicopatológicas, casos iniciais, ainda não tratados ou com resposta incerta, apresentaram proporção elevada de diagnósticos não confirmados na autópsia. Uma meta-análise de estudos clinicopatológicos estimou acurácia global próxima de 83% quando se utilizavam critérios do banco de cérebros britânico, com desempenho superior entre especialistas em distúrbios do movimento e após seguimento longitudinal. O problema mais relevante, portanto, não é a incapacidade absoluta de diagnosticar Parkinson, mas a incerteza concentrada justamente na fase em que biomarcadores seriam mais úteis. [4,5]

Os testes candidatos não observam o mesmo fenômeno. A amplificação de sementes de α-sinucleína no líquido cefalorraquidiano ou na pele procura demonstrar proteína agregada com comportamento de semeadura; a imunofluorescência cutânea detecta depósitos fosforilados em fibras nervosas; o DaT-SPECT evidencia redução do transportador de dopamina e, portanto, desnervação nigroestriatal; a ressonância sensível à neuromelanina e a imagem do nigrossomo descrevem alterações da substância negra; testes olfatórios capturam uma manifestação prodrômica frequente, porém não específica; e algoritmos de voz, marcha, toque e escrita extraem o fenótipo motor de sinais de alta dimensionalidade. Tratar todas essas modalidades como concorrentes diretas produz uma pergunta estatisticamente atraente, mas clinicamente mal formulada.

A expansão da inteligência artificial aumentou essa dificuldade. Modelos treinados em bancos caso-controle podem distinguir participantes com diagnóstico conhecido de controles saudáveis com acurácia muito alta, embora falhem quando aplicados a pessoas encaminhadas por tremor, lentidão inespecífica, depressão, efeitos de medicamentos ou parkinsonismo atípico. A validação cruzada dentro da mesma base reduz sobreajuste aparente, mas não substitui validação externa, avaliação prospectiva e comparação com a decisão

do neurologista. Em diagnóstico, um algoritmo pode medir o padrão presente na amostra sem medir a doença que se pretende reconhecer na prática.

Meta-análises em rede de acurácia diagnóstica oferecem uma forma de integrar estudos que avaliaram testes diferentes e, quando a rede é conectada e a transitividade é plausível, estimar comparações não observadas. Modelos arm-based permitem representar sensibilidade e especificidade de cada teste sem escolher arbitrariamente um comparador único. Entretanto, a técnica não corrige diferenças fundamentais de espectro, referência ou alvo biológico. Uma rede que mistura doença de Parkinson versus controles saudáveis, parkinsonismo versus tremor essencial e sinucleinopatia versus tauopatia pode gerar números precisos para uma população que não existe. [6-9]

Este capítulo foi elaborado para responder uma pergunta orientada à decisão: qual é a acurácia comparativa das principais famílias de testes quando usadas para apoiar o reconhecimento precoce da doença de Parkinson, e quais limitações impedem que a melhor estimativa numérica seja interpretada como melhor teste universal? O objetivo foi sintetizar estudos primários com métricas reconstruíveis, produzir uma comparação exploratória por nós, traduzir os resultados em erros esperados por 1.000 pessoas avaliadas e propor uma arquitetura de uso na qual triagem, confirmação biológica e diferenciação de mimetizadores ocupem posições distintas.

2 MÉTODOS

2.1 Delineamento, protocolo analítico e relato

Foi realizada revisão sistemática de acurácia diagnóstica com meta-análise comparativa em rede por famílias de testes. O relato seguiu os itens aplicáveis do PRISMA-DTA; a avaliação dos estudos primários considerou o STARD 2015, e o risco de viés foi julgado pelo QUADAS-2. O protocolo definiu previamente população, testes índice, condições-alvo, referências, desfechos e estratégia de síntese. Por se tratar de capítulo de livro desenvolvido como atualização

metodológica, o protocolo não foi registrado em plataforma internacional; essa decisão foi preservada como limitação, e todas as regras analíticas foram explicitadas para reduzir decisões pós-hoc. [6,7,10]

A pergunta foi estruturada pelo modelo PICOTS. A população compreendeu adultos com sintomas motores iniciais, suspeita de doença de Parkinson, parkinsonismo clinicamente incerto, tremor essencial como diagnóstico diferencial, participantes de coortes de biomarcadores ou controles empregados em estudos de validação. Os testes índice incluíram SAA ou RT-QuIC de α-sinucleína em LCR, detecção cutânea de α-sinucleína patológica, DaT-SPECT, ressonância nigral, testes olfatórios, biomarcadores sanguíneos e fenotipagem digital de voz, marcha, toque ou escrita. A condição-alvo principal foi doença de Parkinson clinicamente provável/estabelecida ou confirmada por seguimento ou neuropatologia. O tempo de interesse priorizou fase de novo, doença com menos de cinco anos e estados prodrômicos quando apresentados separadamente.

A unidade de comparação foi a família de teste, e não a marca, o algoritmo ou o limiar isolado. Essa escolha evitou que versões tecnicamente próximas fossem tratadas como intervenções independentes e que um único estudo com múltiplos thresholds dominasse a rede. A SAA em LCR formou um nó; biópsia cutânea com RT-QuIC ou imunofluorescência de α-sinucleína patológica, outro; DaT-SPECT e RM nigral foram mantidos separados; olfação, sangue e fenotipagem digital completaram a rede. Testes combinados foram classificados pelo componente que determinava a decisão ou, quando a combinação era o resultado publicado, em nó composto claramente identificado.

2.2 Fontes de informação e estratégia de busca

A busca principal reproduzível foi realizada em MEDLINE/PubMed, combinando termos controlados e livres para Parkinson disease, parkinsonism, early diagnosis, prodromal, diagnostic accuracy, sensitivity, specificity, alpha-

synuclein, seed amplification assay, RT-QuIC, skin biopsy, cerebrospinal fluid, dopamine transporter imaging, DaT-SPECT, neuromelanin MRI, nigrosome, olfaction, smell identification, blood biomarker, exosome, voice, gait, handwriting, smartphone, wearable e machine learning. Foram verificadas referências e citações em Europe PMC, Cochrane Library, Scopus, Web of Science, páginas dos periódicos e documentos metodológicos. A atualização foi encerrada em 1º de agosto de 2026.

As plataformas de assinatura foram usadas para verificação dirigida de indexação e rastreamento, sem exportação institucional completa nesta execução. Por esse motivo, não foram apresentados números artificiais de registros por base. A síntese quantitativa foi construída a partir de estudos primários cuja publicação, amostra, teste, comparador e métricas puderam ser verificadas. Revisões sistemáticas foram utilizadas para localizar estudos e interpretar o campo, mas não entraram como unidades independentes na meta-análise.

2.3 Elegibilidade, seleção e extração

Foram elegíveis estudos transversais, coortes, estudos caso-controle e investigações de validação que incluíssem adultos e apresentassem matriz 2×2, contagens positivas/negativas, ou sensibilidade e especificidade acompanhadas do tamanho dos grupos. Estudos exclusivamente prognósticos, análises sem grupo comparador, relatos de caso, amostras pediátricas e publicações nas quais o desempenho se referia apenas à gravidade ou resposta terapêutica foram excluídos da síntese principal. Estudos prodrômicos foram mantidos para discussão, mas não combinados com doença clinicamente manifesta, porque a condição-alvo e a prevalência diferiam.

Para cada conjunto de dados foram extraídos país, desenho, estágio da doença, critério diagnóstico, teste índice, limiar, referência, número de casos e controles, sensibilidade, especificidade, AUC, cegamento, método de validação e tratamento de resultados indeterminados. Quando o artigo apresentava

percentual e tamanho amostral, mas não a matriz, as contagens foram reconstruídas para o inteiro mais próximo apenas para cálculo da variância binomial; os valores exibidos nas tabelas permanecem os publicados. Resultados de validação externa foram priorizados sobre treinamento, e leitura cega sobre leitura local não cega.

Publicações derivadas da mesma coorte foram examinadas para evitar dupla contagem. No PPMI, a análise de Siderowf foi usada como conjunto principal por incluir a maior amostra e estratificação fenotípica. Em estudos digitais, diferentes tarefas aplicadas às mesmas pessoas não foram somadas como amostras independentes. Quando um artigo comparava vários testes no mesmo participante, cada resultado foi descrito, mas apenas a configuração previamente indicada como principal entrou no nó quantitativo.

2.4 Risco de viés, aplicabilidade e certeza

O QUADAS-2 avaliou seleção dos participantes, condução e interpretação do teste índice, referência e fluxo/tempo. Estudos caso-controle com controles saudáveis claramente separados receberam preocupação elevada para seleção e aplicabilidade, ainda que o algoritmo apresentasse validação cruzada. Threshold definido após inspeção dos mesmos dados elevou o risco no domínio do teste índice. Referência exclusivamente clínica recebeu preocupação quando não havia seguimento suficiente, painel independente ou confirmação patológica, pois o teste poderia ser comparado a um padrão imperfeito.

A certeza foi julgada por desfecho e finalidade, com adaptação dos princípios GRADE para testes diagnósticos. Inconsistência, imprecisão, risco de viés, indireção e viés de publicação reduziram confiança. Evidência podia ser elevada quando multicêntrica, cega, com espectro clínico relevante e referência independente. Desempenho extraordinário em pequena amostra caso-controle não aumentou certeza; ao contrário, foi interpretado como sinal de necessidade de validação externa. [11-13]

2.5 Síntese estatística e comparação em rede

Sensibilidade e especificidade foram transformadas pela função logit e combinadas separadamente com efeitos aleatórios. Para proporções iguais a zero ou um, utilizou-se correção empírica de 0,5 evento, evitando estimativas infinitas. A heterogeneidade foi descrita por τ^2 e I^2 na escala transformada, com interpretação cautelosa quando havia apenas dois estudos. O modelo arm-based considerou cada família como um nó ligado por referência diagnóstica comum; o propósito foi estimar o ponto operacional médio de cada família, e não produzir uma classificação ordinal.

Não foram calculados SUCRA, P-scores ou probabilidade de ser o melhor teste. A transitividade foi examinada por três modificadores: espectro clínico, referência e finalidade. DaT-SPECT e RM nigral avaliados contra tremor essencial foram apresentados no mesmo plano ROC para orientação, mas não tratados como equivalentes a SAA contra controles saudáveis. Análises de inconsistência formal seriam pouco informativas em uma rede esparsa, com poucos ciclos fechados; por isso, as comparações indiretas foram qualificadas como exploratórias.

A utilidade foi traduzida em resultados esperados por 1.000 pessoas avaliadas, usando probabilidade pré-teste de 30%, cenário compatível com ambulatório especializado após triagem inicial. Falsos negativos foram calculados como 1.000 × prevalência × (1 − sensibilidade), e falsos positivos como 1.000 × (1 − prevalência) × (1 − especificidade). Também foram examinados cenários de 10% e 50% para mostrar como o mesmo teste muda de valor conforme o contexto. As análises foram reproduzidas em Python 3.11, com rotinas explícitas para transformação logit, variância binomial e efeitos aleatórios.

Quadro 1. Famílias de testes, alvo medido e posição clínica provável

Família	Alvo	Uso mais plausível	Limitação decisiva
SAA de α-sinucleína no LCR	Semeadura de α-sinucleína agregada	Confirmação de biologia sinucleinopática; enriquecimento de ensaios	Punção lombar; não distingue de forma absoluta PD, DLB e algumas outras sinucleinopatias
α-sinucleína cutânea	Depósito fosforilado ou atividade de semeadura em fibras cutâneas	Confirmação menos centralmente invasiva e possibilidade de repetição	Padronização do sítio, processamento e interpretação; referência clínica imperfeita
DaT-SPECT	Integridade do transportador dopaminérgico estriatal	Diferenciar tremor essencial/causas não degenerativas de déficit nigroestriatal	Não separa de modo confiável PD de MSA/PSP; radiação e custo
RM nigral	Neuromelanina, nigrossomo-1 e depósito de ferro	Apoio não ionizante à diferenciação de tremores e pesquisa de biomarcador	Protocolos e thresholds heterogêneos; dependência de campo e pós-processamento
Olfação	Identificação, discriminação ou limiar olfatório	Triagem, cálculo de risco prodrômico e seleção para teste confirmatório	Influência de idade, cultura, tabagismo e outras doenças
Sangue/exossomos	Espécies de α-sinucleína, DJ-1 e outros analitos periféricos	Triagem escalável e monitoramento futuro	Contaminação periférica, pré-analítica e baixo desempenho independente
Fenotipagem digital	Expressão motora e vocal do fenótipo	Triagem remota, quantificação e seguimento	Viés caso-controle, sobreajuste, dependência de dispositivo e baixa validação externa

Fonte: elaboração dos autores a partir da síntese. SAA: seed amplification assay; LCR: líquido cefalorraquidiano; DLB: demência com corpos de Lewy; MSA: atrofia de múltiplos sistemas; PSP: paralisia supranuclear progressiva.

3 RESULTADOS

3.1 Composição do conjunto analítico

Treze conjuntos de dados primários forneceram métricas comparáveis para a síntese central. Três avaliaram SAA no LCR, dois α-sinucleína cutânea, um

DaT-SPECT, um RM nigral combinada, dois olfação, dois biomarcadores sanguíneos e dois fenotipagem digital. O conjunto reuniu 1.421 participantes com doença de Parkinson ou parkinsonismo e 796 comparadores, sem somar braços secundários nem publicações derivadas. Essa soma descreve observações analíticas, não participantes únicos de todo o campo, pois os estudos tinham critérios e referências diferentes.

A assimetria metodológica foi marcante. Os estudos de SAA incluíram desde uma coorte confirmatória de 20 casos até a análise PPMI com 545 participantes com doença de Parkinson. A biópsia cutânea evoluiu de prova de conceito com 20 casos para estudo multicêntrico com 96 casos de Parkinson e 120 controles. Neuroimagem utilizou tremor essencial como comparador clínico, enquanto olfação, sangue e ferramentas digitais frequentemente selecionaram controles saudáveis. Essa diferença desloca o grau de dificuldade da classificação e explica parte da distância entre acurácia observada e utilidade real.

Tabela 1. Estudos primários incluídos na síntese quantitativa

Estudo	Família	Desenho	Casos/comp.	Comparador	Sens.	Esp.
Fairfoul et al., 2016	SAA no LCR	Coorte confirmatória cega	20/15	Controles clínicos	95,0 %	100,0 %
Rossi et al., 2020	SAA no LCR	Coorte clínica multicêntrica	71/62	Controles clínicos	94,4 %	98,4 %
Siderowf et al., 2023	SAA no LCR	PPMI, transversal	545/163	Controles saudáveis	87,7 %	96,3 %
Wang et al., 2021	α-sinucleína cutânea	Biópsia antemortem, prova de conceito	20/21	Controles sem PD	95,0 %	100,0 %
Gibbons et al., 2024	α-sinucleína cutânea	Multicêntrico cego, 30 centros	96/120	Controles sem sinucleinopatia	92,7 %	96,7 %
Benamer et al., 2000	DaT-SPECT	Multicêntrico com leitura cega	158/27	Tremor essencial	94,9 %	92,6 %
Jin et al., 2019	RM nigral combinada	De novo, sem tratamento	68/25	Tremor essencial	85,3 %	92,0 %

Estudo	Família	Desenho	Casos/comp.	Comparador	Sens.	Esp.
Campabadal et al., 2019	Olfação	Validação espanhola do UPSIT	97/65	Controles saudáveis	81,4 %	84,6 %
Cao et al., 2019	Olfação	Validação do B-SIT na China	163/112	Controles saudáveis	64,1 %	83,9 %
Wang et al., 2015	Biomarcador sanguíneo	ELISA em eritrócitos	100/102	Controles saudáveis	79,0 %	64,7 %
Zheng et al., 2021	Biomarcador sanguíneo	Exossomos plasmáticos	36/36	Controles saudáveis	60,5 %	59,4 %
Arora et al., 2015	Fenotipagem digital	Piloto com validação interna	10/10	Controles	96,2 %	96,9 %
Drotár et al., 2016	Fenotipagem digital	Base PaHaW, SVM	37/38	Controles pareados	87,4 %	80,9 %

Comp.: comparadores; Sens.: sensibilidade; Esp.: especificidade. Percentuais preservam os valores publicados; pequenas diferenças decorrem de arredondamento.

3.2 SAA de α-sinucleína no líquido cefalorraquidiano

Fairfoul e colaboradores demonstraram que o RT-QuIC podia detectar agregação de α-sinucleína em LCR de pessoas com Parkinson com sensibilidade de 95% e especificidade global de 100% em relação a amostras de Alzheimer e controles. A coorte confirmatória clínica incluiu 20 pessoas com Parkinson e 15 controles, sendo analisada de forma codificada. O estudo estabeleceu a viabilidade, mas a amostra pequena e a seleção de grupos bem definidos limitavam a extrapolação para parkinsonismo não selecionado. [14]

Rossi e colaboradores ampliaram o espectro para 439 amostras e encontraram 67 resultados positivos entre 71 casos de Parkinson provável, além de especificidade próxima de 98% em controles clínicos. O ensaio reconheceu patologia relacionada a corpos de Lewy ao longo de Parkinson, DLB, transtorno comportamental do sono REM isolado e falência autonômica pura. A baixa positividade em MSA sob aquelas condições mostrou que parâmetros cinéticos

podem carregar informação de cepa, mas não autorizou concluir que todo SAA separa de forma estável as sinucleinopatias. [15,16]

Na PPMI, Siderowf e colaboradores avaliaram 1.123 participantes, incluindo 545 com Parkinson e 163 controles saudáveis. A sensibilidade global foi 87,7% e a especificidade 96,3%; entre casos esporádicos com déficit olfatório típico, a sensibilidade alcançou 98,6%. Resultados negativos foram mais frequentes em alguns portadores de LRRK2 e em fenótipos sem hiposmia, mostrando que o biomarcador redefine heterogeneidade biológica em vez de apenas confirmar um rótulo clínico. [17]

A síntese dos três conjuntos produziu sensibilidade de 89,5% (IC95% 84,3-93,2) e especificidade de 96,4% (IC95% 93,1-98,2). A heterogeneidade da sensibilidade foi moderada-baixa, mas clinicamente relevante: um resultado negativo reduz a probabilidade de patologia sinucleinopática, porém não exclui Parkinson em variantes genéticas ou fenótipos específicos. O teste é particularmente forte como marcador de inclusão biológica em pesquisa e como confirmação quando a pergunta é presença de α-sinucleína agregada.

3.3 α-sinucleína cutânea

Wang e colaboradores aplicaram RT-QuIC a biópsias cervicais e de perna de 20 pessoas vivas com Parkinson e 21 controles. A sensibilidade foi 95% e a especificidade 100%, com AUC próxima de 1,0. O sinal foi mais forte e mais precoce no sítio cervical, evidenciando que local da coleta e número de amostras são componentes do teste, não detalhes operacionais. A prova de conceito foi importante, mas o diagnóstico clínico funcionou como referência e a validação permaneceu pequena. [18]

O estudo multicêntrico de Gibbons e colaboradores incluiu 428 participantes em 30 centros; 343 compuseram a análise primária. α-sinucleína fosforilada foi detectada em 89 de 96 pessoas com Parkinson e em 4 de 120 controles, equivalendo a sensibilidade de 92,7% e especificidade de 96,7%. A

positividade também foi alta em MSA, DLB e falência autonômica pura. Assim, o resultado confirma uma sinucleinopatia com boa frequência, mas sua interpretação diagnóstica precisa ser combinada ao fenótipo para nomear a síndrome. [19]

A estimativa resumida foi 92,4% de sensibilidade (IC95% 86,0-96,0) e 96,5% de especificidade (IC95% 91,8-98,5). Em comparação descritiva com o SAA no LCR, a biópsia cutânea apresentou ponto operacional semelhante. A diferença prática está menos em alguns pontos percentuais e mais em invasividade, disponibilidade laboratorial, padronização de coleta, capacidade de distinguir distribuições de depósito e experiência do serviço.

3.4 Neuroimagem dopaminérgica e ressonância nigral

No estudo multicêntrico de Benamer, 158 pessoas com diagnóstico clínico de parkinsonismo, 27 com tremor essencial e 35 controles foram submetidas a [123I]-FP-CIT SPECT. A leitura institucional alcançou sensibilidade de 97% e especificidade de 100% contra tremor essencial; a leitura de consenso cega apresentou sensibilidade de 95% e especificidade de 93%. O resultado sustenta o uso do DaT-SPECT quando a questão é se há déficit dopaminérgico pré-sináptico, especialmente diante de tremor essencial, tremor distônico, parkinsonismo induzido por fármacos ou incerteza clínica. [20]

DaT-SPECT não demonstra diretamente α-sinucleína e não separa Parkinson de todas as degenerações com perda nigroestriatal. Um exame anormal em MSA ou PSP pode ser indistinguível do padrão observado no Parkinson, embora distribuição e contexto possam acrescentar informação. Sua força é reduzir a incerteza entre causas com e sem desnervação dopaminérgica; sua fraqueza é ser interpretado como exame etiológico definitivo.

Jin e colaboradores avaliaram 68 pacientes com Parkinson de novo, 25 com tremor essencial não tratado e 34 controles. A inspeção de neuromelanina apresentou sensibilidade de 88,2% e especificidade de 80,0% contra tremor

essencial; a ausência do nigrossomo-1 teve sensibilidade de 79,4% e especificidade de 92,0%. A combinação não linear dos dois sinais alcançou sensibilidade de 85,3%, especificidade de 92,0% e AUC de 0,935. O achado ilustra uma vantagem do multimodal: um marcador mais sensível e outro mais específico podem ocupar posições complementares. [21]

A estimativa representativa foi 94,7%/91,1% para DaT-SPECT e 84,8%/90,4% para RM nigral combinada. Esses valores não autorizam afirmar superioridade do DaT-SPECT, porque os limiares, a amostra e o padrão de referência eram diferentes. A RM evita radiação e pode ser incorporada a exames já indicados, mas exige padronização multicêntrica. Estudos de neuromelanina em doença de novo e seguimento mostram potencial para quantificar perda nigral, ainda sem um protocolo universal. [22,23]

3.5 Testes olfatórios

A hiposmia antecede sintomas motores em parte das pessoas com Parkinson e integra os critérios prodrômicos como marcador de aumento de risco. Entretanto, o teste depende de familiaridade cultural com odores, idade, tabagismo, rinossinusite, infecções, traumatismo e outras doenças neurodegenerativas. A mesma pontuação pode ter significado diferente em populações distintas, e um threshold otimizado localmente raramente permanece ótimo em outro país. [2,3]

Em uma amostra mexicana, Rodríguez-Violante e colaboradores compararam UPSIT, Sniffin' Sticks-16 e B-SIT. O UPSIT alcançou sensibilidade próxima de 82% e especificidade de 66% em um ponto de corte, enquanto SS-16 apresentou sensibilidade de 77,8% e especificidade de 71,2%. Em validação espanhola conduzida por Campabadal e colaboradores com 97 casos e 65 controles, o corte ≤25 do UPSIT produziu sensibilidade de 81,4% e especificidade de 84,6%; o ponto de corte para anosmia (≤18) alcançou especificidade de 100%, mas sensibilidade de apenas 54,6%. [24,25]

Na validação chinesa do B-SIT, 163 pessoas com Parkinson e 112 controles foram avaliadas. A sensibilidade foi 64,1% e a especificidade 83,9%. A estimativa combinada dos dois conjuntos foi 73,0% de sensibilidade e 83,8% de especificidade, com heterogeneidade alta para sensibilidade. O teste olfatório, portanto, é barato e útil para enriquecer probabilidade pré-teste, mas seu valor negativo isolado é insuficiente e seu valor positivo depende do contexto cultural. [26,27]

3.6 Biomarcadores sanguíneos

O sangue é a matriz mais atraente para triagem populacional, porém a α-sinucleína é abundante em eritrócitos e plaquetas, tornando hemólise, processamento e origem celular determinantes. Wang e colaboradores mediram oligômeros de α-sinucleína em eritrócitos de 100 pessoas com Parkinson, 22 com MSA e 102 controles. A razão oligômero/proteína total foi maior no Parkinson, com sensibilidade de 79,0% e especificidade de 64,7%. O resultado separava grupos em média, mas gerava muitos falsos positivos para uso isolado. [28]

Zheng e colaboradores investigaram espécies de α-sinucleína em exossomos plasmáticos de 36 casos e 36 controles. A razão oligômero/α-sinucleína total apresentou AUC de 0,71, sensibilidade de 60,5% e especificidade de 59,4%. Trabalhos com exossomos enriquecidos para origem neural e combinações com DJ-1 aumentaram algumas AUCs, mas a dependência de anticorpo de captura, isolamento vesicular e normalização ainda reduz reprodutibilidade. [29-31]

A síntese produziu 70,9% de sensibilidade e 63,1% de especificidade, com incerteza ampla. O resultado não invalida o sangue como caminho de pesquisa; mostra que analitos isolados de primeira geração não atingem desempenho suficiente para confirmação. Painéis multimoleculares, assinaturas proteômicas ou RNA podem melhorar discriminação, desde que sejam bloqueados antes da

validação, avaliados em amostras consecutivas e comparados com doenças mimetizadoras.

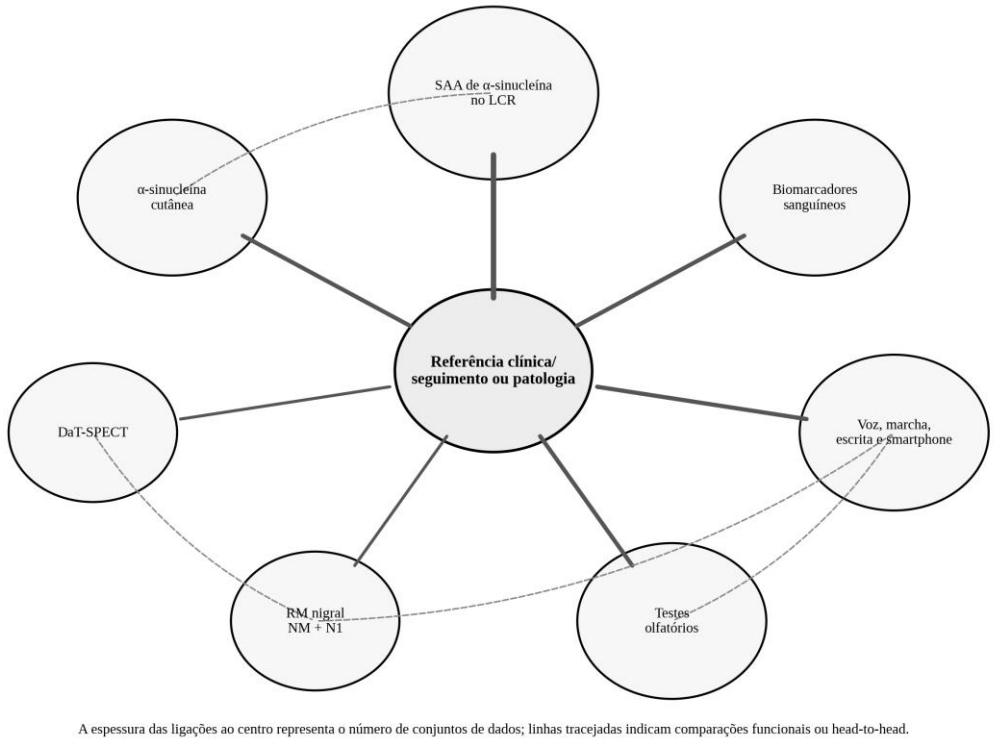
3.7 Fenotipagem digital: voz, marcha, toque e escrita

Arora e colaboradores utilizaram tarefas de voz, postura, marcha, toque e tempo de resposta registradas por smartphone em 10 pessoas com Parkinson e 10 controles. A sensibilidade média foi 96,2% e a especificidade 96,9%, além de boa predição do componente motor da UPDRS. O estudo demonstrou que sensores comuns capturam o fenótipo, mas não que o aplicativo diagnostica Parkinson em população clínica. Amostra muito pequena, controles vinculados aos casos e validação interna tornam o desempenho vulnerável a particularidades do conjunto de dados. [32]

Na base PaHaW, Drotár e colaboradores analisaram oito tarefas de escrita de 37 pessoas com Parkinson e 38 controles. O melhor modelo SVM apresentou sensibilidade de 87,4%, especificidade de 80,9% e acurácia de 81,3%. Características de pressão foram tão ou mais informativas que medidas cinemáticas convencionais, sugerindo que a caneta digital captura micrografia, hesitação, movimento no ar e controle de força. [33]

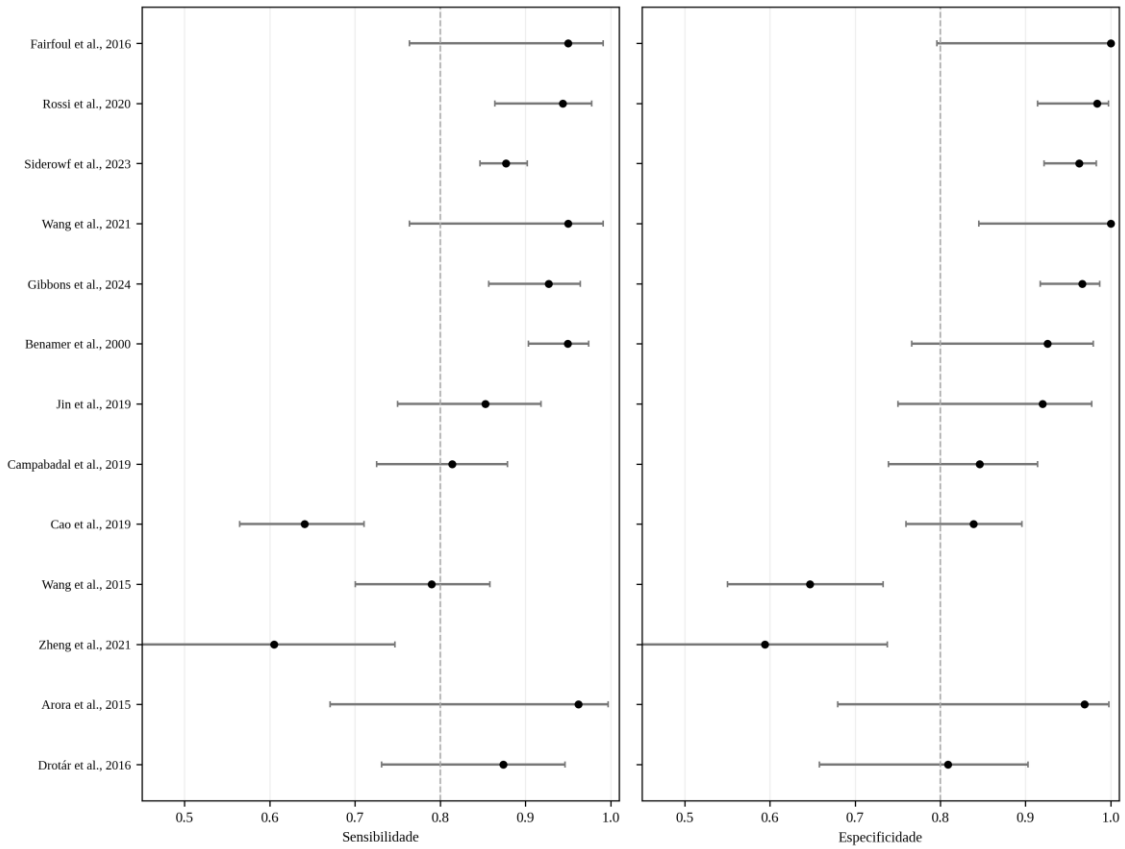
A estimativa resumida foi 87,4% de sensibilidade e 82,0% de especificidade. A proximidade com modalidades estabelecidas no plano ROC é enganosa quando ignorado o desenho. Ferramentas digitais quase sempre foram treinadas para separar casos diagnosticados de controles saudáveis, enquanto o desafio clínico é separar Parkinson inicial de tremor essencial, parkinsonismo vascular, efeitos de medicamentos, depressão, fragilidade e envelhecimento. Estudos de espiral, voz e marcha relatam acurácia elevada em bases fechadas, mas a queda entre validação interna e externa permanece uma das principais ameaças do campo. [34,35]

Figura 1. Estrutura da rede comparativa por famílias de testes



Nota: a rede é conceitual e analítica. As famílias se conectam por referência clínica, seguimento ou patologia; linhas tracejadas indicam comparações funcionais ou head-to-head. Não representa equivalência de condição-alvo.

Figura 2. Sensibilidade e especificidade observadas nos estudos primários



Nota: intervalos de 95% calculados pelo método de Wilson a partir do tamanho dos grupos e das proporções publicadas. Estudos digitais usam desempenho médio informado e devem ser interpretados como validação interna.

3.8 Meta-análise comparativa por famílias

Tabela 2. Estimativas resumidas por nó da rede

Família	k	Sensibilidade (IC95%)	Especificidade (IC95%)	I² sens.	I² esp.
SAA no LCR	3	89,5% (84,3-93,2)	96,4% (93,1-98,2)	24%	0%
α-sinucleína cutânea	2	92,4% (86,0-96,0)	96,5% (91,8-98,5)	0%	0%
DaT-SPECT	1	94,7% (89,9-97,2)	91,1% (73,6-97,4)	0%	0%
RM nigral combinada	1	84,8% (74,3-91,5)	90,4% (71,8-97,2)	0%	0%
Olfação	2	73,0% (53,3-86,5)	83,8% (77,6-88,5)	88%	0%

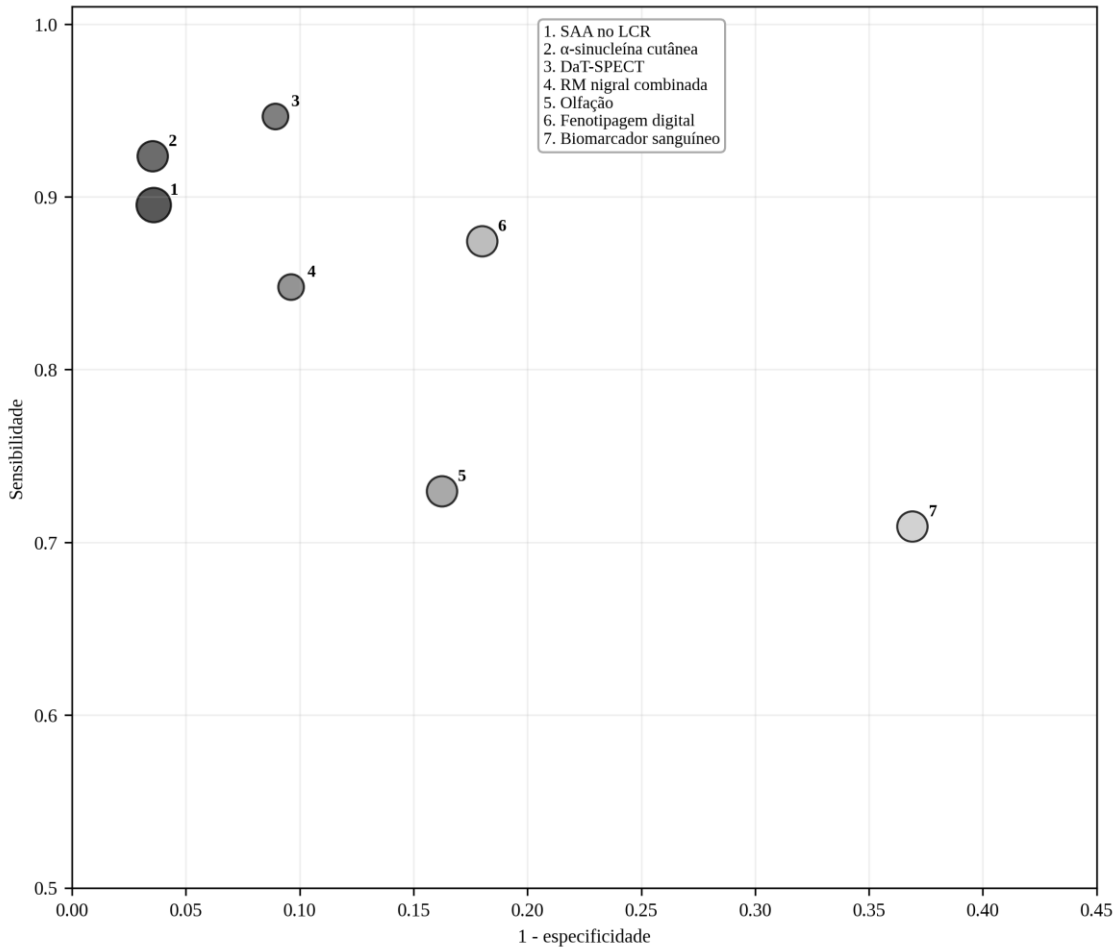
Família	k	Sensibilidade (IC95%)	Especificidade (IC95%)	I^2 sens.	I^2 esp.
Fenotipagem digital	2	87,4% (74,8-94,2)	82,0% (68,4-90,5)	0%	0%
Biomarcador sanguíneo	2	70,9% (50,4-85,4)	63,1% (54,8-70,7)	78%	0%

Modelos de efeitos aleatórios na escala logit. Para nós com um estudo, o intervalo reflete incerteza binomial e não heterogeneidade entre estudos. As comparações são exploratórias e não constituem ranking.

No plano ROC, SAA no LCR e α-sinucleína cutânea ocuparam a região de menor taxa de falsos positivos e sensibilidade elevada. DaT-SPECT apresentou sensibilidade ligeiramente maior, mas foi avaliado sobretudo contra tremor essencial e mede desnervação, não sinucleinopatia. A RM combinada mostrou equilíbrio semelhante em um estudo de Parkinson de novo. Olfação e sangue deslocaram-se para maior taxa de falsos positivos; a fenotipagem digital permaneceu em posição intermediária, com certeza menor que a sugerida pelo ponto estimado.

A rede não continha ciclos fechados suficientes para avaliar inconsistência por métodos tradicionais. O único conjunto claramente head-to-head foi composto por testes olfatórios no mesmo estudo e por sinais de RM na mesma população; comparações entre SAA, pele e imagem eram predominantemente indiretas. Consequentemente, qualquer diferença inferior a aproximadamente cinco pontos percentuais entre famílias de alto desempenho deve ser interpretada como indeterminada diante de seleção, referência e espectro.

Figura 3. Plano ROC das estimativas resumidas por família



O tamanho do ponto reflete o número de estudos no nó. A figura descreve pontos operacionais médios e não uma curva ROC hierárquica completa.

3.9 Consequências clínicas em diferentes probabilidades pré-teste

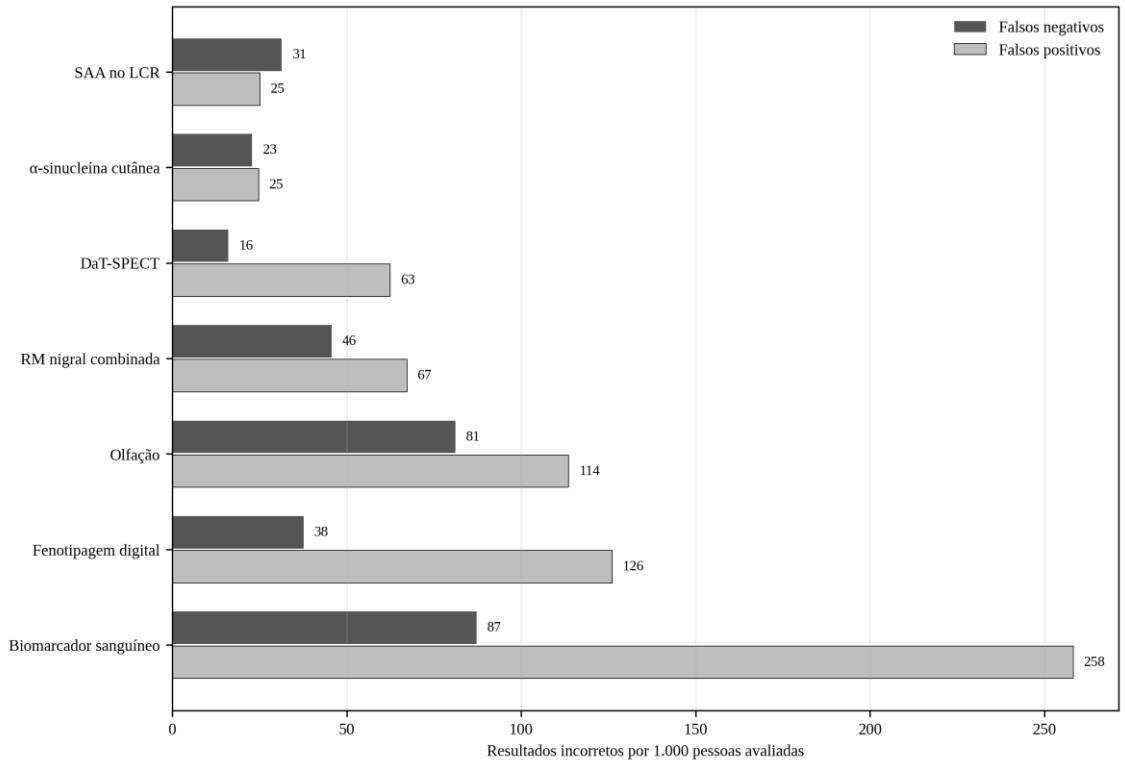
Com probabilidade pré-teste de 30%, a SAA no LCR produziria aproximadamente 31 falsos negativos e 25 falsos positivos por 1.000 pessoas; a α-sinucleína cutânea, 23 e 25; o DaT-SPECT, 16 e 63; a RM nigral combinada, 46 e 67; a olfação, 81 e 114; a fenotipagem digital, 38 e 126; e os biomarcadores sanguíneos, 87 e 258. A comparação demonstra que especificidade modesta tem grande impacto quando 70% dos avaliados não têm a condição-alvo.

Em cenário de triagem com probabilidade de 10%, falsos positivos passam a dominar. Mesmo especificidade de 96,4% gera cerca de 32 resultados positivos

incorretos por 1.000, valor semelhante ou superior ao número de casos verdadeiros identificados. Em ambulatório especializado com probabilidade de 50%, a sensibilidade ganha peso maior e o número de falsos negativos se aproxima de metade do complemento da sensibilidade por 1.000. Assim, um teste confirmatório em pesquisa não deve ser automaticamente transportado para rastreamento populacional.

A utilidade depende também da consequência do erro. Falso positivo de DaT-SPECT pode levar a tratamento, ansiedade e investigação prolongada; falso negativo de SAA pode excluir uma pessoa de ensaio dirigido à α-sinucleína; falso positivo digital pode sobrecarregar encaminhamentos; e falso negativo olfatório pode ser aceitável se o teste for apenas uma etapa de triagem repetível. Sensibilidade e especificidade são propriedades incompletas sem o caminho assistencial que segue o resultado.

Figura 4. Falsos negativos e falsos positivos esperados por 1.000 avaliações



Cenário ilustrativo com probabilidade pré-teste de 30%. Valores arredondados e condicionais às estimativas resumidas; não incorporam testes indeterminados, perdas ou dependência entre testes sequenciais.

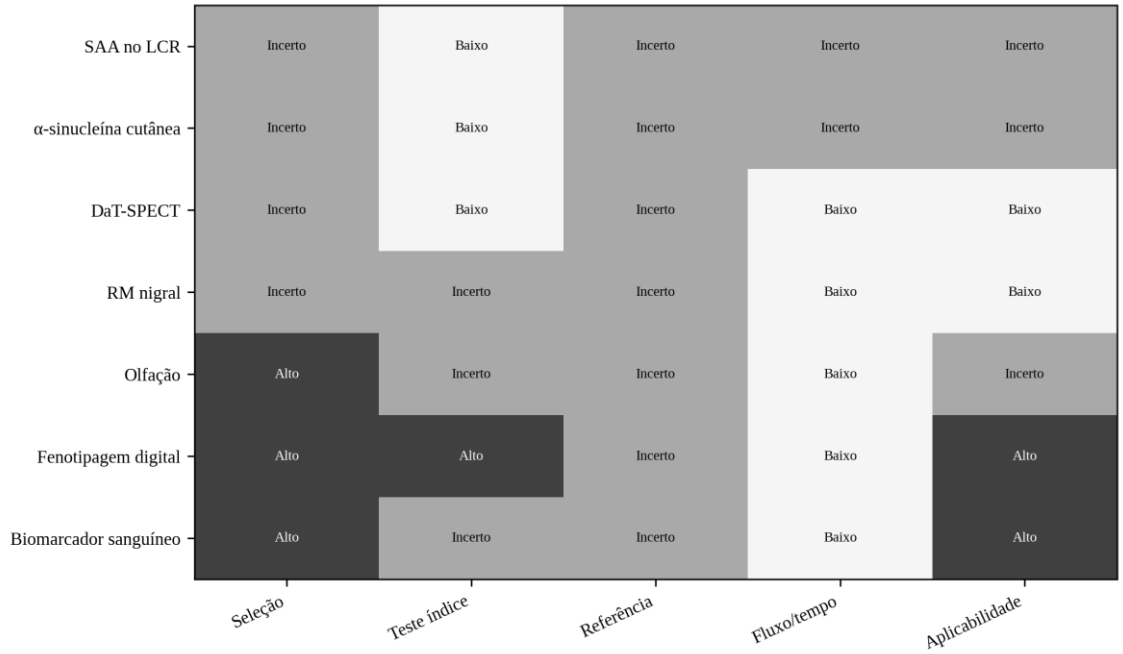
3.10 Risco de viés e certeza

SAA no LCR e α-sinucleína cutânea apresentaram baixo risco na execução do teste e preocupação moderada em seleção e referência. A PPMI ofereceu amostra grande e fenotipagem profunda, mas não reproduz toda a diversidade de pessoas encaminhadas em serviços gerais. O estudo cutâneo multicêntrico contou com painel cego, embora a referência permanecesse clínica e controles excluíssem sintomas prodrômicos, criando separação mais nítida que a encontrada na prática.

DaT-SPECT e RM tiveram boa aplicabilidade à diferenciação de tremor essencial, mas não responderam diretamente à distinção entre Parkinson e parkinsonismos atípicos. Olfação sofreu com threshold e influência cultural.

Estudos sanguíneos e digitais apresentaram maior risco de seleção; nos digitais, a mesma base frequentemente orientou seleção de características, ajuste e validação. A certeza foi considerada moderada para SAA no LCR, moderada-baixa para α-sinucleína cutânea e DaT-SPECT em sua finalidade específica, baixa para RM e olfação, e muito baixa a baixa para sangue e fenotipagem digital como diagnóstico independente.

Figura 5. Risco de viés e aplicabilidade segundo adaptação do QUADAS-2



A classificação resume o padrão predominante da família e não substitui a avaliação individual de cada artigo.

4 DISCUSSÃO

4.1 Interpretação principal

A principal conclusão não é que um teste venceu a rede, mas que a evidência se organiza em camadas. SAA no LCR e α-sinucleína cutânea apresentaram o conjunto mais coerente para demonstrar biologia sinucleinopática. DaT-SPECT foi altamente sensível para identificar déficit

dopaminérgico quando o comparador era tremor essencial. RM nigral ofereceu alternativa não ionizante promissora. Olfação e sinais digitais foram capazes de enriquecer risco e reconhecer fenótipo, enquanto os analitos sanguíneos avaliados isoladamente permaneceram insuficientes.

Essa arquitetura resolve uma aparente contradição. Um exame pode ser excelente e ainda assim não responder à pergunta do paciente. SAA positivo em pessoa com parkinsonismo confirma α-sinucleína agregada, mas não nomeia automaticamente Parkinson em vez de DLB ou outra sinucleinopatia. DaT-SPECT anormal confirma desnervação, mas não prova corpos de Lewy. Um algoritmo de voz pode detectar hipofonia e instabilidade articulatória, mas esses sinais aparecem em outras condições. A interpretação correta acrescenta o resultado a uma hipótese; a interpretação incorreta substitui a hipótese pelo teste.

As estimativas de alta acurácia para SAA e pele são clinicamente relevantes porque se repetiram em estudos maiores e multicêntricos. Ainda assim, os controles foram frequentemente selecionados para não apresentar manifestações sugestivas de sinucleinopatia. Em uma clínica de memória, sono ou disautonomia, parte dos supostos controles pode ter copatologia de Lewy, reduzindo especificidade aparente ou revelando que o padrão clínico de referência estava errado. Um teste biológico pode parecer falso positivo hoje e tornar-se verdadeiro positivo após seguimento.

A heterogeneidade dentro do Parkinson também impede leitura binária. Na PPMI, a sensibilidade da SAA variou conforme genética e olfação. Portadores de variantes LRRK2 podem apresentar patologia menos uniforme, e pessoas sem déficit olfatório típico foram menos frequentemente positivas. O teste, portanto, não apenas diagnostica: ele estratifica. Essa capacidade é valiosa para ensaios de terapias dirigidas, mas exige que critérios de inclusão não eliminem subtipos biologicamente legítimos de Parkinson.

4.2 Diagnóstico precoce e fase prodrômica

O termo precoce precisa ser usado com precisão. Estudos caso-controle com doença já diagnosticada podem utilizar participantes de curta duração, mas não demonstram detecção antes do diagnóstico. Evidência verdadeiramente prodrômica requer coortes de transtorno comportamental do sono REM, hiposmia, disautonomia, mutações ou risco populacional, com seguimento para conversão. SAA já mostrou positividade em parte desses grupos, especialmente RBD, mas sensibilidade e valor preditivo dependem do horizonte temporal e do fenótipo que surgirá.

Olfação é especialmente útil no primeiro estágio, por ser barata e não invasiva. Seu papel mais defensável é selecionar pessoas para uma segunda etapa de maior especificidade, não confirmar doença. Um caminho possível combina questionário de RBD, teste olfatório culturalmente validado e avaliação motora digital; pessoas acima de um limiar de risco seguem para exame neurológico e biomarcador de α-sinucleína. Essa estratégia precisa ser testada prospectivamente, incluindo custos, ansiedade, adesão e consequências dos resultados indeterminados.

A busca por diagnóstico antes dos sintomas também levanta um problema ético. Na ausência de terapia modificadora comprovada para todos os indivíduos, revelar biomarcador positivo pode produzir benefício por planejamento e acesso a pesquisa, mas também estigma, insegurança ocupacional e sofrimento. Estudos de acurácia devem incorporar consentimento específico, aconselhamento, preferência de comunicação e avaliação do impacto psicológico, em vez de tratar o resultado como informação neutra.

4.3 Tecnologias digitais e inteligência artificial

Ferramentas digitais oferecem vantagens que não aparecem em uma tabela de acurácia: repetição frequente, avaliação no ambiente cotidiano, baixo custo marginal e capacidade de medir flutuação. Uma classificação diagnóstica

única pode ser menos valiosa que a trajetória de toque, marcha e voz ao longo de meses. O mesmo dispositivo pode servir à triagem, mensuração de gravidade e monitoramento de resposta, desde que cada finalidade tenha validação própria.

O maior risco é confundir separabilidade da base com generalização clínica. Modelos aprendem idade, idioma, microfone, instruções, medicação, ambiente e centro de recrutamento. Se casos e controles foram gravados em condições diferentes, o algoritmo pode reconhecer o protocolo. Divisão aleatória de registros da mesma pessoa entre treino e teste causa vazamento. A unidade de divisão deve ser o participante; validação externa precisa ocorrer em centro, dispositivo e população diferentes; e o threshold deve ser congelado antes do teste.

A próxima geração deve comparar algoritmos com neurologistas e não apenas com controles. Coortes consecutivas de pessoas encaminhadas por sintomas deveriam incluir tremor essencial, distonia, parkinsonismo vascular, depressão, neuropatia, efeitos farmacológicos, MSA e PSP. O desempenho deve ser relatado por idade, sexo, idioma, escolaridade, raça/cor, deficiência, dispositivo e qualidade do sinal. Sem essa auditoria, uma ferramenta escalável pode ampliar desigualdades com aparência de objetividade.

4.4 Como escolher o teste na prática

Quadro 2. Seleção do teste a partir da decisão clínica

Cenário	Pergunta	Estratégia	Cuidado interpretativo
Pessoa com tremor, exame clínico ainda incerto	Há déficit dopaminérgico nigroestriatal?	DaT-SPECT; RM nigral quando disponível e validada	Evitar interpretar exame anormal como etiologia específica
Parkinson provável, necessidade de confirmação biológica	Há patologia de α-sinucleína detectável?	SAA no LCR ou α-sinucleína cutânea	Escolher conforme acesso, invasividade e experiência do laboratório
RBD, hiposmia ou risco genético sem parkinsonismo	Qual é a probabilidade de fase prodrômica?	Combinação de critérios prodrômicos; SAA em pesquisa	Resultado positivo exige aconselhamento e seguimento

Cenário	Pergunta	Estratégia	Cuidado interpretativo
Triagem remota ou grande escala	Quem merece avaliação especializada?	Olfação + questionários + fenotipagem digital	Maximizar sensibilidade e criar rota para confirmar
Diferenciar PD de MSA/PSP	A síndrome é uma sinucleinopatia ou tauopatia e qual fenótipo?	Clínica longitudinal, imagem e biomarcadores combinados	Nenhum teste isolado resolve todos os atípicos
Monitorar progressão	O marcador muda com o tempo e tratamento?	Medidas digitais, escalas e imagem quantitativa	Acurácia diagnóstica não prova responsividade longitudinal

Na prática, a sequência importa. Um teste barato e sensível pode anteceder um exame específico e mais invasivo. A combinação em série aumenta especificidade, enquanto a combinação em paralelo aumenta sensibilidade. Entretanto, os erros dos testes não são independentes: hiposmia, SAA e depósito cutâneo podem refletir o mesmo processo; voz e marcha compartilham gravidade motora; DaT-SPECT e RM nigral capturam componentes da via nigroestriatal. Modelos de utilidade devem estimar dependência, e não multiplicar razões de verossimilhança como se cada sinal fosse novo.

A escolha também precisa considerar o que acontece depois. Não há benefício em oferecer SAA quando o resultado não muda acompanhamento, aconselhamento, acesso a estudo ou decisão terapêutica. Da mesma forma, triagem digital sem capacidade de avaliação especializada converte tecnologia em fila. A incorporação responsável exige definição de população, threshold, prazo de resposta, responsável por comunicar, manejo de indeterminados e auditoria de resultados por subgrupo.

4.5 Fortalezas e limitações

Este capítulo reuniu modalidades frequentemente discutidas separadamente, preservou seus alvos biológicos e traduziu desempenho em consequências clínicas. A recusa em produzir ranking evitou uma conclusão artificial e tornou explícita a diferença entre detectar sinucleinopatia,

desnervação e fenótipo. A inclusão de estudos primários verificáveis, a avaliação QUADAS-2 e a reconstrução transparente das variâncias aumentaram rastreabilidade.

As limitações são importantes. A busca em bases de assinatura não gerou exportação completa e, portanto, a síntese deve ser entendida como atualização sistemática dirigida, não censo definitivo. A rede foi esparsa, com poucos estudos por nó e quase nenhum ciclo fechado. Parte das matrizes foi reconstruída a partir de percentuais; resultados digitais representaram validação interna; e diferentes condições-alvo foram colocadas no mesmo plano apenas para comparação descritiva. O modelo separado para sensibilidade e especificidade não reproduz integralmente a correlação bivariada ou o efeito de threshold.

Além disso, a referência clínica é imperfeita e pode favorecer ou penalizar biomarcadores. Acompanhamento longo e neuropatologia são raros; critérios atuais incorporam sinais que alguns testes também medem, criando risco de incorporação. Estudos com controles saudáveis superestimam aplicabilidade. Finalmente, a tecnologia evolui rapidamente: versões de algoritmo, reagente, substrato e protocolo podem constituir testes diferentes, mesmo quando recebem o mesmo nome.

4.6 Agenda de pesquisa

A prioridade é construir estudos prospectivos, consecutivos e multicêntricos em pessoas com incerteza diagnóstica real. Todos os participantes devem receber o conjunto de testes planejado, com interpretação cega, threshold pré-especificado e referência por painel longitudinal independente. O seguimento deve durar o suficiente para que diagnósticos atípicos se revelem. Resultados indeterminados e falhas técnicas precisam permanecer no denominador.

Comparações head-to-head entre SAA no LCR e pele são necessárias para estimar concordância, preferência, eventos adversos, tempo de resposta e custo.

Neuroimagem deve ser integrada como marcador complementar, não competidor abstrato. Estudos fatoriais podem avaliar se olfação ou digital reduzem o número de procedimentos invasivos sem perder casos. Meta-análises de dados individuais poderão modelar genética, sexo, idade, duração, RBD e olfação como modificadores.

Para inteligência artificial, protocolos e código devem registrar pré-processamento, divisão de participantes, seleção de características, calibração, threshold e mudanças de versão. O desempenho deve incluir curvas de calibração, decisão clínica, falso encaminhamento e benefício líquido, além de AUC. Modelos federados e validação em dispositivos de baixo custo podem ampliar diversidade, mas não substituem governança, privacidade e supervisão humana.

Por fim, a pesquisa precisa migrar de acurácia para impacto. Um teste é útil quando reduz atraso, melhora certeza, evita tratamento inadequado, seleciona terapia ou pesquisa, e faz isso a custo e risco aceitáveis. Ensaios de estratégia diagnóstica e estudos de implementação devem medir tempo até diagnóstico, mudança de conduta, qualidade de vida, ansiedade, equidade e custo-efetividade. A melhor tecnologia será a que melhora uma decisão relevante, não a que ocupa o canto superior esquerdo de um gráfico isolado.

5 CONCLUSÃO

A evidência comparativa indica que SAA de α-sinucleína no LCR e detecção cutânea de α-sinucleína patológica são as modalidades mais consistentes para confirmar biologia sinucleinopática. DaT-SPECT apresenta desempenho elevado quando a dúvida é presença de desnervação dopaminérgica, e a RM nigral oferece alternativa promissora sem radiação. Testes olfatórios e fenotipagem digital acrescentam valor principalmente como triagem, enriquecimento de risco e mensuração repetida. Biomarcadores sanguíneos isolados ainda não sustentam confirmação clínica.

Nenhuma dessas modalidades constitui teste universal para doença de Parkinson. O resultado deve ser interpretado segundo alvo medido, população, referência e consequência. Uma estratégia racional começa com avaliação clínica e probabilidade pré-teste, seleciona o exame capaz de resolver a incerteza específica e mantém seguimento longitudinal. A próxima etapa do campo é substituir estudos caso-controle pequenos por validação externa em pessoas consecutivas, comparações head-to-head e ensaios de impacto diagnóstico.

REFERÊNCIAS

1. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015;30(12):1591-1601. doi:10.1002/mds.26424.

2. Berg D, Postuma RB, Adler CH, Bloem BR, Chan P, Dubois B, et al. MDS research criteria for prodromal Parkinson's disease. Mov Disord. 2015;30(12):1600-1611. doi:10.1002/mds.26431.

3. Heinzel S, Berg D, Gasser T, Chen H, Yao C, Postuma RB; MDS Task Force on the Definition of Parkinson's Disease. Update of the MDS research criteria for prodromal Parkinson's disease. Mov Disord. 2019;34(10):1464-1470. doi:10.1002/mds.27802.

4. Adler CH, Beach TG, Hentz JG, Shill HA, Caviness JN, Driver-Dunckley E, et al. Low clinical diagnostic accuracy of early vs advanced Parkinson disease: clinicopathologic study. Neurology. 2014;83(5):406-412. doi:10.1212/WNL.0000000000000641.

5. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. Neurology. 2016;86(6):566-576. doi:10.1212/WNL.0000000000002350.

6. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319(4):388-396. doi:10.1001/jama.2017.19163.

7. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009.

8. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. Stat Methods Med Res. 2018;27(6):1766-1784. doi:10.1177/0962280216669182.

9. Veroniki AA, Tsokani S, Rücker G, Mavridis D, Takwoingi Y. Diagnostic test accuracy network meta-analysis methods: a scoping review and empirical assessment. J Clin Epidemiol. 2022;146:86-98. doi:10.1016/j.jclinepi.2022.02.007.

10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi:10.1136/bmj.h5527.

11. Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. J Clin Epidemiol. 2020;122:129-141. doi:10.1016/j.jclinepi.2019.12.020.

12. Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, et al. GRADE guidelines: 21 part 2. Inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence for test accuracy. J Clin Epidemiol. 2020;122:142-152. doi:10.1016/j.jclinepi.2019.12.021.

13. Mustafa RA, Wiercioch W, Santesso N, Cheung A, Prediger B, Baldeh T, et al. Decision making about healthcare-related tests and diagnostic test strategies: GRADE evidence to decision frameworks. BMJ. 2017;359:j4312. doi:10.1136/bmj.j4312.

14. Fairfoul G, McGuire LI, Pal S, Ironside JW, Neumann J, Christie S, et al. Alpha-synuclein RT-QuIC in the CSF of patients with alpha-synucleinopathies. Ann Clin Transl Neurol. 2016;3(10):812-818. doi:10.1002/acn3.338.

15. Rossi M, Candelise N, Baiardi S, Capellari S, Giannini G, Orrù CD, et al. Ultrasensitive RT-QuIC assay with high sensitivity and specificity for Lewy body-associated synucleinopathies. Acta Neuropathol. 2020;140(1):49-62. doi:10.1007/s00401-020-02160-8.

16. Shahnawaz M, Mukherjee A, Pritzkow S, Mendez N, Rabadia P, Liu X, et al. Discriminating α-synuclein strains in Parkinson's disease and multiple system atrophy. Nature. 2020;578(7794):273-277. doi:10.1038/s41586-020-1984-7.

17. Siderowf A, Concha-Marambio L, Lafontant DE, Farris CM, Ma Y, Urenia PA, et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. Lancet Neurol. 2023;22(5):407-417. doi:10.1016/S1474-4422(23)00109-6.

18. Wang Z, Becker K, Donadio V, Siedlak S, Yuan J, Rezaee M, et al. Skin α-synuclein aggregation seeding activity as a novel biomarker for Parkinson disease. JAMA Neurol. 2021;78(1):30-40. doi:10.1001/jamaneurol.2020.3311.

19. Gibbons CH, Levine T, Adler CH, et al. Skin biopsy detection of phosphorylated α-synuclein in patients with synucleinopathies. JAMA. 2024;331(15):1298-1306. doi:10.1001/jama.2024.0792.

20. Benamer HTS, Patterson J, Grosset DG, Booij J, de Bruin K, van Royen E, et al. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. Mov Disord. 2000;15(3):503-510. doi:10.1002/1531-8257(200005)15:3<503::AID-MDS1013>3.0.CO;2-V.

21. Jin L, Wang J, Wang C, Lian D, Zhou Y, Zhang Y, et al. Combined visualization of nigrosome-1 and neuromelanin in the substantia nigra using 3T MRI for the differential diagnosis of essential tremor and de novo Parkinson's disease. Front Neurol. 2019;10:100. doi:10.3389/fneur.2019.00100.

22. Reimão S, Pita Lobo P, Neutel D, Correia Guedes L, Coelho M, Rosa MM, et al. Substantia nigra neuromelanin magnetic resonance imaging in de novo Parkinson's disease patients. Eur J Neurol. 2015;22(3):540-546. doi:10.1111/ene.12613.

23. Fabbri M, Reimão S, Carvalho M, Nunes RG, Abreu D, Guedes LC, et al. Substantia nigra neuromelanin as an imaging biomarker of disease progression in Parkinson's disease. J Parkinsons Dis. 2017;7(3):491-501. doi:10.3233/JPD-171135.

24. Rodríguez-Violante M, Gonzalez-Latapi P, Camacho-Ordoñez A, Martínez-Ramírez D, Morales-Briceño H, Cervantes-Arriaga A. Comparing the accuracy of different smell identification tests in Parkinson's disease: relevance of cultural aspects. Clin Neurol Neurosurg. 2014;123:9-14. doi:10.1016/j.clineuro.2014.04.030.

25. Campabadal A, Segura B, Baggio HC, Abos A, Uribe C, Garcia-Diaz AI, et al. Diagnostic accuracy, item analysis and age effects of the UPSIT Spanish version in Parkinson's disease. Arch Clin Neuropsychol. 2019;34(4):435-444. doi:10.1093/arclin/acy053.

26. Cao M, Li Y, Gu Z, Mi T, Zhang P, Wang L, et al. Validation of the utility of the Brief Smell Identification Test in Chinese patients with Parkinson's disease. J Clin Neurosci. 2019;60:68-72. doi:10.1016/j.jocn.2018.10.049.

27. Morley JF, Duda JE. Optimizing olfactory testing for the diagnosis of Parkinson's disease: item analysis of the University of Pennsylvania Smell Identification Test. NPJ Parkinsons Dis. 2018;4:2. doi:10.1038/s41531-017-0039-8.

28. Wang X, Yu S, Li F, Feng T. Detection of α-synuclein oligomers in red blood cells as a potential biomarker of Parkinson's disease. Neurosci Lett. 2015;599:115-119. doi:10.1016/j.neulet.2015.05.030.

29. Shi M, Liu C, Cook TJ, Bullock KM, Zhao Y, Ginghina C, et al. Plasma exosomal α-synuclein is likely CNS-derived and increased in Parkinson's disease. Acta Neuropathol. 2014;128(5):639-650. doi:10.1007/s00401-014-1314-y.

30. Zheng H, Xie Z, Zhang X, Mao J, Wang M, Wei S, et al. Investigation of α-synuclein species in plasma exosomes and the oligomeric and phosphorylated α-synuclein as potential peripheral biomarker of Parkinson's disease. Neuroscience. 2021;469:79-90. doi:10.1016/j.neuroscience.2021.06.033.

31. Zhao ZH, Chen ZT, Zhou RL, Zhang X, Ye QY, Wang YZ. Increased DJ-1 and α-synuclein in plasma neural-derived exosomes as potential markers for Parkinson's disease. Front Aging Neurosci. 2019;10:438. doi:10.3389/fnagi.2018.00438.

32. Arora S, Venkataraman V, Donohue S, Biglan KM, Dorsey ER, Little MA. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism Relat Disord. 2015;21(6):650-653. doi:10.1016/j.parkreldis.2015.02.026.

33. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med. 2016;67:39-46. doi:10.1016/j.artmed.2016.01.004.

34. Zham P, Kumar DK, Dabnichki P, Poosapadi Arjunan S, Raghav S. Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. Front Neurol. 2017;8:435. doi:10.3389/fneur.2017.00435.

35. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng. 2009;56(4):1015-1022. doi:10.1109/TBME.2008.2005954.

APÊNDICE A — ESTRATÉGIA DE BUSCA REPRODUZÍVEL

("Parkinson Disease"[Mesh] OR Parkinson*[tiab] OR parkinsonism[tiab]) AND (early diagnos*[tiab] OR prodrom*[tiab] OR diagnostic accuracy[tiab] OR

sensitivity[tiab] OR specificity[tiab]) AND ((alpha-synuclein[tiab] OR seed amplification assay[tiab] OR RT-QuIC[tiab] OR skin biopsy[tiab] OR cerebrospinal fluid[tiab]) OR (dopamine transporter[tiab] OR DaT-SPECT[tiab] OR FP-CIT[tiab] OR neuromelanin[tiab] OR nigrosome[tiab] OR MRI[tiab]) OR (olfact*[tiab] OR smell identification[tiab]) OR (blood biomarker*[tiab] OR exosome*[tiab] OR erythrocyte*[tiab]) OR (voice[tiab] OR gait[tiab] OR handwriting[tiab] OR smartphone[tiab] OR wearable*[tiab] OR machine learning[tiab]))

Filtros não foram aplicados inicialmente por idioma ou data. A busca foi refinada por famílias e complementada por DOI, título e autor. Referências de revisões e artigos pivôs foram rastreadas retrospectiva e prospectivamente. A atualização terminou em 1º de agosto de 2026.

Para futuras atualizações, recomenda-se salvar as estratégias traduzidas para Embase, Scopus e Web of Science, exportar os resultados com data e identificador, deduplicar em software documentado e publicar a lista de textos completos excluídos com justificativa.

APÊNDICE B — REGRAS DE RECONSTRUÇÃO E CÁLCULO

Quando a publicação forneceu contagens, elas foram usadas diretamente. Quando forneceu apenas sensibilidade, especificidade e número de casos/controles, o número esperado de positivos e negativos foi obtido por multiplicação e arredondamento para o inteiro mais próximo. Essa reconstrução foi usada apenas para variância e intervalo; o percentual publicado permaneceu como estimativa do estudo.

Para cada proporção p com denominador n, utilizou-se transformação logit após correção de 0,5 evento: $p^*=(np+0,5)/(n+1)$. A variância aproximada foi $1/(np+0,5)+1/(n-np+0,5)$. O efeito médio foi estimado com pesos inversos de variância e τ^2 de DerSimonian-Laird. Sensibilidade e especificidade foram

modeladas separadamente; portanto, os resultados não substituem um modelo hierárquico bivariado quando dados completos estiverem disponíveis.

As consequências clínicas foram calculadas para 1.000 avaliações. Com prevalência π, verdadeiros positivos=$1.000\times\pi\times$sensibilidade; falsos negativos=$1.000\times\pi\times(1-$sensibilidade); verdadeiros negativos=$1.000\times(1-\pi)\times$especificidade; falsos positivos=$1.000\times(1-\pi)\times(1-$especificidade). Valores são ilustrativos e não incluem resultados inconclusivos, perdas, eventos adversos ou dependência entre testes.

CAPÍTULO 4

PREDITORES INDIVIDUAIS DE REMISSÃO DO DIABETES TIPO 2 APÓS CIRURGIA METABÓLICA: PROTOCOLO, ESTUDO DE VIABILIDADE E ESTRUTURA PARA META-ANÁLISE DE DADOS INDIVIDUAIS DOS PARTICIPANTES

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Rafaela Nonato de Menezes; Antônia Araújo Firmo; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Leila Silveira Vieira Bezerra; Davi Pedro Soares Macêdo; Edglê Pedro de Sousa Filho; Paula Andrea Travecedo Ramos; Nélio Barreto Vieira

RESUMO

Objetivo: estabelecer uma estrutura metodológica completa para meta-análise de dados individuais dos participantes (IPD-MA) sobre preditores de remissão do diabetes tipo 2 após cirurgia metabólica, avaliar a viabilidade de harmonização dos principais bancos e sintetizar, como ponte, o desempenho publicado dos modelos prognósticos. Métodos: revisão sistemática orientada por PRISMA-IPD, CHARMS, TRIPOD e PROBAST. Foram considerados adultos com diabetes tipo 2 submetidos a bypass gástrico em Y de Roux, gastrectomia vertical, bypass de uma anastomose ou procedimentos com derivação duodenal. O desfecho principal foi remissão definida por HbA1c <6,5% por pelo menos três meses sem fármacos hipoglicemiantes; definições históricas foram mapeadas para análises de sensibilidade. O protocolo prescreve análise em um estágio com intercepto aleatório por estudo, splines para preditores contínuos, múltipla imputação dentro de cada banco, validação interna-externa e avaliação de calibração, discriminação, benefício líquido e equidade. Resultados: a revisão de referência identificou 38 estudos, 16 de desenvolvimento e 22 de validação externa, correspondendo a 16 modelos. ABCD e DiaRem foram os mais avaliados. Em síntese agregada publicada, o ABCD apresentou AUC de 0,79 em

um ano e 0,80 no longo prazo; o DiaRem apresentou AUC de 0,78 no curto prazo e 0,83 no longo prazo. A calibração variou entre populações, e somente uma validação externa foi classificada como baixo risco de viés. Duração do diabetes, uso de insulina, reserva de célula beta, HbA1c e carga medicamentosa formaram o núcleo prognóstico mais consistente. Não houve, nesta execução, transferência de bancos individuais; portanto, não são atribuídas estimativas de efeito a uma IPD-MA concluída. Conclusão: a área possui coortes suficientemente informativas, mas permanece fragmentada por definições de remissão, medições de C-peptídeo, procedimentos e horizontes temporais. Uma IPD-MA colaborativa pode corrigir essas limitações, testar não linearidade e interações e produzir previsões transportáveis, desde que a governança e a harmonização sejam planejadas antes da análise.

Palavras-chave: cirurgia metabólica; diabetes mellitus tipo 2; remissão; prognóstico; dados individuais dos participantes; modelo de predição; meta-análise.

ABSTRACT

Objective: to establish a complete methodological framework for an individual participant data meta-analysis (IPD-MA) of predictors of type 2 diabetes remission after metabolic surgery, assess the feasibility of harmonizing major cohorts, and summarize published model performance as an aggregate-data bridge. Methods: a systematic review was structured according to PRISMA-IPD, CHARMS, TRIPOD, and PROBAST. Eligible participants are adults with type 2 diabetes undergoing Roux-en-Y gastric bypass, sleeve gastrectomy, one-anastomosis gastric bypass, or duodenal diversion procedures. The primary outcome is remission defined as HbA1c <6.5% for at least three months without glucose-lowering medication; historical definitions are retained for sensitivity analyses. The prespecified IPD analysis uses a one-stage hierarchical model, restricted cubic splines, within-study multiple imputation, internal-external cross-validation, and assessment of calibration, discrimination, net benefit, and

fairness. Results: the reference systematic review identified 38 studies, including 16 model-development and 22 external-validation studies, covering 16 models. ABCD and DiaRem were the most frequently evaluated. Published aggregate meta-analysis reported AUCs of 0.79 at one year and 0.80 at longer follow-up for ABCD, and 0.78 at short-term and 0.83 at longer-term follow-up for DiaRem. Calibration was inconsistent, and only one external validation study was rated at low risk of bias. Diabetes duration, insulin use, beta-cell reserve, baseline HbA1c, and medication burden formed the most consistent prognostic core. No participant-level databases were transferred in the present execution; therefore, no effect estimate is presented as a completed IPD-MA result. Conclusion: informative cohorts exist, but heterogeneity in remission definitions, C-peptide measurement, surgical procedures, and follow-up prevents robust individual prediction from aggregate reports alone. A collaborative IPD-MA can address non-linearity, interactions, transportability, and clinical utility when governance and harmonization are prospectively organized.

Keywords: metabolic surgery; type 2 diabetes; remission; prognosis; individual participant data; prediction model; meta-analysis.

1 INTRODUÇÃO

A remissão do diabetes tipo 2 tornou-se um desfecho clínico legítimo, mas não equivale a cura definitiva. O consenso internacional de 2021 propôs HbA1c inferior a 6,5%, documentada pelo menos três meses após a suspensão da farmacoterapia hipoglicemiante, como critério usual. Esse enquadramento evita a linguagem de cura, reconhece a possibilidade de recaída e preserva a necessidade de vigilância para complicações microvasculares, risco cardiovascular e retorno da hiperglicemia.[1]

A cirurgia metabólica produz controle glicêmico mais intenso e duradouro que o tratamento clínico isolado em ensaios e coortes de longo prazo. No STAMPEDE, os grupos cirúrgicos mantiveram maior controle glicêmico aos cinco anos; no ensaio italiano, parte dos participantes permaneceu em remissão

durante dez anos; e análises colaborativas de ensaios randomizados confirmaram vantagem da cirurgia sobre intervenções médicas e de estilo de vida.[2–6] Entretanto, a remissão diminui com o tempo e apresenta grande variabilidade entre pessoas submetidas ao mesmo procedimento.

A possibilidade de prever essa resposta antes da operação tem implicações para aconselhamento, escolha do procedimento, intensidade do acompanhamento e desenho de estudos. Uma estimativa individual não deve ser usada para negar tratamento a quem tem baixa probabilidade de remissão, pois a cirurgia pode oferecer perda ponderal, redução medicamentosa e proteção cardiovascular mesmo sem normalização glicêmica. O uso correto é apoiar decisão compartilhada, tornar expectativas realistas e identificar quem necessita de terapia combinada desde o início.

Os modelos mais conhecidos foram construídos em contextos distintos. O DiaRem combina idade, HbA1c, uso de insulina e classes de fármacos; o ABCD inclui idade, índice de massa corporal, C-peptídeo e duração do diabetes; o IMS usa duração, número de medicamentos, insulina e controle glicêmico para estratificar gravidade e apoiar escolha entre gastrectomia vertical e bypass; o Ad-DiaRem e o DiaRem2 acrescentam duração e carga terapêutica; e o DiaBetter privilegia HbA1c, duração e medicamentos.[7–12]

Essa proliferação não resolveu a decisão clínica. A revisão sistemática de Singh e colaboradores encontrou 16 modelos, mas poucos foram validados repetidamente, a calibração foi raramente relatada e a maioria dos estudos apresentou risco de viés elevado ou incerto.[13] AUC aceitável em uma amostra não garante que probabilidades absolutas estejam corretas em outro país, procedimento ou período. Um modelo pode ordenar pacientes de modo razoável e ainda superestimar ou subestimar o risco em todos os estratos.

A meta-análise convencional de dados agregados é limitada porque recebe apenas coeficientes selecionados, médias por grupos ou AUCs. Ela não permite padronizar integralmente o desfecho, recuperar variáveis contínuas, investigar relações não lineares, avaliar interações pessoa-procedimento nem imputar

dados ausentes de forma coerente. A IPD-MA solicita os bancos originais, verifica a integridade dos registros, harmoniza variáveis e reanalisa cada participante dentro de uma estrutura hierárquica, sendo especialmente adequada para prognóstico.[14–16]

O objetivo deste capítulo foi construir uma IPD-MA metodologicamente executável e cientificamente honesta. A parte concluída consiste em revisão de viabilidade, mapeamento de coortes, síntese de desempenho publicado e especificação analítica. A etapa que depende de acordos de compartilhamento é explicitamente separada. Assim, o texto não converte dados agregados em supostos registros individuais nem apresenta números simulados como resultados clínicos.

2 MÉTODOS

2.1 Delineamento e padrão de relato

Foi estruturada uma revisão sistemática com plano de IPD-MA de estudos prognósticos e de modelos de predição. O relato segue os itens aplicáveis do PRISMA-IPD, enquanto a identificação, extração e crítica de modelos se baseiam em CHARMS, TRIPOD e PROBAST.[14,17–19] Para fatores prognósticos isolados, os domínios de participação, atrito, mensuração, confundimento e análise foram alinhados ao QUIPS e ao guia de Riley e colaboradores.[15]

A investigação foi dividida em três camadas. A primeira mapeou modelos, coortes e definições. A segunda estabeleceu a ponte quantitativa com resultados agregados já publicados, evitando duplicar participantes. A terceira prescreveu aquisição, harmonização e análise de IPD. Apenas as duas primeiras são apresentadas como executadas; a terceira permanece protocolo operacional até que os controladores dos bancos autorizem a transferência.

2.2 Pergunta, população e estimandos

A população-alvo compreende adultos com diabetes tipo 2 submetidos a cirurgia metabólica primária. Foram contemplados bypass gástrico em Y de Roux, gastrectomia vertical, bypass gástrico de uma anastomose e procedimentos com derivação duodenal, desde que o número por técnica pudesse ser identificado. Revisões, séries exclusivamente pediátricas, transplante, pancreatectomia e operações realizadas por câncer foram excluídos.

O desfecho primário é remissão bioquímica em um ano segundo o consenso atual: HbA1c <6,5% por pelo menos três meses sem medicamento hipoglicemiante. Desfechos secundários são remissão em dois a cinco anos, remissão além de cinco anos, tempo até recaída, HbA1c contínua, redução de medicamentos e remissão estrita conforme critérios históricos. Quando apenas glicemia de jejum está disponível, o estudo permanece na análise de sensibilidade, não na definição principal.

O estimando prognóstico primário é a associação entre cada característica pré-operatória e a probabilidade de remissão, ajustada por procedimento e estudo. O estimando de predição é a probabilidade absoluta individual em cada horizonte. Variáveis pós-operatórias, como perda de peso, não entram no modelo exclusivamente pré-operatório; são analisadas em modelo dinâmico separado por serem mediadores e só estarem disponíveis após o tratamento.

Quadro 1. Estrutura dos estimandos e decisões analíticas

Elemento	Definição principal	Análise de sensibilidade	Razão
População	Adultos com diabetes tipo 2 e cirurgia primária	IMC <35 kg/m²; etnias e países separadamente	Examinar transportabilidade
Desfecho	HbA1c <6,5% ≥3 meses sem fármacos	HbA1c <6,0%; remissão parcial/total histórica	Reduzir erro por definições incompatíveis

Elemento	Definição principal	Análise de sensibilidade	Razão
Horizonte	1 ano	2–5 anos; >5 anos; tempo até recaída	Separar resposta precoce de durabilidade
Preditores	Somente dados pré-operatórios	Modelo dinâmico com perda de peso e resposta precoce	Evitar ajuste por mediadores
Procedimento	Efeito e interação por técnica	Agrupamento por mecanismo anatômico	Avaliar heterogeneidade pessoa-procedimento
Uso clínico	Probabilidade absoluta e benefício líquido	Pontos de corte dos escores existentes	Evitar ranking baseado apenas em AUC

Fonte: elaboração dos autores. HbA1c: hemoglobina glicada; IMC: índice de massa corporal.

2.3 Fontes de informação e busca

A busca reproduzível foi construída para MEDLINE/PubMed com termos para diabetes tipo 2, cirurgia bariátrica ou metabólica, remissão, prognóstico, predição, validação, DiaRem, ABCD, IMS, Ad-DiaRem, DiaBetter, C-peptídeo e duração da doença. A verificação foi complementada por rastreamento de referências, PubMed, PubMed Central, páginas dos periódicos e registros bibliográficos por DOI. A atualização final ocorreu em 1º de agosto de 2026.

Foram priorizados estudos de desenvolvimento ou validação de modelos, coortes com seguimento de pelo menos um ano e ensaios ou registros com variáveis basais potencialmente harmonizáveis. Uma publicação secundária de uma mesma coorte foi retida quando acrescentava horizonte, procedimento, calibração ou variável não disponível no artigo principal. Para a futura IPD-MA, cada coorte, e não cada artigo, constitui a unidade de aquisição.

2.4 Elegibilidade e extração

Dois níveis de elegibilidade foram definidos. O nível A compreende bancos com identificador de participante, procedimento, data da cirurgia, HbA1c, situação medicamentosa e pelo menos uma avaliação pós-operatória. O nível B compreende publicações agregadas úteis para mapear fatores, modelos e desempenho, mas insuficientes para reanálise individual. Bancos sem informação sobre uso de medicamentos podem contribuir para HbA1c contínua, mas não para remissão principal.

Foram extraídos país, centro, período, desenho, amostra, proporção de mulheres, idade, IMC, duração do diabetes, HbA1c, C-peptídeo, insulinoterapia, número e classes de medicamentos, procedimento, definição de remissão, perdas de seguimento, modelos avaliados, AUC, calibração, sensibilidade, especificidade e disponibilidade de equação completa. Também foram registrados responsável pelo banco, existência de comitê de acesso, formato provável e restrições éticas.

2.5 Solicitação, transferência e verificação de IPD

O pacote de solicitação deverá conter protocolo, dicionário mínimo, finalidade, plano de autoria, segurança, prazo de retenção e proibição de reidentificação. Cada centro pode transferir dados pseudonimizados ou executar código federado e devolver estatísticas suficientes. A arquitetura federada é preferível quando legislação ou contrato impede transferência internacional.

Após o recebimento, serão verificadas amplitude de datas, duplicatas, coerência entre idade e data da cirurgia, unidades de HbA1c e C-peptídeo, categorias de procedimento, códigos de medicamentos e consistência entre artigo e banco. Divergências maiores serão discutidas com o investigador responsável. Um relatório de integridade por coorte será congelado antes da modelagem.

2.6 Harmonização das variáveis

HbA1c será convertida entre porcentagem NGSP e mmol/mol IFCC. C-peptídeo será convertido para ng/mL ou nmol/L, mantendo indicador de jejum, estímulo e função renal. Duração do diabetes será contínua e calculada pela diferença entre ano da cirurgia e ano do diagnóstico, com categoria de incerteza quando o início for autorreferido. Medicamentos serão mapeados por classe e contagem, distinguindo insulina, metformina, sulfonilureias, agonistas de GLP-1, inibidores de SGLT2, DPP-4 e outros.

O procedimento será codificado por técnica e por mecanismo: restritivo predominante, derivação gastrointestinal e derivação duodenal. Revisões e conversões serão tratadas como eventos pós-basais. O IMC permanecerá contínuo; categorizações históricas serão usadas apenas para reproduzir escores. Idade, HbA1c, C-peptídeo e duração não serão dicotomizados no modelo novo.

A remissão será rederivada a partir de HbA1c, data, uso de medicamentos e intervalo desde a suspensão. Quando somente a classificação publicada estiver disponível, o registro receberá indicador de desfecho não rederivável. Isso permite separar heterogeneidade biológica de heterogeneidade criada pela definição.

2.7 Risco de viés e aplicabilidade

Estudos de modelos serão avaliados pelo PROBAST nos domínios participantes, preditores, desfecho e análise. Serão considerados problemas críticos a seleção de casos completos, número insuficiente de eventos, seleção de variáveis baseada em p-valor univariado, ausência de tratamento adequado de dados faltantes, avaliação apenas no conjunto de derivação, categorização arbitrária e ausência de calibração.

Para fatores prognósticos isolados, atenção especial será dada ao confundimento entre duração, insulina, HbA1c e escolha do procedimento. A associação bruta entre um procedimento e remissão não será interpretada como

efeito causal quando a técnica foi escolhida segundo gravidade. A aplicabilidade será julgada contra a população contemporânea, incluindo terapias modernas e diversidade de IMC, sexo, raça/etnia e sistemas de saúde.

2.8 Análise estatística prespecificada

A análise principal será de um estágio, por regressão logística hierárquica com intercepto aleatório por estudo e, quando suportado, inclinação aleatória para os preditores centrais. Essa abordagem preserva o agrupamento e estima simultaneamente associação média e heterogeneidade. Uma análise em dois estágios, com coeficientes estimados em cada coorte e combinados por efeitos aleatórios, será usada como verificação.

Idade, IMC, duração, HbA1c e C-peptídeo serão modelados com splines cúbicas restritas. A linearidade será examinada sem escolher pontos de corte a partir dos dados. Interações prespecificadas são duração × procedimento, C-peptídeo × procedimento, IMC × procedimento, insulina × duração e HbA1c × duração. Interações serão retidas apenas quando clinicamente plausíveis, estáveis na validação interna-externa e acompanhadas de ganho de calibração ou utilidade.

Dados ausentes serão imputados dentro de cada estudo por equações encadeadas, preservando o desfecho, o procedimento e variáveis auxiliares. Não se imputará variável inteiramente ausente em uma coorte para a análise principal. Modelos com conjunto mínimo e conjunto estendido permitirão que estudos sem C-peptídeo ainda contribuam. A proporção de informação perdida será documentada.

O desempenho será estimado por AUC/C-statistic, Brier score, calibração no intercepto, inclinação de calibração e análise de curva de decisão. A validação interna-externa excluirá uma coorte por vez, desenvolverá ou atualizará o modelo nas restantes e testará na omitida. O resultado central será a distribuição de desempenho entre contextos, e não uma média que esconda falhas locais.

A utilidade clínica será avaliada por análise de curva de decisão em limiares de probabilidade previamente definidos com pacientes e clínicos. Serão comparadas estratégias de usar o modelo, operar todos, operar nenhum e usar escores existentes. O objetivo não é selecionar candidatos à cirurgia exclusivamente pela remissão, mas indicar aconselhamento, necessidade de combinação terapêutica e escolha de procedimento quando clinicamente apropriada.

A equidade será explorada por calibração e benefício líquido em sexo, faixas etárias, raça/etnia, região, IMC e acesso a terapias modernas. Diferenças de desempenho não serão automaticamente atribuídas à biologia; poderão refletir qualidade dos dados, tratamento prévio, seguimento ou prática clínica.

2.9 Ponte com dados agregados

Enquanto a transferência de IPD não ocorre, o capítulo resume resultados agregados da revisão de Singh e de coortes de longa duração. Essas estimativas informam a viabilidade, mas não substituem o modelo hierárquico individual. AUCs foram reproduzidas com seus intervalos publicados; não foram recalculadas a partir de gráficos nem convertidas em odds ratios.

Uma matriz qualitativa de consistência classificou cada preditor como consistente, moderado, incerto ou não avaliado em cada horizonte, considerando repetição entre modelos, coortes de longo prazo e plausibilidade clínica. Ela é deliberadamente ordinal e não deve ser interpretada como tamanho de efeito.

3 RESULTADOS

3.1 Base de evidências e viabilidade

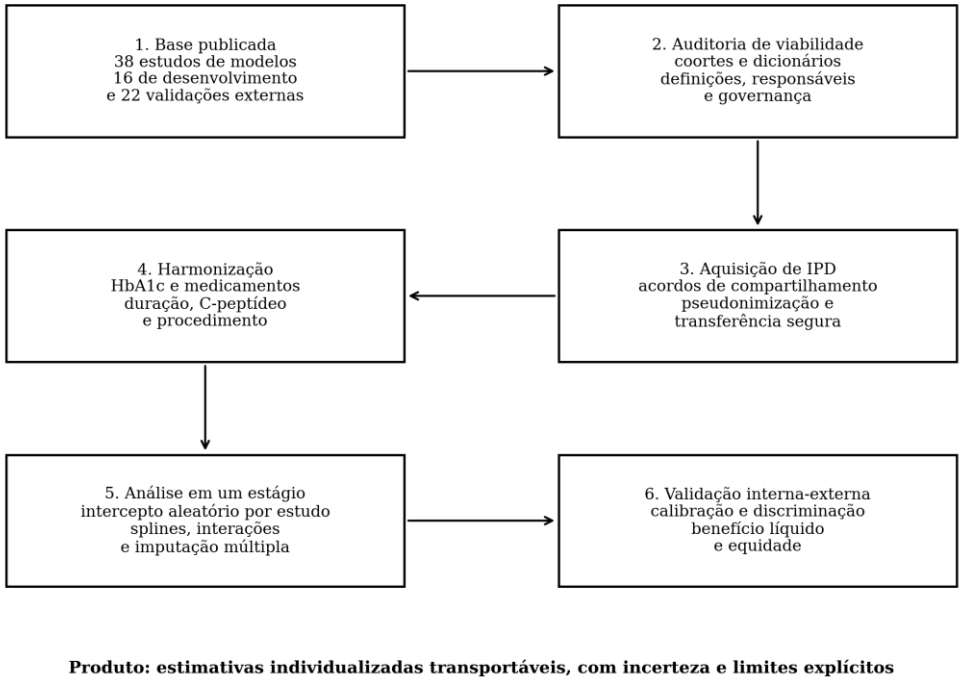
A revisão sistemática de referência rastreou 5.825 registros, manteve 4.165 após remoção de duplicatas e incluiu 38 estudos: 16 de desenvolvimento e 22 de validação externa. Foram identificados 16 modelos, dos quais 11 eram escores e

cinco equações de regressão. Treze dos 16 estudos de desenvolvimento foram julgados de alto risco de viés, dois de baixo e um incerto. Entre as 22 validações, uma foi de baixo risco, três de alto e 18 incertas, principalmente por dados faltantes e relato incompleto do desempenho.[13]

As amostras de desenvolvimento variaram de 46 a 690 participantes; as validações, de 53 a 2.190. Predominaram coortes retrospectivas e seguimento de um ano. Apenas uma validação alcançou dez anos e seis atingiram cinco anos na revisão de 2021. As definições de remissão usaram HbA1c de 5,7%, 6,0% ou 6,5%, sempre com variação adicional na exigência de retirada de medicamentos. Essa diversidade confirma que a rederivação do desfecho é uma das principais vantagens do IPD.

A infraestrutura potencial é favorável: há registros nacionais, coortes prospectivas, ensaios randomizados e centros com longa experiência. O obstáculo não é a inexistência de dados, mas a fragmentação da governança, o uso de dicionários diferentes e a escassez de políticas públicas de compartilhamento no material publicado. O fluxo proposto transforma artigos em coortes elegíveis, coortes em acordos e acordos em bancos harmonizados.

Figura 1. Fluxo da evidência publicada até uma meta-análise de dados individuais



Fonte: elaboração dos autores. Os 38 estudos, 16 estudos de desenvolvimento e 22 validações externas derivam da revisão sistemática de Singh et al.[13]

3.2 Coortes e modelos centrais

Tabela 1. Estudos e coortes prioritários para aquisição de dados individuais

Estudo/coorte	Contexto	Amostra publicada	Contribuição	Preditores-chave	Valor para IPD
Still et al., 2014	Geisinger; RYGB	690 no desenvolvimento	DiaRem; 1 ano	Idade, HbA1c, insulina, medicamentos	Equação simples; forte uso de rotina
Lee et al., 2013	Ásia; bypass	63 na coorte inicial	ABCD	Idade, IMC, C-peptídeo, duração	Reserva beta; ampla validação asiática
Aminian et al., 2017	Cleveland Clinic; RYGB/SG	659; ≥5 anos	IMS; remissão durável	Duração, nº fármacos, insulina, HbA1c	Interação com procedimento

Estudo/coorte	Contexto	Amostra publicada	Contribuição	Preditores-chave	Valor para IPD
Pucci et al., 2018	Reino Unido; RYGB/SG	346 identificados; 210 na ROC	DiaBetter; 2 anos	HbA1c, duração, medicamentos	Inclui perda de peso em análise dinâmica
SOS, Sjöholm et al., 2020	Suécia; cirurgia bariátrica	363 com dados de escores	Remissão durável/complicações	Variáveis de cinco escores e duração	Seguimento prolongado e calibração
LABS, Purnell et al., 2021	Estados Unidos; RYGB/LAGB	Coorte longitudinal; 7 anos	Remissão e recaída	C-peptídeo, HOMA-%B, duração, insulina	Medições metabólicas repetidas
STAMPEDE	Ensaio randomizado; RYGB/SG	150 randomizados	Controle/remissão em 5 anos	Basal amplo e procedimento randomizado	Reduz confusão por seleção de técnica
Ensaio italiano	Ensaio randomizado; RYGB/BPD	60 randomizados	Remissão em 10 anos	Basal e seguimento prolongado	Durabilidade e recaída

Fonte: elaboração dos autores a partir dos estudos originais e revisões. RYGB: bypass gástrico em Y de Roux; SG: gastrectomia vertical; LAGB: banda gástrica ajustável; BPD: derivação biliopancreática; ROC: receiver operating characteristic.

O DiaRem foi derivado de registros eletrônicos de pessoas submetidas a RYGB e tornou-se atraente por utilizar variáveis rotineiras.[7] O ABCD acrescenta C-peptídeo e duração, oferecendo leitura mais direta da reserva de célula beta, mas exige padronização laboratorial.[8] O IMS foi construído em 659 pessoas com pelo menos cinco anos de seguimento e incluiu quatro marcadores de gravidade; além de prever remissão, propôs orientação entre RYGB e SG.[9]

O Ad-DiaRem, DiaBetter e DiaRem2 convergem ao incorporar duração do diabetes, variável ausente no DiaRem original.[10–12] Essa convergência é importante: duração representa tempo de exposição, perda progressiva de função beta e intensidade terapêutica, mas pode conter erro de memória e depender de acesso ao diagnóstico. Em IPD, deve permanecer contínua e acompanhada de indicador de fonte.

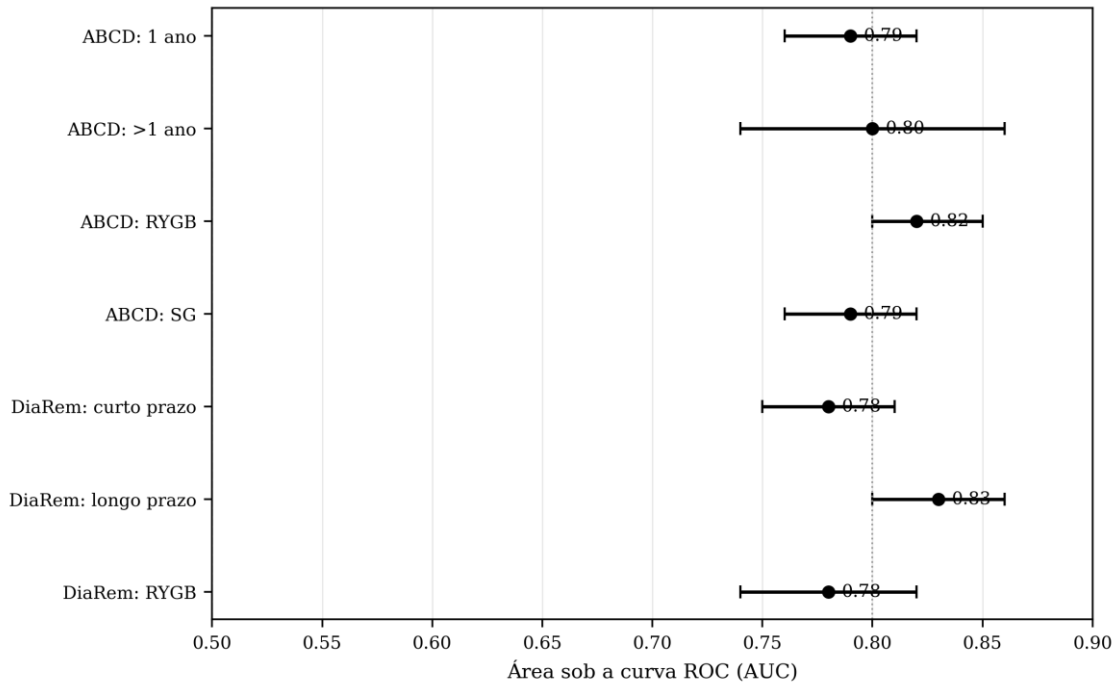
3.3 Desempenho publicado dos escores

Na meta-análise agregada, o ABCD apresentou AUC 0,79 (IC95% 0,76–0,82) em um ano e 0,80 (0,74–0,86) em seguimento mais longo. Para RYGB, a AUC foi 0,82 (0,80–0,85), e para SG, 0,79 (0,76–0,82). O DiaRem apresentou AUC 0,78 (0,75–0,81) no curto prazo, 0,83 (0,80–0,86) no longo prazo e 0,78 (0,74–0,82) para RYGB.[13]

Essas diferenças são pequenas e os intervalos se sobrepõem. Não existe base para declarar um escore universalmente superior. O ponto mais relevante é a instabilidade de calibração. O DiaRem subestimou a remissão em algumas coortes e superestimou em outras, com razões esperado/observado publicadas entre 0,63 e 1,71. O ABCD também alternou entre superestimação e subestimação conforme horizonte.[13]

AUC mede ordenação, não a correção da probabilidade. Dois modelos com AUC semelhante podem produzir aconselhamento muito diferente se um prevê 70% e outro 40% para a mesma pessoa. A IPD-MA deve, portanto, priorizar calibração, atualização local e benefício líquido, e não apenas repetir uma comparação de AUCs.

Figura 2. Desempenho discriminatório agregado publicado para ABCD e DiaRem



Fonte: reconstrução gráfica a partir das AUCs e intervalos de confiança publicados por Singh et al.[13] Não representa meta-análise de IPD.

3.4 Preditores individuais mais consistentes

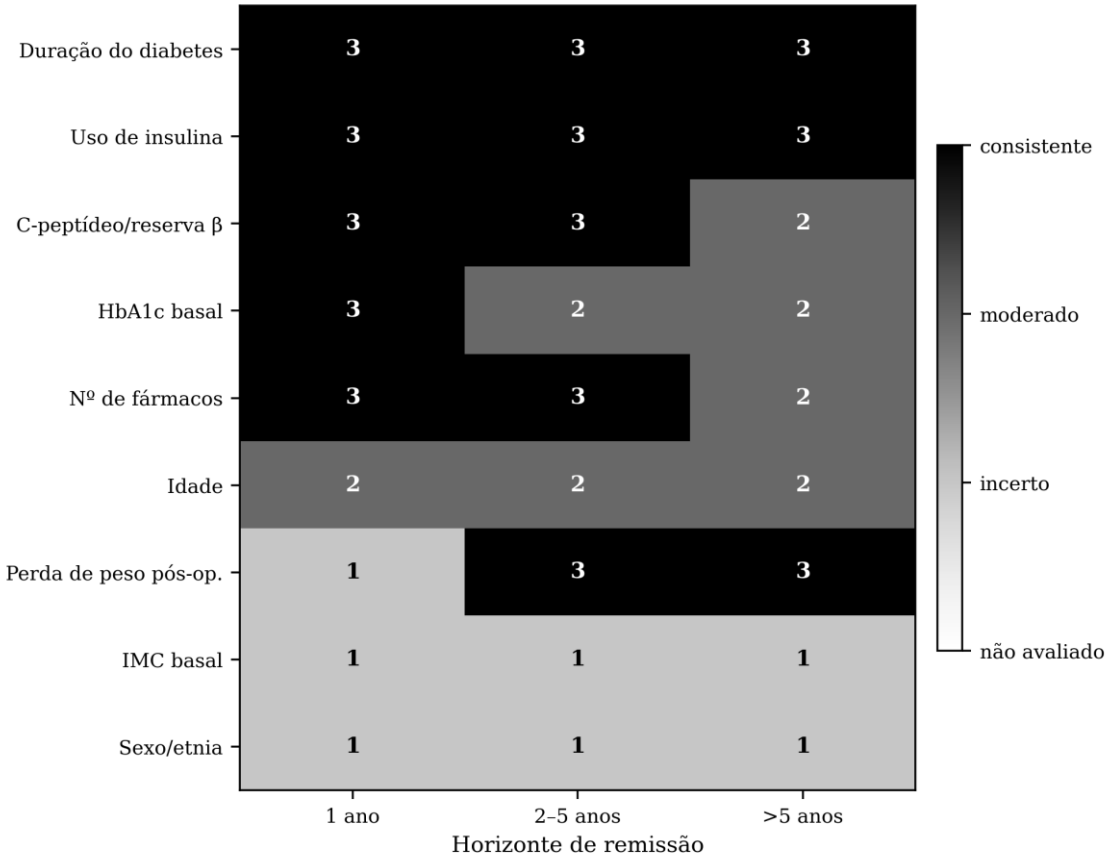
Duração do diabetes, uso de insulina, C-peptídeo ou outra medida de reserva beta, HbA1c e número de medicamentos aparecem repetidamente nos modelos e coortes. No LABS, maior remissão ao longo de sete anos esteve associada a menor idade, menor duração, maior C-peptídeo, maior HOMA-%B, menor uso de insulina e maior perda de peso pós-operatória.[20] O IMS identificou duração, número de medicamentos, insulina e HbA1c <7% como preditores independentes de remissão de longo prazo.[9]

O IMC basal é menos estável. Ele integra o ABCD e alguns modelos, mas meta-análises e ensaios mostram benefício metabólico abaixo e acima de 35 kg/m². O IMC pode modificar a relação entre procedimento, perda de peso e remissão, mas não deve ser usado como substituto de reserva beta. Sexo e raça/etnia são ainda menos claros: diferenças podem refletir acesso, tratamento,

composição corporal, seleção cirúrgica ou seguimento, e devem ser estudadas por calibração e não por coeficientes isolados.

A perda de peso pós-operatória é forte marcador de remissão e durabilidade, mas não é preditor pré-operatório. Misturá-la ao modelo de decisão antes da cirurgia cria um instrumento impossível de usar naquele momento e pode introduzir viés de mediação. Sua melhor função é atualizar a previsão aos três, seis ou 12 meses.

Figura 3. Consistência qualitativa dos principais preditores por horizonte temporal



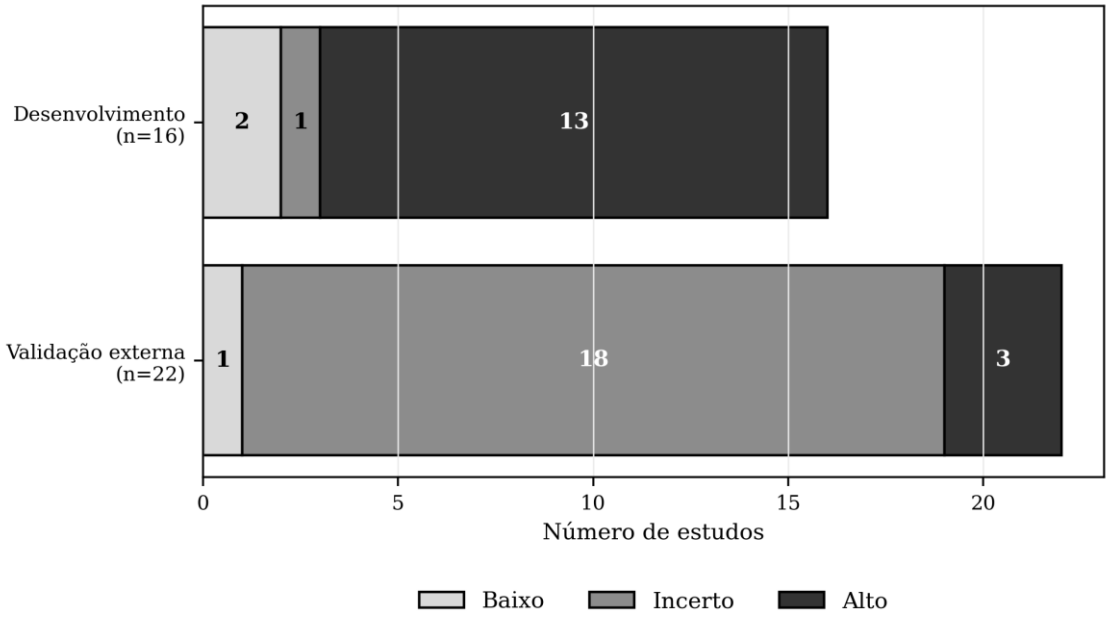
Fonte: síntese qualitativa dos estudos e modelos discutidos. Escala ordinal: 0 = não avaliado; 1 = incerto; 2 = moderado; 3 = consistente. A figura não expressa tamanho de efeito.

3.5 Risco de viés e falhas de relato

O domínio análise concentrou os problemas: poucos eventos por parâmetro, seleção de variáveis por significância, ausência de shrinkage, exclusão de casos incompletos, validação aparente e relato insuficiente de calibração. Em validações, a equação nem sempre era aplicada integralmente e os pontos de corte podiam ser recalibrados no próprio conjunto, misturando validação e atualização.

A predominância de risco alto ou incerto significa que desempenho publicado provavelmente é otimista. Isso não invalida os escores, mas reduz confiança em comparações pequenas. A vantagem da IPD-MA é aplicar todas as equações no mesmo conjunto harmonizado, avaliar a qualidade de cada coorte e realizar validação interna-externa sem depender do que cada artigo escolheu relatar.

Figura 4. Risco de viés dos estudos de modelos na revisão de referência



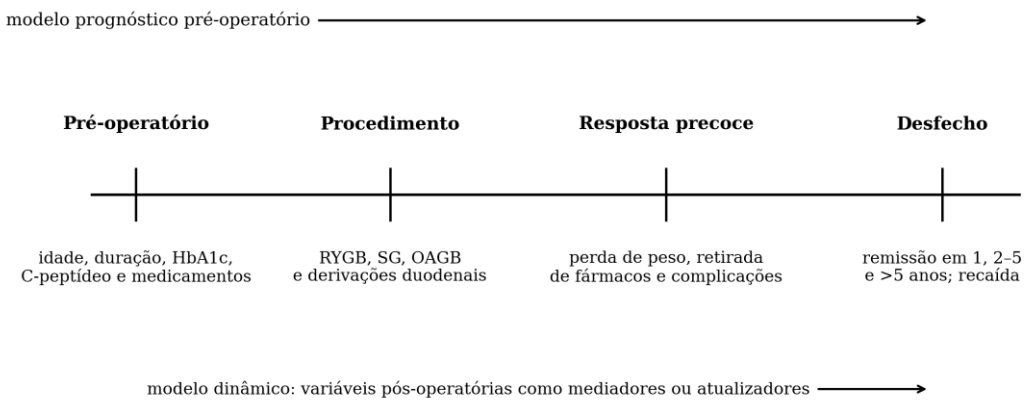
Fonte: elaboração dos autores com base na classificação PROBAST reportada por Singh et al.[13]

3.6 Definição do desfecho e linha temporal

Os estudos misturaram remissão total, parcial, combinada e critérios próprios. Alguns usaram HbA1c <6,0%; outros <6,5%; dois usaram 5,7%. A duração sem medicamentos também variou. Essa heterogeneidade altera prevalência e desempenho dos modelos. Ao rederivar o desfecho, a IPD-MA poderá aplicar uma definição principal e reproduzir as históricas sem descartar bancos.

A linha temporal também modifica interpretação. Em um ano, a previsão captura resposta inicial e retirada de fármacos. Em cinco ou dez anos, incorpora recaída, envelhecimento, ganho de peso, progressão beta e mudanças terapêuticas. Um único modelo estático dificilmente é calibrado em todos os horizontes. O protocolo separa prognóstico pré-operatório, atualização pós-operatória e análise de sobrevivência para recaída.

Figura 5. Linha temporal causal e separação dos modelos pré-operatório e dinâmico



Fonte: elaboração dos autores. Variáveis pós-operatórias devem ser tratadas como mediadores ou atualizadores, não como preditores disponíveis antes da cirurgia.

3.7 Dicionário mínimo harmonizado

Tabela 2. Dicionário mínimo para a IPD-MA

Domínio	Variáveis	Formato	Regra de harmonização	Prioridade
Identificação	study_id, participant_id	Texto/integer	Pseudônimo; unicidade por estudo	Obrigatório
Tempo	data da cirurgia; datas de visitas	AAAA-MM-DD ou intervalo	Permite janelas e tempo até recaída	Obrigatório
Demografia	idade, sexo, raça/etnia, país	Contínua/categórica	Manter categorias originais e mapeadas	Essencial
Antropometria	peso, altura, IMC, cintura	kg, m, kg/m², cm	Derivar IMC quando possível	Essencial
Diabetes	ano diagnóstico, duração, HbA1c	anos; % e mmol/mol	Fonte e unidade registradas	Obrigatório
Reserva beta	C-peptídeo, HOMA-%B	ng/mL/nmol/L; índice	Jejum/estimulado; função renal	Desejável
Tratamento	insulina; classes; número	Binário/categórico/contagem	ATC ou mapa de classes	Obrigatório
Cirurgia	RYGB, SG, OAGB, DS e variantes	Categórica	Primária, revisão e conversão	Obrigatório
Desfecho	HbA1c, fármacos, remissão	Contínuo/binário	Rederivar por horizonte	Obrigatório
Seguimento	perda, óbito, revisão	Categórica/data	Distinguir ausência e evento	Essencial

Fonte: elaboração dos autores. ATC: Anatomical Therapeutic Chemical; OAGB: bypass gástrico de uma anastomose; DS: duodenal switch.

3.8 Plano de análise por modelo

Tabela 3. Modelos que deverão ser reproduzidos e comparados na IPD-MA

Modelo	Preditores	Alvo original	Força	Limitação principal
DiaRem	Idade, HbA1c, insulina, outros fármacos	1 ano; RYGB	Disponível em rotina	Não inclui duração; calibração variável
ABCD	Idade, IMC, C-peptídeo, duração	1–5 anos; várias técnicas	Reserva beta explícita	C-peptídeo heterogêneo
IMS	Duração, nº fármacos, insulina, HbA1c	≥5 anos; RYGB/SG	Gravidade e escolha de técnica	Origem de centro único
Ad-DiaRem	DiaRem + duração + nº fármacos	1 ano; RYGB	Melhor discriminação inicial	Menor validação externa
DiaBetter	HbA1c, duração, medicamentos	2 anos; RYGB/SG	Simples; inclui duração	Validação limitada
DiaRem2	Idade, HbA1c, insulina, duração	1 ano; cirurgia metabólica	Atualiza DiaRem	Menos validações
Modelo IPD mínimo	Idade, duração, HbA1c, insulina, nº fármacos, técnica	1, 2–5 e >5 anos	Aplicável à maioria dos bancos	Menor detalhe metabólico
Modelo IPD estendido	Modelo mínimo + C-peptídeo, IMC e interações	Mesmos horizontes	Melhor individualização potencial	Menos estudos elegíveis

Fonte: elaboração dos autores a partir dos estudos originais. O modelo IPD é prespecificado e não possui desempenho estimado até a aquisição dos dados.

4 DISCUSSÃO

Este capítulo mostra que a pergunta clínica é madura, mas a evidência ainda não está organizada no nível necessário para previsão individual transportável. Há múltiplos modelos, grandes coortes e seguimentos longos, porém as definições de remissão, variáveis e análises permanecem

incompatíveis. A consequência é um paradoxo: a área possui muitos escores e pouca certeza sobre qual probabilidade comunicar a uma pessoa concreta.

O núcleo prognóstico é biologicamente coerente. Menor duração, ausência de insulina, menor carga medicamentosa, melhor HbA1c e maior C-peptídeo apontam para maior reserva funcional de célula beta. Esses fatores não são independentes em sentido causal: duração influencia tratamento e reserva; tratamento depende de acesso e preferência; HbA1c resulta de biologia e cuidado. A IPD-MA deve modelá-los conjuntamente, testar colinearidade e evitar interpretar insulina como marcador puramente biológico.

A duração do diabetes merece cuidado especial. Ela é uma das variáveis mais consistentes em remissão e recaída, mas o momento real de início é desconhecido. Populações com rastreamento regular terão duração registrada maior que populações diagnosticadas tardiamente para a mesma história biológica. A harmonização deve preservar fonte do diagnóstico e explorar erro de mensuração, em vez de usar apenas categorias rígidas como cinco ou dez anos.

O C-peptídeo pode melhorar a previsão, sobretudo em populações com IMC menor, mas cria barreira de transportabilidade. As coortes mediram sangue ou urina, jejum ou estímulo, e raramente padronizaram pela função renal. Um modelo mínimo sem C-peptídeo e um modelo estendido são preferíveis a excluir grande parte dos bancos. O ganho incremental do biomarcador deve ser demonstrado por calibração e benefício líquido, não apenas por aumento pequeno da AUC.

A escolha do procedimento é outro ponto no qual associação e decisão se confundem. Pessoas com diabetes mais grave podem ter sido encaminhadas para RYGB ou derivação duodenal, enquanto outras receberam SG. Comparações observacionais podem subestimar o procedimento mais potente. Ensaios randomizados como STAMPEDE e o estudo italiano são valiosos para avaliar interação, embora suas amostras sejam pequenas. Uma IPD-MA deve separar predição de resposta do efeito causal do procedimento.

O IMS é particularmente relevante porque conecta gravidade e escolha da técnica. Seu desenho de longo prazo é uma força, mas a recomendação de procedimento precisa ser validada em coortes randomizadas e em contextos contemporâneos. A validação conjunta de SLEEVEPASS e SM-BOSS demonstra a viabilidade de análises com dados individuais combinados de ensaios.[21] A extensão para outras técnicas requer cautela, especialmente OAGB e derivações duodenais.

A perda de peso pós-operatória foi associada a remissão no DiaBetter e no LABS. Ela deve alimentar um modelo dinâmico para acompanhamento, não o instrumento pré-operatório. Essa distinção produz dois produtos clínicos: uma previsão basal para aconselhamento e uma previsão atualizada após três a 12 meses para decidir intensificação farmacológica, vigilância e prevenção de recaída.

A remissão não é o único benefício. No SOS, escores de remissão e duração também foram examinados contra complicações, mostrando que previsão de remissão não substitui previsão de eventos cardiovasculares, renais ou microvasculares.[22] Uma pessoa com baixa probabilidade de remissão pode ainda obter benefício substancial. Por isso, nenhum limiar deve ser apresentado como critério isolado de elegibilidade cirúrgica.

O desempenho agregado de ABCD e DiaRem é moderado a bom, mas a calibração instável é o achado mais importante. A superestimação cria expectativa irreal; a subestimação pode afastar pessoas de tratamento efetivo. Atualização de intercepto e inclinação pode resolver parte do problema localmente. Quando relações entre preditores e desfecho mudam, recalibração simples é insuficiente e o modelo precisa ser revisado.

A validação interna-externa em IPD-MA é superior à divisão aleatória de um banco grande. Ao omitir um estudo inteiro, ela simula aplicação em novo contexto. A heterogeneidade de calibração entre estudos informa onde o modelo falha. Um modelo com AUC média alta, mas calibração ruim em vários centros, não deve ser declarado generalizável.

O risco de viés elevado decorre em grande parte de práticas evitáveis. Selecionar variáveis por p-valor, categorizar contínuas, ignorar dados ausentes e avaliar no mesmo conjunto inflaciona desempenho. A análise prespecificada com shrinkage, splines, imputação por estudo e validação interna-externa reduz esses problemas. Também é essencial publicar a equação completa e um calculador transparente, não apenas faixas de escore.

Modelos de inteligência artificial podem capturar interações complexas em registros grandes, e estudos escandinavos já exploraram redes neurais para remissão.[23] Contudo, algoritmos flexíveis não corrigem desfecho mal definido, seguimento seletivo ou calibração ausente. O modelo de referência deve ser penalizado e interpretável; métodos de aprendizado de máquina podem ser comparadores, avaliados nas mesmas coortes e métricas.

A governança é parte do método. Dados de cirurgia metabólica contêm datas, características raras e trajetórias clínicas que podem facilitar reidentificação. A transferência deve ser mínima e segura, com plano de autoria e retorno aos centros. Quando a transferência não for possível, análise federada permite estimar gradientes, matrizes de informação e desempenho sem centralizar registros.

Este trabalho possui limitações. Não foi concluída transferência de IPD, portanto não há coeficientes individuais novos, interações, curvas de calibração ou modelo final. A ponte quantitativa depende de revisão publicada e pode não incluir estudos recentes não indexados. A matriz de consistência é julgamento estruturado, não meta-análise de tamanho de efeito. Essas limitações são explicitadas para impedir que um protocolo seja apresentado como análise concluída.

Por outro lado, a separação entre evidência executada e análise dependente de dados é uma força. O capítulo entrega pergunta, dicionário, estimandos, estratégia de dados ausentes, modelos, validação e utilidade clínica antes do acesso aos bancos. Essa ordem reduz decisões pós-hoc e facilita a formação de um consórcio internacional.

5 IMPLICAÇÕES CLÍNICAS E DE PESQUISA

5.1 Para a consulta pré-operatória

Os escores existentes podem apoiar a conversa, desde que acompanhados de incerteza e da informação de que não predizem todos os benefícios. Duração, insulinoterapia, HbA1c, número de medicamentos e C-peptídeo devem ser interpretados como marcadores de gravidade. Nenhum escore deve ser usado para excluir automaticamente uma pessoa elegível.

Quando o serviço usa um escore, deve verificar se a população e o procedimento se assemelham aos de desenvolvimento, registrar a definição de remissão e, idealmente, auditar calibração local. Informar apenas uma categoria "alta" ou "baixa" sem probabilidade e horizonte é insuficiente.

5.2 Para serviços e registros

Registros devem coletar HbA1c, medicamentos, duração, procedimento e visitas em datas padronizadas. C-peptídeo precisa de unidade, estado de jejum/estímulo e creatinina. Perdas de seguimento devem ser diferenciadas de retorno a outro serviço. Esses elementos são mais importantes para ciência de previsão que dezenas de variáveis pouco completas.

Uma política de compartilhamento, dicionário e identificador de coorte reduzem o custo de futuras IPD-MA. A publicação de equações, interceptos, coeficientes e código de escore é requisito para validação independente.

5.3 Agenda prioritária

A primeira prioridade é formar um consórcio com coortes de derivação e validação, ensaios e registros nacionais. A segunda é aplicar a definição contemporânea e análises históricas em paralelo. A terceira é validar conjuntamente modelos existentes antes de criar outro escore. A quarta é

desenvolver modelos mínimo e estendido e testar benefício líquido. A quinta é incorporar recaída, complicações e qualidade de vida como desfechos complementares.

Análises futuras devem examinar tratamentos contemporâneos, incluindo agonistas de GLP-1 e inibidores de SGLT2, porque a carga medicamentosa atual não é equivalente à de coortes de 2004. Mudanças na técnica cirúrgica e no seguimento também exigem atualização temporal.

6 CONCLUSÃO

A previsão de remissão do diabetes tipo 2 após cirurgia metabólica é sustentada por um conjunto repetido de marcadores de gravidade e reserva beta, mas os modelos publicados não são intercambiáveis. ABCD e DiaRem apresentam discriminação agregada aceitável a boa, enquanto calibração e transportabilidade permanecem incertas. Duração do diabetes, insulina, C-peptídeo, HbA1c e carga medicamentosa devem constituir o núcleo de uma análise individual.

Uma IPD-MA colaborativa é o desenho adequado para padronizar remissão, manter variáveis contínuas, testar interações com procedimento, tratar dados ausentes e realizar validação interna-externa. Até que os bancos sejam obtidos, qualquer coeficiente novo seria especulativo. O produto imediato deste capítulo é um protocolo completo e uma síntese de viabilidade; o produto futuro será um modelo transparente, calibrado e clinicamente útil, que complemente — e não substitua — a decisão compartilhada.

REFERÊNCIAS

1. Riddle MC, Cefalu WT, Evans PH, Gerstein HC, Nauck MA, Oh WK, et al. Consensus report: definition and interpretation of remission in type 2 diabetes. Diabetes Care. 2021;44(10):2438-2444. doi:10.2337/dci21-0034.

2. Schauer PR, Bhatt DL, Kirwan JP, Wolski K, Aminian A, Brethauer SA, et al. Bariatric surgery versus intensive medical therapy for diabetes - 5-year outcomes. N Engl J Med. 2017;376(7):641-651. doi:10.1056/NEJMoa1600869.

3. Mingrone G, Panunzi S, De Gaetano A, Guidone C, Iaconelli A, Capristo E, et al. Metabolic surgery versus conventional medical therapy in patients with type 2 diabetes: 10-year follow-up of an open-label, single-centre, randomised controlled trial. Lancet. 2021;397(10271):293-304. doi:10.1016/S0140-6736(20)32649-0.

4. Courcoulas AP, Belle SH, Neiberg RH, Pierson SK, Eagleton JK, Kalarchian MA, et al. Long-term outcomes of medical management vs bariatric surgery in type 2 diabetes. JAMA. 2024;331(8):654-664. doi:10.1001/jama.2024.0318.

5. Kirwan JP, Courcoulas AP, Cummings DE, Goldfine AB, Kashyap SR, Simonson DC, et al. Diabetes remission in the Alliance of Randomized Trials of Medicine versus Metabolic Surgery in Type 2 Diabetes. Diabetes Care. 2022;45(7):1574-1583. doi:10.2337/dc21-2441.

6. Sjöström L, Peltonen M, Jacobson P, Ahlin S, Andersson-Assarsson J, Anveden Å, et al. Association of bariatric surgery with long-term remission of type 2 diabetes and with microvascular and macrovascular complications. JAMA. 2014;311(22):2297-2304. doi:10.1001/jama.2014.5988.

7. Still CD, Wood GC, Benotti P, Petrick AT, Gabrielsen J, Strodel WE, et al. A probability score for preoperative prediction of type 2 diabetes remission following Roux-en-Y gastric bypass surgery. Lancet Diabetes Endocrinol. 2014;2(1):38-45. doi:10.1016/S2213-8587(13)70070-6.

8. Lee WJ, Hur KY, Lakadawala M, Kasama K, Wong SKH, Chen SC, et al. Predicting success of metabolic surgery: age, body mass index, C-peptide, and duration score. Surg Obes Relat Dis. 2013;9(3):379-384. doi:10.1016/j.soard.2012.07.015.

9. Aminian A, Brethauer SA, Andalib A, Nowacki AS, Jimenez A, Corcelles R, et al. Individualized metabolic surgery score: procedure selection based on diabetes severity. Ann Surg. 2017;266(4):650-657. doi:10.1097/SLA.0000000000002407.

10. Aron-Wisnewsky J, Sokolovska N, Liu Y, Comaneshter DS, Vinker S, Pecht T, et al. The advanced-DiaRem score improves prediction of diabetes remission 1 year post-Roux-en-Y gastric bypass. Diabetologia. 2017;60(10):1892-1902. doi:10.1007/s00125-017-4371-7.

11. Pucci A, Tymoszuk U, Cheung WH, Makaronidis JM, Scholes S, Tharakan G, et al. Type 2 diabetes remission 2 years post Roux-en-Y gastric bypass and

sleeve gastrectomy: the role of weight loss and comparison of DiaRem and DiaBetter scores. Diabet Med. 2018;35(3):360-367. doi:10.1111/dme.13532.

12. Still CD, Benotti P, Mirshahi T, Cook A, Wood GC. DiaRem2: incorporating duration of diabetes to improve prediction of diabetes remission following metabolic surgery. Surg Obes Relat Dis. 2019;15(5):717-724. doi:10.1016/j.soard.2018.12.020.

13. Singh P, Adderley NJ, Hazlehurst J, Price M, Tahrani AA, Nirantharakumar K, et al. Prognostic models for predicting remission of diabetes following bariatric surgery: a systematic review and meta-analysis. Diabetes Care. 2021;44(11):2626-2641. doi:10.2337/dc21-0166.

14. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD statement. JAMA. 2015;313(16):1657-1665. doi:10.1001/jama.2015.3656.

15. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597. doi:10.1136/bmj.k4597.

16. Simmonds M, Stewart G, Stewart L. A decade of individual participant data meta-analyses: a review of current practice. Contemp Clin Trials. 2015;45(Pt A):76-83. doi:10.1016/j.cct.2015.06.012.

17. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744.

18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63. doi:10.7326/M14-0697.

19. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-58. doi:10.7326/M18-1376.

20. Purnell JQ, Selzer F, Wahed AS, Pender J, Pories W, Pomp A, et al. Diabetes remission status during seven-year follow-up of the Longitudinal Assessment of Bariatric Surgery study. J Clin Endocrinol Metab. 2021;106(3):774-788. doi:10.1210/clinem/dgaa849.

21. Salminen P, Helmio M, Ovaska J, Juuti A, Leivonen M, Peromaa-Haavisto P, et al. Validation of the individualized metabolic surgery score for bariatric

procedure selection in merged data from the SLEEVEPASS and SM-BOSS randomized trials. Surg Obes Relat Dis. 2023;19(2):132-140.

22. Sjöholm K, Carlsson LMS, Svensson PA, Andersson-Assarsson JC, Kristensson F, Jacobson P, et al. Comparison of preoperative remission scores and diabetes duration alone as predictors of durable type 2 diabetes remission and risk of diabetes complications after bariatric surgery. Diabetes Care. 2020;43(11):2804-2811. doi:10.2337/dc20-0567.

23. Cao Y, Näslund E, Näslund I, Ottosson J, Montgomery S, Stenberg E. Using a convolutional neural network to predict remission of diabetes after gastric bypass surgery: machine learning study from the Scandinavian Obesity Surgery Register. JMIR Med Inform. 2021;9(8):e25612. doi:10.2196/25612.

24. Wang GF, Yan YX, Xu N, Yin D, Hui Y, Zhang JP, et al. Predictive factors of type 2 diabetes mellitus remission following bariatric surgery: a meta-analysis. Obes Surg. 2015;25(2):199-208. doi:10.1007/s11695-014-1391-y.

25. Wood GC, Horwitz D, Still CD, Mirshahi T, Benotti P, Parikh M, et al. Performance of the DiaRem score for predicting diabetes remission in two health systems following bariatric surgery procedures in Hispanic and non-Hispanic White patients. Obes Surg. 2018;28(1):61-68. doi:10.1007/s11695-017-2805-7.

26. Dicker D, Golan R, Aron-Wisnewsky J, Zucker JD, Sokolovska N, Comaneshter DS, et al. Prediction of long-term diabetes remission after RYGB, sleeve gastrectomy, and adjustable gastric banding using DiaRem and Advanced-DiaRem scores. Obes Surg. 2019;29(3):796-804. doi:10.1007/s11695-018-3573-3.

27. Kam H, Tu Y, Pan J, Han J, Zhang P, Bao Y, et al. Comparison of four risk prediction models for diabetes remission after Roux-en-Y gastric bypass surgery in obese Chinese patients with type 2 diabetes mellitus. Obes Surg. 2020;30(6):2147-2157. doi:10.1007/s11695-020-04455-x.

28. Debédat J, Sokolovska N, Coupaye M, Panunzi S, Chakaroun R, Genser L, et al. Long-term relapse of type 2 diabetes after Roux-en-Y gastric bypass: prediction and clinical relevance. Diabetes Care. 2018;41(10):2086-2095. doi:10.2337/dc18-0567.

29. Park JY. Prediction of type 2 diabetes remission after bariatric or metabolic surgery. J Obes Metab Syndr. 2018;27(4):213-222. doi:10.7570/jomes.2018.27.4.213.

30. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model

for individual prognosis or diagnosis: explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73. doi:10.7326/M14-0698.

APÊNDICE A — ESTRATÉGIA DE BUSCA REPRODUZÍVEL

("Diabetes Mellitus, Type 2"[Mesh] OR type 2 diabetes[tiab] OR T2DM[tiab]) AND ("Bariatric Surgery"[Mesh] OR metabolic surgery[tiab] OR bariatric surgery[tiab] OR gastric bypass[tiab] OR Roux-en-Y[tiab] OR sleeve gastrectom*[tiab] OR one-anastomosis gastric bypass[tiab] OR duodenal switch[tiab]) AND (remission[tiab] OR relapse[tiab] OR recurrence[tiab]) AND (predict*[tiab] OR prognos*[tiab] OR model*[tiab] OR score[tiab] OR validation[tiab] OR DiaRem[tiab] OR ABCD[tiab] OR individualized metabolic surgery[tiab] OR C-peptide[tiab] OR diabetes duration[tiab])

A estratégia deve ser traduzida para Embase, Scopus e Web of Science no protocolo do consórcio, com registro da data, plataforma, número de resultados e processo de deduplicação. Os nomes dos escores devem ser pesquisados separadamente para recuperar validações que não usam termos de prognóstico no título.

APÊNDICE B — FORMULÁRIO DE SOLICITAÇÃO DE IPD

1. Identificação da coorte, investigador responsável e instituição controladora.

2. Base legal e aprovação ética para compartilhamento ou análise federada.

3. Número de participantes com diabetes tipo 2, por procedimento e horizonte.

4. Disponibilidade das variáveis obrigatórias do dicionário mínimo.

5. Definição original de remissão e disponibilidade de HbA1c e medicamentos para rederivação.

6. Método de pseudonimização, formato, canal de transferência e restrições.

7. Termos de autoria, reconhecimento, uso secundário e destruição ao final.

8. Nome do dicionário, versão, unidades e códigos de ausência.

9. Confirmação de que o mesmo banco não foi transferido por outro artigo/centro.

10. Contato para esclarecimento de inconsistências e validação dos resultados específicos da coorte.

APÊNDICE C — PLANO DE SAÍDA ESTATÍSTICA

Tabela C1: características por estudo antes e depois da imputação, sem combinar médias que ocultem diferenças.

Tabela C2: completude de cada variável e motivo de ausência por estudo.

Tabela C3: coeficientes do modelo mínimo e estendido com shrinkage e intervalos.

Figura C1: relação não linear entre duração, HbA1c, C-peptídeo e remissão.

Figura C2: calibração por estudo e por horizonte.

Figura C3: distribuição de AUC, Brier e inclinação na validação interna-externa.

Figura C4: benefício líquido por procedimento e limiar.

Figura C5: desempenho em subgrupos de equidade.

Análise S1: definição contemporânea versus critérios históricos.

Análise S2: casos completos versus imputação múltipla.

Análise S3: exclusão de estudos de alto risco de viés.

Análise S4: modelo mínimo versus estendido e escores existentes.

Análise S5: tempo até recaída com risco basal específico por estudo.

CAPÍTULO 5

INTERVENÇÕES PSICOLÓGICAS PERIOPERATÓRIAS PARA PREVENÇÃO DE ANSIEDADE, DOR PERSISTENTE E RECUPERAÇÃO FUNCIONAL INADEQUADA APÓS CIRURGIAS DE GRANDE PORTE

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: estabelecer um protocolo completo e executável de meta-análise prospectiva sobre intervenções psicológicas perioperatórias para reduzir ansiedade, dor persistente e recuperação funcional inadequada após cirurgias de grande porte, além de testar sua viabilidade com evidências e ensaios reais. Métodos: o plano foi estruturado segundo os princípios de meta-análise prospectiva, PRISMA-P, SPIRIT e PRISMA 2020. Foram considerados elegíveis ensaios randomizados identificados e incluídos antes do conhecimento de seus resultados comparativos, envolvendo cirurgia bariátrica, cardíaca, oncológica, ortopédica, de coluna e abdominal. Os desfechos centrais foram definidos como ansiedade, dor persistente entre três e seis meses e recuperação funcional, com janelas temporais e instrumentos harmonizados. A análise principal foi especificada como hierárquica de um estágio, por intenção de tratar, com efeitos aleatórios por ensaio e interações prespecificadas. Resultados: a síntese de viabilidade encontrou um campo ativo, porém fragmentado. Meta-análise de 2025 reuniu 27 ensaios e 2.990 participantes, com 17 estudos combináveis, e observou efeitos favoráveis para dor (g = −0,45; IC95% −0,77 a −0,13) e ansiedade (g = −0,33; IC95% −0,54 a −0,11). Ensaios publicados variaram de 36 a 203 participantes, utilizaram terapia cognitivo-comportamental, hipnose, mindfulness, psicoeducação e neuroeducação da dor, e apresentaram resultados

positivos, nulos ou restritos a determinados períodos. Conclusão: a área mostrou-se adequada a uma meta-análise prospectiva porque dispunha de programas colaborativos e novos ensaios, embora apresentasse multiplicidade de escalas, tempos e comparadores. A harmonização anterior ao desbloqueio dos resultados foi proposta para reduzir relato seletivo, preservar dados e esclarecer quais componentes funcionaram, para quem e em qual cirurgia.

Palavras-chave: cirurgia; ansiedade; dor persistente pós-cirúrgica; recuperação funcional; intervenção psicológica; meta-análise prospectiva; ensaio clínico.

ABSTRACT

Objective: to establish a complete and operational protocol for a prospective meta-analysis of perioperative psychological interventions to reduce anxiety, persistent pain, and inadequate functional recovery after major surgery, and to evaluate feasibility using real trials and published evidence. Methods: the plan was structured according to prospective meta-analysis principles, PRISMA-P, SPIRIT, and PRISMA 2020. Randomised trials identified and included before their comparative results became known were considered eligible, covering bariatric, cardiac, oncological, orthopaedic, spine, and abdominal surgery. Core outcomes were defined as anxiety, persistent pain between three and six months, and functional recovery, using harmonised time windows and measures. The primary analysis was specified as a one-stage hierarchical intention-to-treat model with random trial effects and prespecified interactions. Results: the feasibility synthesis showed an active but fragmented field. A 2025 meta-analysis included 27 trials and 2,990 participants, with 17 studies contributing to quantitative synthesis, and reported favourable effects on pain ($g = -0.45$; 95% CI -0.77 to -0.13) and anxiety ($g = -0.33$; 95% CI -0.54 to -0.11). Published trials ranged from 36 to 203 participants and evaluated cognitive behavioural therapy, hypnosis, mindfulness, psychoeducation, and pain neuroscience education, with beneficial, null, or time-limited findings. Conclusion: the field was considered suitable for a prospective meta-analysis because new trials and collaborative

programmes were available, although outcome scales, assessment times, and comparators remained inconsistent. Harmonisation before results were unblinded was proposed to reduce selective reporting, preserve information, and clarify which components worked, for whom, and in which surgical setting.
Keywords: surgery; anxiety; chronic postsurgical pain; functional recovery; psychological intervention; prospective meta-analysis; clinical trial.

1 INTRODUÇÃO

A cirurgia de grande porte mobiliza dimensões biológicas e psicológicas que não podem ser separadas na experiência do paciente. A ameaça percebida, a incerteza sobre o diagnóstico, o medo da anestesia, a expectativa de dor e a possibilidade de perda funcional ativam respostas autonômicas, cognitivas e comportamentais antes da incisão. Em parte das pessoas, esse estado diminui quando o procedimento termina; em outras, mantém hipervigilância, evitação do movimento, pior adesão à reabilitação e consumo prolongado de analgésicos. A dor persistente pós-cirúrgica é reconhecida quando se desenvolve ou aumenta após a operação e permanece por pelo menos três meses, além do período esperado de cicatrização, sem explicação melhor por outra causa. Essa definição desloca o foco do episódio agudo para uma trajetória em que ansiedade, catastrofização, depressão e expectativas podem atuar como fatores de risco e alvos modificáveis.[5]

Intervenções psicológicas perioperatórias procuram modificar essa trajetória antes que o padrão de sofrimento se consolide. Elas incluem terapia cognitivo-comportamental, treino de enfrentamento, reestruturação de expectativas, hipnose, mindfulness, relaxamento, entrevista motivacional, ativação comportamental, educação em neurociência da dor e programas digitais. Embora compartilhem a intenção de fortalecer autorregulação, diferem em teoria, intensidade, profissional responsável, momento de início, duração e comparador. Uma sessão de hipnose imediatamente antes da anestesia não é equivalente a seis semanas de terapia cognitivo-comportamental iniciada antes

da operação e continuada durante a reabilitação. Agrupar todas as estratégias como “apoio psicológico” pode produzir uma média pouco interpretável e esconder componentes realmente úteis.

A evidência acumulada sugere benefício, mas também mostra por que uma nova síntese retrospectiva isolada não resolve a questão. A meta-análise de Castaño-Asins e colaboradores incluiu 27 ensaios, com 1.462 participantes nos grupos de intervenção e 1.528 nos controles. Dezessete estudos puderam ser combinados; os efeitos médios favoreceram as intervenções para dor e ansiedade, mas a qualidade foi moderada e muitos ensaios apresentaram alto risco de viés.[6] A revisão Cochrane mais ampla sobre preparação psicológica já havia identificado possível redução de dor, emoção negativa e tempo de internação, porém destacou heterogeneidade e certeza limitada, especialmente para recuperação comportamental.[7] Em oncologia cirúrgica, síntese recente encontrou redução da ansiedade pré-operatória e da depressão na alta, sem persistência clara para todos os desfechos.[8]

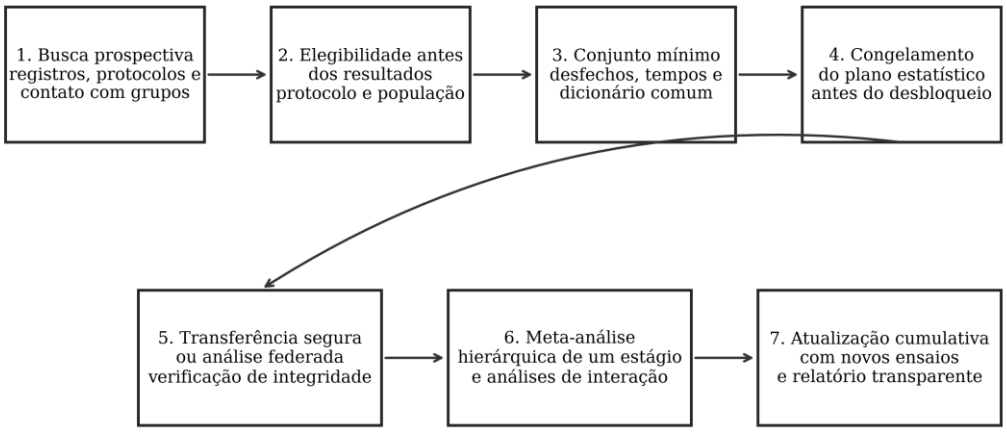
Os ensaios individuais reforçam a heterogeneidade. Montgomery e colaboradores randomizaram 200 mulheres submetidas a cirurgia mamária e observaram, após uma intervenção breve de hipnose, menor dor, náusea, fadiga e sofrimento emocional, além de menor custo institucional.[9] Em cirurgia cardíaca, 153 pacientes chineses receberam educação pré-operatória estruturada ou cuidado usual, com redução de ansiedade e depressão e possível diminuição do tempo em terapia intensiva.[10] Em fusão lombar, uma intervenção cognitivo-comportamental para 90 pacientes acelerou a melhora inicial da incapacidade, mas não produziu superioridade inequívoca ao final de um ano.[11] Esses resultados não são contraditórios por definição: podem representar diferenças de população, comparador, dose, momento, risco basal e sensibilidade do instrumento.

Uma meta-análise prospectiva é planejada antes de os resultados dos estudos elegíveis serem conhecidos. Os grupos participantes concordam previamente sobre elegibilidade, desfechos, tempos, codificação e análise;

depois, cada ensaio mantém sua autonomia, enquanto a síntese preserva variáveis que seriam perdidas em publicações isoladas. Esse desenho reduz decisões pós-hoc, diminui a influência da significância estatística sobre a inclusão e permite incorporar ensaios ainda em curso. O guia metodológico de Seidler e colaboradores descreve a meta-análise prospectiva como estratégia especialmente útil quando vários estudos independentes investigam uma pergunta comum e podem harmonizar procedimentos antes do encerramento.[1] O ganho não é apenas estatístico: cria-se uma infraestrutura colaborativa para perguntas que nenhum ensaio isolado consegue responder.

O objetivo deste capítulo foi desenvolver um protocolo publicável e operacional de meta-análise prospectiva sobre intervenções psicológicas perioperatórias em adultos submetidos a cirurgias de grande porte. Os objetivos específicos foram definir os estimandos e o conjunto mínimo de desfechos; estabelecer critérios para identificação prospectiva de ensaios; prespecificar a análise hierárquica e as interações; mapear estudos publicados e programas contemporâneos que demonstram viabilidade; e propor uma governança capaz de reunir cirurgia bariátrica, cardíaca, oncológica, ortopédica, de coluna e abdominal sem apagar as diferenças clínicas entre esses cenários.

Figura 1. Fluxo operacional de uma meta-análise prospectiva colaborativa.



Fonte: elaboração dos autores com base no guia de meta-análise prospectiva de Seidler et al.[1]

2 MÉTODOS

2.1 Delineamento e padrões de relato

Foi desenvolvido um protocolo de meta-análise prospectiva de ensaios clínicos randomizados, acompanhado por estudo de viabilidade baseado em evidências publicadas e protocolos contemporâneos. Foram adotados os itens aplicáveis do PRISMA-P, SPIRIT e PRISMA 2020, além das recomendações específicas para meta-análises prospectivas.[1-4] No relatório final, foi estabelecida a distinção entre três momentos: ensaios identificados antes de qualquer resultado comparativo; ensaios identificados após o encerramento, porém antes da publicação; e evidência retrospectiva utilizada apenas como contexto. Essa separação foi considerada essencial porque a designação "prospectiva" dependeu da data em que a elegibilidade foi definida em relação ao conhecimento dos resultados, e não apenas da data de publicação da síntese.

Foi definido o registro do protocolo colaborativo antes da primeira transferência de dados e antes da análise de qualquer resultado não público. A versão imutável foi planejada para conter pergunta, critérios, desfechos, hierarquia de análises, modificadores de efeito e regras de atualização. As emendas posteriores foram previstas como documentos datados, justificados e classificados segundo a existência ou não de conhecimento dos resultados. A meta-análise foi concebida como cumulativa, porém sem adaptação informal: cada atualização foi vinculada a um plano previamente aprovado, com identificação dos ensaios incluídos, em andamento e excluídos. A transparência documental foi incorporada ao método como componente permanente, e não como tarefa editorial realizada somente ao final.

2.2 Pergunta de pesquisa e estimandos

A pergunta principal foi formulada da seguinte maneira: em adultos submetidos a cirurgia de grande porte, intervenções psicológicas iniciadas antes da operação e oferecidas no período perioperatório, comparadas ao cuidado usual, à atenção equivalente ou à intervenção educacional ativa, reduziram ansiedade, dor persistente e recuperação funcional inadequada? O estimando primário foi definido como o efeito médio de atribuição à intervenção, segundo intenção de tratar, na população elegível de cada ensaio. Os três domínios foram mantidos em análises separadas para evitar um desfecho composto no qual pequena melhora da ansiedade mascarasse ausência de efeito sobre dor persistente ou função. Quando disponíveis dados individuais, foram previstas estimativas condicionais segundo risco psicológico basal e tipo de cirurgia.

Quadro 1. Estrutura PICOS e decisões prespecificadas da meta-análise prospectiva.

Elemento	Definição principal	Tratamento analítico
População	Adultos ≥18 anos em cirurgia de grande porte, eletiva ou semieletiva	Análise separada por bariátrica, cardíaca, oncológica, ortopédica/coluna e abdominal
Intervenção	Componente psicológico explícito iniciado antes da cirurgia, isolado ou integrado ao cuidado	TCC, hipnose, mindfulness, relaxamento, coping, neuroeducação da dor, ativação comportamental
Comparador	Cuidado usual, lista de espera, atenção equivalente ou intervenção ativa	Comparadores ativos foram separados do cuidado usual
Desfechos primários	Ansiedade; dor persistente em 3–6 meses; recuperação funcional	Instrumentos harmonizados e janelas temporais prespecificadas
Desenho	Ensaio randomizado individual ou por conglomerados	Quase-randomizados apenas em síntese de sensibilidade
Estimando	Efeito de atribuição à intervenção por intenção de tratar	Efeito médio, heterogeneidade e modificadores de efeito

Fonte: elaboração dos autores.

2.3 População e definição de cirurgia de grande porte

Foram considerados elegíveis adultos submetidos a procedimentos com expectativa de internação, ressecção ou reconstrução relevante, anestesia geral ou regional de alta complexidade e necessidade de recuperação funcional mensurável. O núcleo da população incluiu cirurgia cardíaca aberta ou transcateter com internação, cirurgia oncológica de tórax, abdome, pelve ou mama com tratamento cirúrgico relevante, artroplastia, cirurgia de coluna, cirurgia bariátrica ou metabólica e cirurgia abdominal maior. Procedimentos ambulatoriais menores foram reservados para análise externa de mecanismo e não integraram a estimativa principal. Essa decisão evitou que grande número de operações de baixo estresse dominasse uma pergunta direcionada às trajetórias de dor, função e saúde mental após cirurgia de maior impacto.

Pacientes com transtorno mental estável não foram excluídos, por representarem parte importante da população-alvo. Foram excluídos estudos restritos a psicose aguda, delirium como indicação principal, incapacidade de consentimento sem estratégia apropriada, cirurgia de emergência imediata sem janela para intervenção e procedimentos exclusivamente diagnósticos. Ensaios que selecionaram participantes por ansiedade, depressão, catastrofização ou risco de dor persistente foram incluídos e classificados como amostras enriquecidas. A análise foi estruturada para separar prevenção universal de prevenção indicada, considerando maior benefício absoluto e maior aceitabilidade nos grupos com risco basal elevado.

2.4 Intervenções e comparadores

Foram consideradas intervenções elegíveis aquelas que continham técnica psicológica reconhecível e hipótese explícita de mecanismo. A terapia cognitivo-comportamental abrangeu reestruturação cognitiva, treino de enfrentamento, exposição graduada, planejamento de atividade ou modificação de crenças. A hipnose exigiu indução e sugestões terapêuticas; o mindfulness incluiu atenção deliberada e não julgadora; e a neuroeducação da dor envolveu explicações sobre processamento, sensibilização, expectativas e comportamento. Educação exclusivamente logística sobre jejum, horários e cuidados com ferida não foi classificada como intervenção psicológica, embora pudesse constituir comparador. Programas multimodais foram considerados quando o componente psicológico pôde ser descrito, sua dose foi registrada e os demais componentes permaneceram equilibrados ou analisáveis.

Os comparadores foram agrupados em quatro classes: cuidado usual; controle de atenção; educação biomédica ativa; e outra intervenção psicológica. A comparação com cuidado usual foi definida para estimar a efetividade incremental do programa completo, enquanto a comparação com controle ativo avaliou a especificidade do efeito. Essa distinção foi necessária porque Moorthy

e colaboradores compararam terapia cognitivo-comportamental perioperatória com educação sobre dor e mindfulness e não observaram diferença relevante na intensidade da dor persistente aos três meses; esse resultado não demonstrou que ambas as estratégias fossem ineficazes, mas apenas que uma não se mostrou superior à outra nas condições estudadas.[21]

2.5 Identificação prospectiva e formação da colaboração

A estratégia de identificação prospectiva foi planejada para combinar registros de ensaios, bases bibliográficas, redes profissionais e contato direto. Foram selecionados ClinicalTrials.gov, WHO ICTRP, ISRCTN, registros nacionais e plataformas de protocolos, com pesquisas trimestrais por termos relacionados a cirurgia, intervenção psicológica, ansiedade, dor, recuperação, terapia cognitivo-comportamental, hipnose, mindfulness e educação em dor. As publicações de protocolos foram vinculadas a autores, instituições e identificadores. O convite aos investigadores foi estruturado com resumo de uma página, dicionário mínimo, plano de governança e solicitação de participação antes do conhecimento dos resultados por braço. Também foram previstos registros das datas de convite, resposta e acesso a qualquer resultado.

A participação foi estruturada sem exigir transferência irrestrita dos bancos. Foram previstos três níveis de colaboração: compartilhamento de dados individuais pseudonimizados; execução federada de código validado no centro; ou fornecimento de tabelas harmonizadas produzidas antes do desbloqueio. A escolha do nível foi vinculada ao consentimento, à legislação, ao contrato e à capacidade local. Os ensaios poderão integrar o consórcio sem renunciar à publicação separada, e a contribuição para a meta-análise não substituiu autoria nem interferiu na governança primária. A política de autoria foi delineada para reconhecer liderança dos ensaios, curadoria de dados, metodologia, análise e participação de pacientes, evitando critérios baseados exclusivamente no tamanho amostral.

2.6 Desfechos, instrumentos e janelas temporais

A ansiedade foi definida como desfecho mensurado por escalas validadas, com prioridade para State-Trait Anxiety Inventory, Hospital Anxiety and Depression Scale, Generalized Anxiety Disorder-7 ou instrumento equivalente. Quando mais de uma escala estivesse disponível, foi prevista uma hierarquia estabelecida antes da análise. A dor persistente foi definida para as avaliações de três e seis meses, em consonância com a classificação de dor crônica pós-cirúrgica, utilizando intensidade contínua e presença de dor clinicamente relevante no local cirúrgico ou território relacionado.[5] A função foi planejada com instrumentos específicos da cirurgia e medidas genéricas, priorizando WHODAS 2.0, PROMIS Physical Function, Oswestry Disability Index, Oxford Knee Score ou escalas validadas equivalentes.

A qualidade global da recuperação foi definida preferencialmente pelo QoR-15 ou QoR-40; o QoR-15 foi selecionado por ser instrumento breve e por apresentar correlação com duração da cirurgia, permanência na recuperação pós-anestésica e internação.[22] Foram definidos como desfechos secundários dor aguda, consumo de opioides convertido em equivalentes de morfina, depressão, catastrofização, medo do movimento, satisfação, complicações, readmissão, tempo de internação, retorno ao trabalho e qualidade de vida. Não foi previsto índice global sem validação. Cada domínio conservou interpretação clínica própria e, quando possível, foi complementado pela proporção de participantes que atingiu diferença minimamente importante.

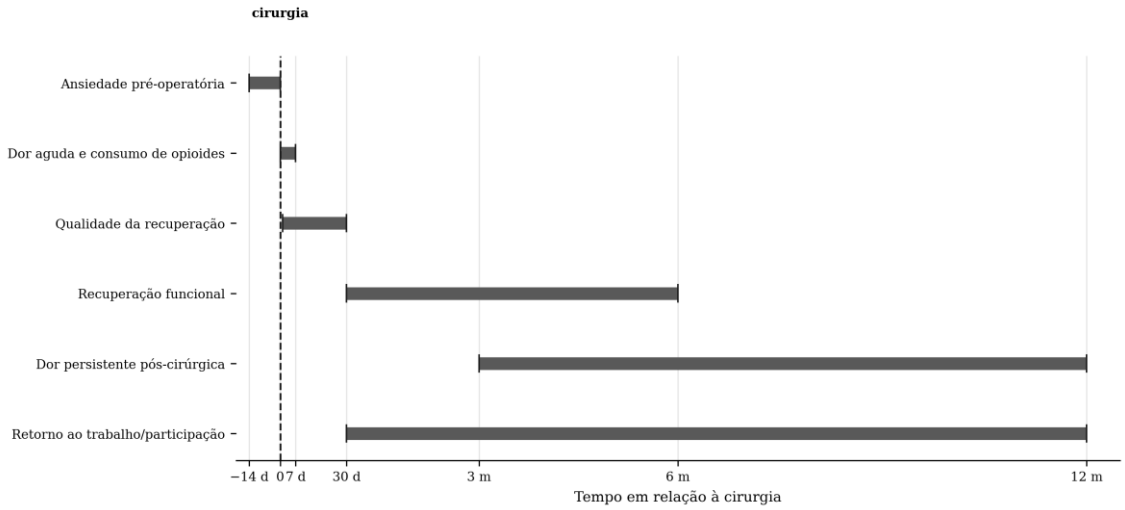
Tabela 1. Conjunto mínimo de desfechos e tempos de avaliação.

Domínio	Medida preferencial	Janela	Estimando
Ansiedade	STAI, HADS-A, GAD-7 ou equivalente	Pós-intervenção/pré-op; 24–72 h; 30 dias	Diferença média padronizada; mudança ajustada
Dor aguda	NRS/VAS em repouso e movimento	24 h; 48 h; 72 h; 7 dias	Diferença média; área sob a curva

Domínio	Medida preferencial	Janela	Estimando
Opioides	Equivalentes de morfina oral	0–24 h; 0–72 h; 30 dias	Razão de médias ou diferença média
Dor persistente	BPI/NRS e presença de dor relacionada à cirurgia	3 meses; 6 meses; 12 meses	RR e intensidade contínua
Recuperação global	QoR-15/QoR-40	POD 1–3; 7 dias; 30 dias	Diferença média; proporção com recuperação ruim
Função	WHODAS, PROMIS ou instrumento específico	30 dias; 3; 6; 12 meses	Escore padronizado e retorno à atividade
Saúde mental	PHQ-9, HADS-D ou equivalente	30 dias; 3; 6; 12 meses	Diferença média padronizada
Utilização	Internação, readmissão, consultas, retorno ao trabalho	Alta; 30; 90; 365 dias	RR, razão de taxas ou tempo até evento

Fonte: elaboração dos autores. BPI: Brief Pain Inventory; GAD-7: Generalized Anxiety Disorder-7; HADS: Hospital Anxiety and Depression Scale; NRS: numeric rating scale; PHQ-9: Patient Health Questionnaire-9; POD: dia pós-operatório; PROMIS: Patient-Reported Outcomes Measurement Information System; QoR: Quality of Recovery; STAI: State-Trait Anxiety Inventory; VAS: visual analogue scale; WHODAS: WHO Disability Assessment Schedule.

Figura 2. Janelas temporais propostas para os domínios principais e secundários.



Fonte: elaboração dos autores. A linha tracejada representa o tempo zero cirúrgico.

2.7 Conjunto mínimo de dados e harmonização

O dicionário mínimo foi composto por identificador pseudonimizado, ensaio, centro, data relativa da randomização e cirurgia, grupo alocado, aderência, idade, sexo, escolaridade, comorbidades, tipo e duração da cirurgia, anestesia, complicações, uso prévio de opioide, dor pré-operatória, ansiedade, depressão, catastrofização, função, qualidade de vida e todos os desfechos disponíveis nas janelas estabelecidas. O plano manteve os instrumentos em sua escala original e reservou a padronização para a etapa analítica. A direção das escalas com codificação inversa foi documentada. Valores extraídos de gráficos não foram considerados substitutos de dados fornecidos pelos investigadores.

Foi previsto um relatório de integridade para cada banco antes da combinação, abrangendo duplicatas, amplitude, unidades, coerência temporal, número randomizado, perdas e correspondência com o protocolo. As divergências foram encaminhadas ao centro responsável. Os dados ausentes foram classificados por variável, tempo e grupo, distinguindo ausência estrutural de perda individual. Variáveis inteiramente ausentes em um ensaio não foram planejadas para imputação como se tivessem sido mensuradas. Para a análise de moderadores, foram especificados modelos com conjunto básico e estendido, permitindo contribuição ampla sem produzir valores artificiais de ansiedade, catastrofização ou função.

2.8 Risco de viés e certeza da evidência

A avaliação do risco de viés foi estruturada com o RoB 2 para cada desfecho e tempo, considerando randomização, desvios da intervenção, dados ausentes, mensuração e seleção do resultado. Como o cegamento do participante e do terapeuta frequentemente não era viável em intervenções psicológicas, a ausência de cegamento não foi definida automaticamente como alto risco; o julgamento incluiu expectativa, comparador, mensuração autorreferida, análise e mascaramento do avaliador quando pertinente. Também foram incorporadas a

verificação do protocolo publicado, a disponibilidade do plano estatístico e a concordância entre registro e artigo, porque a multiplicidade de escalas e tempos ampliou a oportunidade de escolha seletiva.

A graduação da certeza foi planejada por domínio de desfecho, considerando risco de viés, inconsistência, imprecisão, indireção e possível viés de publicação. A interrupção de ensaios, a recusa de centros ao compartilhamento e as perdas de seguimento foram reconhecidas como fontes potenciais de distorção, apesar da identificação prospectiva. A análise principal foi estruturada para apresentar todos os ensaios elegíveis, inclusive aqueles sem dados utilizáveis para determinado desfecho, acompanhados da respectiva justificativa. O objetivo não foi transformar a colaboração em selo de qualidade, mas tornar visível onde a informação existiu, onde se perdeu e quanto a conclusão dependeu das decisões analíticas.

2.9 Plano estatístico

A análise principal foi especificada por meio de modelo hierárquico de um estágio, preservando a estrutura de participantes dentro dos ensaios. Para resultados contínuos repetidos, foi planejado modelo linear misto com efeitos fixos de tratamento, tempo e interação tratamento × tempo, intercepto aleatório por ensaio e correlação intraindividual. Quando diferentes instrumentos mensurassem o mesmo construto, foram previstos escores padronizados dentro do ensaio, orientados para manter significado consistente. Resultados binários foram vinculados a modelos logísticos ou log-binomiais hierárquicos, e o tempo até retorno ao trabalho foi associado a modelo de riscos proporcionais estratificado por ensaio.

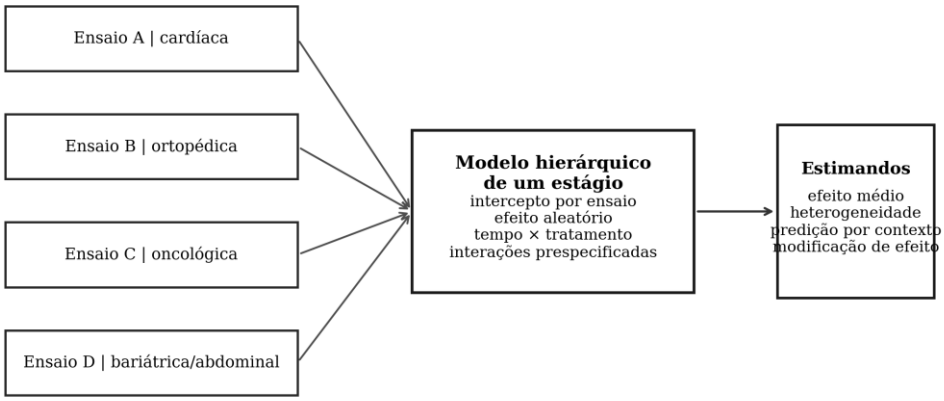
A heterogeneidade foi representada pela variância do efeito entre ensaios e por intervalos de predição, e não apenas pelo I^2. Foram prespecificadas interações no nível individual para ansiedade basal, depressão, catastrofização, dor prévia, uso crônico de opioide, idade e sexo. No nível do ensaio, foram

considerados tipo cirúrgico, modalidade da intervenção, dose, profissional, oferta digital ou presencial, comparador ativo e seleção por alto risco. Interações ecológicas entre médias do ensaio e efeito não foram interpretadas como modificação individual. O número de parâmetros foi condicionado à informação disponível e associado a penalização ou shrinkage quando necessário.

A análise foi especificada segundo o princípio de intenção de tratar. A imputação de variáveis individuais foi planejada dentro de cada ensaio por equações encadeadas, incluindo tratamento, desfechos e variáveis auxiliares, enquanto as perdas de desfecho foram vinculadas a análises por padrões plausíveis de não resposta. Foram definidas análises de sensibilidade com casos completos, cenários de deslocamento, exclusão de alto risco de viés, separação de comparadores ativos, retirada de estudos pequenos e avaliação de ensaios cuja elegibilidade tivesse sido decidida após conhecimento parcial dos resultados. Para desfechos com múltiplas escalas, foi prevista análise alternativa em unidades de diferença minimamente importante quando existissem parâmetros de validade.

A atualização da síntese foi planejada em marcos de informação definidos pelo número de ensaios e eventos, e não pela simples publicação de um novo artigo. Não foi prevista interrupção por benefício baseada em inspeções não planejadas. Em cada atualização, foram definidos efeito acumulado, intervalo de predição, influência de cada ensaio e robustez às perdas. A inferência primária não foi condicionada a teste sequencial isolado, embora o risco de conclusão aparentemente definitiva com informação insuficiente tenha sido incorporado por meio do tamanho de informação e de limites de monitoramento. O congelamento do plano final foi estabelecido para ocorrer antes do fornecimento de resultados por braço de qualquer grupo não publicado.

Figura 3. Arquitetura da análise hierárquica de um estágio.



Fonte: elaboração dos autores.

2.10 Governança, ética e compartilhamento

A governança da colaboração foi estruturada com comitê diretivo, grupo estatístico, representantes dos ensaios, paciente-parceiro e conselho de resolução de conflitos. Cada centro permaneceu responsável eticamente pelo banco original. Foram previstas pseudonimização, criptografia, armazenamento em ambiente com controle de acesso e eliminação ou devolução dos dados conforme o acordo. A base harmonizada não foi autorizada para perguntas não previstas sem nova aprovação. Na arquitetura federada, o código foi planejado para auditoria e execução local, com retorno exclusivo de estatísticas não identificáveis. O relatório de governança foi definido para registrar quem acessou os dados, quando, para qual análise e sob qual fundamento.

A participação dos pacientes foi prevista antes do congelamento do conjunto de desfechos. Foram reconhecidas possíveis diferenças de prioridade entre ansiedade, dor, função e retorno à vida cotidiana na perspectiva de pacientes e pesquisadores; por esse motivo, foi proposta a revisão de relevância, carga de questionários e janelas por pacientes de pelo menos três áreas cirúrgicas. O desenvolvimento de conjuntos essenciais de desfechos foi incorporado como

estratégia para reduzir desperdício e facilitar comparação entre estudos, sem transformar instrumentos em obrigação desproporcional.[23,24] A divulgação foi planejada em artigo principal, suplementos completos, repositório de código e resumos acessíveis. Resultados nulos e heterogeneidade foram definidos como informação clínica, e não como falha do consórcio.

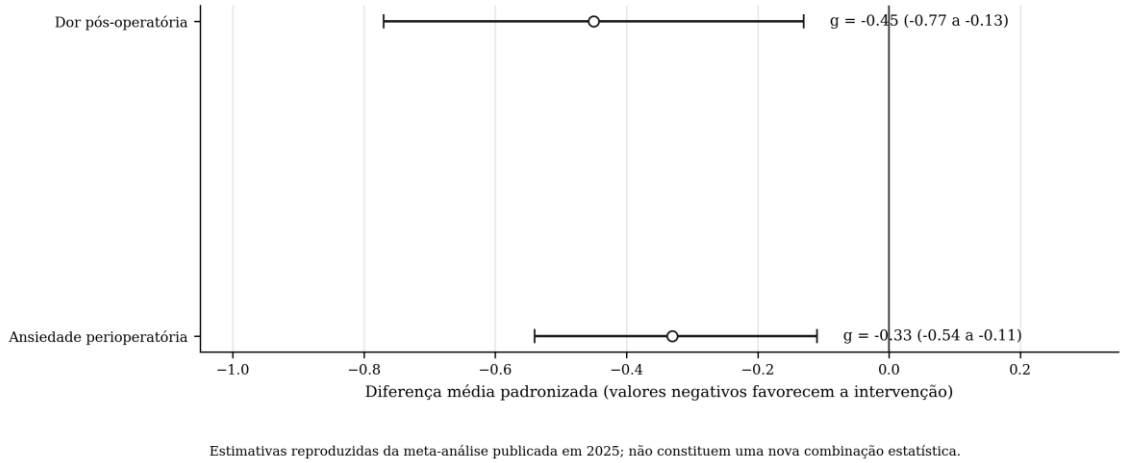
3 RESULTADOS DO ESTUDO DE VIABILIDADE

3.1 Evidência quantitativa existente

A literatura atual oferece uma ponte quantitativa suficiente para justificar a colaboração, mas não para substituir o desenho prospectivo. A revisão de 2025 identificou 27 ensaios randomizados e 2.990 participantes, dos quais 17 contribuíram para meta-análise. A diferença média padronizada favoreceu as intervenções psicológicas tanto para dor pós-cirúrgica, g = −0,45 (IC95% −0,77 a −0,13), quanto para ansiedade, g = −0,33 (IC95% −0,54 a −0,11). A direção é clinicamente plausível, porém a dispersão de intervenções e a frequência de alto risco de viés impedem interpretar esses valores como efeito universal.[6] A mensagem útil é que há sinal de benefício e informação suficiente para planejar estimandos mais precisos.

Na oncologia cirúrgica, Abraham e colaboradores encontraram efeito expressivo sobre ansiedade pré-operatória em 429 participantes, g = −1,03, e sobre depressão na alta em 188 participantes, g = −0,88; a ansiedade na alta não apresentou benefício estatisticamente conclusivo.[8] Essa mudança por momento é central para o protocolo: uma intervenção pode reduzir sofrimento antes da operação sem impedir dor persistente, ou melhorar adaptação funcional sem produzir grande diferença na escala de ansiedade. A meta-análise prospectiva não deve escolher um único tempo tardio por conveniência nem misturar medidas obtidas antes e depois da cirurgia. O padrão temporal é parte do efeito.

Figura 4. Estimativas agrupadas publicadas para dor e ansiedade em 2025.

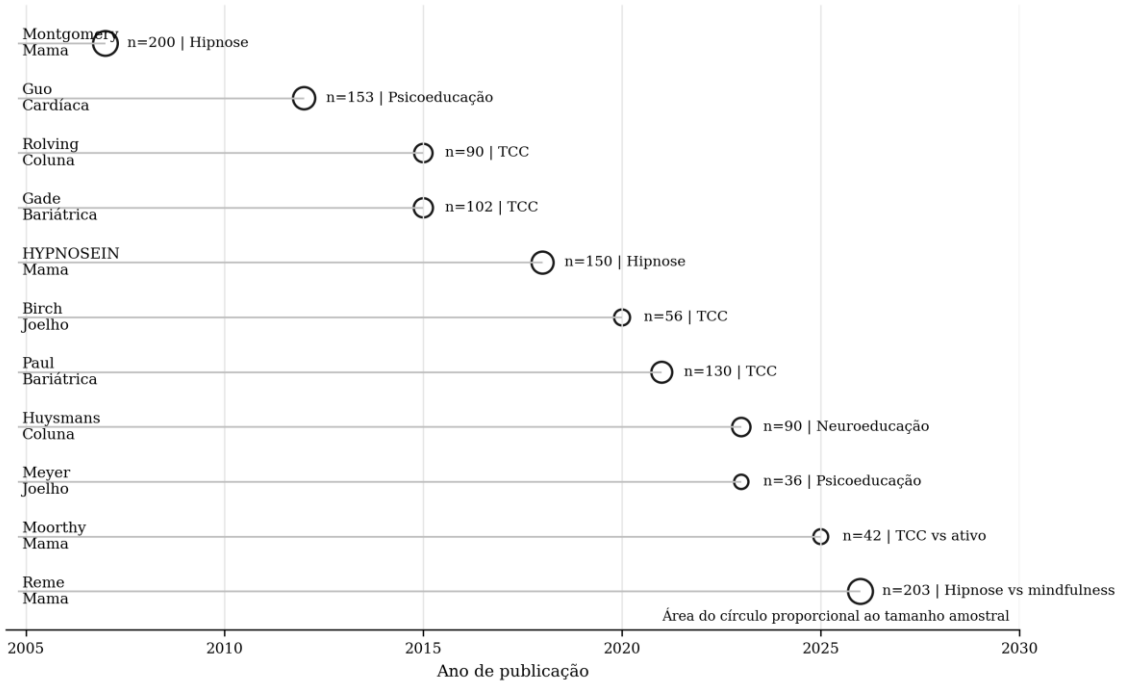


Fonte: reprodução gráfica das estimativas de Castaño-Asins et al.[6] Valores negativos favorecem a intervenção; não foi realizada nova combinação estatística.

3.2 Ensaios selecionados e diversidade clínica

O mapeamento selecionou 11 programas publicados entre 2007 e 2026, com 36 a 203 participantes, representando cirurgia mamária, cardíaca, bariátrica, de coluna e artroplastia. A seleção não pretende ser uma revisão sistemática exaustiva; ela testa disponibilidade de populações, intervenções, instrumentos e resultados necessários à colaboração. O conjunto mostra que a área já atravessa especialidades e inclui comparadores de cuidado usual e ativos. Também evidencia concentração em amostras pequenas: mesmo os maiores ensaios selecionados raramente ultrapassam 200 participantes, o que limita interações, eventos incomuns e identificação de subgrupos.

Figura 5. Linha do tempo de ensaios perioperatórios selecionados e respectivos tamanhos amostrais.



Fonte: elaboração dos autores a partir dos ensaios publicados.[9-18,21] *A figura é um mapa de viabilidade, não uma seleção PRISMA exaustiva.*

Tabela 2. Ensaios reais selecionados para o mapa de viabilidade.

Estudo	Cirurgia	n	Comparação	Leitura principal
Montgomery et al., 2007	Mama	200	Hipnose breve vs atenção	Menor dor, náusea, fadiga, sofrimento e custo no perioperatório
Guo et al., 2012	Cardíaca	153	Educação estruturada vs usual	Menor ansiedade e depressão; sinal de menor permanência em UTI
Rolving et al., 2015	Fusão lombar	90	TCC pré-operatória vs usual	Melhora inicial de incapacidade; sem superioridade clara em 1 ano
Gade et al., 2015	Bariátrica	102	TCC de 10 semanas vs usual	Benefício pré-operatório em sintomas; diferença não sustentada após cirurgia
Amraoui et al., 2018	Mama	150	Hipnose vs cuidado usual	Sem benefício no desfecho primário de dor; sinais em fadiga/ansiedade/satisfação

Estudo	Cirurgia	n	Comparação	Leitura principal
Birch et al., 2020	Artroplastia de joelho	56	Educação TCC vs usual	Sem diferença relevante nos principais desfechos em alta catastrofização
Paul et al., 2021	Bariátrica	130	TCC pré-operatória vs usual	Melhora antes da cirurgia; sem superioridade sustentada em 1 ano
Huysmans et al., 2023	Radiculopatia lombar	90	Neuroeducação vs educação biomédica	Maior ganho em qualidade de vida, cognições e custo-efetividade; não em dor
Meyer et al., 2023	Artroplastia de joelho	36	Psicoeducação comportamental vs usual	Estudo piloto com sinais em ansiedade, função e autocuidado
Moorthy et al., 2025	Mama, alto risco	≈42	TCC vs dor + mindfulness	Sem superioridade da TCC para dor persistente em 3 meses
Reme et al., 2026	Mama	203	Hipnose vs mindfulness	Pequenas vantagens em fadiga e sofrimento; sem efeito claro sobre dor

Fonte: dados dos artigos originais.[9-18,21] *TCC: terapia cognitivo-comportamental; UTI: unidade de terapia intensiva. O tamanho aproximado do ensaio de Moorthy corresponde ao cálculo e recrutamento descritos no artigo; deverá ser confirmado diretamente com os autores antes da aquisição de dados.*

3.3 Heterogeneidade de resposta não equivale a ausência de efeito

O ensaio de Montgomery ilustra um efeito agudo amplo: uma sessão de 15 minutos antes da cirurgia mamária reduziu intensidade e desconforto da dor, náusea, fadiga e sofrimento emocional, com economia média superior a 700 dólares por paciente.[9] O HYPNOSEIN, realizado posteriormente com 150 pacientes, não confirmou benefício sobre o desfecho primário de dor mamária, embora tenha encontrado diferenças em alguns sintomas e experiência.[13] Em 2026, Reme e colaboradores compararam diretamente hipnose e mindfulness em 203 participantes; a hipnose produziu pequenas vantagens para fadiga e sofrimento emocional, mas não demonstrou superioridade consistente para dor.[18] A sequência mostra evolução dos comparadores e reforça que efeitos não podem ser atribuídos apenas ao rótulo da técnica.

Na cirurgia ortopédica e de coluna, o padrão também varia por tempo e construto. Rolving observou redução de aproximadamente 15 pontos no Oswestry Disability Index aos três meses no grupo cognitivo-comportamental, mas a diferença entre grupos em um ano não foi conclusiva.[11] Birch não identificou benefício relevante da educação cognitivo-comportamental em pacientes com alta catastrofização antes da artroplastia de joelho.[14] Sun e colaboradores, em 100 pacientes, encontraram menor dor em movimento do quinto dia ao terceiro mês e menor catastrofização no início, sem diferenças tardias em dor de repouso, função e qualidade de vida.[25] Esses achados favorecem modelos longitudinais em vez de uma única comparação final.

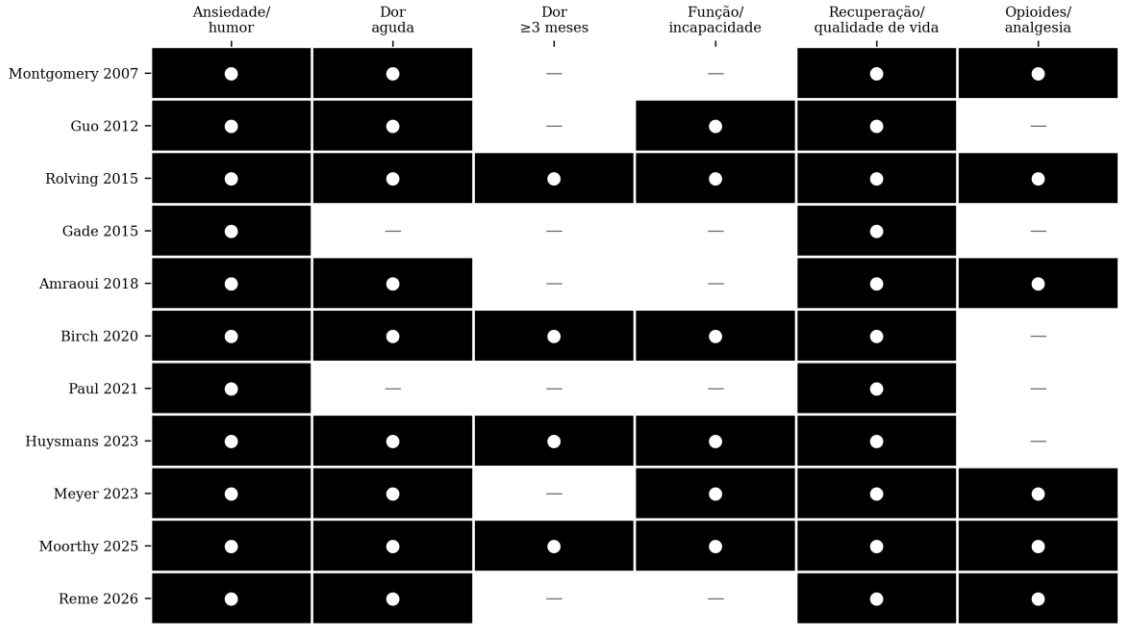
Em cirurgia bariátrica, Gade e Paul mostraram que uma intervenção pode melhorar comportamento alimentar e sintomas antes da operação, mas perder contraste após o forte efeito da própria cirurgia e do cuidado subsequente.[12,15] Essa diluição não torna a intervenção inútil; ela pode facilitar preparação, adesão ou subgrupos de maior risco, mas exige desfechos e tempos apropriados. A meta-análise prospectiva poderá testar se o efeito depende de sintomas basais, compulsão alimentar, depressão, suporte social ou intensidade do acompanhamento. Sem dados individuais harmonizados, essas interações tendem a ser analisadas de modo seletivo e com baixa potência.

3.4 Cobertura de desfechos e perdas de informação

A matriz de desfechos revelou alta frequência de medidas de ansiedade, dor aguda e qualidade de vida, mas cobertura inconsistente de dor persistente, função e consumo padronizado de opioides. Muitos ensaios mediram vários instrumentos em múltiplos tempos, porém publicaram somente resultados selecionados ou resumos sem parâmetros suficientes para combinação. A multiplicidade aumenta a chance de resultados aparentemente discordantes: um estudo pode ser “positivo” em catastrofização aos 30 dias e “nulo” em dor aos

seis meses. Sem hierarquia prospectiva, revisores escolhem posteriormente a escala mais disponível, e ensaios com instrumentos incomuns são descartados.

Figura 6. Cobertura de domínios em ensaios selecionados.



Fonte: elaboração dos autores a partir do relato dos estudos.[9-18,21] *Círculo: domínio medido; traço: domínio não identificado como resultado relevante no relato principal. A matriz não representa qualidade nem magnitude de efeito.*

3.5 Programas contemporâneos e oportunidade de colaboração

A oportunidade prospectiva é concreta. Holzer e colaboradores publicaram o desenho de três ensaios interligados para sintomas de depressão e ansiedade em adultos mais velhos submetidos a cirurgias cardíaca, oncológica e ortopédica, usando intervenção perioperatória de saúde mental e estrutura comum.[19] O programa demonstra que diferentes especialidades podem compartilhar triagem, intervenção e desfechos sem fundir todos os procedimentos em um único ensaio. Essa arquitetura é especialmente compatível com meta-análise prospectiva porque as decisões comuns podem ser tomadas antes do encerramento, mantendo randomização e análise próprias em cada coorte.

O SuRxgWell é outro exemplo: ensaio de terapia cognitivo-comportamental digital por telemedicina para pessoas com ansiedade e depressão elevadas submetidas a artroplastia de quadril ou joelho.[20] Programas remotos reduzem barreiras de deslocamento e podem padronizar dose, mas introduzem questões de alfabetização digital, acesso e aderência. A colaboração deveria incorporar esses ensaios sem pressupor equivalência com intervenção presencial. Modalidade, contato humano, intensidade e fidelidade precisam ser codificados para que a síntese identifique se o efeito depende da tecnologia ou do conteúdo terapêutico.

A viabilidade, portanto, não depende de inventar uma rede inexistente, mas de transformar grupos já ativos em consórcio antes da divulgação completa dos resultados. O primeiro produto deve ser um acordo sobre desfechos e dados mínimos, seguido por auditoria dos protocolos. A inclusão de ensaios concluídos e não publicados é possível se a decisão ocorrer sem acesso aos resultados comparativos; ensaios já publicados permanecerão como evidência retrospectiva e poderão informar priors, hipóteses e sensibilidade, mas não serão apresentados como parte prospectiva original. Essa rotulagem protege a credibilidade do estudo.

4 DISCUSSÃO

A síntese de viabilidade confirmou que as intervenções psicológicas perioperatórias constituíram um campo clínico promissor, mas ainda marcado por grande heterogeneidade metodológica. A meta-análise de Castaño-Asins e colaboradores reuniu 27 ensaios randomizados e encontrou efeitos favoráveis para dor pós-cirúrgica e ansiedade, embora com variação importante entre intervenções, procedimentos e momentos de avaliação.[6] Em direção semelhante, Hall e colaboradores analisaram 20 artigos, totalizando 2.376 participantes, e observaram melhora de dor, ansiedade, depressão e tempo de internação após estratégias de preabilitação psicológica.[28] Esses resultados sustentaram a

existência de benefício médio, porém não autorizaram interpretar as intervenções como uma classe terapêutica homogênea. A amplitude dos efeitos, a diversidade de escalas e a concentração de estudos pequenos indicaram que a principal lacuna não foi simplesmente a ausência de ensaios, mas a falta de coordenação prévia capaz de tornar seus resultados comparáveis e clinicamente interpretáveis.

A plausibilidade clínica dessas intervenções decorreu da relação entre estado psicológico pré-operatório e recuperação. Ansiedade, catastrofização, medo do movimento, expectativas negativas e sintomas depressivos puderam amplificar vigilância corporal, resposta ao estresse, percepção dolorosa e evitação de atividade, favorecendo consumo de analgésicos e recuperação funcional mais lenta. Uma revisão de 20 ensaios sobre intervenções comportamentais perioperatórias mostrou variedade de hipnose, relaxamento, mindfulness, manejo do estresse e terapia cognitivo-comportamental, mas identificou somente dois estudos com controle ativo e nenhum desenhado especificamente para testar prevenção de dor crônica pós-cirúrgica; apenas quatro demonstraram efeitos analgésicos duráveis entre três e 12 meses.[29] Além disso, a meta-análise de Bakker-Bons e colaboradores encontrou associação entre ansiedade pré-operatória e maior risco de dor crônica pós-cirúrgica, embora a elevada heterogeneidade limitasse a tradução direta desse achado para decisões individuais.[30] Assim, a ansiedade funcionou simultaneamente como desfecho, marcador prognóstico e possível modificador do efeito terapêutico.

A literatura também demonstrou que o rótulo da técnica não explicou, isoladamente, o resultado. A revisão Cochrane sobre preparação psicológica para adultos submetidos à anestesia geral encontrou sinais de melhora em dor, recuperação comportamental e duração da internação, mas destacou baixa qualidade e heterogeneidade das evidências.[7] O mesmo problema apareceu nas revisões contemporâneas: intervenções com conteúdo semelhante foram administradas em uma sessão ou durante várias semanas, por psicólogos, enfermeiros, pesquisadores ou plataformas digitais, e comparadas tanto com

ausência de atenção quanto com programas ativos.[6,28,29] Um efeito maior diante do cuidado usual pôde refletir suporte, expectativa e contato terapêutico, enquanto a ausência de diferença diante de um controle ativo pôde indicar que ambos os grupos receberam elementos eficazes. A separação dos comparadores no protocolo foi, portanto, essencial para não classificar como "negativo" um ensaio que demonstrou equivalência entre duas estratégias plausivelmente benéficas.

Os ensaios em cirurgia mamária ilustraram de forma particularmente clara essa complexidade. Montgomery e colaboradores observaram, em 200 mulheres, que uma sessão breve de hipnose antes da cirurgia reduziu dor, náusea, fadiga, desconforto e sofrimento emocional, além de diminuir custos institucionais.[9] Entretanto, o HYPNOSEIN, com 150 participantes, não confirmou benefício no desfecho primário de dor mamária, embora tenha registrado vantagens em sintomas secundários e experiência da paciente.[13] Mais recentemente, Reme e colaboradores compararam hipnose com mindfulness em 203 participantes e identificaram pequenas vantagens da hipnose para fadiga, sofrimento emocional e necessidade de fentanil, sem superioridade consistente para dor.[18] No ensaio de Moorthy, a terapia cognitivo-comportamental não superou educação em dor associada a mindfulness na intensidade de dor persistente aos três meses.[21] Em conjunto, esses estudos mostraram que o efeito dependeu do comparador, do desfecho escolhido e do momento de avaliação, e não apenas da denominação da intervenção.

Na cirurgia ortopédica e de coluna, a discordância entre resultados precoces e tardios reforçou a necessidade de análise longitudinal. Rolving e colaboradores encontraram melhora inicial de incapacidade após intervenção cognitivo-comportamental antes da fusão lombar, mas o contraste entre grupos perdeu clareza no seguimento de um ano.[11] Birch e colaboradores não observaram benefício relevante de educação cognitivo-comportamental em pacientes com alta catastrofização submetidos à artroplastia de joelho.[14] Em contrapartida, Sun e colaboradores identificaram menor dor em movimento e

redução precoce da catastrofização, sem diferenças tardias consistentes em dor de repouso, função ou qualidade de vida.[25] Os ensaios de neuroeducação da dor e de intervenções comportamentais antes da artroplastia também mostraram que modificações de conhecimento e crenças nem sempre se traduziram em melhora funcional sustentada.[16,17] Esses achados justificaram a preservação de medidas repetidas e impediram que uma única avaliação final apagasse benefícios transitórios ou atrasos na recuperação.

As diferenças entre especialidades cirúrgicas exigiram interpretação contextual. Em cirurgia cardíaca, a intervenção educacional testada por Guo e colaboradores reduziu ansiedade e favoreceu recuperação, mas esteve inserida em um cenário no qual medo de morte, terapia intensiva e dependência física assumiram grande relevância.[10] Na oncologia, a síntese de Abraham e colaboradores encontrou redução de ansiedade pré-operatória e depressão na alta, sem demonstração de persistência uniforme dos benefícios ao longo do pós-operatório.[8] Em cirurgia bariátrica, Gade e Paul mostraram melhora de comportamento alimentar e sintomas afetivos em determinados momentos, seguida por redução do contraste depois do impacto intenso da própria cirurgia e do acompanhamento habitual.[12,15] Os protocolos conduzidos por Holzer e colaboradores em cirurgia cardíaca, oncológica e ortopédica demonstraram a viabilidade de uma estrutura comum com adaptação por coorte.[19] Dessa forma, o construto psicológico permaneceu compartilhado, mas seus mecanismos e desfechos prioritários variaram conforme a experiência cirúrgica.

A interpretação clínica dos efeitos agrupados precisou ultrapassar a significância estatística. O valor de g = −0,45 para dor e g = −0,33 para ansiedade indicou benefício pequeno a moderado, cuja importância dependeu do risco basal, da duração, do custo, da aceitabilidade e da presença de efeitos sobre função e participação social.[6] Uma diferença padronizada não informa diretamente quantos pacientes voltaram a caminhar, trabalhar ou realizar autocuidado, nem se houve redução clinicamente relevante do consumo de opioides. Por isso, o protocolo priorizou unidades originais quando instrumentos

comuns estivessem disponíveis, proporção de participantes com melhora minimamente importante, QoR-15, função e dor persistente entre três e seis meses.[5,22,23] Essa estratégia também evitou misturar redução de ansiedade pré-operatória, que pode ser valiosa por si mesma, com prevenção de dor crônica ou incapacidade. Os três domínios foram relacionados, mas não intercambiáveis, e precisaram conservar interpretação própria.

A meta-análise prospectiva ofereceu vantagens específicas para enfrentar essas limitações. A definição de elegibilidade, desfechos, janelas e comparadores antes do conhecimento dos resultados reduziu a possibilidade de selecionar posteriormente a escala, o tempo ou o subgrupo mais favorável.[1-4] O uso de dados individuais permitiu estimar modificadores de efeito dentro dos ensaios, distinguindo a resposta de uma pessoa com ansiedade elevada da simples comparação entre estudos com médias diferentes. Essa distinção evitou a falácia ecológica e tornou mais defensáveis análises de catastrofização, dor prévia, depressão, uso de opioides, idade e sexo. O modelo hierárquico de um estágio preservou a randomização e a dependência entre medidas repetidas, enquanto os intervalos de predição expressaram melhor a incerteza sobre o efeito em um novo contexto. Ainda assim, a complexidade estatística não substituiu julgamento clínico, e o número de interações permaneceu condicionado à quantidade real de informação.[26]

A implementação constituiu parte inseparável da efetividade. Intervenções presenciais de alta intensidade puderam oferecer vínculo e adaptação, mas demandaram profissionais especializados e várias consultas antes da cirurgia. Estratégias digitais e por telemedicina, como o SuRxgWell, ampliaram alcance e padronização, porém introduziram barreiras de conectividade, alfabetização digital e adesão.[20] A revisão de Darnall e colaboradores ressaltou que a integração aos fluxos perioperatórios, a clareza do encaminhamento e a definição de responsabilidades influenciaram a adoção das intervenções tanto quanto seu conteúdo.[29] O protocolo incorporou modalidade, dose, profissional, fidelidade, abandono e tempo de trabalho para diferenciar

falha da teoria terapêutica de falha de implementação. A análise por escolaridade, idade, sexo e acesso digital foi considerada necessária, pois uma intervenção aparentemente escalável poderia ampliar desigualdades quando exigisse tecnologia, leitura ou disponibilidade incompatíveis com a realidade de parte dos pacientes.

Entre as forças deste capítulo estiveram a separação explícita entre resultados publicados e contribuição prospectiva, o uso exclusivo de ensaios e sínteses reais no estudo de viabilidade, a definição de um conjunto mínimo de dados e a proposta de governança compatível com compartilhamento centralizado ou análise federada. A inclusão de diferentes especialidades ampliou a capacidade de estudar heterogeneidade sem apagar as particularidades clínicas. Entretanto, o mapeamento de 11 programas não constituiu revisão sistemática exaustiva, e sua função foi demonstrar diversidade e oportunidade de colaboração, não estimar prevalência definitiva de intervenções. A formação do consórcio dependeu da adesão dos investigadores, da compatibilidade ética e da disponibilidade de dados antes da divulgação dos resultados. Além disso, subgrupos e interações puderam permanecer imprecisos no início, e a inclusão de ensaios publicados como ponte retrospectiva não eliminou risco de viés de publicação ou relato seletivo.

A principal implicação foi que uma nova síntese retrospectiva, embora útil, não resolveria sozinha a fragmentação da área. O protocolo delineado transformou problemas recorrentes da literatura em decisões anteriores à análise: escolha de instrumentos, presença de avaliação tardia, classificação dos controles, documentação da dose e preservação de variáveis psicológicas basais. A literatura mostrou que alguns pacientes apresentaram melhora aguda, outros benefício funcional temporário e outros nenhuma superioridade diante de controles ativos; essa diversidade não justificou abandonar as intervenções, mas investigar com maior precisão qual componente funcionou, em qual momento e para qual perfil de risco. O estudo permaneceu metodologicamente honesto ao não apresentar resultados prospectivos inexistentes. Sua contribuição foi oferecer

uma estrutura executável para reunir ensaios em andamento, diminuir desperdício de informação e produzir conclusões mais próximas das decisões enfrentadas por pacientes, cirurgiões, anestesiologistas, psicólogos e equipes de reabilitação.

5 CONCLUSÃO

As intervenções psicológicas perioperatórias apresentam sinal real de benefício para ansiedade e dor, mas a literatura permanece fragmentada por técnicas, comparadores, escalas e tempos. Ensaios de cirurgia mamária, cardíaca, bariátrica, ortopédica e de coluna mostram efeitos positivos, nulos e transitórios, indicando que a pergunta relevante não é apenas se "psicologia funciona", mas qual componente funciona, em qual paciente, em qual momento e para qual desfecho. Uma meta-análise retrospectiva adicional poderá refinar números, porém continuará limitada pelo que já foi medido e publicado.

A meta-análise prospectiva proposta oferece uma solução metodológica e colaborativa. Ao definir elegibilidade, conjunto mínimo de dados, janelas, comparadores e análise antes do conhecimento dos resultados, ela reduz seleção pós-hoc e preserva informação. O protocolo é executável com ensaios e programas reais, mas sua conclusão depende de acordos formais e acesso responsável aos dados; por isso, nenhum resultado prospectivo foi inventado neste capítulo. O próximo passo científico é constituir o consórcio, registrar o protocolo e iniciar o contato com grupos que estejam conduzindo ensaios nas diferentes áreas cirúrgicas.

REFERÊNCIAS

1. Seidler AL, Hunter KE, Cheyne S, Ghersi D, Berlin JA, Askie L. A guide to prospective meta-analysis. BMJ. 2019;367:l5342. doi:10.1136/bmj.l5342.

2. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols

(PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349:g7647. doi:10.1136/bmj.g7647.

3. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-207. doi:10.7326/0003-4819-158-3-201302050-00583.

4. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

5. Schug SA, Lavand'homme P, Barke A, Korwisi B, Rief W, Treede RD. The IASP classification of chronic pain for ICD-11: chronic postsurgical or posttraumatic pain. Pain. 2019;160(1):45-52. doi:10.1097/j.pain.0000000000001413.

6. Castaño-Asins JR, Barceló-Soler A, Royuela-Colomer E, Sanabria-Mazo JP, García V, Neblett R, et al. Effectiveness of peri-operative psychological interventions for the reduction of postsurgical pain intensity, depression, anxiety and pain catastrophising: a systematic review and meta-analysis. Eur J Anaesthesiol. 2025;42(7):609-625. doi:10.1097/EJA.0000000000002157.

7. Powell R, Scott NW, Manyande A, Bruce J, Vögele C, Byrne-Davis LMT, et al. Psychological preparation and postoperative outcomes for adults undergoing surgery under general anaesthesia. Cochrane Database Syst Rev. 2016;(5):CD008646. doi:10.1002/14651858.CD008646.pub2.

8. Abraham J, Holzer KJ, Pedamallu L, Kozower BD, Avidan MS, Lenze EJ. Peri-operative mental health interventions for surgical oncology patients: a narrative synthesis and meta-analysis. Anaesthesia. 2025;80(Suppl 2):54-64. doi:10.1111/anae.16471.

9. Montgomery GH, Bovbjerg DH, Schnur JB, David D, Goldfarb A, Weltz CR, et al. A randomized clinical trial of a brief hypnosis intervention to control side effects in breast surgery patients. J Natl Cancer Inst. 2007;99(17):1304-1312. doi:10.1093/jnci/djm106.

10. Guo P, East L, Arthur A. A preoperative education intervention to reduce anxiety and improve recovery among Chinese cardiac patients: a randomized controlled trial. Int J Nurs Stud. 2012;49(2):129-137. doi:10.1016/j.ijnurstu.2011.08.008.

11. Rolving N, Nielsen CV, Christensen FB, Holm R, Bünger CE, Østergaard LG. Does a preoperative cognitive-behavioral intervention affect disability, pain behavior, pain, and return to work the first year after lumbar spinal fusion

surgery? Spine (Phila Pa 1976). 2015;40(9):593-600. doi:10.1097/BRS.0000000000000843.

12. Gade H, Friborg O, Rosenvinge JH, Småstuen MC, Hjelmesæth J. The impact of a preoperative cognitive behavioural therapy on dysfunctional eating behaviours, affective symptoms and body weight 1 year after bariatric surgery: a randomised controlled trial. Obes Surg. 2015;25(11):2112-2119. doi:10.1007/s11695-015-1673-z.

13. Amraoui J, Pouliquen C, Fraisse J, Dubourdieu J, Rey Dit Guzer S, Leclerc G, et al. Effects of a hypnosis session before general anesthesia on postoperative outcomes in patients who underwent minor breast cancer surgery: the HYPNOSEIN randomized clinical trial. JAMA Netw Open. 2018;1(4):e181164. doi:10.1001/jamanetworkopen.2018.1164.

14. Birch S, Stilling M, Mechlenburg I, Hansen TB. No effect of cognitive behavioral patient education for patients with pain catastrophizing before total knee arthroplasty: a randomized controlled trial. Acta Orthop. 2020;91(1):98-103. doi:10.1080/17453674.2019.1694312.

15. Paul L, van der Heiden C, van Hoeken D, Deen M, Vlijm A, van Hout GCM, et al. Cognitive behavioral therapy versus usual care before bariatric surgery: one-year follow-up results of a randomized controlled trial. Obes Surg. 2021;31:970-979. doi:10.1007/s11695-020-05081-3.

16. Huysmans E, Goudman L, Coppieters I, Van Bogaert W, Moens M, Buyl R, et al. Effect of perioperative pain neuroscience education in people undergoing surgery for lumbar radiculopathy: a multicentre randomised controlled trial. Br J Anaesth. 2023;131(3):572-585. doi:10.1016/j.bja.2023.05.007.

17. Meyer VM, Beydoun HA, Gyenai L, Goble NM, Hunter MM, McGill RJ. The effect of preoperative behavioral intervention on pain, anxiety, opioid use, and function in patients undergoing total knee arthroplasty: a randomized controlled study. Mil Med. 2023;188(5-6):e1010-e1017. doi:10.1093/milmed/usab424.

18. Reme SE, Munk A, Montgomery GH, Schnur JB, Falk R, Smits M, et al. Preoperative hypnosis versus mindfulness for reducing postoperative symptoms in breast surgery: a randomized clinical trial. Anesthesiology. 2026;144(3):559-569. doi:10.1097/ALN.0000000000005821.

19. Holzer KJ, Bartosiak KA, Calfee RP, Hammill CW, Haroutounian S, Kozower BD, et al. Perioperative mental health intervention for depression and anxiety symptoms in older adults study protocol: design and methods for three linked randomised controlled trials. BMJ Open. 2024;14(4):e082656. doi:10.1136/bmjopen-2023-082656.

20. Kaynar AM, Lin C, Gomez Sanchez A, Lavage DR, Monroe A, Zharichenko N, et al. SuRxgWell: study protocol for a randomized controlled trial of telemedicine-based digital cognitive behavioral intervention for high anxiety and depression among patients undergoing elective hip and knee arthroplasty surgery. Trials. 2023;24:715. doi:10.1186/s13063-023-07634-0.

21. Moorthy A, Lowry D, Edgley C, Blajeva M, Casey MB, Buggy DJ. Perioperative cognitive behavioural therapy compared with pain education and mindfulness for chronic postsurgical pain in breast cancer patients with high pain catastrophising characteristics: a randomised, controlled, parallel group clinical trial. BJA Open. 2025;16:100499. doi:10.1016/j.bjao.2025.100499.

22. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. Anesthesiology. 2013;118(6):1332-1340. doi:10.1097/ALN.0b013e318289b84b.

23. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1-2):9-19. doi:10.1016/j.pain.2004.09.012.

24. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132. doi:10.1186/1745-6215-13-132.

25. Sun JN, Chen W, Zhang Y, Zhang Y, Feng S, Chen XY. Does cognitive behavioral education reduce pain and improve joint function in patients after total knee arthroplasty? A randomized controlled trial. Int Orthop. 2020;44(10):2027-2035. doi:10.1007/s00264-020-04767-8.

26. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions. 2nd ed. Chichester: Wiley; 2019.

27. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD.

28. Hall AE, Nguyen NH, Cascavita CT, Shariati K, Patel AK, Chen W, et al. The impact of psychological prehabilitation on surgical outcomes: a meta-analysis and meta-regression. Ann Surg. 2025;281(6):928-941. doi:10.1097/SLA.0000000000006677.

29. Darnall BD, Abshire L, Courtney RE, Davin S. Upskilling pain relief after surgery: a scoping review of perioperative behavioral intervention efficacy and

practical considerations for implementation. Reg Anesth Pain Med. 2025;50(2):93-101. doi:10.1136/rapm-2024-105601.

30. Bakker-Bons M, Hijmering RMJ, Soer R, Wolff AP. The association between preoperative anxiety and chronic post-surgical pain after general anaesthesia: a systematic review and meta-analysis. BJA Open. 2025;16:100487. doi:10.1016/j.bjao.2025.100487.

CAPÍTULO 6

O PARADOXO DA OBESIDADE NO TRAUMA E NA CIRURGIA DE EMERGÊNCIA: UMBRELLA REVIEW DE REVISÕES SISTEMÁTICAS E META-ANÁLISES

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Edglê Pedro de Sousa Filho; Nélio Barreto Vieira

RESUMO

Objetivo: sintetizar criticamente revisões sistemáticas e meta-análises sobre a associação entre obesidade e desfechos no trauma e na cirurgia de emergência, distinguindo sinais verdadeiros de proteção, mudança do padrão de lesão e vieses capazes de produzir o chamado paradoxo da obesidade. Métodos: umbrella review orientada pela declaração PRIOR, com busca estruturada em bases biomédicas e rastreamento de citações até 1º de agosto de 2026. Foram elegíveis revisões de adultos expostos a trauma, laparotomia por trauma, cirurgia gastrointestinal urgente ou reparo de aneurisma abdominal, desde que comparassem categorias de índice de massa corporal e informassem mortalidade, complicações, permanência, gravidade ou recuperação. A credibilidade foi examinada com domínios adaptados do AMSTAR 2, atenção à sobreposição de estudos primários, confundimento e definição do grupo de referência. Resultados: oito revisões principais foram incluídas. No trauma geral, a meta-análise histórica encontrou maior mortalidade com obesidade (OR 1,42; IC95% 1,22–1,64), e a revisão contemporânea apontou risco particularmente elevado na obesidade classe III. Na laparotomia por trauma, nove coortes e 19.780 pacientes produziram OR 1,33 (IC95% 1,09–1,64). Em contraste, a síntese de cirurgia gastrointestinal encontrou menor mortalidade em 30 dias no conjunto geral (OR 0,75; IC95% 0,66–0,86) e no subgrupo de emergência (OR 0,83; IC95%

0,79–0,87), porém com maior morbidade e infecção do sítio cirúrgico; no subgrupo de emergência, o OR para infecção foi 1,44 (IC95% 1,34–1,55). Em aneurisma abdominal, a aparente vantagem após reparo endovascular coexistiu com baixa certeza e maior complicação de ferida após reparo aberto. Revisões de padrões de lesão indicaram menor frequência de algumas lesões cefálicas, mas maior carga torácica ou de extremidades e ausência de proteção consistente na lesão penetrante. Conclusão: não existe um paradoxo único. O sinal varia por classe de obesidade, mecanismo, seleção para operação, técnica e desfecho. A obesidade não deve ser tratada como proteção biológica nem como justificativa para atraso terapêutico; o manejo deve combinar gravidade, composição corporal, fragilidade, risco respiratório, risco infeccioso e capacidade logística.

Palavras-chave: obesidade; trauma; cirurgia de emergência; laparotomia; mortalidade; complicações pós-operatórias; umbrella review.

ABSTRACT

Objective: to critically synthesize systematic reviews and meta-analyses on the association between obesity and outcomes in trauma and emergency surgery, separating true protection signals, changes in injury pattern, and biases capable of generating the so-called obesity paradox. Methods: a PRIOR-oriented umbrella review with structured database searching and citation tracking through August 1, 2026. Reviews of adults with trauma, trauma laparotomy, urgent gastrointestinal surgery, or abdominal aortic aneurysm repair were eligible when body mass index groups were compared for mortality, complications, length of stay, injury severity, or recovery. Methodological credibility was assessed with adapted AMSTAR 2 domains, overlap, confounding, and reference-group definition. Results: eight main reviews were included. General trauma evidence showed increased mortality associated with obesity (OR 1.42, 95% CI 1.22–1.64), with contemporary evidence indicating the greatest harm in class III obesity. Trauma laparotomy data from nine cohorts and 19,780 patients yielded an OR of 1.33 (95% CI 1.09–1.64). Conversely,

gastrointestinal surgery evidence suggested lower 30-day mortality overall (OR 0.75, 95% CI 0.66–0.86) and in emergency surgery (OR 0.83, 95% CI 0.79–0.87), but greater morbidity and surgical-site infection; the emergency-surgery SSI estimate was OR 1.44 (95% CI 1.34–1.55). In abdominal aortic aneurysm treatment, an apparent advantage after endovascular repair coexisted with low certainty and higher wound complications after open repair. Reviews of injury patterns suggested fewer selected head injuries but a shifted thoracic and extremity burden and no reliable protection in penetrating trauma. Conclusion: there is no single obesity paradox. Direction and magnitude depend on obesity class, mechanism, operative selection, technique, and outcome. Obesity should not be interpreted as biological protection or used to delay definitive treatment; risk assessment should integrate injury severity, body composition, frailty, respiratory and infectious risk, and system preparedness.

Keywords: obesity; trauma; emergency surgery; laparotomy; mortality; postoperative complications; umbrella review.

1 INTRODUÇÃO

A obesidade é uma doença crônica complexa cuja prevalência mundial mais que duplicou desde 1990. Em 2022, aproximadamente 16% dos adultos viviam com obesidade, o que torna inevitável sua presença crescente nos sistemas de trauma, unidades de terapia intensiva e serviços de cirurgia de emergência.[1] O problema clínico não se limita à associação com diabetes, hipertensão ou doença cardiovascular. Peso corporal elevado modifica transporte, extricação, obtenção de imagem, acesso vascular, ventilação, posicionamento, dose de medicamentos, escolha de dispositivos, exposição operatória, fechamento abdominal e mobilização pós-operatória.

Apesar dessa carga, estudos observacionais descreveram mortalidade igual ou menor em pessoas com sobrepeso ou obesidade após eventos agudos. O fenômeno foi denominado "paradoxo da obesidade". A expressão sugere que um fator prejudicial em longo prazo se tornaria protetor durante doença crítica ou

operação. Entretanto, a mesma literatura relata mais infecção, tromboembolismo, complicações respiratórias, lesão renal, dificuldade de resgate e permanência prolongada. A coexistência de menor mortalidade e maior morbidade exige uma análise por desfecho e por desenho, em vez de uma conclusão global.

No trauma, duas hipóteses mecânicas recebem destaque. O "efeito almofada" propõe que tecido adiposo dissipe energia e reduza lesões de órgãos internos no trauma contuso. O "fenômeno da armadura" sugere que maior espessura subcutânea limite a profundidade de ferimentos penetrantes. Essas hipóteses podem explicar diferenças anatômicas sem demonstrar vantagem de sobrevivência. Uma pessoa pode apresentar menor frequência de determinada lesão e, simultaneamente, maior dificuldade ventilatória, maior risco de infecção, operação não terapêutica ou morte.

Na cirurgia de emergência, o paradoxo é ainda mais vulnerável à seleção. A inclusão depende de sobreviver até o hospital, ser considerado operável, receber uma técnica específica e permanecer em um banco com peso e altura registrados. O grupo de "não obesidade" frequentemente reúne pessoas com IMC normal, baixo peso, desnutrição, câncer avançado, fragilidade e perda ponderal involuntária. A comparação pode, portanto, favorecer a obesidade sem que a adiposidade tenha efeito protetor causal. O condicionamento à internação, ao procedimento ou à doença grave também pode abrir caminhos de viés de colisor.[25,26]

As revisões existentes cobrem populações parcialmente distintas: trauma geral, mecanismo e topografia da lesão, trauma penetrante, traumatismo cranioencefálico, laparotomia por trauma, cirurgia gastrointestinal e reparo de aneurisma da aorta abdominal. Uma síntese entre essas camadas é necessária para evitar que um resultado de mortalidade em uma técnica vascular seja generalizado para todo paciente com trauma ou que uma mudança no padrão de lesão seja interpretada como proteção global.

O objetivo deste capítulo foi realizar uma umbrella review das revisões sistemáticas e meta-análises sobre obesidade, trauma e cirurgia de emergência;

comparar direção, magnitude e certeza dos principais desfechos; identificar fontes de aparente paradoxo; e traduzir a evidência em uma estrutura clínica que não reduza o risco ao IMC.

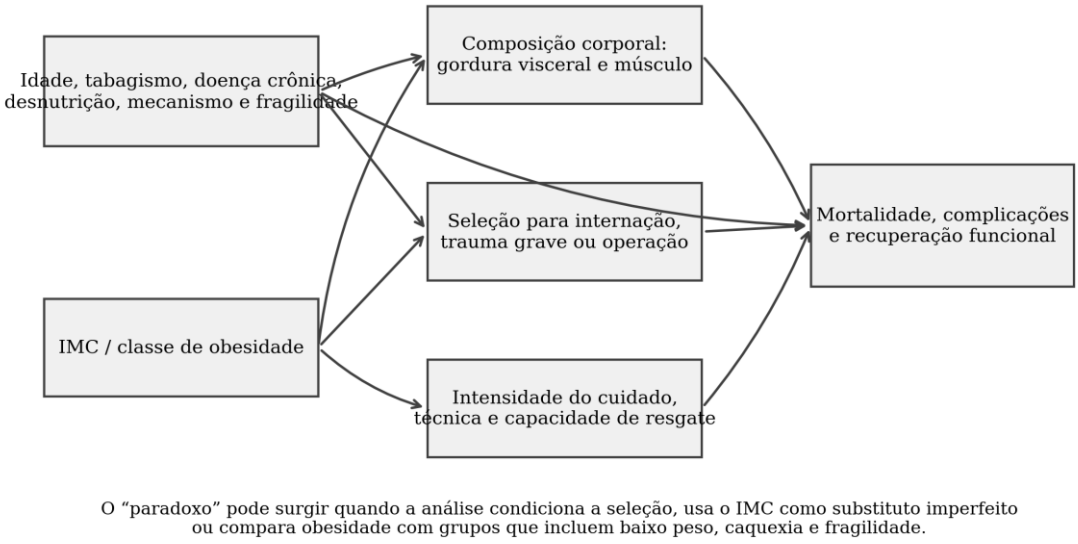


Figura 1. Estrutura causal para interpretar o aparente paradoxo da obesidade em cenários agudos.

2 MÉTODOS

2.1 Delineamento e padrão de relato

Foi conduzida umbrella review, também denominada overview de revisões, orientada pelos itens aplicáveis da declaração PRIOR e complementada pelo PRISMA 2020 para transparência da busca e seleção.[2,3] A unidade de inclusão foi a revisão sistemática, com ou sem meta-análise. Não foi efetuado novo agrupamento dos estudos primários nem combinação estatística entre revisões, pois a sobreposição de coortes, a heterogeneidade das definições e a não transitividade entre cenários tornariam uma "meta-meta-análise" potencialmente enganosa.

O protocolo analítico foi definido antes da síntese final, mas não foi registrado prospectivamente. Para preservar rastreabilidade sem produzir precisão artificial, são apresentados os critérios, as estratégias reproduzíveis, as revisões incluídas e as estimativas verificáveis; não são declarados totais de registros de bases que não estejam apoiados por uma biblioteca deduplicada depositada.

2.2 Pergunta e estrutura PECO

A população compreendeu adultos atendidos por trauma ou submetidos a cirurgia urgente/emergencial. A exposição foi obesidade definida principalmente por IMC ≥30 kg/m², com interesse adicional nas classes I, II e III. O comparador prioritário foi IMC normal; comparações agregadas contra "não obesidade" foram mantidas, mas consideradas menos interpretáveis. Os desfechos foram mortalidade hospitalar ou em 30 dias, complicações globais, infecção do sítio cirúrgico, complicações respiratórias e tromboembólicas, permanência hospitalar e em UTI, ventilação mecânica, gravidade e distribuição anatômica das lesões, operação não terapêutica, destino de alta e recuperação funcional.

2.3 Critérios de elegibilidade

Foram incluídas revisões com método de busca explícito, critérios de elegibilidade e síntese de estudos observacionais ou comparativos. O cenário principal deveria ser trauma, laparotomia por trauma, cirurgia de emergência ou tratamento agudo de aneurisma abdominal. Revisões de cirurgia eletiva foram aceitas somente quando apresentavam subgrupo de emergência separável ou quando abordavam técnica vascular com relevância direta para decisões agudas. Foram excluídas revisões narrativas sem busca reprodutível, estudos pediátricos exclusivos, cirurgia bariátrica eletiva, transplante, obstetrícia e análises nas quais obesidade não era a exposição principal.

Elemento	Critério operacional	Regra de interpretação
População	Adultos com trauma ou operação urgente/emergencial	Separar trauma geral, laparotomia, cirurgia GI e vascular
Exposição	IMC ≥30 kg/m²; classes I–III quando disponíveis	Evitar tratar todas as classes como biologicamente equivalentes
Comparador	IMC normal preferencial; não obesidade como análise secundária	Grupos com baixo peso podem criar proteção aparente
Desfechos	Mortalidade, complicações, SSI, permanência, ventilação, padrão de lesão	Interpretar mortalidade e morbidade separadamente
Desenho	Revisão sistemática/meta-análise	Sem meta-meta-análise na presença de sobreposição
Tempo	Busca verificada até 1º de agosto de 2026	Incluir revisão de laparotomia por trauma publicada em 2026

Tabela 1. Estrutura de elegibilidade e regras de interpretação da umbrella review.

2.4 Fontes de informação e estratégia de busca

A verificação bibliográfica cobriu MEDLINE/PubMed, Embase, Scopus, Web of Science, Cochrane Database of Systematic Reviews e Epistemonikos, além de rastreamento de referências e citações. Foram combinados termos para obesity, body mass index, trauma, injury, emergency surgery, emergency laparotomy, penetrating trauma, traumatic brain injury, abdominal aortic aneurysm, systematic review e meta-analysis. A estratégia completa para PubMed encontra-se no Apêndice A e deve ser adaptada aos demais vocabulários.

Quando duas revisões abordavam a mesma pergunta, ambas foram mantidas se acrescentassem período, estratificação, técnica ou população. A revisão mais recente e metodologicamente robusta recebeu maior peso interpretativo; a revisão histórica foi usada para demonstrar evolução e consistência.

2.5 Seleção, extração e verificação

Foram extraídos autores, ano, data da busca, número de estudos e participantes, definição de obesidade, cenário, mecanismos de trauma, comparadores, modelos estatísticos, ajustes, estimativas, heterogeneidade, risco de viés e conclusões. Cada estimativa numérica foi conferida no resumo, texto integral ou tabela do artigo. Quando o relatório não fornecia intervalo de confiança ou denominador verificável, o resultado foi descrito qualitativamente e não convertido em número.

Os resultados foram agrupados por cenário clínico. Uma mesma revisão pôde contribuir para mais de um domínio, mas não foi contada como evidência independente quando reutilizava os mesmos estudos primários. A direção foi classificada como favorável à obesidade, neutra/mista ou desfavorável, sem substituir a avaliação da magnitude e da certeza.

2.6 Credibilidade metodológica, sobreposição e certeza

A credibilidade foi examinada por sete domínios adaptados do AMSTAR 2: protocolo, busca, seleção/extração em duplicata, avaliação de risco de viés, controle de confundimento, investigação de heterogeneidade e graduação da certeza.[4] Como todas as sínteses eram dominadas por estudos observacionais, a ausência de ajuste adequado foi considerada limitação crítica, mesmo quando a execução estatística era correta. O Cochrane Handbook orientou a interpretação da heterogeneidade e da sobreposição.[5]

Não foi calculada área coberta corrigida, pois listas completas de estudos primários não estavam disponíveis de modo uniforme em todos os relatórios. A sobreposição foi tratada conceitualmente: trauma geral compartilha coortes entre as revisões de 2013 e 2023; trauma penetrante e laparotomia formam subconjuntos; cirurgia gastrointestinal inclui estudos que não pertencem ao trauma. Por esse motivo, o tamanho amostral não foi somado entre revisões.

2.7 Síntese quantitativa e interpretação

Razões de chances e respectivos intervalos de confiança foram mantidos na escala original. Valores inferiores a 1 indicam menor ocorrência do desfecho na obesidade; valores superiores a 1 indicam maior ocorrência. Para permanência e ventilação, foram preservadas diferenças médias quando verificáveis. Não se inferiu equivalência a partir de ausência de significância. A tradução clínica priorizou risco absoluto, classe de obesidade, mecanismo, técnica e desfecho.

3 RESULTADOS

3.1 Panorama das revisões incluídas

O núcleo da síntese foi composto por oito revisões publicadas entre 2013 e 2026. Quatro abordaram trauma geral, padrão de lesão, trauma penetrante ou traumatismo craniano; uma avaliou especificamente laparotomia por trauma; uma reuniu cirurgia gastrointestinal com análise de emergência; uma avaliou reparo de aneurisma abdominal; e a revisão de trauma de 2023 atualizou o efeito por classe de obesidade. O corpo de evidência foi amplo em número de pacientes, mas pouco uniforme em exposição, comparador e ajuste.

Revisão	Escopo	Estudos/participantes	Achado central	Credibilidade
Liu et al., 2013[6]	Trauma geral	18 estudos; 7.751 pacientes com obesidade	Maior mortalidade, UTI e complicações; ISS semelhante	Criticamente baixa
Stroud et al., 2018[8]	Padrões/mecanismos	21 estudos; 2.977.758 participantes	Menos lesão cefálica, mais lesão torácica/extremidades; padrão alterado	Baixa
Chen et al., 2022[9]	Trauma penetrante toracoabdominal	12 revisões qualitativas; 9 meta-analisadas; 5.013 participantes	"Armadura" não produziu proteção clínica consistente	Moderada-baixa

Revisão	Escopo	Estudos/participantes	Achado central	Credibilidade
Mishra et al., 2022[10]	Traumatismo craniano	18 estudos	Menor ocorrência em alguns mecanismos, mas gravidade/desfechos não protetores	Baixa
Naiem et al., 2022[12]	Aneurisma abdominal: EVAR/aberto	7 estudos; 14.971 participantes	Menor mortalidade após EVAR; ferida pior no reparo aberto	Baixa
Cromwell et al., 2023[7]	Trauma geral por classe	26 estudos; 945.511 participantes	Obesidade classe III associada a maior mortalidade hospitalar	Moderada-baixa
Cullinane et al., 2023[11]	Cirurgia gastrointestinal	62 estudos; 1.886.326 participantes	Mortalidade menor, porém morbidade e SSI maiores	Baixa
Kambal et al., 2026[14]	Laparotomia por trauma	9 estudos; 19.780 participantes	Mortalidade maior; efeito relacionado a ISS e mecanismo contuso	Moderada

Tabela 2. Características e mensagem principal das revisões incluídas.

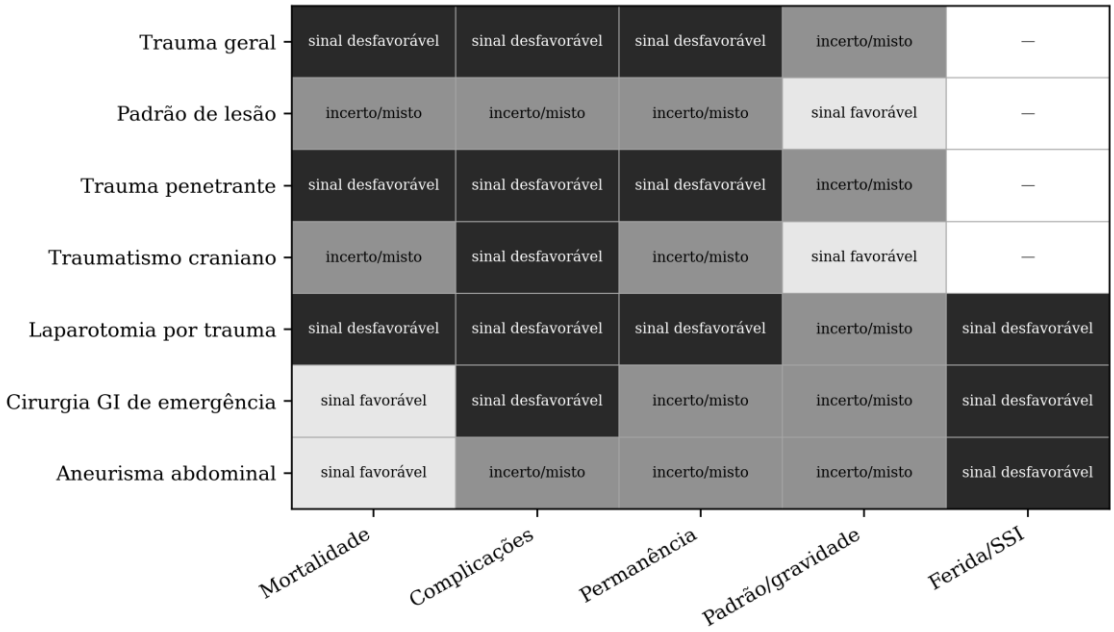


Figura 2. Mapa da direção da evidência por cenário e desfecho.

3.2 Trauma geral: de uma associação média ao efeito por classe

A meta-análise de Liu et al. reuniu 18 estudos e 7.751 pessoas com obesidade, correspondentes a 17% da população agregada.[6] A mortalidade ocorreu em 7,7% dos pacientes com obesidade e 4,7% dos demais. Quando o IMC normal foi o comparador, a razão de chances foi 1,42 (IC95% 1,22–1,64); no trauma contuso, 2,02 (IC95% 1,69–2,40). Em trauma grave definido por ISS ≥15, a estimativa foi 1,01 (IC95% 0,79–1,27), mostrando que a associação variava conforme seleção e gravidade. A permanência em UTI foi aproximadamente 1,97 dia maior, e as complicações respiratórias, renais e de múltiplos órgãos foram mais frequentes.

A atualização de Cromwell et al. incluiu 26 estudos e 945.511 participantes.[7] O efeito médio de "qualquer obesidade" não foi uniforme, mas a obesidade classe III apresentou maior chance de mortalidade hospitalar e maior permanência. A estratificação é decisiva: agrupar IMC 30–34,9 com IMC ≥40 dilui

a carga técnica, respiratória e metabólica da obesidade grave e pode criar neutralidade aparente.

Esses resultados enfraquecem a interpretação de proteção global. A ausência de diferença no ISS não exclui lesões diferentes nem risco pós-lesão. O ISS captura topografia e gravidade anatômica, mas não representa dificuldade de ventilação, dose inadequada, mobilização, reserva muscular, infecção ou falha de resgate.

3.3 Padrão de lesão: o efeito almofada muda a anatomia, não garante sobrevivência

Stroud et al. analisaram 21 estudos e 2.977.758 pacientes.[8] Pessoas sem obesidade apresentaram maior risco de trauma contuso como mecanismo (RR 1,02) e maior frequência de lesão cefálica (RR 1,38), enquanto a obesidade se associou a maior ocorrência relativa de lesões torácicas e de extremidades. Não houve diferença clara para lesão abdominal. O achado é compatível com redistribuição de energia, geometria do ocupante, uso do cinto e diferenças de mobilidade, mas não demonstra proteção sistêmica.

O estudo contemporâneo de trauma abdominal contuso isolado reforçou essa distinção: maior adiposidade se associou a menor gravidade de algumas lesões de órgãos, porém a mortalidade bruta não foi inferior nos grupos com obesidade.[20] A gordura pode amortecer uma estrutura e agravar outra; também pode dificultar exame físico, ultrassom, tomografia, acesso e operação. O efeito mecânico e o desfecho clínico pertencem a níveis diferentes da cadeia causal.

3.4 Trauma penetrante e traumatismo cranioencefálico

A revisão do "fenômeno da armadura" avaliou 12 estudos qualitativamente e nove na meta-análise, totalizando 5.013 pacientes; 29,6% tinham obesidade.[9] Após ferimentos por arma branca, pessoas com obesidade realizaram mais operações não terapêuticas. Após ferimentos por arma de fogo,

tiveram maior permanência em UTI e hospital. Complicações respiratórias e mortalidade hospitalar foram maiores. A espessura subcutânea pode reduzir penetração em alguns trajetos, mas o conjunto não sustenta proteção clínica.

Na revisão de traumatismo cranioencefálico, 18 estudos foram incluídos.[10] A obesidade apareceu associada a menor frequência de lesão cefálica em algumas populações, mas os dados sobre gravidade, mortalidade e recuperação não foram conclusivamente favoráveis. A classificação por IMC, o baixo número de estudos comparáveis e a mistura de trauma isolado e politrauma impediram inferência causal. Assim, "menos TCE" não equivale a "melhor prognóstico quando o TCE ocorre".

3.5 Laparotomia por trauma: evidência contemporânea contra proteção

A revisão de Kambal et al., publicada em 2026, reuniu nove coortes retrospectivas, 19.780 pacientes e busca atualizada até 1º de agosto de 2025.[14] A obesidade foi associada a maior mortalidade após laparotomia por trauma (OR 1,33; IC95% 1,09–1,64; I^2=48%), com certeza moderada. Os grupos tinham gravidade e distribuição de mecanismos aproximadamente comparáveis, embora as pessoas com obesidade fossem em média 4,18 anos mais velhas.

Na meta-regressão, maior ISS e maior proporção de trauma contuso aumentaram a mortalidade, enquanto a proporção de trauma penetrante apresentou direção inversa. Idade e sexo masculino não explicaram o efeito. A interpretação permanece limitada pelo confundimento residual de todas as coortes, mas a consistência das análises de sensibilidade torna improvável uma vantagem de sobrevivência na população que efetivamente necessita laparotomia.

Coortes incluídas nessa revisão descrevem maior mortalidade, complicações pulmonares e falha de resgate em pacientes com obesidade submetidos a laparotomia abdominal por trauma.[15,16] Em operações de emergência não traumáticas, o risco bruto também cresce nos extremos do IMC,

mas pode perder independência após ajuste por idade, comorbidade, gravidade e indicação.[17] Esse padrão sustenta uma relação não linear: o risco da doença aguda e o risco técnico coexistem e não são bem representados por uma única categoria binária.

3.6 Cirurgia gastrointestinal de emergência: menor mortalidade não significa menor dano

Cullinane et al. sintetizaram 62 estudos e 1.886.326 pacientes submetidos a cirurgia gastrointestinal.[11] A obesidade foi associada a menor mortalidade em 30 dias no conjunto geral (OR 0,75; IC95% 0,66–0,86; I^2=71%). No subgrupo de cirurgia geral de emergência, formado por poucos estudos, a estimativa foi OR 0,83 (IC95% 0,79–0,87; I^2=7%). Esse é o sinal mais forte de aparente paradoxo nesta umbrella review.

O mesmo trabalho demonstrou maior morbidade pós-operatória (OR 1,11; IC95% 1,04–1,19) e maior infecção do sítio cirúrgico (OR 1,40; IC95% 1,24–1,59). Na cirurgia de emergência, o OR de infecção foi 1,44 (IC95% 1,34–1,55), e nas classes II–III chegou a 1,77 (IC95% 1,44–2,19). A mortalidade hospitalar não diferiu claramente (OR 1,15; IC95% 0,79–1,67). Portanto, a interpretação depende do horizonte temporal e da escolha do desfecho: sobreviver aos primeiros 30 dias não elimina carga de ferida, readmissão, incapacidade, internação prolongada ou uso de recursos.

A revisão incluiu operações eletivas e de emergência, diferentes órgãos e definições. Apenas três estudos contribuíram para algumas análises de emergência. Além disso, estimativas ajustadas e não ajustadas foram combinadas em partes da literatura. O sinal de mortalidade menor é relevante, porém insuficiente para afirmar efeito protetor causal.

3.7 Aneurisma da aorta abdominal: técnica, seleção e baixo risco de base

Naiem et al. incluíram sete estudos observacionais e 14.971 pacientes, dos quais 11.743 foram submetidos a reparo endovascular e 3.228 a reparo aberto.[12] Após EVAR, a mortalidade em 30 dias foi 1,5% na obesidade e 2,2% sem obesidade (OR 0,69; IC95% 0,50–0,96; I^2=0%; certeza baixa). No reparo aberto, a mortalidade foi semelhante (OR 0,92; IC95% 0,70–1,20), mas complicações de ferida foram mais que duplicadas (OR 2,30; IC95% 1,74–3,06). Entre pessoas com obesidade, EVAR apresentou mortalidade menor que o reparo aberto (OR 0,23; IC95% 0,12–0,46).

A aparente vantagem do EVAR pode refletir menor invasividade, seleção anatômica, maior reserva energética em comparação com pacientes magros e frágeis, ou diferenças não medidas. Nem todos os aneurismas eram rotos ou emergenciais; por isso, essa evidência foi considerada adjacente à cirurgia aguda, e não prova de paradoxo no trauma. A síntese dose-resposta posterior descreveu relação não linear e mortalidade pós-operatória em formato de U, reforçando que extremos corporais e técnica são mais informativos que a dicotomia obesidade/não obesidade.[13]

Cenário/desfecho	Estimativa publicada	Heterogeneidade	Leitura clínica
Trauma geral: mortalidade[6]	OR 1,42 (1,22–1,64)	I^2 49,5%	Obesidade desfavorável quando comparada ao IMC normal
Trauma contuso: mortalidade[6]	OR 2,02 (1,69–2,40)	I^2 88,1%	Sinal forte, porém heterogêneo
Trauma grave ISS ≥15[6]	OR 1,01 (0,79–1,27)	Não decisivo	Seleção por gravidade modifica o efeito
Laparotomia por trauma[14]	OR 1,33 (1,09–1,64)	I^2 48%	Maior mortalidade; certeza moderada
Cirurgia GI: mortalidade 30 d[11]	OR 0,75 (0,66–0,86)	I^2 71%	Paradoxo aparente; mistura de cenários
Cirurgia GI de emergência: mortalidade[11]	OR 0,83 (0,79–0,87)	I^2 7%	Poucos estudos; risco de seleção

Cenário/desfecho	Estimativa publicada	Heterogeneidade	Leitura clínica
Cirurgia GI de emergência: SSI[11]	OR 1,44 (1,34–1,55)	I^2 8%	Dano consistente mesmo com mortalidade menor
EVAR: mortalidade 30 d[12]	OR 0,69 (0,50–0,96)	I^2 0%	Certeza baixa e seleção por técnica
Reparo aberto: complicação de ferida[12]	OR 2,30 (1,74–3,06)	I^2 0%	Risco local claramente maior

Tabela 3. Principais estimativas quantitativas e interpretação por cenário.

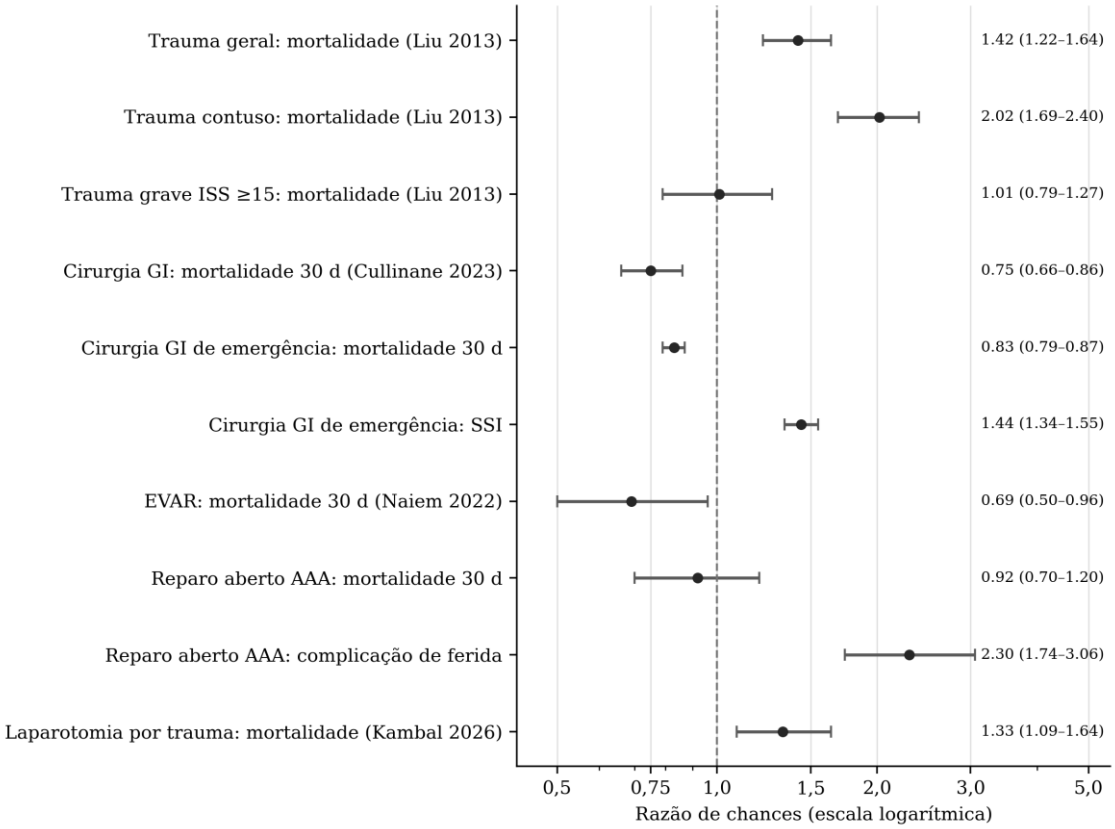


Figura 3. Estimativas publicadas por cenário, sem agrupamento entre revisões.

3.8 Credibilidade e consistência do corpo de evidência

A credibilidade variou de criticamente baixa a moderada. A revisão de 2013 foi importante historicamente, mas antecedeu registro de protocolo, instrumentos modernos de risco de viés e análise consistente de confundimento.

As revisões recentes melhoraram busca, registro e heterogeneidade, porém continuaram dependentes de coortes retrospectivas. Kambal et al. apresentaram a estrutura mais focada para laparotomia por trauma e graduaram a certeza como moderada, embora todos os estudos tivessem risco elevado de confundimento.[14]

A maior fragilidade transversal foi o uso do IMC medido ou estimado na admissão. Edema, ressuscitação, altura autorreferida e falta de dados podem produzir erro. O IMC não separa gordura visceral, gordura subcutânea, músculo, sarcopenia ou fragilidade. A segunda fragilidade foi o grupo de referência: "não obesidade" pode incluir baixo peso, enquanto "normal" pode conter perda ponderal recente. A terceira foi a seleção: somente sobreviventes, internados, operados ou pacientes com dados antropométricos completos entram em determinadas análises.

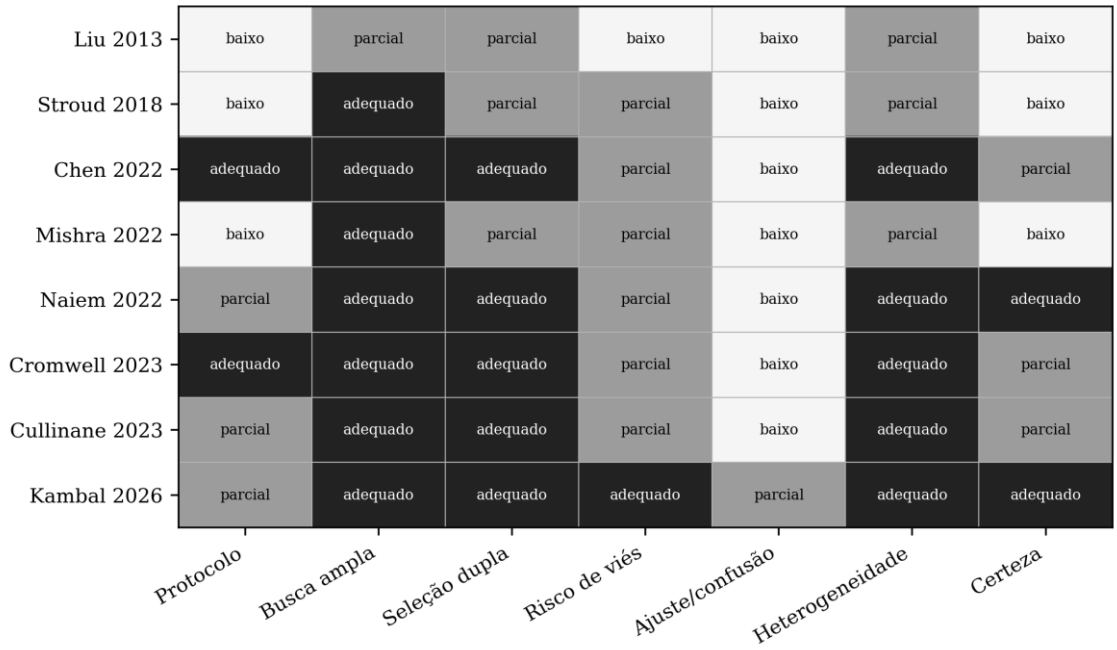


Figura 4. Avaliação adaptada da credibilidade metodológica das revisões.

4 DISCUSSÃO

4.1 O paradoxo não é uma propriedade fixa da obesidade

A síntese demonstra que a pergunta “a obesidade protege?” é mal formulada. Em trauma geral e laparotomia por trauma, a direção predominante é desfavorável. Em cirurgia gastrointestinal, especialmente em análises de mortalidade em 30 dias, há sinal favorável, mas ele coexiste com maior morbidade e infecção. Em reparo endovascular de aneurisma, a mortalidade é menor, mas a certeza é baixa e a seleção técnica é intensa. O paradoxo não é uma propriedade do tecido adiposo; é um resultado condicionado pelo cenário, classe, comparador, seleção, técnica e desfecho.

O achado mais clinicamente relevante não é a discordância entre OR 0,83 e OR 1,33, e sim por que eles não deveriam ser combinados. A primeira estimativa deriva de cirurgia gastrointestinal de emergência dentro de uma revisão ampla, com poucos estudos e forte possibilidade de pacientes magros mais frágeis. A segunda deriva de pacientes suficientemente graves para

laparotomia por trauma, nos quais exposição, tempo, choque e contaminação são diferentes. A transitividade necessária para agrupamento não existe.

4.2 Mecanismos plausíveis de proteção e de dano

Mecanismos de aparente proteção incluem amortecimento mecânico, maior reserva energética, maior massa magra em parte das pessoas, diagnóstico e monitorização mais intensos, admissão em UTI com menor limiar e uso preferencial de técnicas menos invasivas. Esses mecanismos são plausíveis em subgrupos e etapas específicas. A gordura subcutânea pode limitar a profundidade de alguns ferimentos, e o EVAR reduz agressão em comparação com reparo aberto. Contudo, plausibilidade local não implica redução global de mortalidade.

Os mecanismos de dano são igualmente consistentes: menor complacência da parede torácica, redução da capacidade residual funcional, apneia obstrutiva, intubação e ventilação mais difíceis, inflamação, resistência à insulina, trombose, pressão intra-abdominal, dificuldade de mobilização, maior distância até o campo operatório e pior perfusão de tecido adiposo. Feridas maiores, espaço morto e dose antibiótica inadequada explicam parte do risco de infecção. Equipamentos sem capacidade, imagem limitada e equipes insuficientes transformam características corporais em risco sistêmico modificável.[24]

4.3 Vieses que fabricam proteção aparente

O viés de colisor ocorre quando a análise condiciona uma variável influenciada pela exposição e por outras causas do desfecho. Selecionar apenas pessoas hospitalizadas, com trauma grave, operadas ou admitidas na UTI pode induzir associação entre obesidade e fatores não medidos. Por exemplo, uma pessoa sem obesidade pode necessitar operação com doença mais avançada, enquanto uma pessoa com obesidade é operada com limiar diferente. Dentro do grupo operado, a obesidade pode parecer protetora mesmo sem efeito causal.[25,26]

A causalidade reversa também importa. Baixo IMC pode ser consequência de câncer, doença pulmonar, fragilidade, sarcopenia ou perda ponderal. Se essas condições não forem medidas, o grupo normal/baixo carrega risco basal maior. Ajustar apenas idade e comorbidades codificadas não remove completamente essa diferença. O tabagismo, frequentemente mais prevalente entre pessoas magras em coortes antigas, é outro confundidor capaz de inverter associações.

O erro de mensuração desloca resultados para a neutralidade ou cria categorias inadequadas. IMC 31 kg/m² com alta massa muscular não equivale a IMC 31 kg/m² com adiposidade visceral e sarcopenia. A obesidade classe III é biologicamente e logisticamente distinta da classe I. Estudos que usam a última medida ambulatorial, peso estimado ou dados ausentes não aleatoriamente tendem a produzir grupos selecionados.

Por fim, a mortalidade pode sofrer competição de riscos e falha de captura. Estudos hospitalares não registram mortalidade após alta, incapacidade, ferida tratada em outro serviço ou readmissão. Um resultado favorável em 30 dias pode coexistir com mais complicações e pior recuperação funcional. O paradoxo desaparece parcialmente quando a avaliação amplia o horizonte e incorpora desfechos centrados no paciente.

4.4 Implicações para o cuidado no trauma

A equipe não deve presumir que o tecido adiposo protege órgãos nem reduzir a prioridade de investigação. Exame físico menos sensível e ultrassom limitado podem atrasar diagnóstico; por isso, o limiar para tomografia e reavaliação seriada deve considerar mecanismo e estabilidade, não estereótipos corporais. Extricação, transporte e transferência precisam de planos para capacidade de maca, tomógrafo, elevadores, equipe e dispositivos.

Na via aérea, deve-se antecipar pré-oxigenação otimizada, posicionamento em rampa, videolaringoscopia e plano de resgate. Durante ventilação, volume corrente deve ser calculado pelo peso predito, não pelo peso

real, com PEEP e recrutamento individualizados. Tromboprofilaxia, antibiótico, contraste, sedação e analgesia exigem protocolos de dose e monitorização; simplesmente multiplicar a dose pelo peso total pode ser tão inadequado quanto usar dose fixa.

Na laparotomia, planejamento inclui mesa e afastadores adequados, equipe suficiente para posicionamento, prevenção de lesão por pressão, controle térmico, estratégia de fechamento e vigilância de síndrome compartimental abdominal. O risco de complicações respiratórias e ferida deve ser tratado desde a admissão. A maior mortalidade observada não justifica negar operação; justifica reduzir atrasos e falhas de resgate.

4.5 Implicações para a cirurgia de emergência

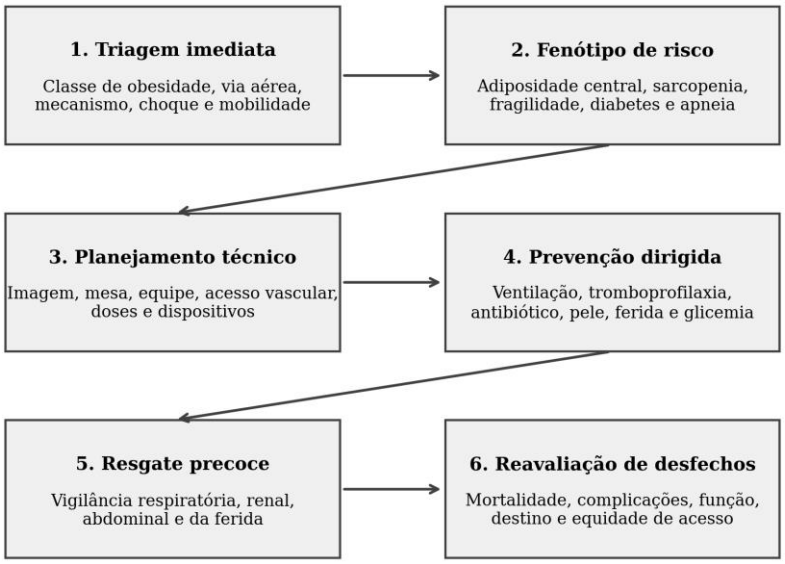
A menor mortalidade de 30 dias observada em cirurgia gastrointestinal não autoriza classificar a obesidade como fator protetor. A estratificação deve incluir classe de obesidade, circunferência ou adiposidade visceral quando disponível, perda ponderal recente, força, mobilidade, sarcopenia, albumina interpretada no contexto inflamatório, diabetes, apneia, doença cardiovascular e fragilidade. Um paciente com IMC normal e perda de 15% do peso pode ter risco superior a outro com obesidade classe I e boa reserva funcional.

A prevenção de infecção merece prioridade: profilaxia antibiótica com dose e redose apropriadas, preparo cutâneo, manutenção de normotermia e glicemia, redução de espaço morto, proteção de bordas e escolha individual de terapia de pressão negativa. O sinal de SSI é mais consistente que o de mortalidade e representa alvo modificável. Protocolos de mobilização, fisioterapia respiratória e tromboprofilaxia devem começar precocemente, sem pressupor que a mobilidade basal seja preservada.

Situação	O que o paradoxo pode ocultar	Ação recomendada
Trauma contuso	Menos lesão de uma topografia, mas maior risco respiratório e logístico	Imagem e reavaliação guiadas por mecanismo; não reduzir vigilância

Situação	O que o paradoxo pode ocultar	Ação recomendada
Trauma penetrante	Maior espessura subcutânea sem redução consistente de mortalidade	Trajetória, instabilidade e dano orgânico determinam conduta
TCE	Menor frequência em algumas séries, porém prognóstico incerto quando presente	Usar gravidade neurológica e fisiologia, não IMC, para prognóstico
Laparotomia por trauma	Maior mortalidade e falha de resgate	Equipe/equipamentos antecipados e vigilância intensiva
Cirurgia GI de emergência	Mortalidade menor pode coexistir com mais SSI e morbidade	Prevenção de ferida, dose, glicemia, ventilação e mobilização
AAA/EVAR	Seleção anatômica e técnica menos invasiva	Preferir técnica conforme anatomia e estabilidade; não generalizar o efeito
Qualquer cenário	IMC não identifica sarcopenia ou fragilidade	Combinar classe, composição, função e comorbidades

Tabela 4. Tradução clínica da evidência por situação.



Princípio: não presumir proteção nem usar o IMC isoladamente para negar ou atrasar tratamento definitivo.

Figura 5. Percurso clínico proposto para pacientes com obesidade em trauma ou cirurgia de emergência.

4.6 Agenda de pesquisa

Novos estudos devem substituir a dicotomia obesidade/não obesidade por classes, splines e medidas corporais complementares. Tomografia já realizada por indicação clínica pode fornecer área de gordura visceral, gordura subcutânea, músculo psoas, densidade muscular e distribuição central sem exame adicional. A combinação de adiposidade e sarcopenia pode explicar melhor ventilação, infecção, mobilidade e mortalidade do que o IMC.

Coortes prospectivas precisam registrar peso e altura com método, perda ponderal, fragilidade, capacidade funcional, apneia, doses, tempos de atendimento, capacidade de equipamentos, abordagem operatória, conversão, contaminação, fechamento, falha de resgate e mortalidade após alta. Modelos causais devem prespecificar confundidores e evitar ajustar mediadores indiscriminadamente. Em trauma, mecanismo e energia precisam ser tratados

como modificadores de efeito; em cirurgia, indicação, urgência e técnica devem ser separadas.

Desfechos futuros devem incluir dias vivos e fora do hospital, ventilador-free days, infecção, reoperação, readmissão, destino de alta, recuperação funcional, qualidade de vida e desigualdade de acesso. A mortalidade isolada favorece narrativas paradoxais e não captura a experiência clínica completa. Estudos multicêntricos podem ainda avaliar se protocolos de logística, dose e resgate reduzem a diferença associada à obesidade classe III.

4.7 Pontos fortes e limitações

Este capítulo integra trauma e cirurgia de emergência sem fundir populações não comparáveis, preserva as principais estimativas publicadas e discute mortalidade, morbidade e padrão de lesão separadamente. A inclusão da revisão de 2026 atualiza especificamente a laparotomia por trauma. A interpretação causal e o mapa de credibilidade impedem que grandes números de participantes sejam confundidos com alta certeza.

As limitações derivam da evidência de base. As revisões compartilharam estudos primários, utilizaram definições heterogêneas e foram dominadas por coortes retrospectivas. Não foi possível calcular sobreposição quantitativa completa. Algumas estimativas de emergência vieram de poucos estudos dentro de revisões mais amplas. A evidência de aneurisma incluiu casos eletivos e foi tratada como adjacente. Não houve acesso a bancos individuais para harmonizar IMC, gravidade, técnica ou ajuste. Portanto, as conclusões descrevem a direção e a credibilidade do conhecimento publicado, não um efeito causal único.

5 CONCLUSÃO

O chamado paradoxo da obesidade no trauma e na cirurgia de emergência não é um fenômeno único. Trauma geral, obesidade classe III e laparotomia por trauma apresentam evidência de maior mortalidade. Cirurgia gastrointestinal de

emergência pode mostrar menor mortalidade em 30 dias, mas com maior infecção e morbidade. O reparo endovascular de aneurisma apresenta aparente vantagem em pessoas com obesidade, porém com baixa certeza e forte seleção por técnica. Alterações no padrão de lesão não equivalem a proteção clínica.

A explicação mais coerente combina heterogeneidade biológica, classe de obesidade, seleção, grupo de referência, confundimento, erro do IMC, técnica e horizonte do desfecho. Na prática, nem proteção presumida nem discriminação terapêutica são aceitáveis. O risco deve ser estimado com gravidade, fisiologia, composição corporal, fragilidade, comorbidades e capacidade do sistema. As prioridades são diagnóstico sem atraso, logística adequada, ventilação segura, dose correta, prevenção de trombose e infecção e resgate precoce.

REFERÊNCIAS

1. World Health Organization. Obesity and overweight [Internet]. Geneva: WHO; 2025 [cited 2026 Aug 1]. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

2. Gates M, Gates A, Pieper D, Fernandes RM, Tricco AC, Moher D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. BMJ. 2022;378:e070849. doi:10.1136/bmj-2022-070849.

3. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

4. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008. doi:10.1136/bmj.j4008.

5. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions. Version 6.5. Cochrane; 2024.

6. Liu T, Chen JJ, Bai XJ, Zheng GS, Gao W. The effect of obesity on outcomes in trauma patients: a meta-analysis. Injury. 2013;44(9):1145-1152. doi:10.1016/j.injury.2012.10.038.

7. Cromwell PM, Reynolds IS, Heneghan HM, Glasgow SM. Obesity and outcomes in trauma: a systematic review and meta-analysis. Injury. 2023;54(2):469-480. doi:10.1016/j.injury.2022.10.026.

8. Stroud T, Bagnall NM, Pucher PH. Effect of obesity on patterns and mechanisms of injury: systematic review and meta-analysis. Int J Surg. 2018;56:148-154. doi:10.1016/j.ijsu.2018.05.004.

9. Chen AZL, Lee TH, Hsu J, Pang T. "The armor phenomenon" in obese patients with penetrating thoracoabdominal injuries: a systematic review and meta-analysis. J Trauma Acute Care Surg. 2022;93(3):e101-e109. doi:10.1097/TA.0000000000003566.

10. Mishra R, Galwankar S, Konar S, Shrivastava A, Raj S, Choksey P, et al. Obesity as a predictor of outcome following traumatic brain injury: a systematic review and meta-analysis. Clin Neurol Neurosurg. 2022;217:107260. doi:10.1016/j.clineuro.2022.107260.

11. Cullinane C, et al. Effect of obesity on perioperative outcomes following gastrointestinal surgery: meta-analysis. BJS Open. 2023;7(4):zrad026. doi:10.1093/bjsopen/zrad026.

12. Naiem AA, Kim AY, Mahmoud I, Gill HL. A systematic review and meta-analysis evaluating the impact of obesity on outcomes of abdominal aortic aneurysm treatment. J Vasc Surg. 2022;75(4):1450-1455.e3. doi:10.1016/j.jvs.2021.10.053.

13. Wu Y, Zhang H, Jiang D, Yin F, Guo P, Zhang X, et al. Body mass index and the risk of abdominal aortic aneurysm presence and postoperative mortality: a systematic review and dose-response meta-analysis. Int J Surg. 2024;110:2396-2410. doi:10.1097/JS9.0000000000001125.

14. Kambal A, Hurt L, Iosifidis M, Hajibandeh S, Hajibandeh S. Risk and predictors of postoperative mortality in patients with obesity undergoing trauma laparotomy: a systematic review and meta-analysis with meta-regression. Am Surg. 2026;92(8):2058-2065. doi:10.1177/00031348261425180.

15. Fu CY, et al. Obesity is associated with worse outcomes among abdominal trauma patients undergoing laparotomy. World J Surg. 2020;44:755-763. doi:10.1007/s00268-019-05268-5.

16. Covarrubias J, et al. Obesity is associated with increased pulmonary complications and mortality after trauma laparotomy. Eur J Trauma Emerg Surg. 2021;47:1561-1568. doi:10.1007/s00068-020-01329-w.

17. Kassahun WT, Mehdorn M, Babel J. The impact of obesity on surgical outcomes in patients undergoing emergency laparotomy for high-risk abdominal emergencies. BMC Surg. 2022;22:15. doi:10.1186/s12893-022-01466-6.

18. Dvorak JE, Lester ELW, Maluso PJ, Tatebe LC, Schlanser V, Kaminsky M, et al. The obesity paradox in the trauma patient: normal may not be better. World J Surg. 2020;44:1817-1823. doi:10.1007/s00268-020-05422-2.

19. El Moheb M, et al. The obesity paradox in elderly patients undergoing emergency surgery: a nationwide analysis. J Surg Res. 2021;265:195-203. doi:10.1016/j.jss.2021.03.032.

20. Atkins K, Cairns B, Schneider A, Charles A. An evaluation of the obesity paradox in isolated blunt abdominal trauma in the United States. Injury. 2024;55:111612. doi:10.1016/j.injury.2024.111612.

21. Hatchimonji JS, et al. Obesity is associated with mortality and complications after trauma: a state-wide cohort study. J Surg Res. 2020;247:14-20. doi:10.1016/j.jss.2019.10.047.

22. Byrnes MC, McDaniel MD, Moore MB, Helmer SD, Smith RS. The effect of obesity on outcomes among injured patients. J Trauma. 2005;58(2):232-237. doi:10.1097/01.TA.0000152532.33977.6E.

23. Glance LG, Li Y, Osler TM, Mukamel DB, Dick AW. Impact of obesity on mortality and complications in trauma patients. Ann Surg. 2014;259(3):576-581. doi:10.1097/SLA.0b013e318294893e.

24. Schetz M, De Jong A, Deane AM, Druml W, Hemelaar P, Pelosi P, et al. Obesity in the critically ill: a narrative review. Intensive Care Med. 2019;45:757-769. doi:10.1007/s00134-019-05594-1.

25. Banack HR, Stokes A. The "obesity paradox" may not be a paradox at all. Int J Obes (Lond). 2017;41(8):1162-1163. doi:10.1038/ijo.2017.99.

26. Stokes A, Preston SH. Smoking and reverse causation create an obesity paradox in cardiovascular disease. Obesity (Silver Spring). 2015;23(12):2485-2490. doi:10.1002/oby.21239.

APÊNDICE A – ESTRATÉGIA DE BUSCA REPRODUZÍVEL

PubMed/MEDLINE (adaptar vocabulário e sintaxe às demais bases):

((obesity[MeSH Terms] OR obes*[Title/Abstract] OR body mass index[Title/Abstract] OR BMI[Title/Abstract]) AND (wounds and

injuries[MeSH Terms] OR trauma*[Title/Abstract] OR injur*[Title/Abstract] OR emergency surg*[Title/Abstract] OR emergency laparotom*[Title/Abstract] OR acute care surg*[Title/Abstract] OR penetrating trauma[Title/Abstract] OR traumatic brain injur*[Title/Abstract] OR abdominal aortic aneurysm[Title/Abstract]) AND (systematic review[Publication Type] OR meta-analysis[Publication Type] OR systematic review[Title/Abstract] OR meta-analy*[Title/Abstract] OR umbrella review[Title/Abstract]))

Filtros: nenhum filtro de idioma na busca; humanos e adultos aplicados na seleção. Data final: 1º de agosto de 2026. A atualização deve incluir busca por título, rastreamento das referências das revisões e busca de citações dos artigos mais recentes.

APÊNDICE B – FORMULÁRIO MÍNIMO DE EXTRAÇÃO

Tabela suplementar B1. Estrutura recomendada para atualização e auditoria da síntese.

Domínio	Itens a registrar
Identificação	Autores, ano, periódico, DOI, protocolo, data da busca
População	Países, cenário, mecanismo, urgência, gravidade, critérios de operação
Exposição	Fonte de peso/altura, classes de IMC, adiposidade ou composição corporal
Comparador	Normal, não obesidade, baixo peso incluído ou excluído
Desfechos	Definição, tempo, ajuste, perdas, denominador e efeito
Método	Modelo, heterogeneidade, subgrupos, sensibilidade, publicação
Viés	Confundimento, seleção, colisor, causalidade reversa, erro de mensuração
Aplicabilidade	Equipamentos, dose, ventilação, ferida, mobilidade, resgate

CAPÍTULO 7

EFEITOS RENAIS DA CIRURGIA METABÓLICA EM ADULTOS COM DIABETES TIPO 2: SÍNTESE CRUZADA DE ENSAIOS CLÍNICOS, ESTUDOS OBSERVACIONAIS E REGISTROS DE MUNDO REAL

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Leila Silveira Vieira Bezerra; Paula Andrea Travecedo Ramos; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: integrar evidências de diferentes desenhos sobre os efeitos renais da cirurgia metabólica em adultos com diabetes tipo 2, examinando remissão da albuminúria, progressão da doença renal crônica, declínio da taxa de filtração glomerular, necessidade de terapia renal substitutiva, eventos cardiovasculares e mortalidade. Métodos: foi realizada uma síntese cruzada de desenhos com busca estruturada em PubMed/MEDLINE e rastreamento de referências até 1º de agosto de 2026. Ensaios randomizados, coortes prospectivas, estudos pareados e registros foram analisados separadamente e, depois, integrados segundo convergência, temporalidade, precisão, risco de viés e proximidade clínica do desfecho. Não foi calculada estimativa única entre desenhos incompatíveis. Resultados: um ensaio randomizado com 100 participantes mostrou maior remissão da albuminúria após bypass gástrico em 24 meses, mas a diferença deixou de ser estatisticamente demonstrável em cinco anos. Coortes e registros indicaram reduções consistentes de nefropatia, doença renal crônica incidente, doença renal avançada e eventos cardiorrenais. A interpretação da eGFR permaneceu limitada pela perda de massa muscular e pela resolução da hiperfiltração. Registros mais amplos também apontaram maior risco de lesão renal aguda e nefrolitíase, exigindo prevenção específica. Conclusão: a

triangulação sustentou benefício renal provável da cirurgia metabólica, mais convincente para albuminúria, doença renal incidente e desfechos renais graves do que para mudanças isoladas da eGFR. A decisão clínica deve considerar risco operatório, estado nutricional, trajetória renal e tratamento farmacológico contemporâneo.
Palavras-chave: cirurgia metabólica; cirurgia bariátrica; diabetes mellitus tipo 2; doença renal diabética; albuminúria; doença renal crônica; diálise; mundo real.

ABSTRACT

Objective: to integrate evidence from different study designs on the renal effects of metabolic surgery in adults with type 2 diabetes, focusing on albuminuria remission, chronic kidney disease progression, glomerular filtration rate decline, kidney replacement therapy, cardiovascular events, and mortality. Methods: a cross-design evidence synthesis was conducted using a structured PubMed/MEDLINE search and reference tracking through August 1, 2026. Randomized trials, prospective cohorts, matched observational studies, and registries were first assessed separately and then integrated according to convergence, temporality, precision, risk of bias, and clinical proximity of the outcome. No single pooled estimate was calculated across incompatible designs. Results: one randomized trial including 100 participants showed greater albuminuria remission after gastric bypass at 24 months, although the difference was no longer statistically demonstrable at five years. Cohorts and registries consistently suggested lower risks of nephropathy, incident chronic kidney disease, advanced kidney disease, and cardiorenal events. Interpretation of estimated glomerular filtration rate remained limited by loss of muscle mass and resolution of glomerular hyperfiltration. Broader registries also identified increased risks of acute kidney injury and nephrolithiasis, supporting targeted prevention. Conclusion: triangulation supported a probable renal benefit of metabolic surgery, with stronger evidence for albuminuria, incident kidney

disease, and severe renal outcomes than for isolated changes in estimated glomerular filtration rate.

Keywords: metabolic surgery; bariatric surgery; type 2 diabetes mellitus; diabetic kidney disease; albuminuria; chronic kidney disease; dialysis; real-world evidence.

1 INTRODUÇÃO

A doença renal diabética não decorre apenas da hiperglicemia persistente. Excesso de adiposidade visceral, resistência à insulina, hiperfiltração glomerular, ativação do sistema renina-angiotensina-aldosterona, inflamação de baixo grau e sobrecarga hemodinâmica renal formam uma rede de agressões que acelera albuminúria e perda funcional. Em pessoas com obesidade e diabetes tipo 2, esse processo costuma coexistir com hipertensão, apneia do sono, dislipidemia e doença cardiovascular, o que torna a progressão renal um fenômeno sistêmico. A cirurgia metabólica modifica vários desses determinantes simultaneamente: produz perda ponderal duradoura, reduz a resistência à insulina, altera sinalização incretínica, melhora pressão arterial e perfil lipídico e, em parte dos pacientes, induz remissão do diabetes. A questão central, contudo, não é se a cirurgia melhora marcadores metabólicos, mas se essa transformação se converte em proteção renal sustentada e clinicamente relevante.[1-4]

A literatura disponível foi construída por desenhos que respondem a perguntas diferentes. Ensaios randomizados oferecem melhor controle de confundimento, mas incluem poucos participantes e, em geral, concentram-se em albuminúria e remissão precoce da doença renal. Coortes prospectivas, como o Swedish Obese Subjects, ampliaram o tempo de observação e alcançaram doença renal estágio 4 e doença renal terminal, embora sem randomização. Estudos pareados em sistemas integrados de saúde e bases de atenção primária acrescentaram milhares de pessoas com diabetes tipo 2, permitindo estimar nefropatia, doença renal crônica, eventos cardiovasculares e mortalidade em condições de prática clínica. Registros nacionais, por sua vez, capturaram

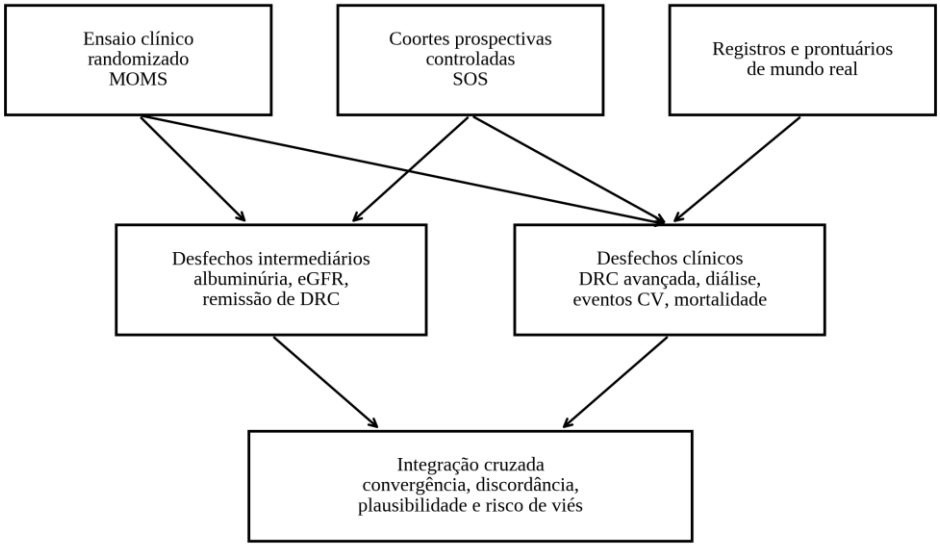
desfechos raros e efeitos adversos renais que dificilmente apareceriam em um ensaio pequeno. Isoladamente, nenhum desenho é suficiente; em conjunto, eles podem oferecer uma inferência mais robusta.[5-13]

A síntese cruzada de desenhos não consiste em misturar indiscriminadamente estimativas. Ela preserva a lógica de cada fonte, avalia a proximidade entre população, intervenção e desfecho e procura padrões de convergência ou discordância. Um sinal observado em ensaio randomizado ganha credibilidade quando reaparece em coortes com maior duração e em registros com desfechos duros. Em sentido contrário, uma associação forte em dados administrativos precisa ser interpretada com cautela quando não se reproduz em ensaio ou quando depende de definições diagnósticas pouco específicas. Essa abordagem é especialmente útil para a cirurgia metabólica, na qual randomização prolongada é difícil, cruzamento entre tratamentos é frequente e mudanças terapêuticas contemporâneas, como inibidores de SGLT2 e agonistas do receptor de GLP-1, alteram o comparador clínico ao longo do tempo.[14-17]

O objetivo deste capítulo foi integrar criticamente ensaios clínicos, estudos observacionais e registros de mundo real sobre os efeitos renais da cirurgia metabólica em adultos com diabetes tipo 2. Foram examinados seis domínios: remissão e incidência de albuminúria; progressão da doença renal crônica; trajetória da taxa de filtração glomerular; doença renal avançada e necessidade de diálise; eventos cardiovasculares; e mortalidade. Também foram analisados os riscos de lesão renal aguda e nefrolitíase, porque uma avaliação clinicamente útil não pode apresentar benefícios renais sem discutir danos específicos do procedimento.

Antes de apresentar os métodos, a Figura 1 resume a lógica da síntese. O desenho foi estruturado para impedir que a grande dimensão dos registros anulasse a força causal do ensaio randomizado ou que a precisão de um banco administrativo fosse confundida com ausência de viés. Cada nível de evidência contribuiu com uma parte distinta da resposta clínica.

Figura 1. Arquitetura da síntese cruzada de evidências sobre cirurgia metabólica e desfechos renais.



Fonte: elaboração dos autores. DRC: doença renal crônica; eGFR: taxa de filtração glomerular estimada; CV: cardiovascular.

A Figura 1 mostra que a integração ocorreu somente após a análise separada dos desenhos. O ensaio randomizado sustentou a comparação direta para remissão da albuminúria; as coortes controladas contribuíram com seguimento prolongado; e os registros forneceram poder para desfechos raros. Essa organização evitou uma média artificial entre medidas de natureza distinta e permitiu que discordâncias, como a atenuação do efeito do bypass aos cinco anos, fossem tratadas como informação clínica e não como falha a ser ocultada.

2 MÉTODOS

2.1 Delineamento, pergunta e elegibilidade

Foi realizada uma síntese cruzada de evidências quantitativas, orientada pelos princípios do PRISMA 2020 para identificação e relato e por uma estratégia de integração convergente. A pergunta foi formulada da seguinte maneira: em

adultos com diabetes tipo 2 e obesidade, a cirurgia metabólica, em comparação com tratamento clínico ou cuidado usual, modificou desfechos renais e cardiorrenais? Foram elegíveis ensaios clínicos randomizados, estudos prospectivos controlados, coortes pareadas, estudos de prontuários eletrônicos e registros populacionais. Estudos de braço único foram mantidos apenas quando trouxeram informação mecanística ou trajetória de albuminúria e hiperfiltração não disponível nos comparadores. Foram excluídos relatos de caso, séries com menos de 20 participantes, estudos exclusivamente pediátricos e publicações sem resultado renal identificável.[18-20]

Foram considerados procedimentos metabólicos o bypass gástrico em Y de Roux, a gastrectomia vertical, a banda gástrica e operações predominantemente disabsortivas. O comparador incluiu tratamento clínico intensivo, cuidado usual ou indivíduos pareados sem cirurgia. Os desfechos primários foram remissão de albuminúria, incidência de nefropatia ou doença renal crônica, progressão para estágio 4 ou doença renal terminal e necessidade de terapia renal substitutiva. Os desfechos secundários incluíram mudança da eGFR, resolução de hiperfiltração, eventos cardiovasculares, insuficiência cardíaca, mortalidade cardiovascular, mortalidade por todas as causas, lesão renal aguda e nefrolitíase. A cirurgia não foi tratada como exposição homogênea quando o artigo permitiu distinguir técnicas.

A Tabela 1 apresenta as decisões de elegibilidade e a função atribuída a cada desenho. A explicitação dessa estrutura foi necessária porque o mesmo desfecho podia aparecer como medida laboratorial em um ensaio e como código diagnóstico em um registro, com validade e sensibilidade diferentes.

Tabela 1. Estrutura de elegibilidade e integração dos desenhos.

Elemento	Decisão adotada	Papel na inferência
População	Adultos com diabetes tipo 2 e obesidade; subgrupos com albuminúria ou DRC foram priorizados.	Reduz indireção e permite análise de maior risco basal.

Elemento	Decisão adotada	Papel na inferência
Intervenção	Bypass em Y de Roux, gastrectomia vertical, banda e outras cirurgias metabólicas.	A técnica foi registrada para avaliar diferenças de mecanismo e dano.
Comparador	Tratamento clínico intensivo, cuidado usual ou controle pareado sem cirurgia.	A qualidade do comparador definiu a força da inferência.
RCT	Comparação direta, principalmente para remissão de albuminúria.	Maior validade interna; baixo poder para eventos raros.
Prospectivo controlado	Seguimento prolongado de albuminúria, DRC avançada e DRT.	Boa temporalidade; risco de seleção por ausência de randomização.
Coortes pareadas	Nefropatia, DRC, eventos CV e mortalidade em prática real.	Maior generalização; confundimento residual permaneceu possível.
Registros nacionais	Eventos raros, segurança renal e terapia substitutiva.	Grande precisão; menor granularidade clínica.
Integração	Convergência de direção, temporalidade, magnitude e plausibilidade.	Não foi calculado efeito único entre métricas incompatíveis.

Fonte: elaboração dos autores. RCT: ensaio clínico randomizado; DRC: doença renal crônica; DRT: doença renal terminal; CV: cardiovascular.

A Tabela 1 evidencia que a síntese não atribuiu o mesmo peso a todas as fontes. O ensaio randomizado foi dominante para a pergunta de eficácia sobre albuminúria, enquanto os registros foram decisivos para doença renal terminal e eventos adversos. A coerência entre esses níveis foi considerada mais importante do que a quantidade isolada de participantes, porque bases grandes podem produzir intervalos estreitos em torno de associações ainda vulneráveis a confundimento.

2.2 Busca, seleção e extração

Foram pesquisados PubMed/MEDLINE e os textos completos indexados no PubMed Central desde o início das bases até 1º de agosto de 2026. Foram combinados termos relacionados a “metabolic surgery”, “bariatric surgery”,

“gastric bypass”, “sleeve gastrectomy”, “type 2 diabetes”, “diabetic kidney disease”, “albuminuria”, “eGFR”, “chronic kidney disease”, “end-stage renal disease”, “dialysis”, “cardiovascular outcomes” e “mortality”. A busca foi complementada pelo rastreamento das referências dos estudos principais e por pesquisa dos títulos de coortes conhecidas, como MOMS, STAMPEDE e Swedish Obese Subjects. A seleção priorizou artigos originais com denominadores, medidas de efeito e seguimento claramente apresentados.

Foram extraídos ano, país, desenho, tamanho da amostra, critério de diabetes, técnica cirúrgica, comparador, duração, definição do desfecho e estimativa ajustada. Quando uma mesma coorte gerou publicações em tempos diferentes, elas foram mantidas como seguimentos complementares e não como amostras independentes. Para o ensaio MOMS, os resultados de 24 e 60 meses foram apresentados em conjunto. Para o SOS, publicações sobre albuminúria incidente, complicações microvasculares e doença renal terminal foram tratadas como análises de uma coorte comum com subconjuntos distintos.

2.3 Avaliação do risco de viés e integração

O ensaio randomizado foi avaliado segundo randomização, desvios da intervenção, dados ausentes, mensuração e seleção do resultado. Estudos não randomizados foram examinados quanto a confundimento, seleção, classificação da exposição, desvios, perdas, mensuração e relato. A classificação final foi expressa como baixo risco, algumas preocupações, moderado ou sério. Em estudos de eGFR, foi adicionada uma avaliação específica do viés de mensuração decorrente da redução de massa muscular, pois a creatinina pode diminuir após a cirurgia sem aumento real da filtração.

A integração foi conduzida em quatro etapas. Primeiro, a direção do efeito foi descrita dentro de cada desenho. Segundo, foi verificada a concordância entre marcadores intermediários e desfechos clínicos. Terceiro, foram examinadas explicações para discordância, incluindo tempo de seguimento, risco basal,

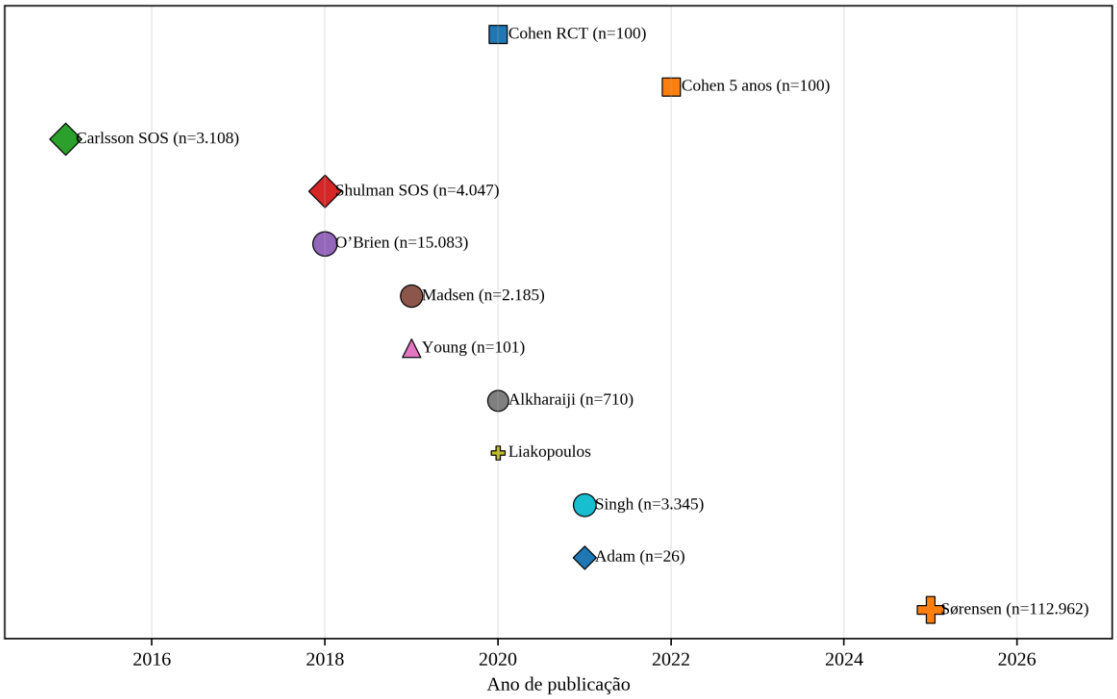
definição do desfecho, técnica operatória e tratamento clínico contemporâneo. Quarto, foi atribuído um julgamento de confiança por domínio: convincente, provável, sugestivo ou inconclusivo. Não foi realizada meta-análise única, porque hazard ratios, proporções de remissão, mudanças contínuas de eGFR e incidências administrativas não representavam o mesmo estimando.

3 RESULTADOS

3.1 Composição e evolução da evidência

A síntese incluiu um ensaio randomizado com dois seguimentos, duas grandes linhas prospectivas controladas, cinco estudos pareados ou baseados em prontuários e dois estudos prospectivos de braço único que esclareceram albuminúria e hiperfiltração. Também foi incorporado um registro populacional dinamarquês mais amplo para segurança e desfechos renais tardios. A Figura 2 apresenta a evolução temporal e a amplitude amostral dessas fontes, permitindo visualizar a transição de estudos pequenos e mecanísticos para bases com dezenas de milhares de participantes.

Figura 2. Linha do tempo e dimensão amostral dos principais estudos incluídos na síntese.



Fonte: elaboração dos autores a partir dos estudos originais.[5-13,21-23] O estudo de Liakopoulos foi representado sem tamanho amostral no rótulo quando o resumo consultado não forneceu o denominador de forma inequívoca.

A Figura 2 demonstra duas características importantes. A primeira é que a evidência randomizada permaneceu pequena: o MOMS reuniu 100 participantes, número suficiente para albuminúria, mas não para diálise ou mortalidade. A segunda é que o salto de precisão ocorreu por meio de coortes pareadas e registros. Esse aumento de escala ampliou a capacidade de detectar eventos raros, porém também aumentou a distância entre a exposição cirúrgica e variáveis clínicas finas, como adesão, uso de bloqueadores do sistema renina-angiotensina e início de inibidores de SGLT2.

A Tabela 2 detalha as populações, os comparadores e os resultados centrais. As publicações foram organizadas por desenho, evitando que seguimentos da mesma coorte fossem interpretados como confirmações independentes.

Tabela 2. Características dos principais estudos e resultados renais ou cardiorrenais.

Estudo	Desenho e amostra	Comparação/seguimento	Resultado principal
Cohen et al., 2020/2022	RCT; Brasil; n=100	RYGB vs melhor tratamento clínico; 2 e 5 anos	Remissão de albuminúria 82% vs 55% em 2 anos; 69,7% vs 59,6% em 5 anos.
Carlsson et al., 2015	SOS prospectivo; n=3.108 sem albuminúria	Cirurgia vs cuidado usual; mediana 10 anos	Albuminúria incidente: HR ajustado 0,37 (IC95% 0,30–0,47).
Shulman et al., 2018	SOS prospectivo; n=4.047	Cirurgia vs cuidado usual; mediana 18 anos	DRT: HR 0,27; DRC4/DRT: HR 0,33.
Carlsson et al., 2017	SOS, análise glicêmica; n=4.032	Cirurgia vs cuidado usual; até 20 anos	Menor incidência microvascular; no diabetes estabelecido, HR 0,54.
O'Brien et al., 2018	Coorte pareada; 4.024 vs 11.059	Cirurgia vs cuidado usual; 5 anos	Nefropatia: HR 0,41 (IC95% 0,29–0,58).
Madsen et al., 2019	Dinamarca; 1.111 vs 1.074	RYGB vs não operados; 5,3 anos	Complicações microvasculares: HR 0,53; macrovasculares: HR 0,76.
Young et al., 2019	Coorte de braço único; n=101	RYGB/SG; média 61 meses	Albuminúria melhorou em 77% e resolveu em 51%.
Alkharaiji et al., 2020	THIN; 131 vs 579	Cirurgia vs controle pareado; até 10 anos	DRC: HR 0,46; diferença média de eGFR 4,1 mL/min/1,73 m².
Liakopoulos et al., 2020	Registro nacional; diabetes tipo 2	Bypass vs controles	Renal grave/queda de 50% da eGFR: HR 0,56; IC 0,44–0,71.
Singh et al., 2021	IMRD; 1.126 vs 2.219	Cirurgia vs cuidado usual; mediana 3,9 anos	DRC: HR 0,63; albuminúria incidente reduzida em 40%.
Adam et al., 2021	Prospectivo; n=26	Antes e 12 meses após cirurgia	Hiperfiltração resolveu em 42%; sem mudança significativa de uACR.
Sørensen et al., 2025	Registro; 18.827 vs 94.135	RYGB/SG vs obesidade sem cirurgia; 8,1 anos	Menor DRC e terapia renal substitutiva, mas maior LRA e nefrolitíase.

Fonte: dados dos artigos originais.[5-13,21-23] RYGB: bypass gástrico em Y de Roux; SG: gastrectomia vertical; DRC: doença renal crônica; DRT: doença renal terminal; LRA: lesão renal aguda; uACR: razão albumina/creatinina urinária.

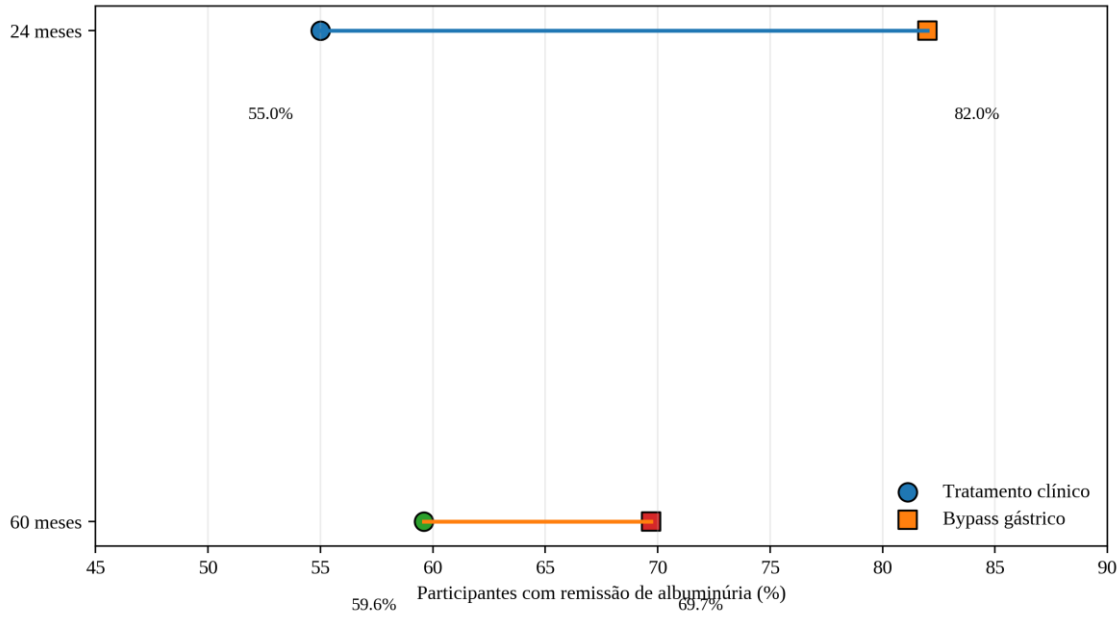
A Tabela 2 mostra uma convergência relevante, mas não perfeita. Os efeitos mais consistentes ocorreram para albuminúria, nefropatia e doença renal crônica incidente. A eGFR apresentou comportamento menos uniforme, sobretudo em pessoas com hiperfiltração basal. A redução de desfechos renais graves apareceu em seguimentos longos, enquanto os sinais de dano — lesão renal aguda e nefrolitíase — foram identificados somente em registros suficientemente grandes. A síntese, portanto, favoreceu uma interpretação de benefício líquido condicionado à prevenção de complicações específicas.

3.2 Albuminúria e remissão inicial da doença renal

O ensaio MOMS forneceu a comparação causal mais direta. Em 24 meses, a remissão da albuminúria ocorreu em 82% dos participantes submetidos ao bypass e em 55% dos que receberam melhor tratamento clínico. A remissão combinada de doença renal crônica inicial ocorreu em 82% e 48%, respectivamente, e a média geométrica da uACR foi 55% menor após a cirurgia. Aos cinco anos, contudo, a remissão estimada foi 69,7% após bypass e 59,6% com tratamento clínico, diferença de 10 pontos percentuais cujo intervalo incluiu ausência de efeito. A melhora glicêmica, ponderal, lipídica e da qualidade de vida permaneceu maior no grupo cirúrgico.[5,6]

A atenuação do contraste aos cinco anos não anulou o resultado inicial. Ela indicou que o tratamento clínico também evoluiu e que a albuminúria é dinâmica. Intensificação de bloqueio do sistema renina-angiotensina, melhor controle pressórico e adoção de fármacos renoprotetores podem ter reduzido a distância entre grupos. Além disso, o estudo teve perda amostral e poder limitado para uma diferença menor. A Figura 3 apresenta essa mudança temporal sem combinar os dois momentos, porque cada seguimento representa um contexto terapêutico e uma população observada diferentes.

Figura 3. Remissão de albuminúria no ensaio MOMS após 24 e 60 meses.



Fonte: reprodução gráfica das proporções publicadas por Cohen et al.[5,6] Não foi realizada nova análise ajustada dos dados individuais.

A Figura 3 evidencia que o benefício absoluto foi amplo aos dois anos e mais estreito aos cinco. Clinicamente, isso sugere uma janela de vantagem renal precoce, seguida de aproximação parcial entre estratégias. A manutenção de quase sete remissões em cada dez pessoas operadas continuou relevante, mas a diferença tardia deixou de sustentar superioridade estatística. Esse padrão reforça que a cirurgia não elimina a necessidade de terapia renal e metabólica continuada.

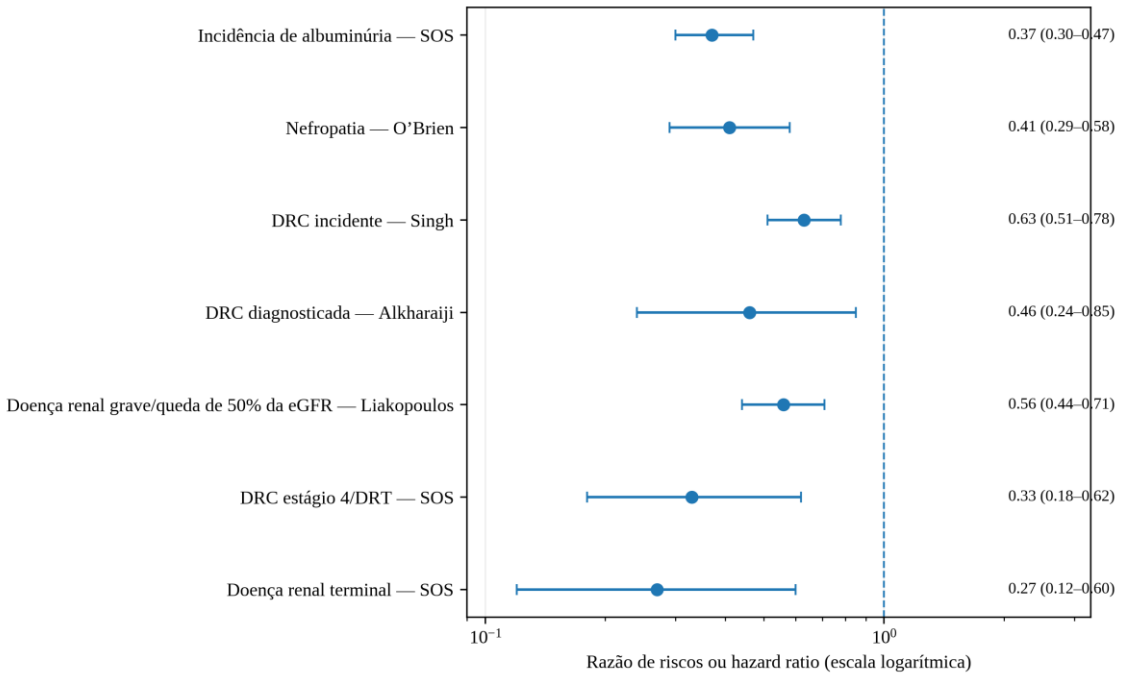
3.3 Doença renal incidente, progressão e desfechos graves

As coortes controladas sustentaram uma redução de risco ao longo do continuum renal. No SOS, entre participantes sem albuminúria, a cirurgia foi associada a incidência ajustada 63% menor ao longo de até 15 anos. Em análise com mediana de 18 anos, ocorreram 13 casos de doença renal terminal no grupo cirúrgico e 26 no controle, com HR ajustado de 0,27. Para o composto DRC estágio 4 ou doença renal terminal, o HR foi 0,33. O benefício foi maior em participantes

com hiperinsulinemia e macroalbuminúria basais, um padrão biologicamente plausível porque maior risco metabólico oferece maior margem absoluta de redução.[7,8]

Nas coortes específicas de diabetes tipo 2, O'Brien observou nefropatia em 4,9% dos operados e 10,0% dos controles em cinco anos, HR 0,41. Singh encontrou HR 0,63 para DRC incidente e redução de 40% da albuminúria, enquanto Alkharaiji identificou HR 0,46 para DRC em pessoas tratadas com insulina. Liakopoulos relatou HR 0,56 para doença renal grave ou redução pela metade da eGFR. Esses estudos foram conduzidos em países e sistemas de saúde diferentes e usaram definições distintas, mas a direção foi estável. A Figura 4 organiza as estimativas sem produzir efeito agrupado.

Figura 4. Estimativas publicadas para albuminúria, nefropatia, doença renal crônica e doença renal terminal.



Fonte: elaboração dos autores com estimativas dos estudos originais.[7-13] Valores abaixo de 1 favorecem a cirurgia. As medidas não foram agrupadas por representarem estimandos e definições distintas.

A Figura 4 mostra que todos os intervalos, com exceção de resultados não incluídos por ausência de estimativa comparável, permaneceram do lado de

proteção. A maior redução apareceu para doença renal terminal no SOS, mas esse resultado derivou de poucos eventos e de população não restrita ao diabetes tipo 2. A consistência de nefropatia e DRC em coortes específicas de diabetes aumentou a credibilidade, embora não elimine o risco de que pacientes escolhidos para cirurgia apresentassem melhor acesso, adesão e capacidade de autocuidado.

3.4 eGFR, hiperfiltração e interpretação fisiológica

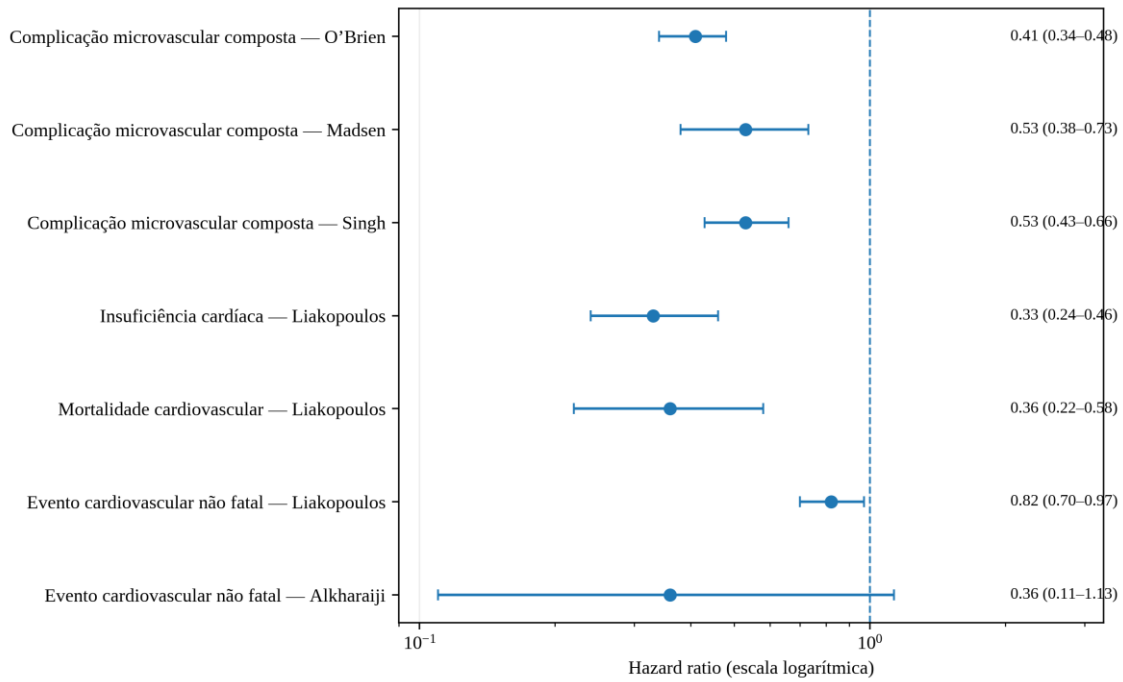
A trajetória da eGFR exigiu interpretação diferente. Em pessoas com obesidade e diabetes, valores altos podem refletir hiperfiltração patológica, e sua redução após cirurgia pode representar normalização, não perda de função. Adam e colaboradores observaram queda da eGFR combinada por creatinina e cistatina C de 128 para 120 mL/min, acompanhada de resolução de hiperfiltração em 42% dos participantes que a apresentavam. Não houve mudança significativa da uACR em 12 meses. Alkharaiji encontrou vantagem média de 4,1 mL/min/1,73 m² para a cirurgia ao longo do seguimento, mas Young mostrou que a melhora da albuminúria foi mais evidente do que qualquer conclusão simples sobre filtração.[11,12,21]

A creatinina sérica perde confiabilidade após grande perda de peso porque a massa muscular diminui. Uma eGFR aparentemente melhor pode resultar de menor geração de creatinina, enquanto uma queda moderada pode significar correção da hiperfiltração. Medidas com cistatina C, depuração medida ou equações combinadas são preferíveis, embora também possam ser influenciadas por inflamação e composição corporal. Por isso, a síntese atribuiu confiança menor à eGFR isolada e maior à combinação de albuminúria, trajetória sustentada e desfechos clínicos.

3.5 Eventos cardiovasculares, mortalidade e segurança renal

A proteção renal ocorreu dentro de uma mudança cardiorrenal mais ampla. Madsen observou HR 0,53 para complicações microvasculares e estimativa não significativa de 0,76 para eventos macrovasculares. Liakopoulos encontrou HR 0,33 para insuficiência cardíaca, 0,36 para mortalidade cardiovascular e 0,82 para eventos cardiovasculares não fatais. Em contraste, Alkharaiji apresentou estimativa imprecisa para evento não fatal, com intervalo atravessando a nulidade. A Figura 5 permite comparar a robustez e a precisão dessas associações sem sugerir equivalência entre desfechos.

Figura 5. Estimativas de complicações microvasculares, eventos cardiovasculares e mortalidade.



Fonte: elaboração dos autores a partir de O'Brien, Madsen, Singh, Liakopoulos e Alkharaiji.[9,10,12-14] Valores abaixo de 1 favorecem a cirurgia.

A Figura 5 mostra que os efeitos mais expressivos foram observados para insuficiência cardíaca e mortalidade cardiovascular, resultados coerentes com redução de volume circulante, pressão arterial, resistência à insulina e inflamação. O benefício foi menor para eventos ateroscleróticos não fatais, que

dependem de carga vascular acumulada e podem responder mais lentamente. A estimativa ampla de Alkharaiji demonstra como amostras menores não conseguem excluir benefício ou ausência de efeito, mesmo quando a direção é favorável.

O registro dinamarquês de 2025 acrescentou um contraponto essencial. Em 18.827 operados e 94.135 controles, a cirurgia foi associada a menor DRC e menor terapia renal substitutiva, mas a maior lesão renal aguda e nefrolitíase. A elevação de risco provavelmente refletiu desidratação, alterações de absorção, hiperoxalúria entérica e mudanças urinárias, sobretudo após bypass. Esses eventos não contradizem a proteção crônica; mostram que benefícios e danos ocorrem em horizontes diferentes e exigem protocolos de hidratação, monitoramento e prevenção de cálculos.[23]

A Tabela 3 resume o risco de viés e a contribuição de cada estudo. O julgamento não foi usado para excluir automaticamente evidências, mas para moderar a linguagem e impedir que uma associação observacional fosse descrita como efeito causal definitivo.

Tabela 3. Julgamento do risco de viés e contribuição para a síntese.

Fonte	Risco global	Principal limitação	Contribuição
MOMS RCT	Algumas preocupações	Pequena amostra, ensaio aberto e perdas no seguimento tardio.	Base causal para remissão de albuminúria.
SOS	Moderado	Alocação não randomizada e técnicas cirúrgicas antigas.	Seguimento longo e desfechos renais graves.
O'Brien	Moderado	Confundimento residual e classificação por prontuário.	Grande coorte específica de diabetes e nefropatia.
Madsen	Moderado	Seleção para cirurgia e desfecho microvascular composto.	Registro nacional com pareamento e seguimento completo.
Singh	Moderado	Desfechos codificados e duração relativamente curta.	Confirmação de DRC e albuminúria em atenção primária.

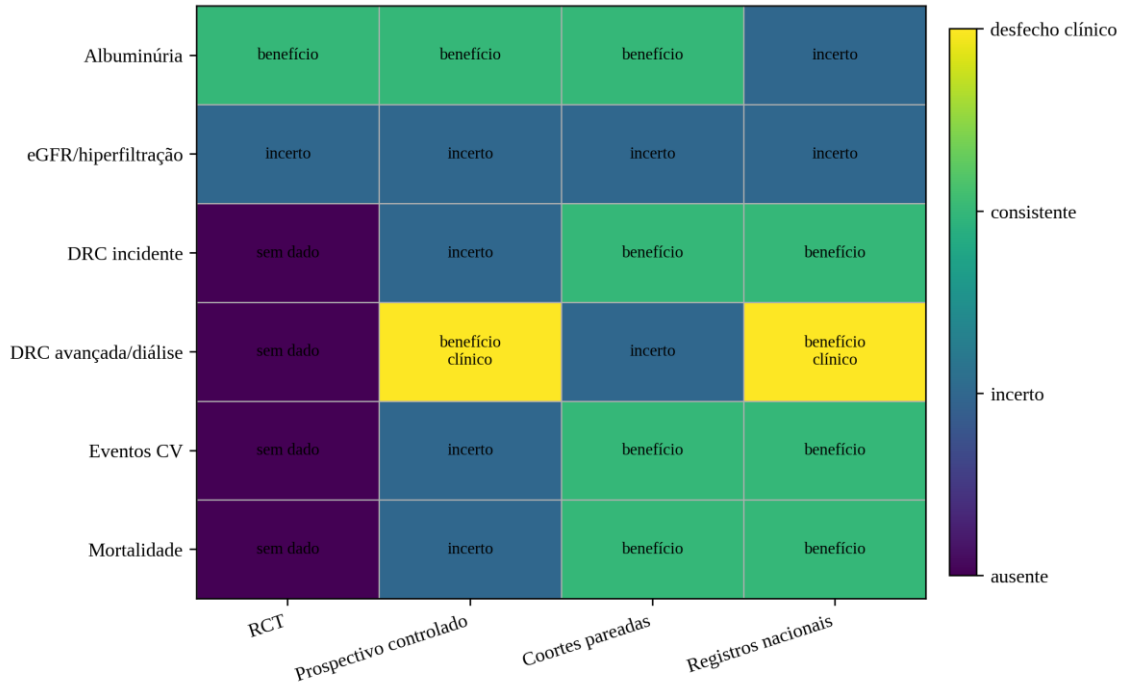
Fonte	Risco global	Principal limitação	Contribuição
Alkharaiji	Sério	Amostra cirúrgica pequena e intervalos amplos.	População com insulina e análise de eGFR.
Liakopoulos	Moderado	Exposição não randomizada e dependência de registro.	Desfechos cardiorrenais e estratos de função renal.
Young/Adam	Sério	Braço único e amostras pequenas.	Informação mecanística sobre albuminúria e hiperfiltração.
Sørensen	Moderado	População ampla de obesidade, não apenas diabetes.	Segurança renal e eventos raros em escala nacional.

Fonte: avaliação dos autores com base nos relatórios publicados e nos domínios do RoB 2 e ROBINS-I.[19,20]

A Tabela 3 esclarece por que a conclusão foi formulada como benefício provável, e não como certeza absoluta para todos os desfechos. A remissão de albuminúria teve suporte randomizado, mas a proteção contra diálise e mortalidade dependeu de estudos não randomizados. A convergência reduziu a possibilidade de que todo o sinal fosse artefato, porém o confundimento por seleção permaneceu relevante, sobretudo porque candidatos à cirurgia são avaliados, acompanhados e tratados de modo mais intensivo.

A Figura 6 integra os domínios e os desenhos em uma matriz de evidência. Ela não representa pontuação estatística, mas uma síntese qualitativa estruturada da presença, consistência e proximidade clínica dos resultados.

Figura 6. Matriz de convergência da evidência por desenho e domínio de desfecho.



Fonte: elaboração dos autores. "Benefício clínico" indica evidência em desfecho renal ou cardiovascular maior; "incerto" indica sinal limitado, heterogêneo ou vulnerável à mensuração.

A Figura 6 demonstra que a albuminúria apresentou a cobertura mais completa entre desenhos. A DRC incidente e os desfechos graves foram sustentados principalmente por coortes e registros, enquanto a eGFR permaneceu o domínio mais incerto. Eventos cardiovasculares e mortalidade mostraram benefícios em dados de mundo real, mas sem confirmação randomizada desenhada para esses eventos. Essa assimetria define a principal agenda de pesquisa: ampliar o seguimento de ensaios e padronizar medidas renais independentes de creatinina.

4 DISCUSSÃO

A síntese cruzada indicou que a cirurgia metabólica provavelmente modifica a história renal do diabetes tipo 2, mas o benefício não foi uniforme entre marcadores, tempos e desenhos. A evidência mais direta ocorreu para

remissão da albuminúria: o bypass foi claramente superior ao tratamento clínico aos 24 meses no MOMS. Aos cinco anos, a diferença diminuiu e perdeu significância, apesar de a maioria dos operados permanecer em remissão. Esse achado é importante porque impede uma narrativa simplista de superioridade permanente. Ele sugere que a cirurgia produz uma vantagem precoce robusta, mas que o tratamento clínico intensivo pode recuperar parte da diferença, especialmente quando incorpora terapias renoprotetoras modernas. Também mostra que a operação deve ser entendida como componente de uma estratégia longitudinal, não como substituto do cuidado nefrológico e diabetológico.[5,6]

O sinal de albuminúria foi reforçado por evidências de maior duração. No SOS, a incidência foi substancialmente menor após cirurgia, e a necessidade estimada de tratar para evitar um caso em dez anos foi nove. Young observou resolução em aproximadamente metade dos pacientes com albuminúria preexistente, enquanto Singh encontrou menor incidência em prática real. A coerência entre prevenção de novos casos, remissão de doença estabelecida e menor nefropatia torna improvável que o resultado decorra apenas de flutuação laboratorial. Entretanto, a albuminúria pode regredir por melhora pressórica, glicêmica e hemodinâmica sem que toda a lesão estrutural seja revertida. Por isso, sua redução deve ser valorizada como marcador prognóstico, mas não tratada como prova isolada de cura renal.[7,9,11,13]

Os desfechos renais graves ofereceram uma segunda linha de confirmação. O SOS mostrou menor doença renal terminal e menor composto de estágio 4 ou terminal em quase duas décadas. Liakopoulos encontrou menor risco de doença renal grave ou redução pela metade da eGFR em pessoas com diabetes tipo 2, e o registro dinamarquês de 2025 observou menor DRC e terapia renal substitutiva. Esses resultados são clinicamente mais importantes do que pequenas diferenças de creatinina, porque se aproximam daquilo que pacientes e sistemas de saúde desejam evitar: perda irreversível de função, diálise e mortalidade. A limitação é que nenhum desses resultados veio de ensaio randomizado com poder adequado. A magnitude pode estar superestimada por

seleção, mas a repetição em fontes independentes e a existência de gradiente de risco basal sustentam plausibilidade.[8,12,23]

A interpretação da eGFR foi o ponto mais delicado. Após cirurgia, a creatinina pode cair pela perda de massa muscular, elevando artificialmente a estimativa de filtração. Ao mesmo tempo, uma redução de eGFR em pessoas hiperfiltrantes pode representar normalização benéfica. O estudo de Adam, ao combinar creatinina e cistatina C, mostrou queda discreta acompanhada de resolução de hiperfiltração, sem redução significativa da uACR em 12 meses. Esse padrão confirma que "aumento" e "queda" não têm significado fixo. Estudos futuros devem priorizar cistatina C, medidas combinadas, declive anual após estabilização ponderal e, em subamostras, filtração medida. Sem essa padronização, comparações de eGFR continuarão vulneráveis a conclusões opostas a partir do mesmo fenômeno fisiológico.[21,24]

Os mecanismos de proteção são provavelmente múltiplos e interdependentes. A redução de pressão intraglomerular pode ocorrer rapidamente com menor hiperinsulinemia, natriurese, melhora da pressão arterial e mudanças em hormônios intestinais. A perda de tecido adiposo visceral reduz citocinas inflamatórias e melhora sensibilidade à insulina. A remissão do diabetes diminui exposição glicêmica cumulativa, mas não explica tudo, porque benefícios renais podem aparecer antes da perda ponderal máxima e mesmo quando a remissão não é completa. Madsen mostrou que a remissão no primeiro ano se associou a menor complicação microvascular posterior, sugerindo um efeito de legado metabólico. Ao mesmo tempo, a proteção pode depender de manutenção ponderal, controle pressórico e continuidade medicamentosa; recorrência do diabetes e reganho de peso provavelmente reduzem a vantagem.[10,15,16]

A redução de insuficiência cardíaca e mortalidade cardiovascular foi mais intensa do que a redução de eventos ateroscleróticos não fatais. Esse contraste é coerente com um efeito rápido sobre volume, pressão, função diastólica e metabolismo miocárdico, enquanto placas ateroscleróticas acumuladas podem

persistir. Para a doença renal, essa melhora cardíaca é relevante porque congestão venosa, baixo débito e hospitalizações aceleram perda funcional. Assim, parte da renoproteção pode ser mediada pela prevenção de eventos cardiorrenais, e não apenas por mudanças intrarrenais. Essa visão integrada evita separar artificialmente rim, metabolismo e coração em uma população na qual os três sistemas adoecem em conjunto.[12]

A segurança renal modifica a decisão clínica. O maior risco de lesão renal aguda e nefrolitíase observado em registro dinamarquês não deve ser minimizado. Desidratação, vômitos, uso de diuréticos, exposição a anti-inflamatórios, hiperoxalúria e baixa ingestão de cálcio podem criar dano precoce ou tardio. O bypass parece biologicamente mais associado à hiperoxalúria do que a gastrectomia vertical, embora a escolha do procedimento dependa de controle metabólico, refluxo, risco nutricional e experiência do centro. A prevenção deve incluir orientação hídrica, revisão de medicamentos, monitoramento de creatinina e eletrólitos, ingestão adequada de cálcio com refeições e investigação de cálculo em pacientes sintomáticos. Benefício crônico não dispensa vigilância aguda.[23,25]

A aplicabilidade contemporânea também exige reconhecer que parte das coortes começou antes da ampla utilização de inibidores de SGLT2, agonistas de GLP-1 e antagonistas não esteroidais do receptor mineralocorticoide. Esses fármacos reduzem albuminúria e eventos renais e podem diminuir a diferença absoluta entre cirurgia e cuidado clínico, como possivelmente ocorreu no seguimento tardio do MOMS. Por outro lado, pessoas com obesidade grave, longa duração do diabetes, necessidade de múltiplos medicamentos ou dificuldade de alcançar perda ponderal sustentada podem continuar obtendo benefício absoluto elevado. A comparação relevante em 2026 não é cirurgia versus tratamento clínico antigo, mas cirurgia integrada ao tratamento moderno versus tratamento moderno isolado. Estudos emulados e pragmáticos devem incorporar essa mudança.

As diferenças entre procedimentos permanecem incompletamente resolvidas. O bypass apresentou a maior quantidade de evidência e efeitos metabólicos intensos, mas também maior potencial de deficiência nutricional e hiperoxalúria. A gastrectomia vertical é mais recente e menos representada em desfechos renais de longo prazo. Banda e gastroplastia vertical dominaram parte do SOS, reduzindo a correspondência com a prática atual. Não é adequado extrapolar automaticamente o benefício médio para todas as técnicas. A seleção deve considerar risco renal basal, histórico de cálculos, refluxo, capacidade de suplementação e necessidade de potência metabólica. Registros com padronização de técnica e dose de exposição são necessários para comparar efetividade renal real.

Entre as forças deste capítulo estiveram a separação entre desenhos, a apresentação de discordâncias, a inclusão de desfechos benéficos e adversos e o uso de estudos primários reais. A principal limitação foi a impossibilidade de recuperar dados individuais e de harmonizar definições de DRC, nefropatia e terapia renal substitutiva. Várias estimativas vieram de códigos diagnósticos, e o tratamento clínico variou entre décadas. A sobreposição de coortes exigiu cuidado para não contar publicações como amostras independentes. Além disso, registros amplos de obesidade foram usados apenas como triangulação de segurança e desfechos raros, pois nem todos restringiram a população ao diabetes tipo 2.

Em conjunto, a evidência sustentou quatro conclusões práticas. Primeiro, albuminúria e doença renal inicial podem melhorar após cirurgia, com vantagem mais clara nos primeiros anos. Segundo, coortes de longa duração sugerem menor progressão para doença renal avançada e diálise. Terceiro, eGFR isolada é insuficiente para medir benefício, especialmente durante perda rápida de peso. Quarto, a cirurgia introduz riscos renais próprios que precisam ser prevenidos. A decisão deve ser individualizada em equipe multidisciplinar, considerando duração do diabetes, albuminúria, estágio da DRC, histórico de cálculos, estado

nutricional, risco operatório, preferência do paciente e acesso a tratamento farmacológico contemporâneo.

5 CONCLUSÃO

A integração de ensaio randomizado, coortes prospectivas, estudos pareados e registros mostrou um padrão predominantemente convergente de proteção renal após cirurgia metabólica em adultos com diabetes tipo 2 e obesidade. O benefício foi mais convincente para remissão ou prevenção de albuminúria, redução de nefropatia e menor incidência de doença renal crônica. Estudos de longa duração também sugeriram menor progressão para doença renal estágio 4, doença renal terminal e terapia renal substitutiva, embora esses resultados ainda dependam de evidência não randomizada. A eGFR permaneceu um marcador ambíguo por influência de massa muscular e hiperfiltração. A redução de insuficiência cardíaca e mortalidade cardiovascular reforçou a natureza cardiorrenal do efeito, enquanto o aumento de lesão renal aguda e nefrolitíase destacou a necessidade de prevenção ativa. A cirurgia deve ser considerada uma intervenção metabólica sistêmica integrada ao cuidado renal moderno, e não uma solução isolada.

REFERÊNCIAS

1. de Boer IH, Khunti K, Sadusky T, Tuttle KR, Neumiller JJ, Rhee CM, et al. Diabetes management in chronic kidney disease: a consensus report by the American Diabetes Association and Kidney Disease: Improving Global Outcomes. Diabetes Care. 2022;45(12):3075-3090. doi:10.2337/dci22-0027.

2. Schauer PR, Bhatt DL, Kirwan JP, Wolski K, Aminian A, Brethauer SA, et al. Bariatric surgery versus intensive medical therapy for diabetes: 5-year outcomes. N Engl J Med. 2017;376(7):641-651. doi:10.1056/NEJMoa1600869.

3. Mingrone G, Panunzi S, De Gaetano A, Guidone C, Iaconelli A, Capristo E, et al. Metabolic surgery versus conventional medical therapy in patients with type 2 diabetes: 10-year follow-up of an open-label, single-centre, randomised

controlled trial. Lancet. 2021;397(10271):293-304. doi:10.1016/S0140-6736(20)32649-0.

4. Martin WP, Docherty NG, Le Roux CW. Impact of bariatric surgery on cardiovascular and renal complications of diabetes: a focus on clinical outcomes and putative mechanisms. Expert Rev Endocrinol Metab. 2018;13(5):251-262. doi:10.1080/17446651.2018.1518130.

5. Cohen RV, Pereira TV, Aboud CM, Caravatto PP, Petry TBZ, Correa JLL, et al. Effect of gastric bypass vs best medical treatment on early-stage chronic kidney disease in patients with type 2 diabetes and obesity: a randomized clinical trial. JAMA Surg. 2020;155(8):e200420. doi:10.1001/jamasurg.2020.0420.

6. Cohen RV, Pereira TV, Aboud CM, Petry TBZ, Lopes Correa JL, Schiavon CA, et al. Gastric bypass versus best medical treatment for diabetic kidney disease: 5 years follow up of a single-centre open label randomised controlled trial. EClinicalMedicine. 2022;54:101725. doi:10.1016/j.eclinm.2022.101725.

7. Carlsson LMS, Romeo S, Jacobson P, Burza MA, Maglio C, Sjöholm K, et al. The incidence of albuminuria after bariatric surgery and usual care in Swedish Obese Subjects: a prospective controlled intervention trial. Int J Obes. 2015;39(1):169-175. doi:10.1038/ijo.2014.72.

8. Shulman A, Peltonen M, Sjöström CD, Andersson-Assarsson JC, Taube M, Sjöholm K, et al. Incidence of end-stage renal disease following bariatric surgery in the Swedish Obese Subjects Study. Int J Obes. 2018;42(5):964-973. doi:10.1038/s41366-018-0045-x.

9. O'Brien R, Johnson E, Haneuse S, Coleman KJ, O'Connor PJ, Fisher DP, et al. Microvascular outcomes in patients with diabetes after bariatric surgery versus usual care: a matched cohort study. Ann Intern Med. 2018;169(5):300-310. doi:10.7326/M17-2383.

10. Madsen LR, Baggesen LM, Richelsen B, Thomsen RW. Effect of Roux-en-Y gastric bypass surgery on diabetes remission and complications in individuals with type 2 diabetes: a Danish population-based matched cohort study. Diabetologia. 2019;62(4):611-620. doi:10.1007/s00125-019-4816-2.

11. Young LA, Nor Hanipah Z, Brethauer SA, Schauer PR, Aminian A. Long-term impact of bariatric surgery in diabetic nephropathy. Surg Endosc. 2019;33(5):1654-1660. doi:10.1007/s00464-018-6458-z.

12. Liakopoulos V, Franzén S, Svensson AM, Zethelius B, Ottosson J, Näslund I, et al. Renal and cardiovascular outcomes after weight loss from gastric bypass surgery in type 2 diabetes: cardiorenal risk reductions exceed atherosclerotic benefits. Diabetes Care. 2020;43(6):1276-1284. doi:10.2337/dc19-1703.

13. Alkharaiji M, Anyanwagu U, Donnelly R, Idris I. Effect of bariatric surgery on diagnosed chronic kidney disease and cardiovascular events in patients with insulin-treated type 2 diabetes: a retrospective cohort study from a large UK primary care database. Obes Surg. 2020;30(5):1685-1694. doi:10.1007/s11695-019-04201-y.

14. Singh P, Adderley N, Subramanian A, Gokhale K, Singhal R, Toulis KA, et al. The impact of bariatric surgery on incident microvascular complications in patients with type 2 diabetes: a matched controlled population-based retrospective cohort study. Diabetes Care. 2021;44(1):116-124. doi:10.2337/dc20-0571.

15. Sjöström L, Peltonen M, Jacobson P, Ahlin S, Andersson-Assarsson J, Anveden Å, et al. Association of bariatric surgery with long-term remission of type 2 diabetes and with microvascular and macrovascular complications. JAMA. 2014;311(22):2297-2304. doi:10.1001/jama.2014.5988.

16. Carlsson LMS, Sjöholm K, Karlsson C, Jacobson P, Andersson-Assarsson JC, Svensson PA, et al. Long-term incidence of microvascular disease after bariatric surgery or usual care in patients with obesity, stratified by baseline glycaemic status: a post-hoc analysis of participants from the Swedish Obese Subjects study. Lancet Diabetes Endocrinol. 2017;5(4):271-279. doi:10.1016/S2213-8587(17)30061-X.

17. Coleman KJ, Haneuse S, Johnson E, Bogart A, Fisher D, O'Connor PJ, et al. Long-term microvascular disease outcomes in patients with type 2 diabetes after bariatric surgery: evidence for the legacy effect of surgery. Diabetes Care. 2016;39(8):1400-1407. doi:10.2337/dc16-0194.

18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. London: Cochrane; 2019.

20. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi:10.1136/bmj.i4919.

21. Adam S, Ho JH, Liu Y, Soran H, Syed AA, Malik RA. Improvements in diabetic neuropathy and nephropathy after bariatric surgery: a prospective cohort study. Obes Surg. 2021;31(2):554-563. doi:10.1007/s11695-020-04951-2.

22. Mirajkar N, Bellary S, Ahmed M, Singhal R, Daskalakis M, Tahrani AA. The impact of bariatric surgery on estimated glomerular filtration rate in patients with type 2 diabetes: a retrospective cohort study. Surg Obes Relat Dis. 2016;12(10):1883-1889. doi:10.1016/j.soard.2016.02.024.

23. Sørensen CG, Jensen SK, Thomsen RW, Jespersen B, Gribsholt SB, Christiansen CF. Kidney outcomes after bariatric surgery: a population-based cohort study. BMC Nephrol. 2025;26(1):458. doi:10.1186/s12882-025-04378-8.

24. Friedman AN, Moe S, Fadel WF, Inman M, Mattar SG, Shihabi Z, et al. Predicting the glomerular filtration rate in bariatric surgery patients. Am J Nephrol. 2014;39(1):8-15. doi:10.1159/000357231.

25. Lieske JC, Mehta RA, Milliner DS, Rule AD, Bergstralh EJ, Sarr MG. Kidney stones are common after bariatric surgery. Kidney Int. 2015;87(4):839-845. doi:10.1038/ki.2014.352.

CAPÍTULO 8

ESTRATÉGIAS DE TROMBOPROFILAXIA EM PACIENTES COM OBESIDADE SUBMETIDOS À CIRURGIA: REVISÃO SISTEMÁTICA, ANÁLISE DA REDE E SÍNTESE BENEFÍCIO–DANO

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Aline Pinto Monteiro Costa Sousa; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

A obesidade aumenta simultaneamente o risco trombótico, a variabilidade farmacocinética e as consequências clínicas do sangramento, tornando inadequada a escolha automática de uma dose única para todos os pacientes cirúrgicos. O objetivo foi comparar estratégias farmacológicas e mecânicas de tromboprofilaxia, examinar a possibilidade de metanálise em rede e traduzir a evidência em uma síntese de benefício e dano. Foi realizada revisão sistemática orientada pelo PRISMA 2020 e pelos itens aplicáveis do PRISMA-NMA, com busca em bases biomédicas e rastreamento de citações até 1º de agosto de 2026. Ensaios randomizados e estudos comparativos em adultos com obesidade submetidos principalmente à cirurgia metabólica e bariátrica foram elegíveis. Dezoito estudos clínicos centrais compuseram a síntese. A rede foi desconectada, sem ciclos fechados e com violação de transitividade por diferenças de procedimento, rastreamento de trombose, dose, início e duração; por isso, não foram calculados rankings ou SUCRA. Na síntese ampliada de dose intensificada versus dose padrão de heparina de baixo peso molecular, apenas dois estudos apresentaram eventos de tromboembolismo venoso informativos, produzindo RR 0,18 (IC95% 0,03–0,96; I^2=30%), resultado dominado por uma coorte histórica. Para sangramento, seis comparações produziram RR 1,00 (IC95% 0,45–2,22; I^2=0%), com grande imprecisão clínica. A revisão Cochrane restrita a ensaios

encontrou pouca ou nenhuma diferença para dose elevada, enquanto a associação entre profilaxia mecânica e farmacológica apresentou possível benefício com baixa certeza. Estudos de extensão pós-alta mostraram risco concentrado após a internação, porém o benefício absoluto foi pequeno em populações contemporâneas de baixo risco. Em coorte nacional de 275.843 operações, enoxaparina 40 mg duas vezes ao dia, apixabana e rivaroxabana associaram-se a menor tromboembolismo, ao passo que enoxaparina 60 mg duas vezes ao dia elevou sangramento sem demonstrar redução correspondente de trombose. Conclui-se que a melhor estratégia é multimodal e estratificada. A evidência não sustenta intensificação universal, nem equivalência automática entre atingir alvo anti-Xa e prevenir evento clínico.

Palavras-chave: Obesidade; Cirurgia bariátrica; Tromboembolismo venoso; Enoxaparina; Tromboprofilaxia; Sangramento; Metanálise.

ABSTRACT

Obesity simultaneously increases thrombotic risk, pharmacokinetic variability, and the clinical consequences of bleeding, making a one-size-fits-all prophylactic dose inappropriate. This study compared pharmacological and mechanical thromboprophylaxis strategies, assessed the feasibility of network meta-analysis, and translated the evidence into a benefit–harm synthesis. A systematic review was conducted according to PRISMA 2020 and applicable PRISMA-NMA items, with biomedical database searches and citation tracking through August 1, 2026. Randomized trials and comparative studies of adults with obesity undergoing mainly metabolic and bariatric surgery were eligible. Eighteen core clinical studies were synthesized. The treatment network was disconnected, lacked closed loops, and violated transitivity because of differences in procedure, thrombosis screening, dose, timing, and duration; therefore, no ranking or SUCRA was calculated. In an expanded direct synthesis of intensified versus standard low-molecular-weight heparin, only two studies had informative venous thromboembolism events, yielding RR 0.18 (95% CI

0.03–0.96; I^2=30%), dominated by a historical cohort. Six bleeding comparisons yielded RR 1.00 (95% CI 0.45–2.22; I^2=0%), with substantial clinical imprecision. The trial-only Cochrane review found little or no difference with higher-dose heparin, whereas combined mechanical and pharmacological prophylaxis showed a possible low-certainty benefit. Post-discharge studies suggested that risk clusters after hospitalization, although the absolute benefit was small in contemporary low-risk populations. In a national cohort of 275,843 operations, enoxaparin 40 mg twice daily, apixaban, and rivaroxaban were associated with lower thromboembolism, whereas enoxaparin 60 mg twice daily increased bleeding without a corresponding reduction in thrombosis. The most defensible strategy is multimodal and risk-stratified. Current evidence does not support universal dose intensification or treating anti-Xa target attainment as equivalent to clinical event prevention.

Keywords: Obesity; Bariatric surgery; Venous thromboembolism; Enoxaparin; Thromboprophylaxis; Bleeding; Meta-analysis.

1 INTRODUÇÃO

O tromboembolismo venoso (TEV), formado pela trombose venosa profunda e pela embolia pulmonar, permanece entre as complicações potencialmente evitáveis de maior gravidade no período perioperatório. A obesidade agrega estase venosa, inflamação de baixo grau, ativação endotelial, resistência à fibrinólise e maior frequência de apneia obstrutiva do sono, insuficiência venosa e limitação funcional. Em cirurgias abdominais prolongadas, esses mecanismos se somam ao pneumoperitônio, à imobilização, à lesão tecidual e à resposta inflamatória operatória. O resultado não é apenas um risco trombótico maior: é também uma população em que a exposição ao anticoagulante, a mobilização, a técnica e o tempo de recuperação variam amplamente.[1-3]

A prevenção exige equilibrar dois eventos de baixa frequência, mas de grande impacto. Doses insuficientes podem deixar o paciente desprotegido;

doses excessivas podem causar hematoma, sangramento da linha de grampeamento, transfusão, reoperação e infecção da ferida. O problema é agravado porque muitos ensaios foram desenhados para atividade anti-Xa, e não para TEV sintomático. A melhora de um marcador farmacodinâmico pode demonstrar maior exposição à heparina, mas não prova redução de embolia pulmonar nem ausência de dano hemorrágico.

As estratégias disponíveis não diferem apenas pela molécula. Heparina de baixo peso molecular (HBPM) pode ser prescrita em dose fixa, ampliada, estratificada por índice de massa corporal (IMC) ou calculada pelo peso; heparina não fracionada (HNF) pode ser preferida quando há disfunção renal ou necessidade de reversão rápida; fondaparinux e anticoagulantes orais diretos oferecem alternativas; compressão pneumática e mobilização precoce atuam por mecanismos distintos. O início pode ocorrer antes ou depois da operação, e a duração pode terminar na alta ou prosseguir por 7, 10, 14, 28 ou 30 dias. Assim, "tromboprofilaxia" não é uma intervenção única.

A fase pós-alta é central. A internação tornou-se curta, a mobilização precoce é incentivada e a maioria das operações bariátricas contemporâneas é minimamente invasiva. Ao mesmo tempo, parte dos eventos ocorre quando a pessoa já está fora do hospital. Estratégias estendidas podem deslocar o equilíbrio para o benefício em pacientes de alto risco, mas expõem milhares de pacientes de baixo risco a injeções, custo, hematomas ou sangramento para prevenir poucos eventos absolutos. Modelos de risco e decisões compartilhadas são, portanto, mais coerentes que intensificação universal.[24,28]

Uma metanálise em rede seria atraente porque permitiria comparar simultaneamente doses, moléculas, componentes mecânicos e durações. Contudo, essa abordagem exige uma rede conectada e transitividade: pacientes que poderiam, em princípio, ter sido randomizados para qualquer intervenção da rede, desfechos equivalentes e modificadores de efeito distribuídos de forma comparável. Quando um estudo detecta trombose assintomática por rastreamento e outro registra apenas eventos clínicos; quando uma coorte

histórica inclui cirurgia aberta e outra inclui gastrectomia vertical laparoscópica; ou quando dose, início e duração mudam simultaneamente, a inferência indireta perde validade.[4-5]

Este capítulo adota uma regra de rastreabilidade: os números apresentados derivam de estudos identificáveis e estão separados conforme desenho, desfecho e período. A visualização foi planejada para mostrar estrutura da evidência, benefício absoluto e incerteza, e não para produzir uma aparência de precisão. O objetivo foi comparar a eficácia e a segurança das estratégias farmacológicas e mecânicas de tromboprofilaxia em pacientes cirúrgicos com obesidade; avaliar formalmente a viabilidade da metanálise em rede; realizar sínteses diretas quando clinicamente justificadas; e construir uma matriz prática de decisão que integre risco trombótico, risco hemorrágico, função renal, procedimento e fase pós-alta.

2 MÉTODOS

2.1 Delineamento e padrão de relato

Foi conduzida revisão sistemática com avaliação de rede e metanálises diretas. O relato foi orientado pelo PRISMA 2020 e pelos itens aplicáveis da extensão para metanálises em rede.[4-5] O protocolo analítico definiu previamente que efeitos indiretos, probabilidades de melhor tratamento e SUCRA somente seriam calculados se houvesse um componente conectado, ciclos fechados e plausibilidade de transitividade. A ausência dessas condições determinaria conversão para sínteses diretas e narrativa estruturada.

A pergunta seguiu a estrutura PICO. A população compreendeu adultos com obesidade submetidos a cirurgia, com prioridade para cirurgia metabólica e bariátrica. As intervenções foram profilaxia mecânica, HNF, HBPM em dose padrão ou intensificada, fondaparinux, apixabana, rivaroxabana e extensão após a alta. Os comparadores foram outra dose, outra molécula, intervenção mecânica isolada, ausência de profilaxia pós-alta ou duração diferente. Os desfechos

principais foram TEV, trombose venosa profunda, embolia pulmonar e sangramento maior; desfechos secundários incluíram trombose portomesentérica, hematoma, transfusão, reexploração, atividade anti-Xa e mortalidade.

2.2 Fontes de informação e elegibilidade

A busca principal foi atualizada até 1º de agosto de 2026 no MEDLINE/PubMed, com complementação por Cochrane Library, PubMed Central, páginas dos periódicos, referências de revisões e rastreamento de citações. Foram combinados termos para obesity, bariatric surgery, surgical procedures, thromboprophylaxis, enoxaparin, heparin, fondaparinux, apixaban, rivaroxaban, pneumatic compression, venous thromboembolism, bleeding e post-discharge. A estratégia reproduzível encontra-se no Apêndice A.

Foram incluídos ensaios randomizados, quase randomizados, estudos prospectivos comparativos, coortes retrospectivas com comparador e grandes registros com ajuste multivariável. Estudos exclusivamente farmacocinéticos ou farmacodinâmicos foram mantidos quando ajudavam a interpretar dose e anti-Xa, mas não receberam o mesmo peso de estudos com eventos clínicos. Séries sem comparador, tratamento de TEV estabelecido e trabalhos sem identificação da população com obesidade foram excluídos da síntese comparativa.

2.3 Extração, risco de viés e certeza

Foram extraídos desenho, procedimento, tamanho amostral, IMC, intervenção, comparador, início, duração, método de rastreamento, eventos e estimativas ajustadas. Trombose assintomática detectada por imagem foi distinguida de TEV clínico. Desfechos de segurança foram mantidos conforme a definição original: sangramento maior, clinicamente relevante, transfusão, hematoma ou reoperação não foram combinados indiscriminadamente.

Ensaios foram examinados pelos domínios do RoB 2 e estudos não randomizados pelos princípios do ROBINS-I.[6-7] A certeza foi interpretada segundo o GRADE, considerando risco de viés, inconsistência, indireção, imprecisão e viés de publicação.[8] Grandes registros não foram automaticamente classificados como alta certeza: tamanho de amostra reduz erro aleatório, mas não elimina confundimento por indicação.

2.4 Síntese estatística e benefício–dano

Razões de risco (RR) foram calculadas a partir de eventos e denominadores. Estudos com zero evento em ambos os braços foram exibidos nas tabelas e no plano benefício–dano, mas não contribuíram para o logaritmo da RR de TEV. Quando apenas um braço apresentou zero evento, utilizou-se correção de continuidade de 0,5. As sínteses ampliadas utilizaram modelo de efeitos aleatórios de DerSimonian–Laird, mantendo separadas as conclusões de revisões restritas a ensaios.

Para tornar a decisão clinicamente legível, diferenças absolutas de risco foram calculadas como taxa na dose intensificada menos taxa na dose padrão, em pontos percentuais. Valores negativos representam menos eventos com intensificação. Essa representação não substitui a metanálise, mas mostra se a mesma estratégia deslocou simultaneamente trombose e sangramento. Estudos de extensão foram também traduzidos para eventos por 1.000 pacientes, preservando o alerta sobre desenho não randomizado.

Regra de interpretação
Um alvo anti-Xa mais frequente não foi tratado como desfecho substituto validado para TEV. Comparações com rastreamento de trombose assintomática não foram fundidas com estudos de eventos clínicos. A rede desconectada não recebeu ranking.

3 RESULTADOS

3.1 Arquitetura da base de evidências

Dezoito estudos clínicos centrais foram organizados por domínio. A produção inicial concentrou-se em diferentes doses de heparina; a partir de 2008 surgiram comparações de duração, anti-Xa e molécula; nos anos recentes, a pesquisa migrou para extensão pós-alta e anticoagulantes orais diretos. A maior parte da evidência veio de cirurgia bariátrica, especialmente bypass gástrico e gastrectomia vertical, o que limita extrapolação para cirurgia aberta, oncologia, reconstrução de parede abdominal e cirurgia plástica extensa.

A Figura 1 mostra que o aumento de amostra não ocorreu de forma uniforme. Ensaios randomizados continuaram pequenos, enquanto grandes registros passaram a dominar os dados contemporâneos. Essa assimetria explica por que alguns intervalos ajustados são estreitos, embora a certeza causal permaneça limitada.

Figura 1. Linha do tempo e escala amostral dos estudos clínicos centrais, por domínio de comparação.



Fonte: elaboração própria com dados de Kalfarentzos et al., Scholten et al., Hamad et al., Kothari et al., Borkgren-Okonek et al., Raftopoulos et al., Simone et al., Birkmeyer et al., Imberti et al., Steele et al., Steib et al., Gelikas et al., Ahmad et al., Kröll et al., Swartz et al., Kongsawat et al., Brown et al. e Gillikin et al.[[9-21,23,25,28-29]]

Tabela 1. Características das comparações clínicas prioritárias.

Estudo	Desenho/amostra	Estratégia	Desfechos
Kalfarentzos 2001	ECR; n=60	Nadroparina 5.700 vs 9.500 UI/d	TEV; sangramento
Scholten 2002	Coortes consecutivas; n=481	Enoxaparina 30 vs 40 mg 2x/d	TEV; sangramento
Hamad 2005	Multicêntrico; n=668	Regimes de enoxaparina entre centros	TEV após cessação
Kothari 2007	Coortes consecutivas; n=476	HNF vs enoxaparina	TEV; reexploração
Borkgren-Okonek 2008	Prospectivo; n=223	40 ou 60 mg por IMC + 10 dias	Anti-Xa; TEV; transfusão
Raftopoulos 2008	Prospectivo histórico; n=308	Hospitalar vs extensão por 10 dias	TEV pós-alta; sangramento

Estudo	Desenho/amostra	Estratégia	Desfechos
Birkmeyer 2012	Registro ajustado; n=24.775	HNF/HNF, HNF/HBPM, HBPM/HBPM	TEV; hemorragia
Imberti 2014	ECR piloto; n=250	Parnaparina 4.250 vs 6.400 UI/d	TEV; sangramento
Steele 2015	ECR duplo-cego; n=198	Enoxaparina vs fondaparinux	TVP por ressonância
Ahmad 2021	ECR; n=150	Mecânica vs mecânica + enoxaparina	TVP silenciosa
Kröll 2023	ECR; n=269	Rivaroxabana 7 vs 28 dias	TEV; sangramento
Gillikin 2026	Registro ajustado; n=275.843	Múltiplos regimes após alta	TEV e sangramento 30/90 dias

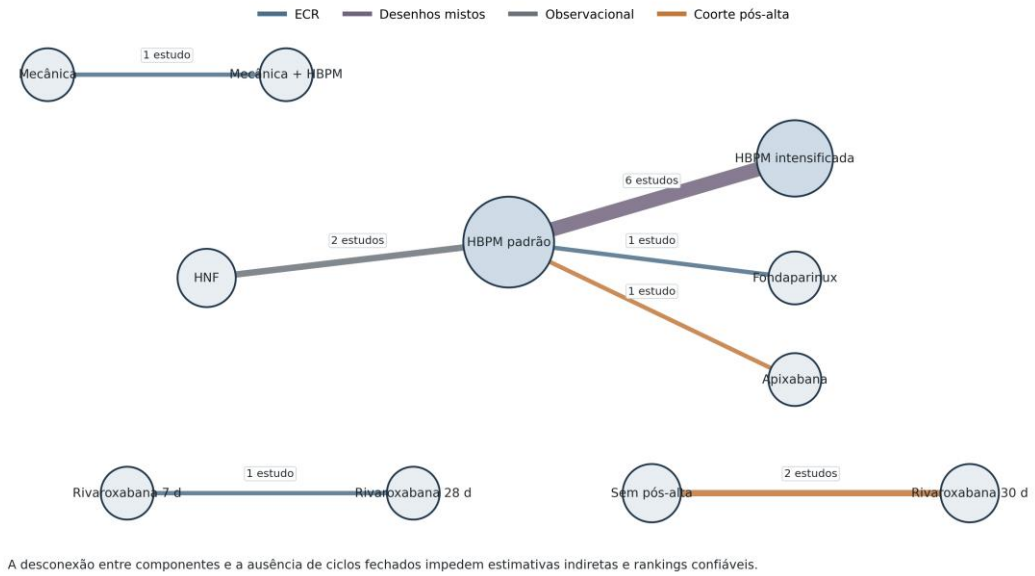
Fonte: estudos primários identificados na revisão.[[9-14,16-18,21,23,28]] HNF: heparina não fracionada; HBPM: heparina de baixo peso molecular; ECR: ensaio clínico randomizado.

3.2 Conectividade: por que a rede não sustentou ranking

O componente principal conectou HNF, HBPM padrão, HBPM intensificada, fondaparinux e apixabana, mas os elos foram sustentados por desenhos e épocas diferentes. A comparação de profilaxia mecânica formou um pequeno ramo. Rivaroxabana por 7 versus 28 dias constituiu componente próprio, e estudos de rivaroxabana por 30 dias contra ausência de pós-alta formaram outro componente. Não houve ciclo fechado capaz de confrontar evidência direta e indireta.

A transitividade foi igualmente frágil. Ahmad et al. rastrearam TVP silenciosa; Steele et al. empregaram ressonância; coortes históricas registraram eventos sintomáticos; registros contemporâneos dependeram de codificação e readmissão. Procedimento, duração de internação, técnica laparoscópica, risco basal e mobilização mudaram ao longo de 25 anos. Produzir uma classificação numérica nessas condições confundiria diferenças de contexto com diferenças de tratamento.

Figura 2. Rede das comparações diretas identificadas e componentes desconectados.



Fonte: elaboração própria com base nas comparações clínicas identificadas.[[9-21,23,26-29]] A largura do elo representa o número de estudos, não a magnitude do efeito.

3.3 Dose intensificada de HBPM versus dose padrão

Seis comparações forneceram eventos clínicos de TEV e sangramento. Quatro apresentaram zero TEV em ambos os braços e, portanto, não informaram a razão relativa de risco. Nos dois estudos com eventos, a síntese ampliada favoreceu a intensificação (RR 0,18; IC95% 0,03–0,96; I^2=30%). O resultado foi dominado por Scholten et al., que comparou períodos consecutivos e encontrou 5/92 eventos com 30 mg duas vezes ao dia e 2/389 com 40 mg duas vezes ao dia. O BAFLUX isolado foi impreciso.

Essa estimativa não deve ser confundida com a revisão Cochrane restrita a sete ensaios. Para heparina em dose elevada versus padrão, a Cochrane encontrou RR 0,55 (IC95% 0,05–5,99) para TEV e RR 1,19 (IC95% 0,48–2,96) para sangramento maior, ambos com baixa certeza.[22] A diferença entre as sínteses decorre principalmente da inclusão da coorte histórica de Scholten na análise ampliada.

Para sangramento, as seis comparações produziram RR 1,00 (IC95% 0,45–2,22; I^2=0%). A aparente ausência de heterogeneidade estatística não elimina a heterogeneidade clínica: definições, doses, duração e gravidade diferiram; vários estudos tiveram um ou nenhum evento. Os dados permitem excluir apenas aumentos ou reduções muito grandes, não diferenças clinicamente importantes de magnitude moderada.

A atividade anti-Xa foi, em geral, maior com 60 mg ou regimes duas vezes ao dia. A metanálise de Albashier et al., com 10 estudos e 1.276 participantes, confirmou maior nível médio e menos resultados subprofiláticos com 60 mg, mas não encontrou diferença significativa em TEV, sangramento, mortalidade ou transfusão.[30] A revisão de Kennedy et al. chegou a conclusão convergente: maior dose altera exposição farmacodinâmica, sem demonstração de benefício clínico proporcional.[31]

Tabela 2. Eventos publicados em comparações de dose intensificada versus dose padrão de HBPM.

Estudo	TEV intensificada	TEV padrão	Sangramento intensificada	Sangramento padrão
Kalfarentzos 2001	0/30	0/30	2/30	0/30
Scholten 2002	2/389	5/92	1/389	1/92
Imberti 2014	1/119	2/131	6/119	8/131
Steib 2016	0/44	0/44	2/44	1/44
Gelikas 2017	0/23	0/31	1/23	0/31
Simone 2008	0/16	0/24	0/16	1/24

Fonte: dados dos estudos primários.[[9-10,15,17,19-20]] Zero evento nos dois braços foi preservado, mas não informa RR de TEV.

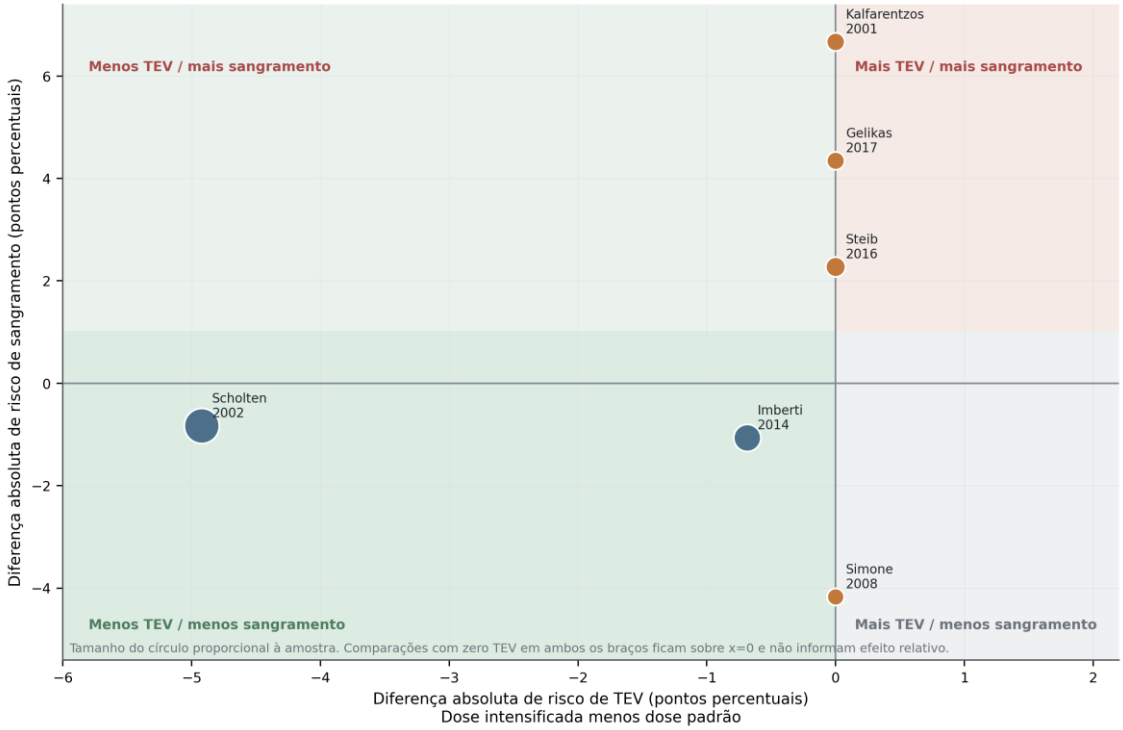
3.4 Plano benefício–dano em escala absoluta

A Figura 3 desloca a análise do "significativo ou não" para o efeito absoluto observado. Scholten aparece no quadrante de menos TEV e menos

sangramento, mas seu desenho por períodos consecutivos permite confundimento por técnica, experiência e mobilização. Imberti apresenta pequenas reduções absolutas nos dois eixos. Kalfarentzos e Gelikas não observaram TEV, mas registraram sangramento apenas na dose mais intensa. Simone mostrou direção oposta para sangramento, baseada em um único evento no braço padrão.

O gráfico evidencia duas propriedades do campo. Primeiro, a maioria dos estudos não teve poder para eventos clínicos; círculos sobre a linha vertical de zero não significam equivalência, apenas ausência de TEV observado. Segundo, a média relativa pode ser dominada por um estudo enquanto as diferenças absolutas permanecem pequenas. Essa leitura justifica selecionar pacientes nos quais o risco basal é suficientemente alto para tornar o benefício absoluto plausível.

Figura 3. Plano benefício–dano das comparações de dose intensificada versus padrão.



Fonte: cálculos próprios a partir de eventos e denominadores publicados.[[9-10,15,17,19-20]] Eixo x: diferença absoluta de TEV; eixo y: diferença absoluta de sangramento.

3.5 HNF, HBPM e fondaparinux

No registro do Michigan Bariatric Surgery Collaborative, com 24.775 operações, esquemas que incorporavam HBPM associaram-se a menor TEV que HNF antes e depois da cirurgia. A OR ajustada foi 0,34 (IC95% 0,19–0,62) para HBPM/HBPM e 0,43 (IC95% 0,21–0,91) para HNF/HBPM, sem aumento estatisticamente significativo de hemorragia.[16] Como quase todos os pacientes também receberam compressão sequencial, o estudo comparou regimes farmacológicos dentro de um pacote multimodal.

Kothari et al. não observaram TVP em duas coortes de 238 pacientes; houve uma embolia pulmonar com HNF e quatro reexplorações por sangramento com enoxaparina.[12] Uma coorte mais recente também levantou sinal de maior sangramento após mudança institucional de HNF para

enoxaparina, demonstrando que a escolha da molécula deve considerar dose, momento e protocolo completo, e não apenas classe farmacológica.

O ensaio EFFORT randomizou 198 participantes para enoxaparina ou fondaparinux. Entre 175 avaliados para TEV, a estimativa foi RR 0,83 (IC95% 0,19–3,61), sem evidência de superioridade clínica.[18,22] O fondaparinux atingiu atividade anti-Xa esperada, mas o número de eventos foi insuficiente. Disfunção renal, reversibilidade, experiência institucional e risco hemorrágico permanecem determinantes práticos.

3.6 Componente mecânico e mobilização

No ensaio de Ahmad et al., 150 participantes receberam compressão mecânica isolada ou compressão associada à enoxaparina iniciada antes da cirurgia e mantida por duas semanas. Nove TVP silenciosas foram identificadas, com direção favorável à combinação. A revisão Cochrane estimou RR 0,05 (IC95% 0,00–0,89) e número necessário para tratar de nove, com baixa certeza.[21-22]

O tamanho do efeito deve ser interpretado no contexto do rastreamento ativo: eventos assintomáticos elevam a taxa basal e não equivalem a embolia pulmonar ou TEV clínico. Ainda assim, os mecanismos são complementares. Compressão reduz estase, mobilização restaura a bomba muscular e anticoagulação atua na coagulação. Na ausência de contraindicação, o conjunto é mais plausível que substituir um componente por outro.

3.7 Profilaxia estendida: risco depois da alta

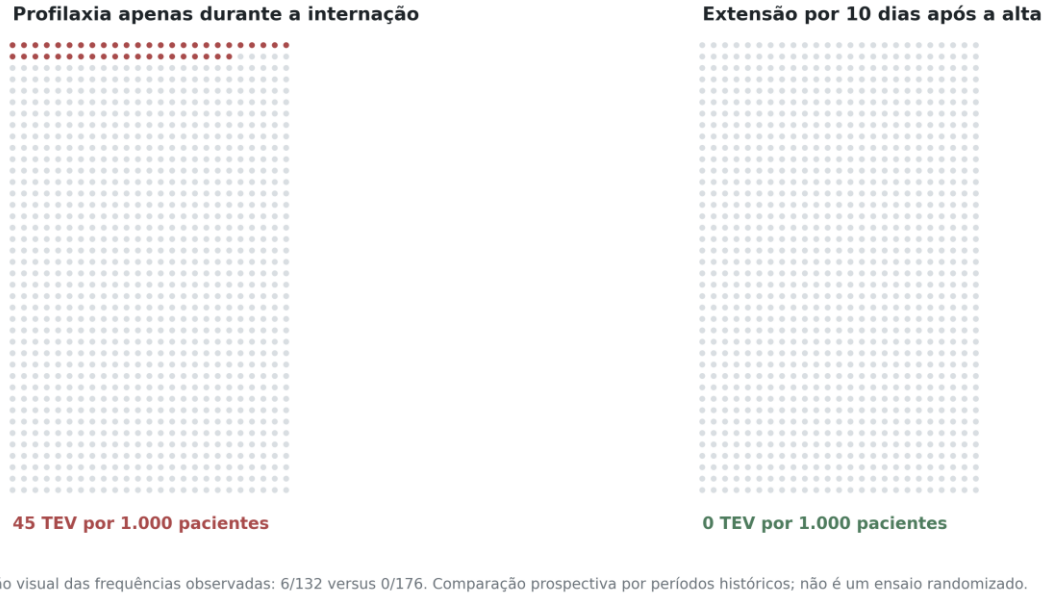
Raftopoulos et al. compararam 132 pacientes com profilaxia limitada à internação e 176 com extensão por 10 dias. Ocorreram seis TEV no primeiro grupo e nenhum no segundo; sangramento significativo foi relatado em 5,3% versus 0,56%.[14] A direção favorece a extensão, mas a comparação por períodos históricos pode incorporar aprendizado cirúrgico, mudança de seleção e maior

mobilização. A Figura 4 traduz a diferença observada para uma população hipotética de 1.000 pacientes, sem convertê-la em efeito causal.

O ensaio BARIVA randomizou 269 participantes para rivaroxabana por 7 ou 28 dias. O evento primário foi extremamente raro: ocorreu um TEV assintomático no grupo de 28 dias, e os sangramentos maiores ou clinicamente relevantes foram poucos.[23] O estudo demonstra que populações contemporâneas selecionadas podem apresentar risco absoluto muito baixo, cenário em que mesmo uma estratégia eficaz exigiria amostra enorme para demonstrar vantagem.

Swartz et al. avaliaram 292 gastrectomias verticais: não ocorreu trombose portomesentérica entre os pacientes que receberam rivaroxabana por 30 dias, contra quatro eventos sem profilaxia pós-alta; sangramento foi semelhante.[24] Guzman-Pruneda et al. também relataram ausência de TEV pós-alta com apixabana em uma coorte contemporânea e poucos sangramentos autolimitados.[26] Esses resultados são encorajadores, mas não randomizados e específicos de protocolos institucionais.

Figura 4. Frequência observada de TEV com profilaxia hospitalar e extensão por 10 dias, traduzida para 1.000 pacientes.



Fonte: conversão visual de 6/132 e 0/176 publicados por Raftopoulos et al.[[14]] Cada ponto representa um paciente; a comparação foi histórica, não randomizada.

3.8 Evidência nacional contemporânea e anticoagulantes orais

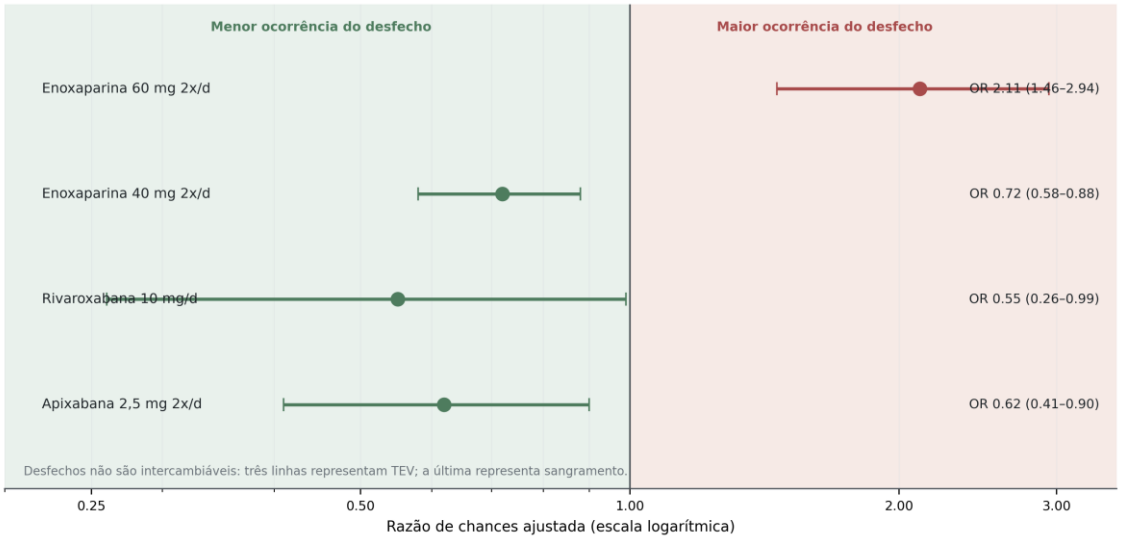
A coorte de Gillikin et al. incluiu 275.843 operações realizadas entre 2018 e 2025. Aproximadamente 23% receberam quimioprofilaxia após a alta. As taxas globais de TEV foram baixas, e os números necessários para tratar ultrapassaram 400 em 90 dias, indicando que a seleção determina o valor clínico.[28]

Após ajuste, apixabana 2,5 mg duas vezes ao dia, rivaroxabana 10 mg/dia e enoxaparina 40 mg duas vezes ao dia associaram-se a menor TEV. Em contraste, enoxaparina 60 mg duas vezes ao dia associou-se a sangramento (OR 2,11; IC95% 1,46–2,94) sem demonstrar redução proporcional de TEV. A Figura 5 apresenta os resultados em uma bússola de efeito e separa explicitamente TEV de sangramento.

Brown et al. compararam apixabana e enoxaparina por 30 dias após gastrectomia vertical e observaram composto de trombose ou sangramento de 5,6% versus 1,7%, favorecendo enoxaparina.[29] Como a alocação não foi

randomizada e a diferença de tamanho entre grupos foi grande, o resultado pode refletir seleção e centro. Anticoagulantes orais oferecem conveniência, mas sua adoção requer considerar tipo de operação, absorção, função renal, adesão, interações e reversão.

Figura 5. Bússola de efeitos ajustados para esquemas de profilaxia pós-alta na coorte nacional de 2026.



Fonte: estimativas ajustadas publicadas por Gillikin et al.[[28]] Linhas de TEV e sangramento representam desfechos distintos e não constituem ranking entre medicamentos.

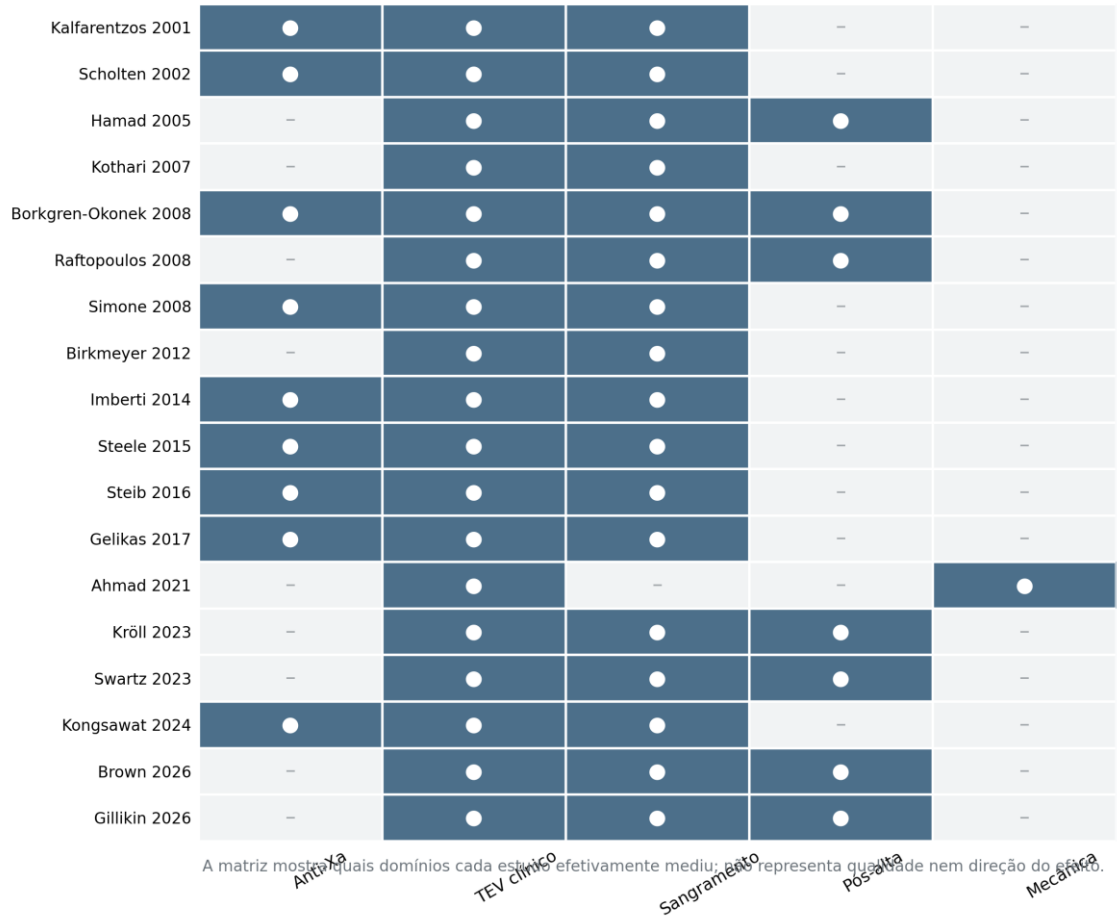
3.9 O que foi realmente medido

A Figura 6 funciona como impressão digital da evidência. Estudos de dose frequentemente mediram anti-Xa e sangramento, mas foram pequenos para TEV. Estudos pós-alta avaliaram eventos clínicos, porém quase sempre observacionais. O componente mecânico foi testado em um único ensaio de 150 participantes. Essa distribuição impede que um desfecho substituto preencha automaticamente a lacuna de outro.

O risco de viés foi menor em ensaios de duração ou molécula com randomização clara, mas a imprecisão permaneceu alta. Em estudos não randomizados, as principais ameaças foram confundimento por indicação,

mudanças temporais, definição de sangramento, detecção incompleta de eventos fora do sistema e seleção para profilaxia. A certeza foi considerada baixa para a maioria das comparações clínicas e muito baixa quando um efeito dependia de um único estudo histórico.

Figura 6. Matriz de cobertura dos desfechos e componentes avaliados em cada estudo.



Fonte: elaboração própria a partir dos métodos e desfechos publicados.[[9-21,23-25,28-29]] Célula preenchida indica que o domínio foi medido, não que houve benefício.

Tabela 3. Síntese de efeito, interpretação e certeza da evidência.

Comparação	Desfecho	Estimativa	Certeza	Leitura clínica
HBPM intensificada vs padrão	TEV	RR 0,18 (0,03–0,96)	Muito baixa	Dois estudos informativos; dominado por coorte histórica

Comparação	Desfecho	Estimativa	Certeza	Leitura clínica
HBPM intensificada vs padrão	Sangramento	RR 1,00 (0,45–2,22)	Baixa	Eventos raros; definições heterogêneas
Dose alta vs padrão, ECRs Cochrane	TEV	RR 0,55 (0,05–5,99)	Baixa	Compatível com grande benefício ou dano
Mecânica + enoxaparina vs mecânica	TVP rastreada	RR 0,05 (0,00–0,89)	Baixa	Um ECR; eventos silenciosos
Enoxaparina vs fondaparinux	TEV	RR 0,83 (0,19–3,61)	Baixa	Ensaio piloto, poucos eventos
Rivaroxabana 7 vs 28 dias	TEV/sangramento	Eventos muito raros	Moderada para segurança; baixa para superioridade	Sem demonstração de vantagem da duração longa
Pós-alta, grande registro	TEV e sangramento	ORs ajustadas por regime	Baixa	Grande precisão, mas confundimento residual

Fonte: síntese do capítulo e revisão Cochrane.[[22-23,28]] A certeza foi apreciada por princípios GRADE; não substitui tabela Summary of Findings formal.

4 DISCUSSÃO

Esta revisão mostra que a tromboprofilaxia no paciente cirúrgico com obesidade não pode ser reduzida à pergunta “40 ou 60 mg?”. A evidência é um mosaico de dose, molécula, momento, duração e componente mecânico. Ensaios pequenos demonstram que a exposição anti-Xa muda com a dose; registros mostram associações clínicas; estudos pós-alta revelam que parte do risco migrou para fora do hospital. Nenhum desses fragmentos, isoladamente, justifica um esquema universal.

O resultado ampliado de RR 0,18 para TEV com intensificação parece expressivo, mas sua base é frágil. Quatro comparações tiveram zero evento em ambos os braços, e a estimativa foi determinada por Scholten, estudo por

períodos consecutivos realizado quando técnica e permanência diferiam da prática contemporânea. A análise de ensaios da Cochrane, metodologicamente mais protegida contra confundimento, foi compatível com nenhuma diferença e apresentou intervalo extremamente amplo. A conclusão correta não é que a dose alta "funciona" ou "não funciona"; é que o efeito clínico médio ainda não está definido.

A discordância entre anti-Xa e eventos clínicos é particularmente instrutiva. A dose de 60 mg aumenta níveis e reduz proporção subprofilática, mas também aumenta resultados supraprofiláticos. A ausência de redução demonstrada em TEV pode refletir baixa potência, risco basal pequeno ou inadequação do alvo como substituto. Anti-Xa pode ser útil em pacientes selecionados — peso extremo, função renal limítrofe, história de TEV ou preocupação com exposição —, mas não deve se transformar em objetivo autônomo desconectado de sangramento e desfechos clínicos.

O risco basal define benefício absoluto. Em uma população com TEV de 0,5%, uma redução relativa de 30% evita aproximadamente 1,5 evento por 1.000, antes de considerar sangramento. Em uma população com TEV de 5%, a mesma redução evita 15 por 1.000. Por isso, extensão pós-alta e intensificação têm maior plausibilidade quando se acumulam história de TEV, imobilidade, operação revisional, tempo operatório prolongado, complicação, idade, insuficiência venosa, superobesidade ou permanência hospitalar prolongada.[24,28,32]

A fase pós-alta concentra a principal oportunidade e o principal risco de excesso de tratamento. Raftopoulos apresentou diferença grande, mas histórica; BARIVA mostrou eventos muito raros; registros de 2026 sugerem benefício com alguns regimes, porém número necessário para tratar elevado. A decisão deve separar pacientes com risco excepcional de populações rotineiras. Prescrever 28 ou 30 dias para todos pode ampliar custo e sangramento sem ganho proporcional.

Os anticoagulantes orais diretos são atraentes pela facilidade de uso, mas não formam uma classe intercambiável com HBPM. Gastrectomia vertical e

bypass modificam anatomia e absorção de maneira distinta; adesão oral não garante exposição; função renal e interação medicamentosa influenciam segurança. O ensaio BARIVA oferece suporte à viabilidade da rivaroxabana em pacientes selecionados, enquanto coortes de apixabana e rivaroxabana apresentam resultados divergentes. A decisão deve permanecer protocolizada e acompanhada.

A profilaxia mecânica não deve ser vista como concorrente da farmacológica. Compressão pneumática, mobilização precoce e anticoagulação atuam em vias diferentes. A evidência randomizada para combinação é limitada, mas a plausibilidade e o baixo custo marginal apoiam abordagem multimodal quando não há contraindicação. O benefício real depende de aplicação correta do dispositivo, tempo de uso, tamanho adequado e mobilização factível.

As visualizações deste capítulo foram escolhidas para impedir atalhos interpretativos. A linha do tempo mostra mudança de desenho; a rede mostra que tratamentos não são comparáveis por uma única hierarquia; o quadrante benefício–dano torna visível a coexistência de trombose e sangramento; a matriz de 1.000 pessoas mostra magnitude absoluta; a bússola separa desfechos; e a matriz de cobertura mostra lacunas de mensuração. Nenhuma figura substitui avaliação de risco de viés, mas todas reduzem o risco de transformar uma porcentagem isolada em recomendação.

4.1 Implicações para a prática

A avaliação deve começar antes da prescrição. História pessoal ou familiar de TEV, trombofilia conhecida, idade, IMC e distribuição corporal, mobilidade, apneia, insuficiência venosa, câncer, terapia hormonal, função renal, anemia, plaquetas, uso de antiagregantes, procedimento, duração esperada e possibilidade de sangramento devem ser registrados. O risco não é estático: complicação, conversão, reoperação, internação prolongada ou mobilização limitada podem alterar a estratégia na alta.

Na maioria dos pacientes sem contraindicação, o núcleo do cuidado inclui mobilização precoce, compressão pneumática durante o período de menor mobilidade e profilaxia farmacológica conforme protocolo institucional. Dose intensificada ou monitorização anti-Xa deve ser reservada a situações em que o risco trombótico ou a incerteza farmacocinética supere o risco de sangramento. Função renal reduzida pode favorecer HNF ou ajuste cuidadoso de HBPM.

A extensão deve ser seletiva. O plano de alta precisa registrar medicamento, dose, duração, técnica de aplicação, sinais de sangramento e TEV, interações e canal de retorno. Injeção domiciliar pode ser uma barreira; DOAC pode facilitar adesão, mas exige considerar absorção e reversibilidade. A decisão deve ser documentada como benefício–dano, não como regra baseada apenas no IMC.

Tabela 4. Matriz clínica para seleção da estratégia de tromboprofilaxia.

Cenário	Estratégia plausível	Salvaguardas
Risco trombótico baixo; cirurgia laparoscópica não complicada; mobilidade precoce	Mecânica + dose farmacológica institucional durante internação	Extensão rotineira pode ter benefício absoluto mínimo
História de TEV, trombofilia, imobilidade, revisão, tempo operatório prolongado	Considerar intensificação ou extensão seletiva; reavaliar na alta	Documentar risco e sangramento; considerar anti-Xa em casos selecionados
Função renal reduzida	Ajustar HBPM ou considerar HNF	Monitorar função renal, plaquetas e sangramento
Sangramento ativo, hemostasia incerta, reoperação	Priorizar mecânica temporariamente; iniciar fármaco após controle	Reavaliar repetidamente; evitar atraso desnecessário
Gastrectomia vertical com opção de DOAC	Pode ser considerada em protocolo validado e paciente selecionado	Verificar absorção, função renal, adesão e interações

Cenário	Estratégia plausível	Salvaguardas
Bypass ou anatomia com possível alteração de absorção	Preferir opção com exposição mais previsível quando houver dúvida	DOAC não deve ser escolhido apenas por conveniência
Alta com mobilidade limitada ou complicação	Recalcular risco e considerar extensão	Fornecer orientação escrita e canal de retorno

Fonte: síntese clínica baseada nos estudos e revisões incluídos.[[22-24,28-31]] Não substitui protocolo institucional, avaliação individual ou diretriz local.

4.2 Agenda de pesquisa

Ensaios futuros precisam ser multicêntricos e dimensionados para TEV clínico e sangramento maior, não apenas anti-Xa. Um desenho adaptativo poderia comparar dose padrão, dose estratificada e extensão em estratos de risco. A randomização deve preservar início e duração, de modo que apenas um componente varie por comparação.

Conjuntos padronizados de desfechos devem distinguir TVP sintomática, embolia pulmonar, trombose portomesentérica, eventos rastreados, sangramento maior, clinicamente relevante, hematoma, transfusão e reoperação. Seguimento de pelo menos 30 e preferencialmente 90 dias é necessário, com captura fora do hospital e adjudicação cega.

Estudos de efetividade devem medir adesão, custo, preferência, capacidade de autoinjeção e desigualdades. Pacientes com superobesidade, cirurgia revisional, doença renal, história de TEV e cirurgia não bariátrica precisam ser representados. Modelos preditivos devem publicar calibração, benefício líquido e validação externa, evitando que grandes bancos produzam escores pouco transportáveis.

5 LIMITAÇÕES

A maior limitação é a raridade dos eventos. A maioria dos ensaios foi pequena e muitos foram desenhados para anti-Xa. Estudos com zero evento não

demonstram equivalência e reduzem a capacidade de estimar heterogeneidade. A síntese ampliada incluiu um estudo não randomizado dominante; por isso, sua certeza foi classificada como muito baixa e contrastada com a análise restrita a ensaios.

Procedimentos, doses, tempos e definições mudaram ao longo de 25 anos. A cirurgia aberta e a internação prolongada das coortes históricas não são diretamente comparáveis à gastrectomia vertical contemporânea. Rastreamento ativo detectou eventos assintomáticos que não devem ser equiparados a TEV clínico. Sangramento foi relatado como transfusão, hematoma, reexploração ou definição composta.

A busca foi estruturada e atualizada em fontes clássicas, porém este capítulo não apresenta contadores artificiais de registros quando não havia biblioteca deduplicada auditável. O conjunto central privilegia estudos comparativos rastreáveis e revisões recentes. Não houve acesso a dados individuais para harmonizar risco, função renal, dose por quilograma, adesão e eventos pós-alta.

As estimativas de 2026 provenientes de grandes registros são contemporâneas e precisas, mas permanecem observacionais. Ajuste estatístico não elimina seleção para profilaxia. Finalmente, as visualizações usam dados publicados e cálculos transparentes, porém algumas representam frequência observada e não efeito causal.

6 CONCLUSÃO

A prevenção do tromboembolismo em pacientes cirúrgicos com obesidade deve ser multimodal, dinâmica e estratificada. A rede disponível não é suficientemente conectada ou transitiva para rankings. Doses maiores de HBPM aumentam exposição anti-Xa, mas a superioridade clínica não foi demonstrada de forma robusta. A síntese ampliada sugere menos TEV com intensificação,

porém depende de uma coorte histórica; a evidência randomizada permanece imprecisa.

Compressão pneumática, mobilização e profilaxia farmacológica devem ser combinadas quando possível. Extensão após a alta pode ser útil em pacientes de maior risco, mas o baixo risco absoluto da prática contemporânea torna inadequada a prescrição universal. Enoxaparina 60 mg duas vezes ao dia não deve ser adotada automaticamente, diante de sinal de sangramento em grande coorte sem redução correspondente de TEV.

A decisão mais segura integra história de TEV, procedimento, mobilidade, função renal, hemostasia, risco de sangramento, anatomia gastrointestinal e capacidade de adesão. O objetivo não é maximizar anti-Xa, e sim evitar trombose sem produzir dano hemorrágico evitável.

REFERÊNCIAS

1. World Health Organization. Obesity and overweight [Internet]. Geneva: WHO; 2025 [cited 2026 Aug 1]. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

2. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: a meta-analysis. Circulation. 2008;117(1):93-102. doi:10.1161/CIRCULATIONAHA.107.709204.

3. Eisenberg D, Shikora SA, Aarts E, Aminian A, Angrisani L, Cohen RV, et al. 2022 American Society for Metabolic and Bariatric Surgery and International Federation for the Surgery of Obesity and Metabolic Disorders: indications for metabolic and bariatric surgery. Surg Obes Relat Dis. 2022;18(12):1345-1356. doi:10.1016/j.soard.2022.08.013.

4. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

5. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions. Ann Intern Med. 2015;162(11):777-784. doi:10.7326/M14-2385.

6. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi:10.1136/bmj.l4898.

7. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi:10.1136/bmj.i4919.

8. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD.

9. Kalfarentzos F, Stavropoulou F, Yarmenitis S, Kehagias I, Karamesini M, Dimitrakopoulos A, et al. Prophylaxis of venous thromboembolism using two different doses of low-molecular-weight heparin in bariatric surgery: a prospective randomized trial. Obes Surg. 2001;11(6):670-676. PMID:11775562.

10. Scholten DJ, Hoedema RM, Scholten SE. A comparison of two different prophylactic dose regimens of low molecular weight heparin in bariatric surgery. Obes Surg. 2002;12(1):19-24. PMID:11868291.

11. Hamad GG, Choban PS, Encinosa W, et al. Enoxaparin for thromboprophylaxis in morbidly obese patients undergoing bariatric surgery: findings of the PROBE study. Obes Surg. 2005;15(10):1368-1374. PMID:16354513.

12. Kothari SN, Lambert PJ, Mathiason MA. A comparison of thromboembolic and bleeding events following laparoscopic gastric bypass in patients treated with prophylactic regimens of unfractionated heparin or enoxaparin. Am J Surg. 2007;194(6):709-711. PMID:18005758.

13. Borkgren-Okonek MJ, Hart RW, Pantano JE, Rantis PC Jr, Guske PJ, Kane JM Jr, et al. Enoxaparin thromboprophylaxis in gastric bypass patients: extended duration, dose stratification, and antifactor Xa activity. Surg Obes Relat Dis. 2008;4(5):625-631. PMID:18261965.

14. Raftopoulos I, Martindale C, Cronin A, Steinberg J. The effect of extended post-discharge chemical thromboprophylaxis on venous thromboembolism rates after bariatric surgery: a prospective comparison trial. Surg Endosc. 2008;22(11):2384-2391. doi:10.1007/s00464-008-0031-9.

15. Simone EP, Madan AK, Tichansky DS, Kuhl DA, Lee MD. Comparison of two low-molecular-weight heparin dosing regimens for patients undergoing laparoscopic bariatric surgery. Surg Obes Relat Dis. 2008;4(5):625-629. PMID:18594915.

16. Birkmeyer NJO, Finks JF, English WJ, Carlin AM, Hawasli AA, Genaw JA, et al. Comparative effectiveness of unfractionated and low-molecular-weight heparin for prevention of venous thromboembolism following bariatric surgery. Arch Surg. 2012;147(11):994-998. doi:10.1001/archsurg.2012.2298.

17. Imberti D, Baldini E, Pierfranceschi MG, Nicolini A, Cartelli C, De Paoli M, et al. Prophylaxis of venous thromboembolism with low molecular weight heparin in bariatric surgery: a prospective, randomised pilot study evaluating two doses of parnaparin (BAFLUX Study). Obes Surg. 2014;24(2):284-291. doi:10.1007/s11695-013-1105-x.

18. Steele KE, Canner J, Prokopowicz G, Verde F, Beselman A, Wyse R, et al. The EFFORT trial: preoperative enoxaparin versus postoperative fondaparinux for thromboprophylaxis in bariatric surgical patients: a randomized double-blind pilot trial. Surg Obes Relat Dis. 2015;11(3):672-683. doi:10.1016/j.soard.2014.10.003.

19. Steib A, Degirmenci SE, Junke E, Asehnoune K, Figier M, Amedro P, et al. Once versus twice daily injection of enoxaparin for thromboprophylaxis in bariatric surgery: effects on antifactor Xa activity and procoagulant microparticles. Surg Obes Relat Dis. 2016;12(3):613-621. doi:10.1016/j.soard.2015.08.505.

20. Gelikas S, Eldar SM, Lahat G. Anti-factor Xa levels in patients undergoing laparoscopic sleeve gastrectomy: 2 different dosing regimens of enoxaparin. Surg Obes Relat Dis. 2017;13(10):1753-1759. doi:10.1016/j.soard.2017.07.027.

21. Ahmad KS, Zayed ME, Faheem MH, Essa MS. Incidence of silent deep venous thrombosis after laparoscopic bariatric surgery in patients who received combined mechanical and chemical thromboprophylaxis compared to patients who received mechanical thromboprophylaxis only. Surg Innov. 2021;28(2):144-150. doi:10.1177/1553350620965812.

22. Amaral FC, Baptista-Silva JCC, Nakano LCU, et al. Pharmacological interventions for preventing venous thromboembolism in people undergoing bariatric surgery. Cochrane Database Syst Rev. 2022;(11):CD013683. doi:10.1002/14651858.CD013683.pub2.

23. Kröll D, Stirnimann G, Borbély Y, et al. Efficacy and safety of rivaroxaban for postoperative thromboprophylaxis in patients after bariatric surgery: a randomized clinical trial. JAMA Netw Open. 2023;6(6):e2315241. doi:10.1001/jamanetworkopen.2023.15241.

24. Swartz DE, Hood L, Swartz DR, Gupta S. 30-day post-discharge prophylaxis with rivaroxaban prevents porto-mesenteric venous thrombosis following

laparoscopic sleeve gastrectomy. Obes Surg. 2023;33(4):1026-1031. doi:10.1007/s11695-023-06471-z.

25. Kongsawat K, et al. Comparison of enoxaparin 40 mg versus 60 mg dosage for venous thromboembolism prophylaxis in bariatric surgery: randomized controlled trial. Asian J Surg. 2024;47:2985-2990. PMID:38514281.

26. Guzman-Pruneda FA, et al. Extended postoperative prophylaxis using apixaban after bariatric surgery: a single-center experience. Obes Surg. 2024;34(4):1217-1223. PMID:38409623.

27. Cuva C, et al. Routine extended thromboprophylaxis with rivaroxaban after sleeve gastrectomy and portomesenteric venous thrombosis. Surg Obes Relat Dis. 2024;20:527-531. doi:10.1016/j.soard.2023.12.006.

28. Gillikin AP, et al. Post-discharge VTE prophylaxis after bariatric surgery: balancing bleeding risk and thrombosis prevention in 275,843 patients. Surg Endosc. 2026;40:6211-6221. doi:10.1007/s00464-026-12954-8.

29. Brown A, et al. Efficacy and safety of direct oral anticoagulants versus enoxaparin for extended thromboprophylaxis following sleeve gastrectomy. Surg Obes Relat Dis. 2026. PMID:41813545.

30. Albashier M, Moqbel I, Saleh AO, Mohammed RK, Mahadin M, Ezz MH, et al. Enoxaparin 40 mg versus 60 mg for venous thromboembolism prophylaxis after bariatric surgery: a systematic review and meta-analysis of pharmacologic and clinical outcomes. Obes Surg. 2026;36(7):3914-3931. doi:10.1007/s11695-026-08765-4.

31. Kennedy C, et al. Prophylactic enoxaparin dosing after bariatric surgery: a systematic review and meta-analysis of pharmacodynamic efficacy and safety. Obes Surg. 2026. PMID:42479412.

32. Aminian A, Andalib A, Khorgami Z, Cetin D, Burguera B, Bartholomew J, et al. Who should get extended thromboprophylaxis after bariatric surgery? A risk assessment tool to guide indications for post-discharge pharmacoprophylaxis. Ann Surg. 2017;265(1):143-150. PMID:28009739.

APÊNDICE A – ESTRATÉGIA DE BUSCA REPRODUZÍVEL

PubMed/MEDLINE (adaptar vocabulário e sintaxe às demais bases):

Estratégia principal

((obesity[Title/Abstract] OR obese[Title/Abstract] OR morbid obesity[Title/Abstract] OR bariatric[Title/Abstract]) AND (surgery[Title/Abstract] OR surgical[Title/Abstract] OR bariatric surgery[Title/Abstract]) AND (thromboprophylaxis[Title/Abstract] OR venous thromboembolism prophylaxis[Title/Abstract] OR enoxaparin[Title/Abstract] OR heparin[Title/Abstract] OR fondaparinux[Title/Abstract] OR apixaban[Title/Abstract] OR rivaroxaban[Title/Abstract] OR pneumatic compression[Title/Abstract] OR post-discharge[Title/Abstract]) AND (venous thromboembolism[Title/Abstract] OR deep vein thrombosis[Title/Abstract] OR pulmonary embolism[Title/Abstract] OR bleeding[Title/Abstract] OR anti-Xa[Title/Abstract]))

Sem filtro inicial de idioma. Humanos e adultos foram aplicados na seleção. A atualização incluiu busca por título, rastreamento de referências, citações dos estudos mais recentes e conferência de DOI/PMID. Data final: 1º de agosto de 2026.

APÊNDICE B – DICIONÁRIO MÍNIMO PARA ATUALIZAÇÃO

Domínio	Itens a registrar
Identificação	Autores, ano, periódico, DOI/PMID, país, protocolo
População	Procedimento, técnica, IMC, peso, idade, sexo, história de TEV, mobilidade
Intervenção	Molécula, dose, frequência, início, duração, ajuste por IMC/peso, componente mecânico
Comparador	Dose, molécula, ausência de pós-alta, duração ou componente
Desfecho	TEV sintomático, rastreado, EP, TVP, PMVT, sangramento, transfusão, reoperação
Tempo	Hospital, 30 dias, 90 dias; método de captura pós-alta
Viés	Randomização, confundimento por indicação, mudanças temporais, perdas, definição
Aplicabilidade	Função renal, anatomia gastrointestinal, adesão, custo, logística e reversão

CAPÍTULO 9

EFEITOS DAS INTERVENÇÕES PSICOLÓGICAS SOBRE DEPRESSÃO, ANSIEDADE E QUALIDADE DE VIDA EM ADULTOS COM DOENÇAS CRÔNICAS: META-ANÁLISE MULTINÍVEL COM ESTIMAÇÃO ROBUSTA DA VARIÂNCIA

Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Sheyla Cristiane Xenofonte de Almeida; Antônia Araújo Firmo; Aline Pinto Monteiro Costa Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Uilna Natércia Soares Feitosa; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: examinar os efeitos de intervenções psicológicas sobre depressão, ansiedade e qualidade de vida em adultos com doenças crônicas, demonstrando uma estratégia multinível capaz de preservar estimativas dependentes. Métodos: foi realizada busca estruturada em PubMed/MEDLINE, seguida de seleção de revisões sistemáticas contemporâneas com estimativas comparáveis. Foram extraídas 15 estimativas padronizadas provenientes de cinco revisões, que abrangeram intervenções cognitivo-comportamentais, autogerenciamento, cuidado organizacional e modalidades digitais. Os sinais foram recodificados para que valores positivos indicassem benefício. Foi ajustado modelo de efeitos aleatórios com variância robusta por revisão, correção para pequeno número de conglomerados e análise de exclusão de uma revisão por vez. Resultados: a síntese exploratória produziu diferença média padronizada de 0,38 (IC95% robusto 0,20–0,56), com heterogeneidade entre fontes (tau^2=0.015). As estimativas permaneceram favoráveis nas cinco análises de influência. Os efeitos foram maiores para depressão e ansiedade em intervenções cognitivo-comportamentais e autogerenciamento guiado, enquanto qualidade de vida apresentou ganhos menores e mais sensíveis ao desenho. Conclusão:

intervenções psicológicas produziram benefício pequeno a moderado, mas a magnitude variou por domínio, formato e tempo. A modelagem multinível e a estimação robusta reduziram a perda de informação; ainda assim, a reanálise de estimativas agregadas não substituiu uma meta-análise com dados extraídos diretamente dos ensaios.

Palavras-chave: doença crônica; depressão; ansiedade; qualidade de vida; psicoterapia; meta-análise multinível; variância robusta.

ABSTRACT

Objective: to examine the effects of psychological interventions on depression, anxiety, and quality of life in adults with chronic diseases while demonstrating a multilevel strategy that retains dependent estimates. Methods: a structured PubMed/MEDLINE search was conducted and contemporary systematic reviews reporting comparable estimates were selected. Fifteen standardized estimates from five reviews were extracted, covering cognitive behavioral therapy, self-management, organizational care, and digital modalities. Signs were recoded so that positive values indicated benefit. A random-effects model with review-clustered robust variance, small-sample correction, and leave-one-review-out analyses was fitted. Results: the exploratory synthesis yielded a standardized mean difference of 0.38 (robust 95% CI 0.20–0.56), with between-source heterogeneity (tau^2=0.015). All influence analyses remained favorable. Effects were larger for depression and anxiety in cognitive behavioral and guided self-management interventions, whereas quality-of-life gains were smaller and more design-sensitive. Conclusion: psychological interventions produced small-to-moderate benefits, but magnitude varied by outcome, format, and follow-up. Multilevel modeling and robust variance estimation reduced unnecessary data loss; however, reanalysis of aggregate estimates did not replace a trial-level meta-analysis.

Keywords: chronic disease; depression; anxiety; quality of life; psychotherapy; multilevel meta-analysis; robust variance estimation.

1 INTRODUÇÃO

A coexistência de doença crônica e sofrimento psíquico é frequente e clinicamente relevante. Depressão e ansiedade interferem na adesão medicamentosa, no autocuidado, na atividade física, no comparecimento a consultas e na percepção de sintomas, ao mesmo tempo em que dor, incapacidade, incerteza prognóstica e carga terapêutica alimentam sofrimento emocional. Essa relação bidirecional aparece em diabetes, doenças cardiovasculares, câncer, doença renal, doenças respiratórias, condições neurológicas e enfermidades autoimunes. Por isso, intervenções psicológicas deixaram de ser complemento periférico e passaram a integrar modelos de cuidado para pessoas que precisam manejar simultaneamente sintomas físicos, decisões terapêuticas e mudanças de vida.[1-4]

A literatura aponta benefícios de terapia cognitivo-comportamental, técnicas mente-corpo, autogerenciamento, intervenções pela internet e cuidado colaborativo, mas a comparação direta é difícil. Um mesmo ensaio pode relatar depressão, ansiedade, estresse, qualidade de vida física e mental, em múltiplos instrumentos e tempos de seguimento. Quando o revisor seleciona apenas um resultado de cada estudo, reduz dependência estatística, porém descarta informação clinicamente útil e pode introduzir escolha seletiva. Quando inclui todos os resultados como se fossem independentes, subestima a incerteza e atribui peso excessivo aos estudos mais prolíficos.[5-8]

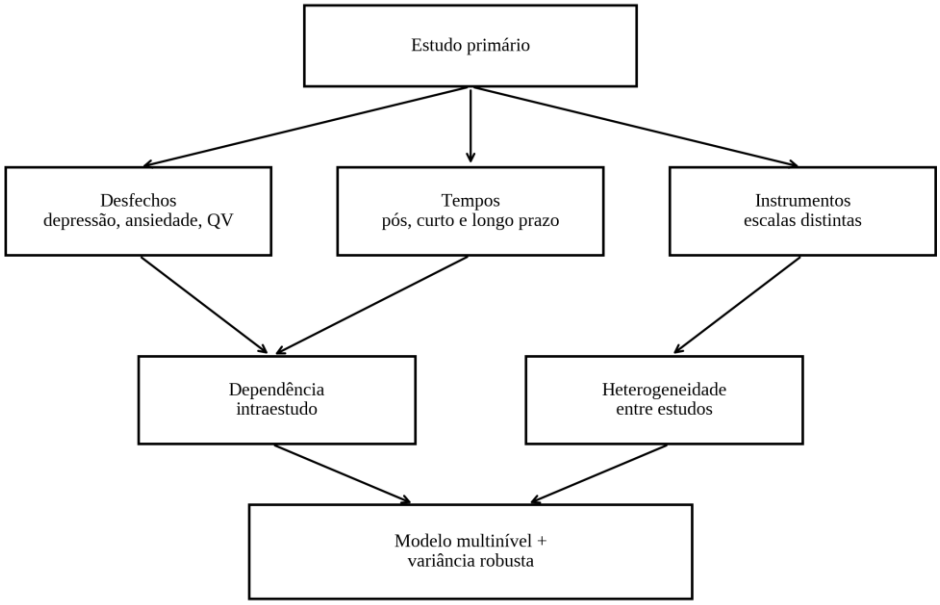
Modelos multiníveis e estimação robusta da variância foram desenvolvidos para esse problema. O modelo multinível representa a estrutura em que estimativas estão aninhadas em estudos e, eventualmente, em revisões, serviços ou populações. A variância robusta permite inferência menos dependente da correlação exata entre efeitos do mesmo conglomerado, desde que o número de conglomerados e as correções de pequena amostra sejam considerados. Essas abordagens não eliminam a necessidade de julgamento

clínico: resultados produzidos por escalas, comparadores e horizontes diferentes continuam exigindo avaliação de compatibilidade.[9-14]

Este capítulo teve dois objetivos complementares. O primeiro foi sintetizar a direção e a magnitude das evidências recentes sobre depressão, ansiedade e qualidade de vida em adultos com doenças crônicas. O segundo foi demonstrar, com estimativas reais publicadas, como uma estrutura multinível com variância robusta pode reter resultados dependentes sem produzir falsa precisão. A análise foi explicitamente tratada como reanálise exploratória de segunda ordem; portanto, suas estimativas não foram apresentadas como substitutas de uma futura base construída diretamente a partir de cada ensaio.

A Figura 1 apresenta a estrutura de dependência que motivou a escolha analítica. Ela foi inserida antes dos procedimentos para deixar claro que a unidade estatística não correspondeu simplesmente ao número de linhas da base, pois várias linhas pertenciam à mesma revisão e, nas revisões originais, ao mesmo estudo.

Figura 1. Estrutura hierárquica das estimativas em intervenções psicológicas para doenças crônicas.



Fonte: elaboração dos autores com base nos princípios de meta-análise multinível e estimação robusta da variância.[9-14]

A Figura 1 mostra que a dependência surgiu de três fontes principais: múltiplos desfechos, múltiplos tempos e múltiplos instrumentos. O modelo robusto foi usado para impedir que uma revisão com maior número de estimativas dominasse a síntese. A representação também evidencia que a correlação estatística e a comparabilidade clínica são problemas relacionados, mas não idênticos: a primeira foi tratada pelo modelo; a segunda, por estratificação e interpretação.

2 MÉTODOS

2.1 Delineamento, busca e elegibilidade

Foi realizada uma reanálise quantitativa exploratória de estimativas meta-analíticas publicadas, acompanhada de síntese metodológica. A busca foi conduzida em PubMed/MEDLINE até 1º de agosto de 2026 com combinações de

termos referentes a chronic disease, depression, anxiety, quality of life, psychological intervention, cognitive behavioral therapy, self-management, internet intervention e meta-analysis. As referências dos estudos elegíveis foram rastreadas. Foram priorizadas revisões sistemáticas recentes, com ensaios randomizados em adultos portadores de condições físicas crônicas, que apresentaram diferença média padronizada e intervalo de confiança para pelo menos um dos três domínios. A estratégia seguiu os princípios de transparência do PRISMA 2020, embora não tenha sido registrada como revisão sistemática de novo.[15]

Foram incluídas cinco revisões que forneceram estimativas numericamente recuperáveis e suficientemente próximas para uma demonstração transdiagnóstica: terapia cognitivo-comportamental em doenças crônicas; intervenções digitais mente-corpo; autogerenciamento em pessoas com doença física e sintomas depressivos; intervenções de atenção primária ou comunitária para multimorbidade envolvendo depressão ou ansiedade; e terapia cognitivo-comportamental pela internet em doença cardiovascular.[1-5] Revisões sem intervalo de confiança utilizável, populações exclusivamente psiquiátricas ou intervenções farmacológicas isoladas foram excluídas. Quando uma revisão apresentou mais de uma classe de intervenção ou tempo de seguimento, cada estimativa foi mantida e vinculada ao mesmo conglomerado.

A unidade da base foi a estimativa publicada, e o conglomerado foi a revisão. Essa decisão permitiu demonstrar dependência em uma base verificável, mas implicou sobreposição potencial de ensaios entre revisões. Por esse motivo, a análise foi classificada como segunda ordem, e nenhuma contagem agregada de participantes foi calculada. Os tamanhos amostrais e números de estudos foram descritos apenas para caracterizar cada fonte, sem soma entre revisões.

2.2 Extração, harmonização e avaliação crítica

Foram extraídos autor, ano, intervenção, população, número de estudos, tamanho amostral informado, domínio, tempo, estimativa e intervalo de confiança. As diferenças médias padronizadas foram recodificadas para que valores positivos indicassem redução de sintomas ou melhora da qualidade de vida. O erro-padrão foi reconstruído pela amplitude do intervalo de confiança, dividida por 3,92. Não foram convertidas razões de risco nem medidas não padronizadas. Efeitos corrigidos por Hedges e outras diferenças médias padronizadas foram tratados como aproximações compatíveis, e essa simplificação foi considerada na interpretação.

A qualidade das revisões foi apreciada por clareza da pergunta, busca, duplicidade de seleção, risco de viés, modelo estatístico, exploração da heterogeneidade e certeza da evidência. Não foi calculado escore somatório. A dependência clínica foi examinada por domínio e tempo, e a dependência estatística foi tratada por agrupamento da matriz de variância no nível da revisão. A presença de apenas cinco conglomerados exigiu correção de pequeno número e uso de distribuição t com quatro graus de liberdade.[10-13]

2.3 Análise estatística

Foi estimada a heterogeneidade por método de DerSimonian-Laird para fornecer um componente entre fontes. Em seguida, foi ajustado modelo de intercepto com pesos inversos da soma da variância amostral e tau^2. A covariância do coeficiente foi recalculada por estimador sanduíche agrupado por revisão, com correção $G/(G-1)$, em que G representou o número de revisões. O intervalo de confiança foi construído com quantil t. Foram produzidas sínteses descritivas por desfecho, sem interpretação confirmatória, porque cada estrato continha poucos conglomerados.

A robustez foi examinada por exclusão de uma revisão por vez. Também foram construídos mapa de evidências e decomposição aproximada de variância.

Não foi realizado teste de assimetria de funil, pois 15 estimativas dependentes provenientes de somente cinco fontes não atenderam às condições para interpretação confiável. Todas as análises foram reproduzidas em Python, e os princípios corresponderam aos implementados em metafor, robumeta e clubSandwich.[11-14,16,17]

3 RESULTADOS

3.1 Características da base

A Tabela 1 apresenta as cinco fontes e as estimativas retidas. A descrição foi colocada antes da tabela porque os números não representam cinco populações independentes: algumas revisões incluíram dezenas de ensaios e houve possibilidade de sobreposição. Assim, o objetivo da tabela foi documentar a proveniência dos efeitos, e não produzir uma soma artificial de participantes.

Tabela 1. Revisões e estimativas publicadas incluídas na reanálise exploratória.

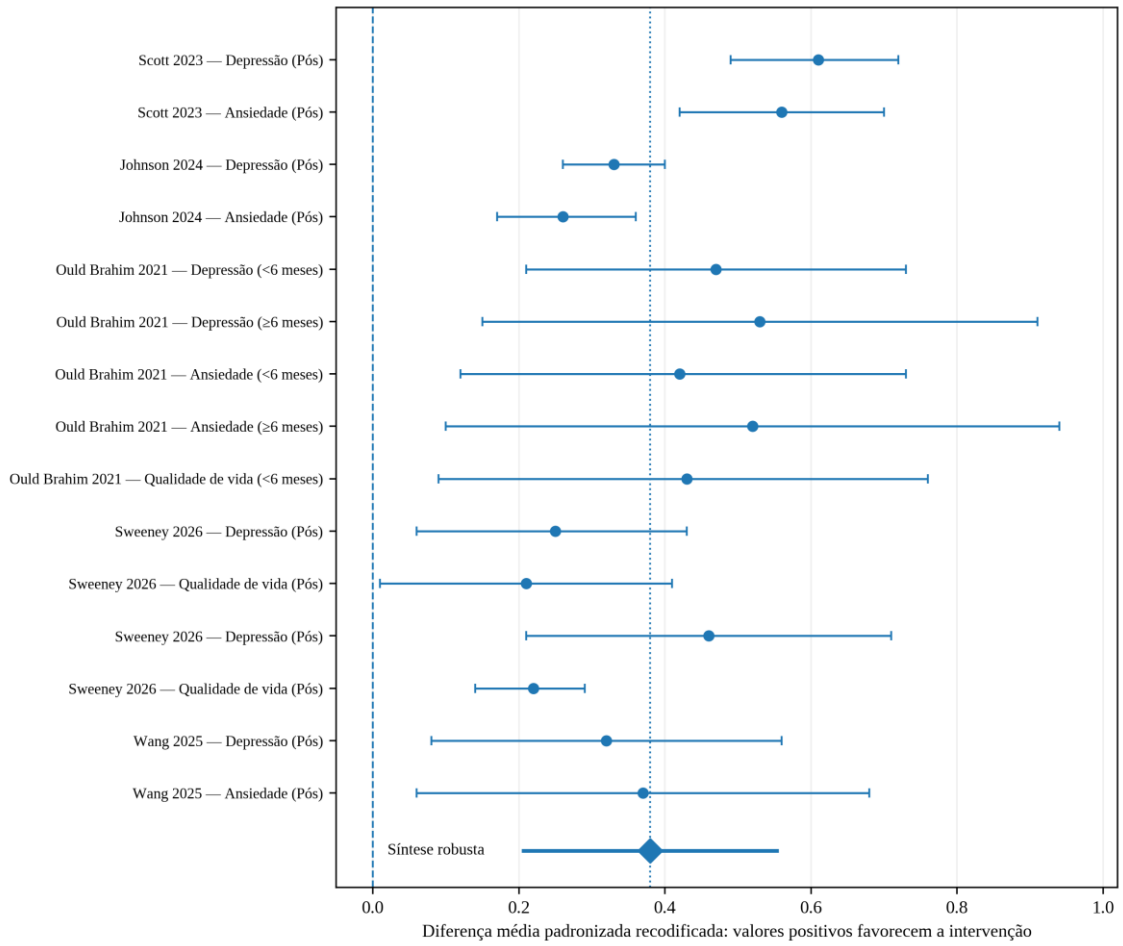
Fonte	Intervenção	Estudos; n informado	Estimativas favoráveis recodificadas
Scott et al., 2023	TCC em doenças crônicas	56; 7.691	Depressão: 0,61 (0,49–0,72); ansiedade: 0,56 (0,42–0,70)
Johnson et al., 2024	Intervenções digitais mente-corpo	56; 7.691	Depressão: 0,33 (0,26–0,40); ansiedade: 0,26 (0,17–0,36)
Ould Brahim et al., 2021	Autogerenciamento guiado	12; 2.064	Depressão: 0,47 e 0,53; ansiedade: 0,42 e 0,52; QV: 0,43
Sweeney et al., 2026	Intervenções organizacionais e orientadas ao paciente	29; 9.487	Depressão: 0,25 e 0,46; QV: 0,21 e 0,22
Wang et al., 2025	TCC pela internet em doença cardiovascular	8; 1.177	Depressão: 0,32 (0,08–0,56); ansiedade: 0,37 (0,06–0,68)

Fonte: dados publicados nas revisões incluídas.[1-5] QV: qualidade de vida; TCC: terapia cognitivo-comportamental. Os participantes não foram somados devido à possível sobreposição.

A Tabela 1 demonstra amplitude de modalidades e populações. Os maiores efeitos apareceram na revisão de terapia cognitivo-comportamental, enquanto intervenções digitais de baixa intensidade e mudanças organizacionais apresentaram efeitos menores. Essa diferença não deve ser atribuída apenas à potência terapêutica, pois comparadores, gravidade basal, suporte profissional e composição das doenças variaram entre fontes.

A Figura 2 mostra todas as 15 estimativas com seus intervalos. Sua finalidade foi permitir inspeção da dispersão antes de qualquer resumo global, evitando que a estimativa combinada escondesse a diferença entre domínios, tempos e formatos.

Figura 2. Estimativas publicadas de depressão, ansiedade e qualidade de vida mantidas na análise multinível.



Fonte: elaboração dos autores a partir das cinco revisões.[1-5] Valores positivos indicam benefício. A linha pontilhada representa a estimativa robusta global.

A Figura 2 revela que todas as estimativas apresentaram direção favorável, mas a precisão variou amplamente. Os intervalos mais estreitos vieram de revisões com maior volume de ensaios; os mais largos corresponderam a seguimentos tardios e subgrupos de autogerenciamento. A ausência de efeito negativo não elimina viés de publicação ou seletividade, sobretudo porque a unidade da reanálise foi o resultado agregado publicado.

3.2 Síntese robusta e resultados por domínio

A Tabela 2 apresenta o resultado do modelo e as decisões analíticas. A estimativa global de 0,38 foi interpretada como benefício pequeno a moderado, mas o intervalo robusto foi calculado com apenas cinco conglomerados. Por isso, a tabela explicita simultaneamente o resultado e suas restrições.

Tabela 2. Resultados do modelo multinível exploratório com variância robusta.

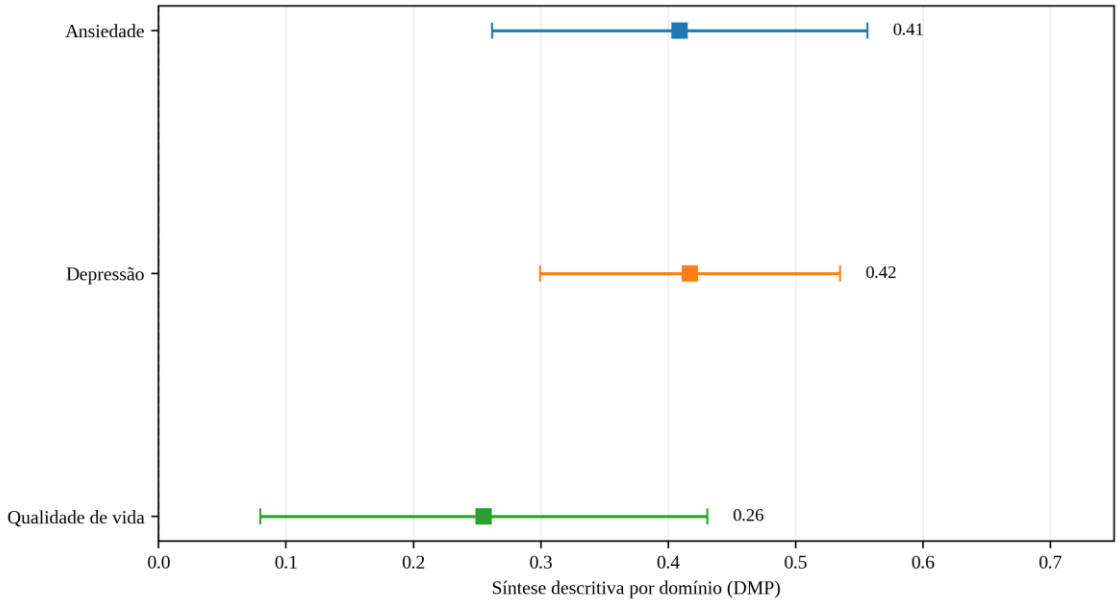
Parâmetro	Resultado	Interpretação
Estimativas / conglomerados	15 / 5	Múltiplos desfechos e tempos foram mantidos.
Efeito global robusto	0.38 (IC95% 0.20–0.56)	Valores positivos favoreceram intervenção.
Heterogeneidade	tau^2=0.015; Q=49,94	Variabilidade relevante entre fontes.
Graus de liberdade	4	Intervalo baseado em distribuição t.
Teste de assimetria	Não realizado	Número de conglomerados insuficiente.
Natureza da inferência	Exploratória de segunda ordem	Não substituiu extração de ensaios primários.

Fonte: cálculos dos autores com estimativas publicadas.[1-5]

A Tabela 2 mostra que o sinal favorável persistiu após correção da variância, mas também evidencia a incerteza metodológica. O intervalo robusto foi mais largo do que seria obtido tratando as 15 linhas como independentes. Essa diferença é precisamente a contribuição do método: reconhecer que novas linhas de um mesmo conglomerado acrescentam informação, porém não equivalem a novos estudos independentes.

A Figura 3 organiza as estimativas por domínio. Ela foi construída como síntese descritiva, e não como comparação formal, porque os estratos continham números diferentes de revisões e intervenções.

Figura 3. Magnitude descritiva dos efeitos por domínio psicológico.



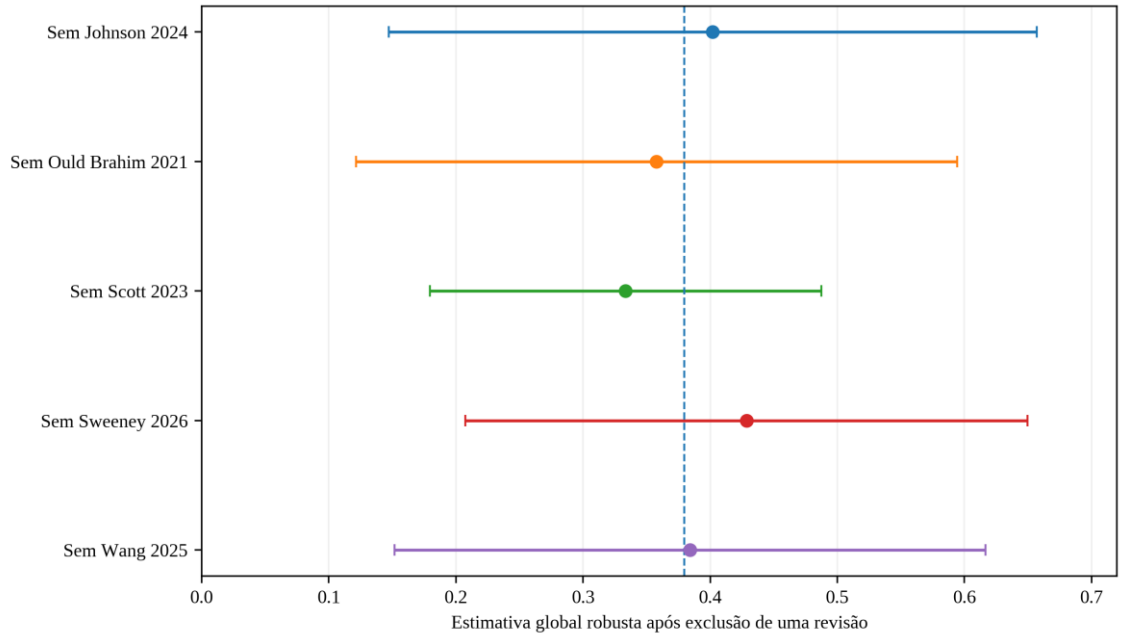
Fonte: cálculos dos autores. Intervalos por efeitos aleatórios dentro de cada domínio, sem correção confirmatória entre domínios.

A Figura 3 sugere maior magnitude para depressão, seguida por ansiedade, enquanto qualidade de vida apresentou efeito menor. Essa hierarquia é plausível: intervenções psicológicas foram desenhadas para modificar sintomas e habilidades de enfrentamento, ao passo que qualidade de vida depende também de dor, capacidade funcional, progressão orgânica, renda, apoio social e carga terapêutica. Portanto, melhora emocional não se converte automaticamente em mudança equivalente da qualidade de vida global.

3.3 Influência, cobertura e estrutura da evidência

A Figura 4 apresenta a análise de exclusão de uma revisão por vez. Essa etapa foi essencial porque, com cinco conglomerados, uma fonte de grande efeito poderia deslocar desproporcionalmente o resultado.

Figura 4. Análise de influência por exclusão de uma revisão.

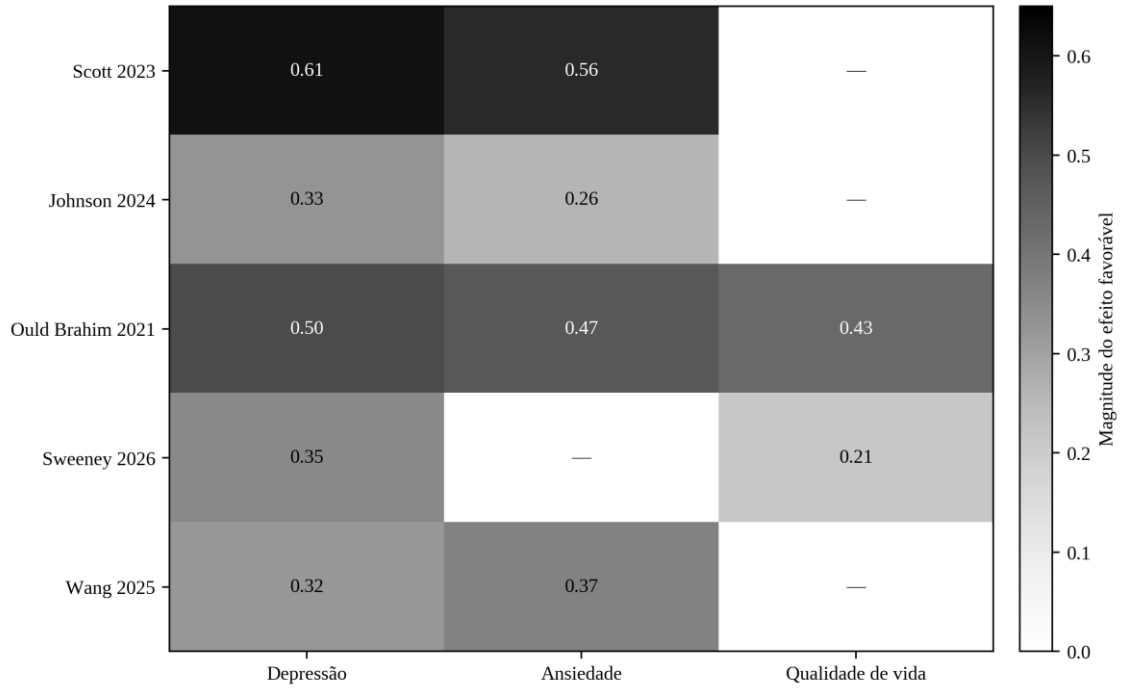


Fonte: cálculos dos autores. A linha tracejada representa a estimativa do modelo com todas as fontes.

A Figura 4 mostra que a estimativa permaneceu entre aproximadamente 0,30 e 0,43 após cada exclusão. A retirada de Scott reduziu o efeito, como esperado, porque essa revisão apresentou as maiores magnitudes. A retirada de Sweeney aumentou o resultado, pois aquela revisão incluiu intervenções organizacionais de menor intensidade. A estabilidade direcional foi favorável, mas os intervalos continuaram largos devido ao pequeno número de conglomerados.

A Figura 5 foi inserida para mostrar onde a evidência se concentrou. Uma meta-análise global pode parecer abrangente mesmo quando alguns domínios dependem de poucas fontes; o mapa torna essa assimetria visível.

Figura 5. Mapa de cobertura e magnitude das estimativas por revisão e domínio.



Fonte: elaboração dos autores. Células com travessão indicam ausência de estimativa elegível.

A Figura 5 demonstra que depressão foi o desfecho mais estudado e apareceu em todas as fontes. Ansiedade teve cobertura intermediária, enquanto qualidade de vida dependeu de duas revisões. Assim, a confiança na direção do efeito é maior para depressão e menor para qualidade de vida. O mapa também mostra que a literatura digital contribuiu principalmente para sintomas, e não para desfechos funcionais ou sociais mais amplos.

A Tabela 3 resume a interpretação clínica e metodológica. Ela foi apresentada após os resultados para distinguir o que a análise sustentou do que permaneceu como hipótese para estudos futuros.

Tabela 3. Interpretação integrada dos achados e prioridades de pesquisa.

Domínio	Síntese	Implicação
Depressão	Benefício consistente pequeno a moderado.	Selecionar pessoas com sintomas clinicamente relevantes; monitorar resposta.

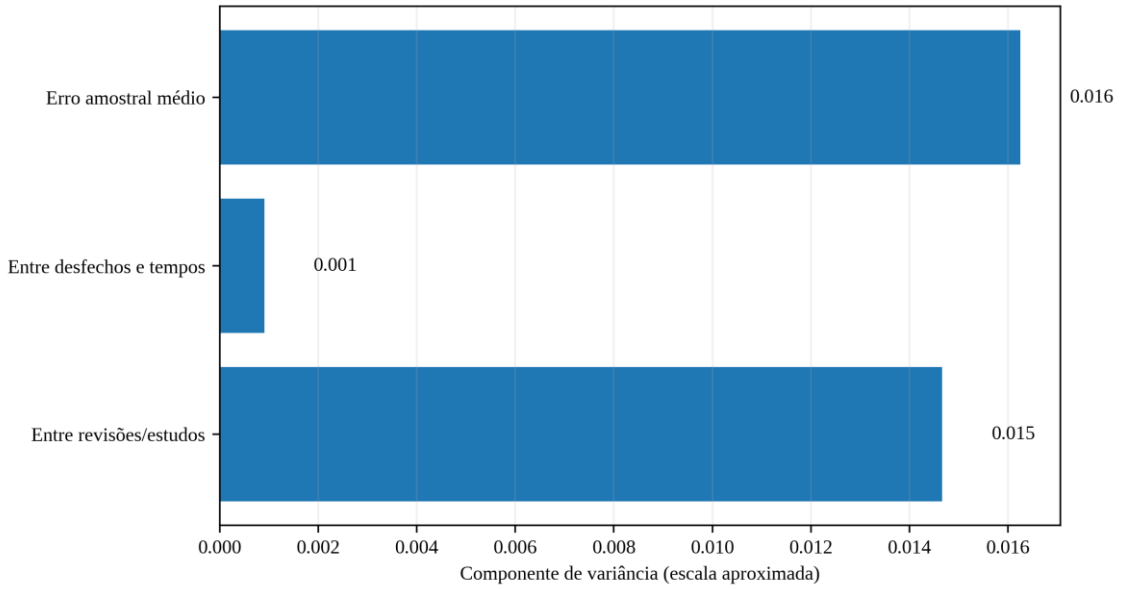
Domínio	Síntese	Implicação
Ansiedade	Benefício favorável, com menor número de fontes.	Adaptar conteúdo a incerteza, medo de sintomas e procedimentos.
Qualidade de vida	Ganho menor e dependente do contexto.	Combinar psicoterapia com reabilitação e manejo da doença.
Formato digital	Escalável, geralmente com efeitos menores.	Manter suporte humano proporcional à complexidade.
Longo prazo	Evidência mais imprecisa e possível atenuação.	Planejar reforços, manutenção e seguimento.
Método	RVE evitou falsa independência.	Construir base de ensaios com matrizes de correlação e IPD quando possível.

Fonte: interpretação dos autores com base nos resultados e na literatura metodológica.[1-14]

A Tabela 3 reforça que a indicação clínica não deve ser guiada apenas pelo tamanho médio. Intervenções de baixa intensidade podem ser adequadas para sintomas leves e acesso ampliado, enquanto sofrimento intenso, multimorbidade e risco funcional exigem maior integração com profissionais de saúde mental. A prioridade metodológica é coletar todos os desfechos e tempos dos ensaios, acompanhados de informações que permitam estimar suas correlações.

A Figura 6 encerra os resultados ao representar os componentes aproximados de variância. Ela não correspondeu a uma decomposição identificável de três níveis com dados individuais, mas serviu para ilustrar por que erro amostral e heterogeneidade não podem ser tratados como a mesma fonte de incerteza.

Figura 6. Componentes aproximados de variabilidade na reanálise.



Fonte: cálculos dos autores. A parcela entre desfechos e tempos foi apresentada de modo descritivo e não como estimativa causal.

A Figura 6 mostra que uma fração relevante da dispersão permaneceu entre fontes além do erro amostral. Isso é esperado em intervenções psicológicas aplicadas a doenças diferentes, com intensidades e comparadores distintos. A mensagem principal não é o valor exato de cada barra, mas a necessidade de modelos que reconheçam simultaneamente variabilidade clínica, dependência interna e imprecisão amostral.

4 DISCUSSÃO

A síntese indicou benefício pequeno a moderado das intervenções psicológicas para adultos com doenças crônicas, com maior consistência para depressão e ansiedade do que para qualidade de vida. A estimativa robusta global permaneceu favorável mesmo após a exclusão alternada de cada revisão. Esse resultado converge com revisões de terapia cognitivo-comportamental, autogerenciamento e intervenções digitais, mas não deve ser lido como uma constante aplicável a todas as doenças, pessoas e formatos. A heterogeneidade

observada expressa diferenças reais de mecanismo, dose, gravidade e contexto, e não apenas ruído estatístico.[1-5]

A maior magnitude encontrada para terapia cognitivo-comportamental é coerente com uma intervenção diretamente orientada a pensamentos, comportamentos, evitação, catastrofização e resolução de problemas. Pessoas com doença crônica frequentemente interpretam sintomas físicos por meio de ameaça e imprevisibilidade; a TCC pode reduzir ciclos de medo, inatividade e perda de autonomia. Entretanto, revisões de TCC também tendem a incluir programas mais estruturados e maior contato profissional do que intervenções digitais autoguiadas. Assim, parte da diferença pode refletir intensidade, seleção de participantes e qualidade de implementação, não apenas a escola psicoterapêutica.[1,3,6]

Os efeitos digitais foram menores, porém clinicamente relevantes diante da possibilidade de ampliar acesso. Intervenções pela internet reduzem deslocamento, permitem ritmo flexível e podem alcançar pessoas que vivem longe de centros especializados. Em contrapartida, exigem alfabetização digital, conexão, motivação e capacidade de transformar conteúdo em prática. Estudos que oferecem suporte humano, lembretes e monitoramento costumam apresentar maior adesão. A escalabilidade, portanto, não deve ser confundida com ausência de trabalho clínico; modelos híbridos podem preservar alcance sem abandonar personalização.[2,5,18]

O autogerenciamento apresentou efeitos favoráveis inclusive em seguimentos mais longos, mas com intervalos amplos. Sua utilidade provavelmente decorre da combinação entre educação, definição de metas, monitoramento de sintomas, resolução de problemas e apoio à autoeficácia. Esses componentes conectam saúde mental ao cotidiano da doença física. A melhora emocional pode surgir porque a pessoa recupera senso de controle, e não apenas porque recebe técnicas psicoterapêuticas formais. A distinção entre psicoterapia e autogerenciamento é, por isso, menos rígida na prática do que nas classificações das revisões.[3]

A qualidade de vida apresentou ganhos menores porque é um desfecho distal e multideterminado. Uma intervenção pode reduzir depressão sem restaurar função pulmonar, mobilidade ou renda. Também pode melhorar a capacidade de lidar com a doença sem alterar pontuações globais que agregam dimensões físicas pouco responsivas. O uso de medidas específicas da condição, participação social e alcance de metas pessoais pode captar benefícios que instrumentos genéricos diluem. Estudos futuros devem evitar tratar qualidade de vida como simples extensão dos sintomas emocionais.

A análise de Sweeney e colaboradores acrescentou uma distinção importante entre intervenções organizacionais e orientadas ao paciente. Mudanças organizacionais, como cuidado colaborativo e gestão de casos, produziram efeitos menores, mas podem atingir pessoas com maior complexidade e sustentar seguimento. Intervenções orientadas ao paciente produziram efeitos mais fortes, porém tenderam a atenuar ao longo do tempo. O contraste sugere que conteúdo terapêutico e arquitetura do serviço são complementares: a técnica inicia a mudança, enquanto coordenação, monitoramento e continuidade ajudam a mantê-la.[4]

Do ponto de vista estatístico, a principal contribuição foi mostrar como resultados dependentes alteram a incerteza. Se as 15 estimativas fossem analisadas como observações independentes, o intervalo seria artificialmente estreito. A variância robusta ampliou o intervalo e atribuiu a cada revisão o papel de conglomerado. Essa correção é particularmente importante em saúde mental, na qual ensaios usam várias escalas e tempos. Excluir resultados para obter independência simplifica a análise, mas empobrece a pergunta; ignorar a dependência produz confiança indevida. Modelos multiníveis oferecem um caminho intermediário.[9-14]

A robustez estatística, contudo, não corrigiu sobreposição de ensaios entre revisões nem substituiu dados primários. A base de segunda ordem foi útil para demonstrar o método com números verificáveis, mas revisões podem compartilhar estudos, classificar intervenções de forma diferente e aplicar

modelos distintos. Além disso, estimativas agregadas escondem correlações entre desfechos e não permitem examinar moderadores no nível do participante. Uma meta-análise definitiva deveria extrair todos os efeitos diretamente dos ensaios, identificar múltiplas publicações e construir matrizes de variância plausíveis.

O pequeno número de conglomerados foi outra limitação. Correções de pequena amostra e distribuição t reduziram a falsa precisão, mas cinco revisões continuam insuficientes para testar moderadores complexos ou assimetria de publicação. O tau² também foi estimado com incerteza. Por isso, as análises por domínio e os componentes de variância foram interpretados como descritivos. A estabilidade do leave-one-out sustentou a direção geral, não uma magnitude definitiva.

As implicações clínicas são pragmáticas. Serviços que cuidam de doenças crônicas devem rastrear sofrimento psíquico e oferecer intervenção proporcional à necessidade. Sintomas leves podem responder a programas digitais ou autogerenciamento apoiado; depressão persistente, ansiedade intensa, risco de suicídio, dor incapacitante ou multimorbidade exigem avaliação especializada e integração com a equipe clínica. O monitoramento deve incluir sintomas, função, adesão, eventos adversos e preferência do paciente, com possibilidade de intensificação quando a resposta for insuficiente.

Para pesquisa, a prioridade não é produzir mais uma média geral, mas explicar heterogeneidade. Ensaios devem registrar componentes, dose, treinamento dos profissionais, fidelidade, suporte digital, condição crônica, gravidade basal, desfechos funcionais e manutenção. Repositórios de dados e protocolos de compartilhamento poderiam permitir meta-análise de dados individuais, examinando quem se beneficia, em qual intensidade e por quanto tempo. A combinação de modelos multiníveis, RVE e análise de mediadores pode aproximar a síntese das decisões reais.

A gravidade basal provavelmente foi um dos moderadores mais importantes, embora não pudesse ser testada adequadamente na base agregada.

Pessoas selecionadas por depressão clinicamente relevante dispõem de maior margem para melhora do que amostras recrutadas apenas pela presença de doença física. Em contrapartida, sintomas muito graves podem indicar maior complexidade, risco, comorbidade psiquiátrica e necessidade de tratamento combinado. A relação entre gravidade e efeito, portanto, não deve ser presumida como linear. Ensaios futuros precisam relatar distribuição basal, proporção com diagnóstico formal, uso de psicofármacos e mudança clinicamente significativa, além da média do grupo. Esses dados permitiriam distinguir redução estatística de sintomas de recuperação funcional percebida pelo paciente.[4,6,18]

O tipo de doença crônica também pode modificar mecanismos e resultados. Em câncer, doença renal e insuficiência cardíaca, fadiga, dor, restrição funcional e incerteza prognóstica podem limitar a aplicação de tarefas comportamentais. Em diabetes e doença cardiovascular, a intervenção pode atuar diretamente sobre adesão, alimentação, atividade física e automonitoramento. Em condições neurológicas, alterações cognitivas e motoras podem exigir adaptação do material e participação de cuidadores. Uma abordagem transdiagnóstica é útil para ampliar acesso, mas precisa preservar módulos específicos para sintomas, tratamentos e decisões próprias de cada condição.

A dose terapêutica não foi representada por uma única medida. Número de sessões, duração, contato síncrono, suporte profissional, treinamento do terapeuta e atividades entre sessões compõem dimensões diferentes. Programas digitais de baixa intensidade podem produzir efeito menor por participante e, ainda assim, gerar maior impacto populacional quando alcançam pessoas que permaneceriam sem atendimento. Já intervenções intensivas podem oferecer maior mudança, porém com custos, filas e abandono. A escolha deve considerar eficiência, equidade e possibilidade de escalonamento, e não apenas a maior diferença média padronizada.

A relação terapêutica e a qualidade da implementação raramente foram incorporadas às meta-análises, embora possam explicar parte da

heterogeneidade. Protocolos semelhantes produzem resultados distintos quando profissionais recebem supervisão, quando há continuidade ou quando o conteúdo é adaptado à doença. Fidelidade não significa repetição rígida: envolve preservar mecanismos centrais enquanto se ajusta linguagem, ritmo e formato. Estudos devem medir competência, aliança, exposição efetiva, conclusão de módulos e uso das habilidades, permitindo que a síntese diferencie falha da intervenção de falha de implementação.

A segurança também merece maior visibilidade. Intervenções psicológicas são geralmente consideradas de baixo risco, mas podem produzir desconforto transitório, ativação emocional, frustração, sensação de culpa por metas não alcançadas e atraso na busca de cuidado especializado quando programas autoguiados são usados além de sua indicação. Ensaios devem relatar piora de sintomas, eventos adversos, crises, encaminhamentos e abandono relacionado ao tratamento. Em pessoas com doença avançada, a comunicação deve evitar transformar adaptação psicológica em responsabilidade individual por limitações determinadas pela própria condição e pelo sistema de saúde.

Uma agenda de síntese mais informativa deveria combinar três níveis. O primeiro estimaria o efeito médio com todos os desfechos dependentes. O segundo examinaria moderadores de estudo, como condição, formato, intensidade, comparador e tempo. O terceiro usaria dados individuais para testar gravidade, idade, sexo, vulnerabilidade social e multimorbidade. Modelos multiníveis, variância robusta e meta-análise de dados individuais não competem entre si; respondem a partes complementares da heterogeneidade. A atualização contínua da base, com protocolos prospectivos e código aberto, reduziria escolhas pós-hoc e permitiria incorporar novos ensaios sem reconstruir toda a revisão.

5 CONCLUSÃO

Intervenções psicológicas apresentaram efeitos favoráveis sobre depressão e ansiedade em adultos com doenças crônicas, enquanto os ganhos em qualidade de vida foram menores e dependentes do contexto. A reanálise robusta preservou múltiplos resultados e mostrou que reconhecer dependência amplia a incerteza de forma metodologicamente apropriada. O resultado global sustentou benefício pequeno a moderado, mas não autorizou uma prescrição uniforme. A escolha deve considerar gravidade, doença física, formato, suporte profissional e continuidade. Estudos futuros devem partir dos ensaios primários, registrar todos os desfechos e tempos e usar modelos multiníveis com correções robustas para responder não apenas se as intervenções funcionam, mas para quem, em quais condições e com qual durabilidade.

REFERÊNCIAS

1. Scott AJ, Bisby MA, Heriseanu AI, et al. Cognitive behavioral therapies for depression and anxiety in people with chronic disease: a systematic review and meta-analysis. Clin Psychol Rev. 2023;106:102353. doi:10.1016/j.cpr.2023.102353.

2. Johnson E, Corrick S, Isley S, et al. Mind-body internet and mobile-based interventions for depression and anxiety in adults with chronic physical conditions: a systematic review of RCTs. PLOS Digit Health. 2024;3(1):e0000435. doi:10.1371/journal.pdig.0000435.

3. Ould Brahim L, Lambert SD, Feeley N, et al. The effects of self-management interventions on depressive symptoms in adults with chronic physical disease(s) experiencing depressive symptomatology: a systematic review and meta-analysis. BMC Psychiatry. 2021;21:584. doi:10.1186/s12888-021-03504-8.

4. Sweeney K, Gilarova M, Ng L, et al. Primary care and community interventions for multimorbidity involving depression or anxiety: systematic review with meta-analysis. BMJ Med. 2026;5:e002400. doi:10.1136/bmjmed-2025-002400.

5. Wang S, et al. Effects of internet-based cognitive behavioral therapy on depression and anxiety in patients with cardiovascular disease: a systematic

review and meta-analysis. Front Psychiatry. 2025;15:1433558. doi:10.3389/fpsyt.2024.1433558.

6. White V, Linardon J, Stone JE, Holmes-Truscott E, Olive L, Mikocka-Walus A. Online psychological interventions to reduce symptoms of depression, anxiety, and general distress in those with chronic health conditions: a systematic review and meta-analysis of randomized controlled trials. Psychol Med. 2022;52(3):548-573. doi:10.1017/S0033291720002251.

7. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester: Wiley; 2009.

8. Cheung MWL. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. Psychol Methods. 2014;19(2):211-229. doi:10.1037/a0032968.

9. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. Res Synth Methods. 2011;2(1):61-76. doi:10.1002/jrsm.35.

10. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. 2010;1(1):39-65. doi:10.1002/jrsm.5.

11. Tanner-Smith EE, Tipton E. Robust variance estimation with dependent effect sizes: a practical consideration and a software tutorial in Stata and SPSS. Res Synth Methods. 2014;5(1):13-30. doi:10.1002/jrsm.1091.

12. Tipton E. Small sample adjustments for robust variance estimation with meta-regression. Psychol Methods. 2015;20(3):375-393. doi:10.1037/met0000011.

13. Tipton E, Pustejovsky JE. Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. J Educ Behav Stat. 2015;40(6):604-634. doi:10.3102/1076998615606099.

14. Pustejovsky JE, Tipton E. Meta-analysis with robust variance estimation: expanding the range of working models. Prev Sci. 2022;23:425-438. doi:10.1007/s11121-021-01246-3.

15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

16. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1-48. doi:10.18637/jss.v036.i03.

17. Fisher Z, Tipton E. robumeta: an R-package for robust variance estimation in meta-analysis. arXiv. 2015:1503.02220.

18. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry. 2021;78(4):361-371. doi:10.1001/jamapsychiatry.2020.4364.

19. Higgins JPT, Thomas J, Chandler J, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester: Wiley; 2019.

20. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. London: Cochrane; 2019.

21. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629-634. doi:10.1136/bmj.315.7109.629.

CAPÍTULO 10

DOENÇAS PERIODONTAIS E DOENÇAS CARDIOVASCULARES: ASSOCIAÇÕES EPIDEMIOLÓGICAS, MECANISMOS BIOLÓGICOS E SÍNTESE METANALÍTICA

Aline Pinto Monteiro Costa Sousa; Bruno Xavier de Araújo; Danilo Ferreira de Sousa; Conceição Soraya Morais Marques; Antônia Araújo Firmo; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Regina Petrola Bastos Rocha; Paula Andrea Travecedo Ramos; Nélio Barreto Vieira

RESUMO

As doenças periodontais compreendem condições inflamatórias que afetam os tecidos de proteção e de suporte dos dentes. Entre elas, a periodontite apresenta especial interesse sistêmico por combinar disbiose microbiana, inflamação crônica, ulceração do epitélio periodontal e exposição recorrente da circulação a mediadores inflamatórios e componentes bacterianos. A hipótese de associação com doença cardiovascular aterosclerótica vem sendo investigada há décadas, mas sua interpretação exige separar associação epidemiológica, plausibilidade biológica e causalidade. Este capítulo revisou criticamente a literatura e realizou uma reanálise metanalítica de estimativas ajustadas recuperáveis em coortes longitudinais. Foram formados dois conjuntos quantitativos: oito coortes para desfechos cardiovasculares predominantemente coronarianos e oito coortes para acidente vascular cerebral (AVC). As medidas relativas ajustadas foram transformadas para a escala logarítmica e combinadas por modelo de efeitos aleatórios com estimativa REML. Nos eventos cardiovasculares/coronarianos, a razão combinada foi 1,11 (IC95% 1,07–1,14), com heterogeneidade baixa (I^2=17,0%) e intervalo de predição de 1,06–1,16. Para AVC, a razão combinada foi 1,24 (IC95% 1,11–1,38), com heterogeneidade elevada (I^2=74,7%) e intervalo de predição de 0,95–1,62. As análises de exclusão sucessiva de estudos mantiveram os efeitos combinados acima da nulidade. A evidência é, portanto, consistente com maior risco cardiovascular entre pessoas

com periodontite, sobretudo para eventos ateroscleróticos e cerebrovasculares, mas permanece insuficiente para afirmar que a periodontite seja causa independente em todos os contextos ou que seu tratamento reduza eventos cardiovasculares maiores. O manejo periodontal deve integrar a promoção global da saúde, com controle do tabagismo, diabetes, hipertensão e demais fatores de risco, sem substituir estratégias cardiovasculares comprovadas.

Palavras-chave: periodontite; doença periodontal; doenças cardiovasculares; doença arterial coronariana; acidente vascular cerebral; metanálise.

1 INTRODUÇÃO

A periodontite deixou de ser compreendida apenas como uma infecção localizada da cavidade oral. A classificação contemporânea a define como uma doença inflamatória crônica multifatorial associada a biofilme disbiótico, caracterizada por destruição progressiva do aparato de inserção periodontal e perda de osso alveolar. Essa definição é importante porque desloca a interpretação do simples acúmulo de bactérias para uma interação entre microbiota, resposta imune do hospedeiro, fatores ambientais e condições sistêmicas. A gengivite, embora também seja inflamatória, não implica necessariamente perda de inserção e é geralmente reversível quando o biofilme é controlado; por isso, não deve ser utilizada de forma intercambiável com periodontite em estudos sobre risco cardiovascular (Papapanou et al., 2018; Sanz et al., 2020).

O interesse pelas repercussões cardiovasculares surgiu da constatação de que pessoas com periodontite apresentam, em diferentes populações, maior frequência de doença arterial coronariana, infarto do miocárdio, acidente vascular cerebral e mortalidade cardiovascular. O problema científico, entretanto, é mais complexo do que uma simples comparação entre indivíduos com e sem doença periodontal. Tabagismo, diabetes, idade, obesidade, padrão alimentar, acesso a serviços de saúde, renda e escolaridade influenciam simultaneamente a saúde periodontal e o risco vascular. A associação pode,

portanto, refletir em parte mecanismos biológicos compartilhados, causalidade bidirecional indireta ou confusão residual, mesmo após ajustes estatísticos extensos.

A primeira declaração científica da American Heart Association dedicada ao tema, publicada em 2012, reconheceu a existência de associação observacional entre doença periodontal e doença vascular aterosclerótica, mas considerou insuficiente a evidência de causalidade ou de prevenção de eventos cardiovasculares pela terapia periodontal. Desde então, o corpo de evidências aumentou de forma expressiva. Revisões sistemáticas, coortes populacionais, estudos de imagem vascular, análises genéticas e ensaios de tratamento periodontal permitiram uma leitura mais refinada. A atualização da American Heart Association publicada em 2026 manteve a distinção entre associação e causalidade e reforçou a necessidade de interpretar conjuntamente dados epidemiológicos, mecanismos e intervenção (Lockhart et al., 2012; Tran et al., 2026).

Em paralelo, a literatura periodontal avançou na padronização diagnóstica e no entendimento da chamada carga inflamatória periodontal. Bolsas profundas, sangramento à sondagem, perda de inserção clínica e perda óssea não representam exatamente o mesmo componente da doença. Algumas medidas refletem atividade inflamatória atual, enquanto outras expressam dano acumulado ao longo da vida. Essa heterogeneidade de exposição ajuda a explicar por que estimativas de risco variam entre estudos e por que perdas dentárias isoladas, apesar de frequentemente associadas a eventos cardiovasculares, não devem ser tomadas automaticamente como marcador específico de periodontite.

A síntese mais recente da literatura também exige cautela com o tamanho do efeito. Associações relativas de 10% a 30% podem ser relevantes em condições muito prevalentes, mas são suscetíveis a alterações quando o controle de tabagismo, diabetes e posição socioeconômica é aprimorado. Por outro lado, a repetição de associações em coortes de diferentes países e a existência de vias biológicas plausíveis tornam inadequado descartar o fenômeno como mero

artefato. O ponto central passa a ser determinar quanto do excesso de risco é independente, em quais grupos ele é mais pronunciado e se o controle periodontal modifica desfechos vasculares clinicamente importantes.

Diante dessas questões, este capítulo apresentou uma revisão crítica e uma reanálise metanalítica atualizada, priorizando estudos longitudinais com estimativas ajustadas e evitando equiparar periodontite a indicadores inespecíficos de saúde oral. A estratégia buscou preservar a utilidade clínica dos resultados sem extrapolar o que os desenhos observacionais podem demonstrar.

2 DOENÇAS PERIODONTAIS: CONCEITOS CLÍNICOS RELEVANTES PARA A INTERPRETAÇÃO SISTÊMICA

A expressão "doença periodontal" abrange um conjunto de condições distintas. A gengivite induzida por biofilme é limitada aos tecidos gengivais e não produz, por definição, a perda de inserção característica da periodontite. A periodontite envolve destruição do ligamento periodontal e do osso alveolar, formação de bolsas e, em formas avançadas, mobilidade e perda dentária. Na classificação de 2017, incorporada à prática a partir de 2018, o estágio expressa gravidade e complexidade do manejo, enquanto o grau procura estimar velocidade de progressão e efeitos de modificadores de risco, como tabagismo e diabetes (Papapanou et al., 2018).

A distinção entre condição atual e dano acumulado é especialmente relevante em epidemiologia cardiovascular. Sangramento à sondagem e profundidade de bolsa tendem a captar inflamação presente; perda de inserção clínica e perda óssea radiográfica registram parte da história destrutiva; perda dentária e edentulismo podem resultar de periodontite, mas também de cárie, trauma, acesso tardio ao cuidado e decisões terapêuticas. Quanto mais indireta a medida de exposição, maior a possibilidade de erro de classificação e de influência socioeconômica. Por esse motivo, a metanálise principal deste capítulo

não incluiu estudos em que a única exposição disponível era número de dentes ou edentulismo.

A Tabela 1 organiza esses marcadores e explicita como cada um deve ser interpretado quando o objetivo é relacionar a condição periodontal a desfechos cardiovasculares. A leitura dessa tabela evita um problema frequente em revisões: agrupar indicadores com significados biológicos diferentes como se fossem equivalentes.

Tabela 1 – Marcadores periodontais e implicações para estudos cardiovasculares

Marcador/condição	Significado clínico	Implicação epidemiológica
Gengivite	Inflamação gengival induzida por biofilme, sem perda de inserção periodontal atribuível à periodontite.	Marcador de inflamação oral, mas não representa destruição periodontal; não foi equiparada a periodontite na síntese principal.
Periodontite	Doença inflamatória crônica multifatorial associada a biofilme disbiótico, com perda de inserção e osso alveolar.	Exposição principal de interesse por combinar inflamação crônica, superfície ulcerada e dano tecidual.
Profundidade de bolsa / sangramento	Medidas clínicas relacionadas à inflamação e ao ambiente subgengival atual.	Úteis para estimar carga inflamatória presente; variam conforme protocolo de sondagem e tratamento prévio.
Perda de inserção / perda óssea	Indicadores do dano periodontal acumulado ao longo do tempo.	Podem refletir exposição crônica, porém não distinguem atividade atual de sequela histórica.
Perda dentária / edentulismo	Desfechos orais multifatoriais e tardios.	Podem acompanhar periodontite grave, mas são inespecíficos; foram tratados como evidência complementar, não como definição automática de periodontite.
Periodontite grave/estágios avançados	Maior destruição e complexidade, frequentemente com bolsas profundas, perda óssea extensa e perda dentária por periodontite.	Apresenta plausibilidade para maior carga sistêmica; comparações entre estudos dependem da definição usada em cada época.

A tabela mostra que a qualidade da evidência não depende apenas do tamanho da amostra, mas também da precisão com que a exposição periodontal foi caracterizada. Estudos com exame periodontal padronizado tendem a oferecer melhor especificidade diagnóstica, enquanto autorrelato favorece grandes coortes e longo seguimento, mas introduz classificação incorreta. Estudos baseados em códigos hospitalares identificam, em geral, casos mais graves e podem não representar a periodontite atendida rotineiramente na atenção odontológica.

3 BASES BIOLÓGICAS DA ASSOCIAÇÃO COM DOENÇAS CARDIOVASCULARES

A plausibilidade biológica da relação periodontite–doença cardiovascular pode ser organizada em duas vias que se sobrepõem. A primeira é uma via direta, associada à disseminação transitória de microrganismos periodontais e de seus componentes. A escovação, mastigação e procedimentos periodontais podem produzir bacteremia transitória, especialmente quando há inflamação e ulceração do epitélio de bolsas. DNA bacteriano e componentes de microrganismos orais já foram identificados em tecidos vasculares, mas sua presença não demonstra, isoladamente, que a periodontite seja causa suficiente de aterosclerose. O achado é mais adequadamente interpretado como evidência de acesso potencial ao compartimento vascular (Sanz et al., 2020; Tran et al., 2026).

A segunda é a via inflamatória sistêmica. A periodontite ativa pode aumentar a liberação de citocinas e mediadores pró-inflamatórios, contribuir para elevação de proteína C reativa e interleucina 6 e favorecer disfunção endotelial. Inflamação sistêmica persistente participa de etapas centrais da aterogênese, desde ativação endotelial e recrutamento leucocitário até instabilidade de placa. A convergência desses processos torna biologicamente razoável que uma carga periodontal inflamatória prolongada possa contribuir,

em magnitude variável, para um ambiente vascular pró-aterogênico e pró-trombótico.

Também foram propostos mecanismos relacionados a estresse oxidativo, alteração do metabolismo lipídico, ativação plaquetária, respostas imunes cruzadas e mudanças na composição de células inflamatórias circulantes. A relevância de cada mecanismo em humanos permanece desigual. Ensaios de tratamento fornecem evidência de que a redução da inflamação periodontal pode melhorar alguns marcadores vasculares intermediários, mas os efeitos sobre biomarcadores e função endotelial não são uniformes e ainda não há base para afirmar prevenção de infarto ou AVC como consequência do tratamento periodontal.

A Figura 1 resume as rotas mais aceitas e, ao mesmo tempo, destaca os fatores compartilhados que podem produzir associação sem causalidade direta. Esse componente é essencial: a mesma pessoa que apresenta tabagismo persistente, diabetes mal controlado e menor acesso ao cuidado preventivo pode acumular simultaneamente maior destruição periodontal e maior risco cardiovascular.

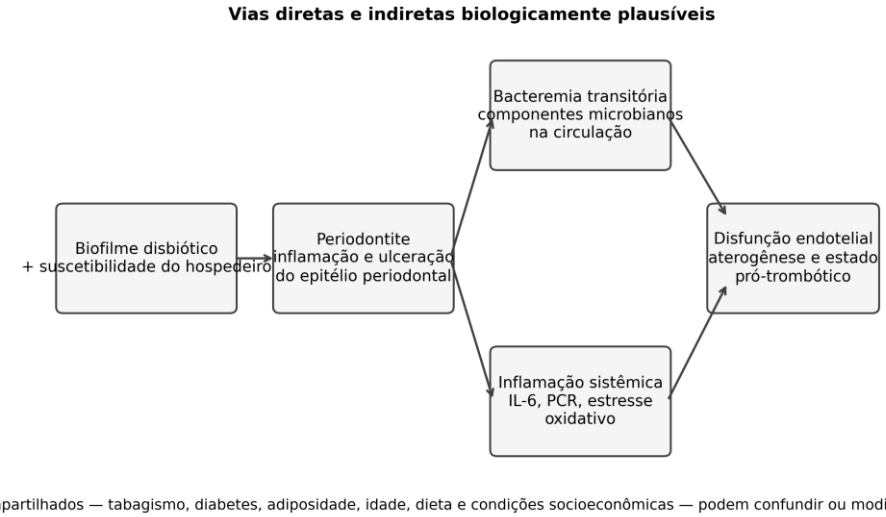


Figura 1 – Modelo conceitual das vias que podem ligar periodontite e aterosclerose.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

A figura evidencia que não existe uma única "via periodontal" para doença cardiovascular. A interpretação mais compatível com a literatura atual é a de múltiplas rotas convergentes, nas quais a inflamação periodontal pode somar-se a fatores cardiovasculares tradicionais. Essa leitura é coerente com os consensos da European Federation of Periodontology e com a declaração científica da American Heart Association de 2026, que reconhecem a consistência da associação, mas não transformam a periodontite em fator causal estabelecido no mesmo nível de hipertensão, tabagismo ou dislipidemia (Herrera et al., 2023; Tran et al., 2026).

4 OBJETIVO

Analisar criticamente a associação entre doenças periodontais, com ênfase em periodontite, e doenças cardiovasculares, integrando fundamentos clínicos e mecanismos biológicos a uma síntese metanalítica de estudos longitudinais sobre eventos cardiovasculares/coronarianos e acidente vascular cerebral, além de

discutir a evidência disponível sobre o efeito do tratamento periodontal em desfechos vasculares intermediários e clínicos.

5 MÉTODO DA SÍNTESE METANALÍTICA

5.1 Estratégia de identificação e seleção dos estudos

Foi realizada uma atualização bibliográfica estruturada em PubMed/MEDLINE, complementada pela verificação das referências de revisões sistemáticas, metanálises, consensos e declarações científicas relevantes. A atualização contemplou publicações disponíveis até agosto de 2026. Foram utilizados, em combinações, os termos "periodontitis", "periodontal disease", "coronary heart disease", "coronary artery disease", "myocardial infarction", "cardiovascular disease", "stroke", "ischemic stroke", "cohort" e "prospective". A busca foi orientada para recuperar estudos primários longitudinais e permitir conferência das estimativas usadas na síntese quantitativa.

A análise quantitativa foi definida como uma reanálise metanalítica dirigida e atualizada, e não como uma nova revisão sistemática exaustiva com contagem PRISMA própria. Essa decisão evitou apresentar um fluxo de seleção artificial quando a finalidade do capítulo era integrar criticamente estudos já identificados e verificáveis na literatura. Revisões de alta abrangência, como as de Humphrey et al. (2008), Larvin et al. (2021), Leng et al. (2023), Arbildo-Vega et al. (2024) e Meng e Chen (2025), foram utilizadas para checagem de cobertura, mas as estimativas da metanálise deste capítulo foram extraídas de estudos primários sempre que estavam recuperáveis.

5.2 Critérios de elegibilidade

Foram considerados elegíveis para a metanálise principal estudos de coorte ou seguimento longitudinal em adultos que compararam participantes com e sem periodontite/doença periodontal e relataram evento cardiovascular

incidente, doença coronariana/infarto ou AVC, com medida relativa ajustada e intervalo de confiança de 95%. Foram priorizadas medidas ajustadas pelo maior conjunto pertinente de confundidores apresentado pelos autores. Quando um estudo relatou estimativas para doença periodontal prevalente e incidente, foi selecionada a estimativa prevalente para manter comparabilidade com as coortes que classificaram a exposição no início do seguimento.

Foram excluídos da combinação principal estudos exclusivamente transversais, caso-controle, séries clínicas, análises de mortalidade sem eventos incidentes distinguíveis e estudos cuja única exposição fosse perda dentária ou edentulismo. Estudos de recorrência após infarto ou AVC também não foram combinados com coortes de primeiro evento. Quando uma mesma população originou publicações relacionadas, foi escolhida a estimativa mais diretamente compatível com o desfecho e a exposição da análise, reduzindo a possibilidade de dupla contagem.

5.3 Extração e tratamento das estimativas

Foram extraídos autor, ano, população, método de avaliação periodontal, desfecho cardiovascular, duração aproximada do seguimento, medida de associação e intervalo de confiança. Hazard ratios (HR) e riscos relativos (RR) ajustados foram mantidos como medidas relativas e transformados para o logaritmo natural. O erro-padrão de cada log-estimativa foi reconstruído a partir dos limites do intervalo de confiança de 95%, usando a relação SE = [ln(limite superior) − ln(limite inferior)]/(2×1,96). As medidas não foram convertidas artificialmente em odds ratios.

Foram formados dois blocos quantitativos. O primeiro reuniu oito coortes com desfechos cardiovasculares ateroscleróticos predominantemente coronarianos, incluindo doença coronariana, infarto e desfechos cardiovasculares maiores quando essa era a estimativa longitudinal diretamente reportada. O segundo reuniu oito coortes com AVC, prioritariamente isquêmico quando

especificado. A heterogeneidade clínica residual entre definições foi mantida explícita e considerada na interpretação.

5.4 Análise estatística

As estimativas foram combinadas pelo modelo de efeitos aleatórios com variância entre estudos estimada por restricted maximum likelihood (REML). Foram calculados efeito combinado, intervalo de confiança de 95%, estatística Q de Cochran, I^2, τ^2 e intervalo de predição de 95%. O intervalo de predição foi incluído porque responde a uma pergunta diferente do intervalo de confiança: indica a faixa plausível para o verdadeiro efeito de um novo estudo semelhante, incorporando a heterogeneidade observada. As análises foram recalculadas em Python 3 com rotinas numéricas baseadas em NumPy e SciPy.

Foi realizada análise de sensibilidade por exclusão sucessiva de um estudo (leave-one-out). Também foi produzida síntese cumulativa por ano de publicação para examinar a estabilidade temporal do efeito. Testes formais de assimetria de funil não foram empregados, pois cada bloco continha menos de dez estudos; nessa situação, a baixa potência e a influência da heterogeneidade podem produzir conclusões enganosas. A interpretação estatística foi acompanhada de avaliação clínica e metodológica, sem considerar significância estatística como equivalente a causalidade ou relevância clínica.

6 RESULTADOS

6.1 Estudos de eventos cardiovasculares e coronarianos

O primeiro conjunto incluiu coortes publicadas entre 1993 e 2024, cobrindo desde estudos clássicos com exame odontológico até grandes bases contemporâneas com autorrelato e integração genética. A Tabela 2 apresenta as estimativas efetivamente utilizadas. O conjunto foi intencionalmente restrito a exposições descritas como doença periodontal ou periodontite, evitando

adicionar estudos de perda dentária isolada apenas para aumentar o número de participantes.

Tabela 2 – Coortes utilizadas na síntese de eventos cardiovasculares/coronarianos

Estudo	População	Exposição periodontal	Desfecho/seguimento	Estimativa ajustada
DeStefano et al., 1993	Estados Unidos; n=9.760	Exame periodontal	Doença coronariana; seguimento ~16 anos	RR 1,25 (1,06–1,48)
Hujoel et al., 2000	Estados Unidos; n=8.032	Exame odontológico NHANES I	Doença coronariana	HR 1,14 (0,96–1,36)
Howell et al., 2001	Estados Unidos; n=22.037 homens	História autorreferida	Eventos cardiovasculares; ~12 anos	RR 1,01 (0,88–1,15)
Dietrich et al., 2008	Estados Unidos; n=1.203 homens	Exame clínico	Doença coronariana; seguimento prolongado	RR 1,31 (0,96–1,81)
Yu et al., 2015	Estados Unidos; n=39.863 mulheres	Periodontite prevalente autorreferida	Infarto do miocárdio; média ~15,7 anos	HR 1,27 (1,04–1,56)
Hansen et al., 2016	Dinamarca; coorte nacional	Diagnóstico hospitalar de periodontite	Infarto do miocárdio; até 15 anos	RR/IRR 1,16 (1,04–1,30)
LaMonte et al., 2017	Estados Unidos; n=57.001 mulheres	História de periodontite	Doença cardiovascular incidente; ~12 anos	HR 1,11 (1,03–1,20)
Sanghvi et al., 2024	UK Biobank; n=481.915	Doença periodontal autorreferida	Doença arterial coronariana; mediana ~13,8 anos	HR 1,09 (1,07–1,13)

Apesar das diferenças de época e mensuração, a maior parte das estimativas ficou acima de 1,00. Os estudos de Hujoel, Howell e Dietrich apresentaram intervalos que incluíram a nulidade individualmente, enquanto DeStefano, Yu, Hansen, LaMonte e Sanghvi indicaram excesso de risco após

ajuste. A grande coorte do UK Biobank forneceu a estimativa mais precisa e, por consequência, contribuiu fortemente para a ponderação estatística; por isso, a análise de sensibilidade sem esse estudo foi examinada separadamente.

A Figura 2 apresenta o forest plot do primeiro bloco. Cada ponto representa a estimativa ajustada de um estudo e a linha horizontal mostra seu intervalo de confiança. O losango inferior resume o efeito pelo modelo de efeitos aleatórios. A linha vertical em 1,00 representa ausência de associação.

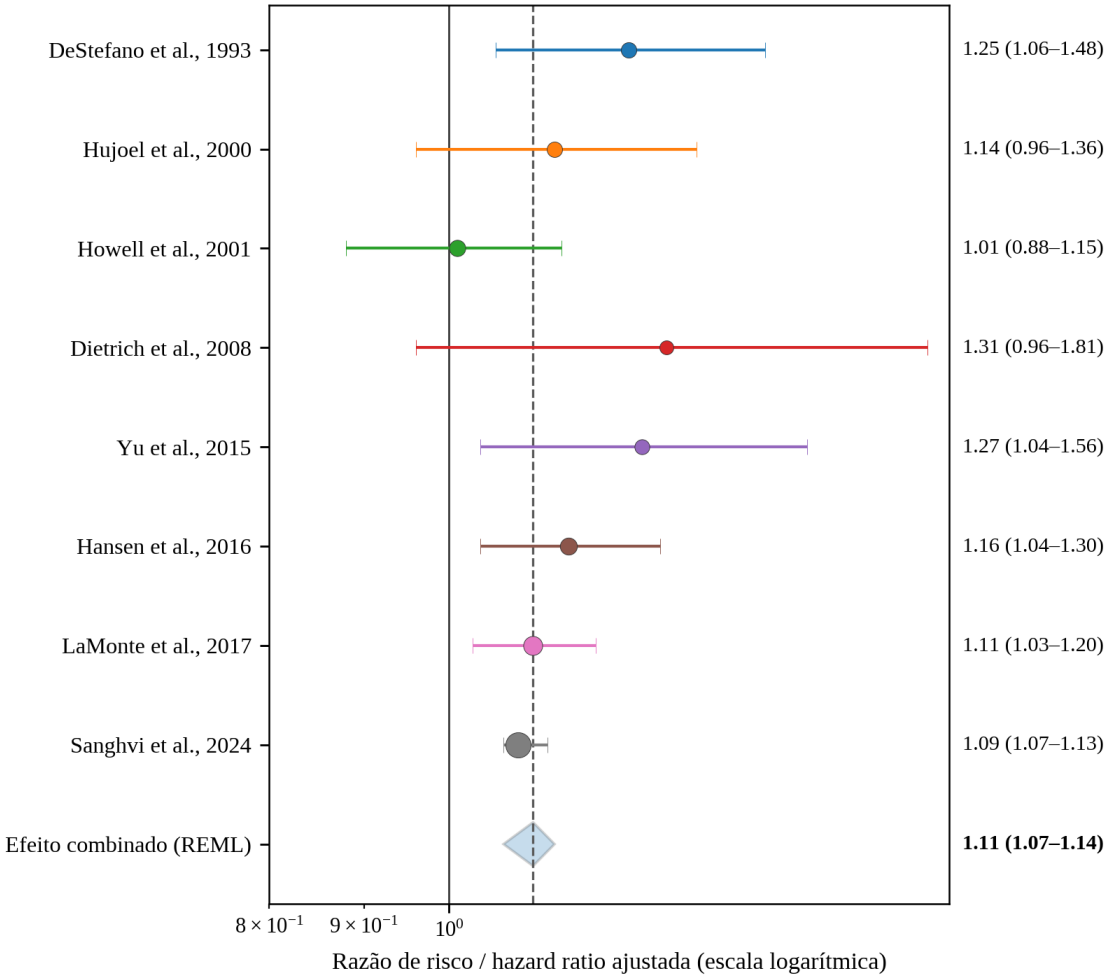


Figura 2 – Metanálise de periodontite/doença periodontal e eventos cardiovasculares predominantemente coronarianos.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

O efeito combinado foi 1,11 (IC95% 1,07–1,14; p<0,001), correspondente a um aumento relativo médio de aproximadamente 11% nos eventos cardiovasculares/coronarianos entre participantes com doença periodontal, dentro das definições utilizadas nas coortes. A heterogeneidade foi baixa (I^2=17,0%; Q=8,43; p=0,296; τ^2=0,0003). O intervalo de predição, de 1,06 a 1,16, permaneceu acima da nulidade, sugerindo que, para estudos semelhantes aos incluídos, o efeito verdadeiro esperado tenderia a continuar positivo. Esse resultado não significa que 11% dos eventos sejam causados pela periodontite; expressa uma associação relativa residual após os ajustes empregados em cada coorte.

6.2 Estudos de acidente vascular cerebral

A associação com AVC apresentou maior variação entre estudos. Foram incluídas coortes com homens e mulheres, bases administrativas e estudos com avaliação periodontal clínica ou autorreferida. A Tabela 3 mostra que as estimativas variaram de valores próximos da nulidade a aumentos superiores a 50%, o que antecipa a heterogeneidade observada na combinação.

Tabela 3 – Coortes utilizadas na síntese de acidente vascular cerebral

Estudo	População	Exposição periodontal	Desfecho/seguimento	Estimativa ajustada
Howell et al., 2001	Estados Unidos; n=22.037 homens	História de doença periodontal	AVC	HR/RR 1,01 (0,81–1,27)
Joshipura et al., 2003	Estados Unidos; n=41.380 homens	História de periodontite	AVC isquêmico; 12 anos	HR 1,33 (1,03–1,70)
Jimenez et al., 2009	Estados Unidos; n=1.231 homens	Exame periodontal/história de periodontite	Doença cerebrovascular; ~34 anos	HR 1,07 (0,90–1,27)
Lee et al., 2013	Taiwan; grande coorte populacional	Diagnóstico/tratamento periodontal	AVC isquêmico; ~11 anos	HR 1,33 (1,29–1,37)

Estudo	População	Exposição periodontal	Desfecho/seguimento	Estimativa ajustada
Yu et al., 2015	Estados Unidos; n=39.863 mulheres	Periodontite prevalente autorreferida	AVC isquêmico	HR 1,12 (0,91–1,37)
Hansen et al., 2016	Dinamarca; coorte nacional	Diagnóstico hospitalar de periodontite	AVC isquêmico; até 15 anos	IRR 1,51 (1,38–1,65)
LaMonte et al., 2017	Estados Unidos; n=57.001 mulheres	História de periodontite	AVC; ~12 anos	HR 1,11 (0,95–1,30)
Wood et al., 2025	ARIC/DARIC; n=5.986	Doença periodontal	AVC isquêmico; ~20 anos	HR 1,44 (1,09–1,91)

A tabela reforça dois aspectos. Primeiro, a direção predominante do efeito foi positiva. Segundo, a magnitude variou consideravelmente conforme definição periodontal, população e método de identificação do AVC. A coorte de Hansen, baseada em diagnóstico hospitalar de periodontite, provavelmente selecionou formas mais graves do que estudos de autorrelato; a coorte de Lee, por sua vez, utilizou dados administrativos em escala populacional. Essas diferenças são clinicamente plausíveis fontes de heterogeneidade.

A Figura 3 apresenta a combinação dos oito estudos de AVC. O modelo de efeitos aleatórios foi particularmente importante aqui, pois a dispersão entre estimativas foi maior do que no bloco cardiovascular/coronariano.

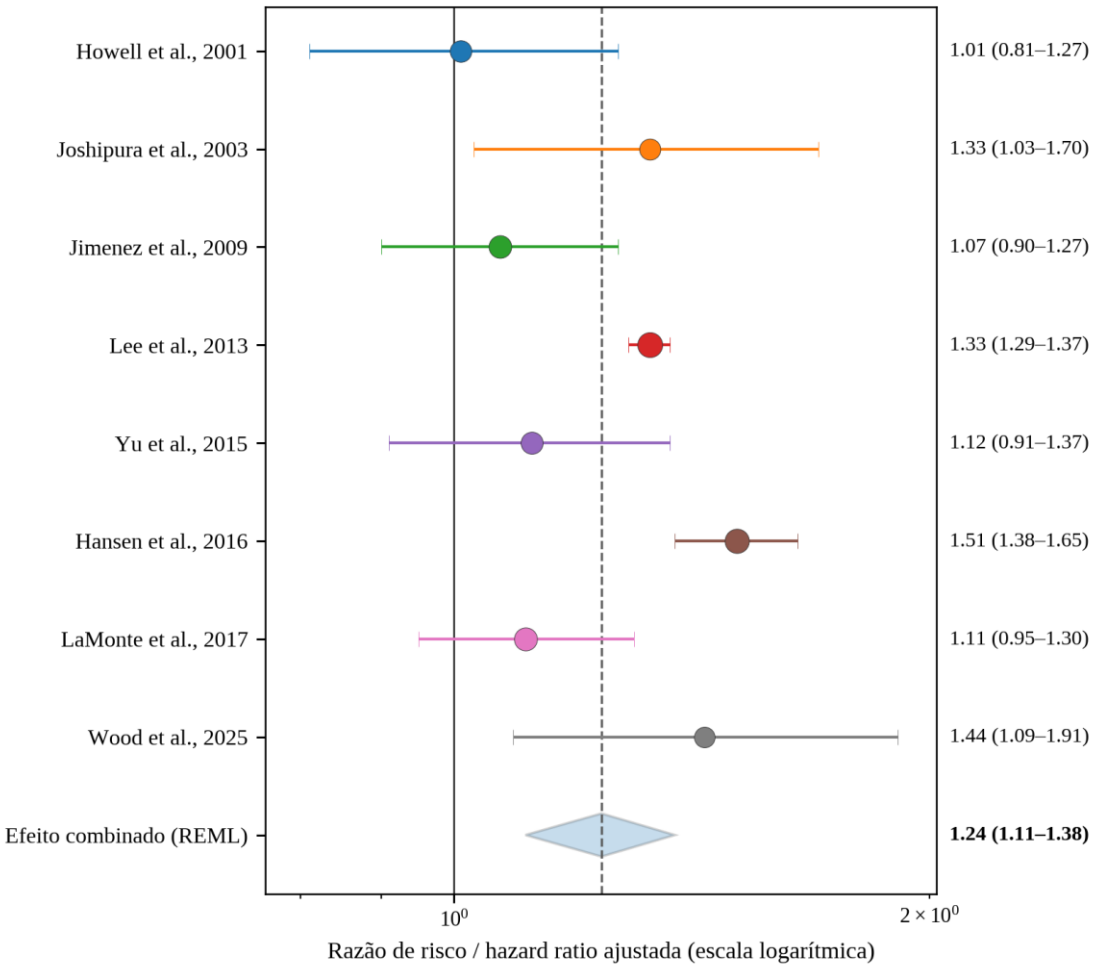


Figura 3 – Metanálise de periodontite/doença periodontal e acidente vascular cerebral.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

A razão combinada foi 1,24 (IC95% 1,11–1,38; p<0,001), indicando associação com aumento relativo médio de cerca de 24%. A heterogeneidade foi elevada (I^2=74,7%; Q=27,62; p<0,001; τ^2=0,0158). O intervalo de predição foi amplo, de 0,95 a 1,62, e cruzou 1,00. Assim, embora a média dos estudos seja claramente positiva, um novo estudo realizado em contexto diferente pode produzir estimativa próxima da ausência de associação. Esse contraste entre intervalo de confiança significativo e intervalo de predição que inclui a nulidade é central para uma interpretação adequada do resultado.

6.3 Síntese estatística e heterogeneidade

A Tabela 4 compara diretamente os dois blocos quantitativos. O objetivo não é afirmar que o risco de AVC seja "duas vezes maior" que o risco coronariano, pois os desfechos, populações e definições diferem, mas mostrar que a consistência entre estudos foi distinta.

Tabela 4 – Resultados principais das metanálises por desfecho

Desfecho	k	Efeito combinado (IC95%)	I^2	τ^2	Q; p	Predição 95%	Leitura principal
Eventos cardiovasculares/coronarianos	8	1,11 (1,07–1,14)	17,0%	0,0003	8,43; p=0,296	1,06–1,16	Associação pequena, precisa e relativamente consistente.
AVC	8	1,24 (1,11–1,38)	74,7%	0,0158	27,62; p<0,001	0,95–1,62	Associação média positiva, porém com heterogeneidade importante.

Os resultados mostram que a estimativa média de AVC foi maior, mas menos previsível entre cenários. Para os eventos cardiovasculares/coronarianos, a baixa heterogeneidade e o intervalo de predição acima de 1,00 aumentam a confiança na existência de uma associação pequena e consistente. Para AVC, a evidência continua favorável a uma associação, mas a amplitude entre estudos impede tratar o valor de 1,24 como uma constante universal. Esse padrão é semelhante ao observado em metanálises recentes de AVC, que também relataram heterogeneidade elevada e enfatizaram que a evidência sustenta associação, não causalidade (Meng; Chen, 2025).

A Figura 4 apresenta, em uma única escala, os efeitos combinados e os respectivos intervalos de predição. As barras claras e espessas mostram a incerteza adicional esperada ao transportar o resultado para um novo contexto.

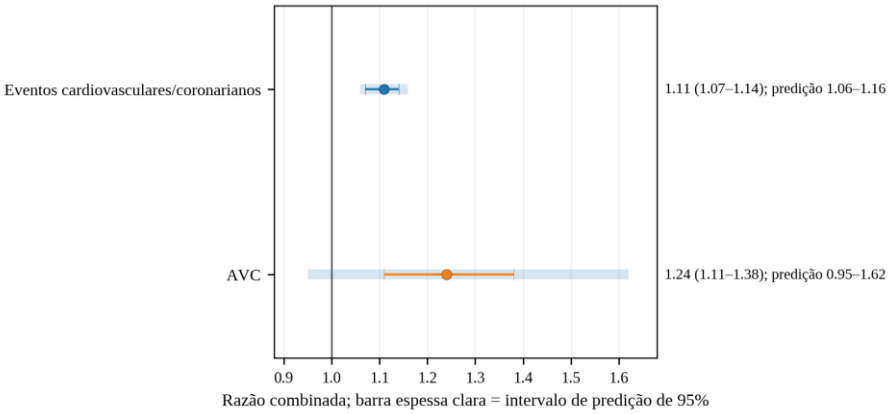


Figura 4 – Comparação dos efeitos combinados e dos intervalos de predição entre os dois blocos.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

A comparação visual deixa claro por que heterogeneidade não deve ser tratada apenas como um número secundário. O efeito coronariano concentra-se em faixa relativamente estreita; o de AVC apresenta maior dispersão e intervalo de predição que alcança a nulidade. Portanto, a conclusão mais sólida é a presença de um sinal epidemiológico global, com maior incerteza na magnitude cerebrovascular em populações específicas.

6.4 Análises de sensibilidade

Para avaliar se os efeitos combinados dependiam de um único estudo, cada coorte foi removida sucessivamente e a metanálise foi recalculada. A Figura 5 mostra os resultados para eventos cardiovasculares/coronarianos. Essa análise é especialmente relevante devido à grande precisão da estimativa do UK Biobank de 2024.

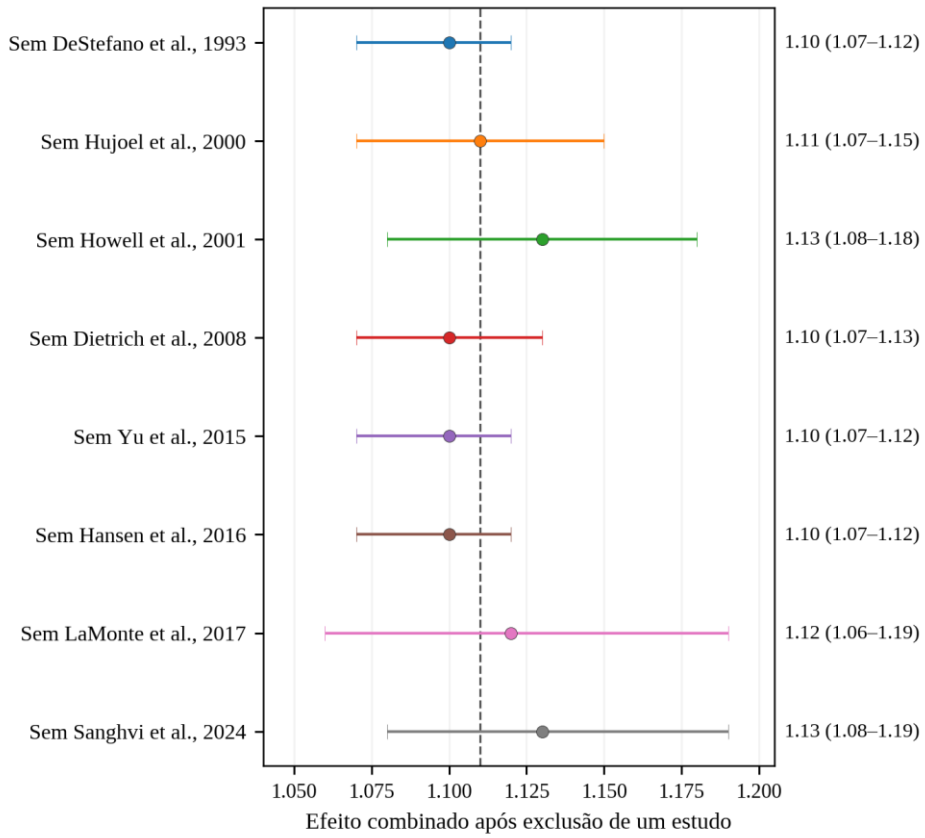


Figura 5 – Análise leave-one-out para eventos cardiovasculares/coronarianos.
Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

A associação permaneceu acima de 1,00 em todas as exclusões. Quando Sanghvi et al. (2024) foi retirado, o efeito combinado aumentou, em vez de desaparecer, mostrando que a grande coorte contemporânea não criou artificialmente o sinal; na verdade, sua estimativa mais próxima de 1,00 reduziu a magnitude global. As demais exclusões também produziram variações discretas, sem reversão da direção do resultado.

A Figura 6 apresenta o mesmo procedimento para AVC. Considerando a heterogeneidade elevada, era possível que um estudo populacional muito grande, como o de Lee, ou uma coorte com diagnóstico hospitalar, como a de Hansen, determinasse isoladamente o resultado.

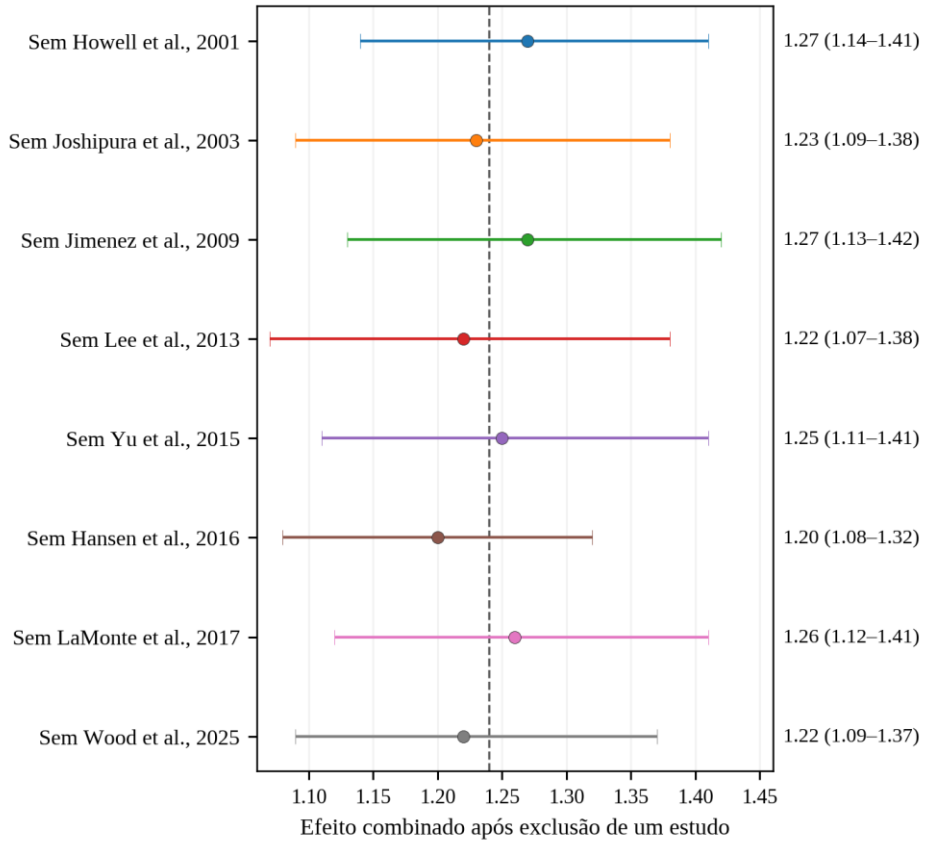


Figura 6 – Análise leave-one-out para acidente vascular cerebral.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

O efeito combinado de AVC também permaneceu estatisticamente acima da nulidade após a retirada de qualquer estudo, variando aproximadamente entre 1,20 e 1,27. A heterogeneidade, porém, continuou moderada ou elevada em várias exclusões. Isso indica que a robustez da média não elimina as diferenças entre contextos; o resultado agregado é estável em direção, mas não em magnitude.

6.5 Evolução temporal da evidência

Uma metanálise cumulativa permite observar como a estimativa agregada mudou à medida que novos estudos foram publicados. Essa abordagem não

substitui a análise principal, mas ajuda a identificar se resultados mais recentes romperam ou apenas refinaram a tendência observada inicialmente.

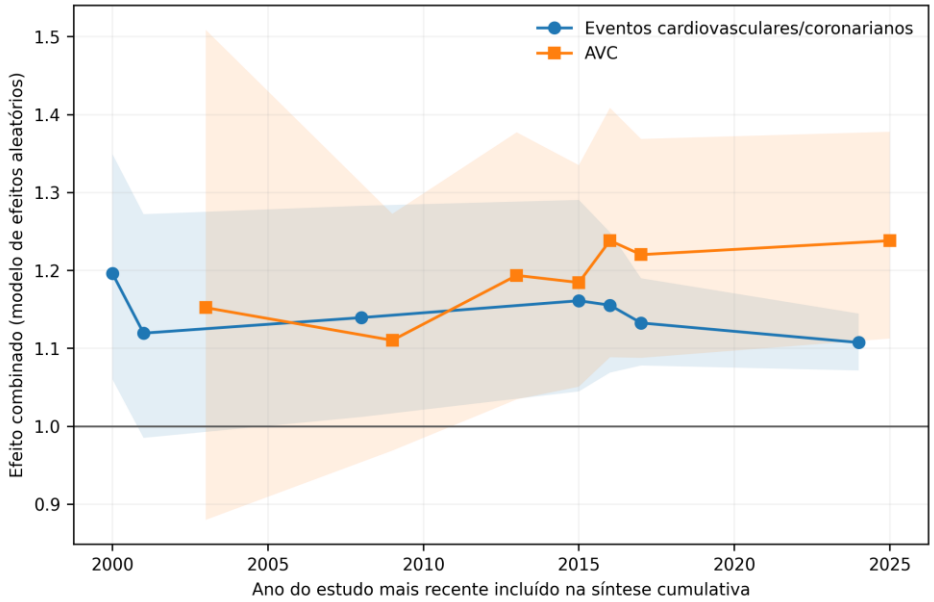


Figura 7 – Evolução cumulativa dos efeitos combinados segundo o ano de publicação.

Fonte: elaboração própria a partir das estimativas e estudos citados neste capítulo.

Nos eventos cardiovasculares/coronarianos, os estudos iniciais sugeriram associações um pouco maiores e menos precisas; com o crescimento das amostras, a estimativa convergiu para um excesso de risco mais modesto, em torno de 10%. Para AVC, a estimativa tornou-se positiva após a incorporação de grandes coortes, mas permaneceu mais variável. A evolução temporal é compatível com um fenômeno comum em epidemiologia: estudos iniciais podem superestimar efeitos, enquanto bases maiores e melhor ajuste de confundidores tendem a produzir associações menores e mais precisas.

7 DISCUSSÃO

7.1 O que a metanálise acrescenta

A reanálise deste capítulo encontrou associação independente entre doença periodontal/periodontite e eventos cardiovasculares em dois domínios. Para desfechos predominantemente coronarianos, o aumento relativo foi pequeno, porém consistente e com baixa heterogeneidade. Para AVC, o aumento médio foi maior, mas acompanhado de heterogeneidade elevada e intervalo de predição que incluiu a nulidade. O conjunto é coerente com revisões anteriores e recentes, sem reproduzir automaticamente suas escolhas de exposição. Humphrey et al. (2008) já identificavam aumento de risco coronariano em categorias de doença periodontal; Larvin et al. (2021) concluíram que o risco cardiovascular incidente é modestamente maior em pessoas com doença periodontal; Leng et al. (2023) encontrou associação em ambos os sexos; e a umbrella review de Arbildo-Vega et al. (2024) mostrou que revisões de maior confiança global sustentam associação entre doença periodontal e doença cardiovascular.

O valor desta reanálise está também na restrição dos marcadores. Muitos estudos históricos incluíram perda de dentes, edentulismo ou índices de saúde oral que não distinguem periodontite. Esses marcadores têm relevância prognóstica, mas podem carregar informação sobre pobreza, acesso à odontologia, dieta, tabagismo e carga cumulativa de doenças ao longo da vida. Quando perda dentária é tratada como sinônimo de periodontite, o mecanismo atribuído à inflamação periodontal torna-se menos específico. A decisão de manter esses estudos fora da combinação principal provavelmente reduz o número total de participantes, mas melhora a validade conceitual da exposição.

A magnitude pequena do efeito coronariano merece interpretação equilibrada. Uma razão em torno de 1,11 não transforma periodontite em determinante cardiovascular de força semelhante ao tabagismo, hipertensão ou

hipercolesterolemia. Por outro lado, periodontite é comum e frequentemente coexistente com esses fatores; um efeito relativo modesto pode ter importância populacional se for causal e se estiver presente por longos períodos. O desafio é que a mesma prevalência elevada amplia também o potencial de confusão residual.

7.2 Confundimento, causalidade e evidência genética

Tabagismo é provavelmente o confundidor clássico mais importante porque aumenta fortemente tanto o risco periodontal quanto o cardiovascular e é difícil de medir sem erro. Classificações simples como “fumante atual”, “ex-fumante” e “nunca fumou” não captam completamente intensidade, duração, tempo desde cessação e exposição passiva. Diabetes apresenta problema semelhante: presença ou ausência do diagnóstico não representa controle glicêmico, duração da doença ou complicações. Ajustes estatísticos reduzem, mas não eliminam essas fontes de confusão.

Condição socioeconômica acrescenta outra camada. Escolaridade, renda, insegurança alimentar, padrão de uso de serviços odontológicos e acesso a medicações cardiovasculares podem influenciar os dois desfechos. A associação entre atendimento odontológico regular e menor incidência de eventos pode refletir benefício do cuidado periodontal, mas também um comportamento global de prevenção conhecido como efeito do usuário saudável. Por isso, estudos observacionais de tratamento ou frequência de consultas devem ser lidos com mais cautela do que ensaios randomizados.

Estudos genéticos e de randomização mendeliana têm sido usados para testar causalidade com menor suscetibilidade a alguns confundidores. Os resultados, contudo, não são uniformemente fortes. Sanghvi et al. (2024) encontraram associação entre escore poligênico relacionado à doença periodontal e doença coronariana de pequena magnitude. A declaração da American Heart Association de 2026 considera que a genética acrescenta plausibilidade, mas

ainda não encerra a questão causal. Limitações incluem instrumentos genéticos de efeito fraco, fenótipos periodontais heterogêneos e possíveis pleiotropias.

Assim, o conjunto atual se encaixa melhor em uma relação de contribuição potencial do que em uma cadeia causal simples. A periodontite pode funcionar como fonte adicional de inflamação em indivíduos suscetíveis, somando-se a fatores tradicionais, ao mesmo tempo em que compartilha determinantes comportamentais e metabólicos com a aterosclerose. Esse modelo explica por que a associação persiste em muitos estudos, mas diminui após ajuste e varia entre populações.

7.3 Por que o resultado de AVC é mais heterogêneo

AVC não é um único desfecho fisiopatológico. AVC isquêmico pode resultar de aterosclerose de grandes artérias, doença de pequenos vasos, cardioembolismo e outras causas; AVC hemorrágico tem mecanismos diferentes. Estudos que agregam todos os subtipos podem diluir ou misturar associações. Trabalhos mais recentes, como Sen et al. (2018) e Wood et al. (2025), sugerem que determinados subtipos isquêmicos podem ter relações mais fortes com doença periodontal. Entretanto, subdividir eventos reduz o número de casos e aumenta imprecisão, razão pela qual ainda não há uma estimativa definitiva por mecanismo de AVC.

A própria definição periodontal também contribui. Diagnóstico hospitalar tende a identificar doença grave; autorrelato pode perder casos leves ou não diagnosticados; exames clínicos diferem em número de sítios sondados e critérios de gravidade. A heterogeneidade de I^2 observada neste capítulo deve, portanto, ser lida como expressão de variação clínica e metodológica, não como falha estatística que possa ser "corrigida" apenas escolhendo outro estimador.

A metanálise de Meng e Chen (2025), com conjunto mais amplo de estudos observacionais, igualmente concluiu que a periodontite está associada a maior risco de AVC, com sinal mais claro para eventos isquêmicos e coortes de longo

seguimento, mas ressaltou heterogeneidade alta e possibilidade de confusão residual. Esse alinhamento reforça a conclusão de associação e, simultaneamente, a cautela com causalidade.

7.4 Tratamento periodontal e marcadores cardiovasculares

Uma questão clinicamente mais importante do que a simples associação é saber se tratar periodontite melhora a saúde cardiovascular. Ensaios randomizados avaliaram principalmente desfechos intermediários, como função endotelial, espessura médio-intimal carotídea e biomarcadores inflamatórios. O ensaio de Tonetti et al. (2007), com 120 participantes com periodontite grave, mostrou piora inflamatória aguda após tratamento intensivo, seguida de melhora da função endotelial em meses posteriores. O achado foi relevante por demonstrar que modificar a inflamação periodontal pode alterar fisiologia vascular, mas a função endotelial continua sendo um marcador substituto, não um evento clínico maior.

Em pacientes com doença coronariana estabelecida, os resultados foram menos uniformes. Saffi et al. (2018) não encontrou melhora superior de vasodilatação no curto prazo após terapia periodontal, embora tenha observado manutenção de parâmetros vasculares e melhora periodontal. Estudos menores e mais recentes continuam a investigar espessura carotídea, inflamação e função endotelial. Molina et al. (2025), em ensaio piloto com doença cardiovascular estabelecida e periodontite avançada, obteve melhora periodontal clara, mas diferenças intergrupos nos principais marcadores vasculares foram limitadas pelo tamanho amostral e pela natureza exploratória.

A Tabela 5 sintetiza o que estudos de intervenção permitem concluir. Ela separa explicitamente melhora periodontal de benefício cardiovascular comprovado, evitando uma extrapolação comum na divulgação do tema.

Tabela 5 – Evidência de tratamento periodontal e repercussões cardiovasculares

Estudo	Desenho/população	Intervenção/exposição	Desfecho vascular	Resultado principal	Interpretação
Tonetti et al., 2007	Ensaio randomizado; periodontite grave	Terapia periodontal intensiva vs. controle	Função endotelial (FMD) e inflamação	Melhora tardia de FMD após resposta inflamatória aguda inicial.	Demonstra modificabilidade de marcador vascular; não avaliou redução de infarto/AVC.
Saffi et al., 2018	Ensaio randomizado; doença coronariana + periodontite	Terapia periodontal	Função endotelial	Não mostrou vasodilatação superior no curto seguimento.	Resultados não uniformes em população coronariana estabelecida.
Lobo et al., 2020	Ensaio clínico após infarto	Tratamento periodontal	Parâmetros clínicos/inflamatórios	Benefícios periodontais e avaliação sistêmica; estudo de porte limitado.	Insuficiente para estabelecer prevenção secundária de eventos maiores.
Molina et al., 2025	Ensaio piloto; doença cardiovascular + periodontite estágio III/IV	Terapia periodontal	FMD, cIMT e biomarcadores	Melhora periodontal; diferenças vasculares intergrupos limitadas.	Evidência exploratória, com baixa potência para eventos clínicos.
Wagner et al., 2023	Coorte nacional após primeiro infarto	Periodontite tratada/estado periodontal	Eventos recorrentes	Associação observacional com recorrência cardiovascular.	Útil para hipótese; não substitui randomização por possível confusão de cuidado/adesão.

O conjunto de intervenção sustenta que o tratamento periodontal melhora a condição oral e pode modificar alguns marcadores sistêmicos, mas ainda não demonstra que raspagem, cirurgia periodontal ou programas de manutenção reduzam infarto, AVC ou mortalidade cardiovascular. Essa distinção aparece tanto nos consensos periodontais quanto na atualização da American Heart Association. A recomendação clínica de tratar periodontite é plenamente

justificada pelos benefícios odontológicos; eventual benefício cardiovascular deve ser considerado potencial adicional, não promessa terapêutica.

7.5 Implicações para a prática clínica integrada

A associação entre periodontite e doença cardiovascular tem utilidade clínica mesmo sem causalidade definitivamente provada. Na odontologia, a identificação de periodontite em paciente com múltiplos fatores de risco deve estimular investigação de tabagismo, hipertensão conhecida, diabetes, obesidade e adesão ao acompanhamento médico. O cirurgião-dentista não precisa calcular risco cardiovascular complexo para reconhecer sinais de vulnerabilidade e recomendar avaliação adequada. Da mesma forma, profissionais médicos e de enfermagem podem incorporar perguntas simples sobre sangramento gengival, mobilidade dentária, perda de dentes por doença periodontal e tempo desde a última avaliação odontológica.

A integração é particularmente pertinente em pessoas com diabetes e doença cardiovascular estabelecida, nas quais a coexistência de periodontite é frequente e o controle de fatores sistêmicos beneficia simultaneamente vários desfechos. Cessação do tabagismo, controle glicêmico, alimentação equilibrada e adesão à higiene oral têm vantagens que transcendem a hipótese periodontal–cardiovascular. Essa abordagem evita a falsa dicotomia de decidir se a periodontite "causa" ou "não causa" doença cardiovascular antes de promover cuidados que já são recomendados por razões próprias.

A Tabela 6 traduz a evidência para decisões práticas e delimita o que pode ser afirmado com segurança. O objetivo é prevenir tanto a minimização da saúde periodontal quanto a divulgação exagerada de causalidade.

Tabela 6 – Tradução da evidência para a prática clínica

Afirmação	Estado da evidência	Conduta interpretativa
"Periodontite está associada a maior risco cardiovascular."	Apoiado por múltiplas coortes e metanálises.	Pode ser comunicado como associação, principalmente para doença aterosclerótica e AVC isquêmico.
"Periodontite causa infarto e AVC."	Não estabelecido de forma definitiva.	Evitar formulação causal; persistem confusão residual e heterogeneidade fenotípica.
"Tratar periodontite previne eventos cardiovasculares maiores."	Não demonstrado por ensaios com poder adequado.	Tratar pelos benefícios periodontais; não prometer prevenção de infarto ou AVC.
"Perda de dentes significa periodontite."	Incorreto como regra.	Investigar causa; cárie, trauma e acesso a cuidado também produzem perda dentária.
"Paciente cardiovascular não precisa de atenção periodontal específica."	Inadequado.	Promover avaliação periodontal e manutenção da saúde oral como parte do cuidado integral.
"Antibiótico sistêmico deve ser usado para reduzir risco aterosclerótico."	Sem suporte.	Não indicar antibiótico com finalidade de prevenir aterosclerose; seguir indicações odontológicas e diretrizes específicas.

Na prática, a mensagem mais defensável é simples: doença periodontal deve ser diagnosticada e tratada adequadamente; pessoas com periodontite merecem atenção aos fatores cardiovasculares tradicionais; e pessoas com doença cardiovascular devem manter cuidado odontológico regular. Nenhuma dessas ações exige assumir causalidade definitiva. Ao contrário, elas decorrem de prevenção integrada e do reconhecimento de que doenças crônicas compartilham determinantes biológicos e sociais.

8 LIMITAÇÕES DA EVIDÊNCIA E DA METANÁLISE

A principal limitação da literatura é a heterogeneidade de exposição. Estudos de diferentes décadas utilizaram classificações periodontais que não são diretamente equivalentes. Autorrelato, códigos diagnósticos, profundidade de bolsa, perda de inserção e avaliação radiográfica possuem sensibilidade e

especificidade distintas. Mesmo dentro de exames clínicos, protocolos parciais podem subestimar doença. Essa variação tende a reduzir comparabilidade e pode explicar parte da heterogeneidade entre estimativas.

Também existe heterogeneidade de desfecho. Doença cardiovascular total, doença coronariana, infarto e AVC não são intercambiáveis. O primeiro bloco desta metanálise foi descrito deliberadamente como predominantemente coronariano porque algumas coortes publicaram estimativas cardiovasculares agregadas. O segundo bloco reuniu AVC com predominância isquêmica, mas nem todos os estudos possuíam classificação etiológica detalhada. A combinação fornece uma visão quantitativa útil, porém não substitui metanálises específicas de cada fenótipo cardiovascular.

Outra limitação é a mistura de HR e RR como medidas relativas no mesmo eixo logarítmico. Essa prática é frequente em sínteses de incidência e é razoável quando os efeitos são modestos e os eventos são avaliados ao longo do seguimento, mas as medidas não são matematicamente idênticas. Por isso, o resultado foi denominado "razão combinada" e não convertido artificialmente para uma métrica única. A estimativa deve ser interpretada como síntese de efeitos relativos ajustados, e não como risco absoluto.

Os estudos observacionais permaneceram suscetíveis a confusão residual. Tabagismo, diabetes, escolaridade e acesso a serviços foram ajustados de formas diferentes, e nenhum modelo garante eliminação completa do viés. Além disso, a periodontite pode mudar durante o seguimento, enquanto muitas coortes classificaram a exposição apenas no início. Tratamento periodontal recebido posteriormente pode reduzir contraste entre grupos e tender a diluir associações.

A reanálise não foi apresentada como revisão sistemática exaustiva com protocolo prospectivo e fluxo PRISMA próprio. Ela foi construída como síntese quantitativa dirigida, com verificação de estimativas em estudos primários e comparação com revisões recentes. Essa escolha deve ser considerada ao interpretar os números. O objetivo foi oferecer um capítulo tecnicamente

transparente e clinicamente coerente, não substituir uma revisão sistemática registrada dedicada exclusivamente a um único desfecho.

Por fim, a ausência de testes formais de viés de publicação decorreu do pequeno número de estudos por bloco. Com menos de dez estudos, testes de assimetria apresentam baixa potência e podem confundir heterogeneidade com viés de publicação. A avaliação foi, portanto, qualitativa e baseada na amplitude de amostras, diversidade de resultados e existência de estudos nulos entre os incluídos.

9 PERSPECTIVAS PARA PESQUISA

O avanço do campo depende menos de novas associações genéricas e mais de estudos capazes de responder perguntas causais. Coortes futuras deveriam usar a classificação periodontal contemporânea, registrar atividade inflamatória e dano acumulado separadamente, repetir o exame periodontal ao longo do tempo e distinguir subtipos de eventos cardiovasculares. A integração com marcadores inflamatórios, microbioma, imagem vascular e dados genéticos pode ajudar a identificar trajetórias específicas em vez de tratar todos os pacientes com periodontite como um único grupo.

Ensaios cardiovasculares de grande porte continuam sendo a lacuna principal. Demonstrar redução de eventos maiores exigiria amostras grandes, acompanhamento prolongado e padronização da terapia periodontal, o que torna esses estudos caros e operacionalmente complexos. Desfechos intermediários validados podem continuar úteis em fases iniciais, mas não devem substituir indefinidamente infarto, AVC, hospitalização e mortalidade. Estudos pragmáticos em sistemas de saúde integrados podem ser uma alternativa para testar estratégias de cuidado periodontal intensivo em populações de alto risco.

Outra prioridade é estudar equidade. Doença periodontal e doença cardiovascular concentram-se em grupos socialmente vulneráveis, nos quais

acesso tardio ao cuidado, tabagismo, diabetes e insegurança alimentar se acumulam. Intervenções que apenas acrescentem consultas especializadas podem ampliar desigualdades. Programas de atenção primária, integração saúde bucal–saúde cardiovascular e estratégias de cessação do tabagismo podem oferecer impacto mais abrangente do que intervenções isoladas.

10 CONCLUSÃO

A literatura acumulada sustenta uma associação entre periodontite e doenças cardiovasculares, especialmente eventos ateroscleróticos e AVC isquêmico. Na reanálise apresentada neste capítulo, a doença periodontal esteve associada a aumento relativo pequeno e consistente de eventos cardiovasculares predominantemente coronarianos e a aumento médio maior de AVC, embora com heterogeneidade substancial no componente cerebrovascular. A estabilidade das análises de sensibilidade mostra que os efeitos combinados não dependeram de um único estudo.

Esses resultados não devem ser convertidos em afirmação causal automática. Periodontite e doença cardiovascular compartilham determinantes importantes, e a mensuração imperfeita de tabagismo, diabetes, condições socioeconômicas e acesso a cuidados deixa espaço para confusão residual. A plausibilidade biológica é consistente, mas plausibilidade e associação não equivalem a demonstração de que a eliminação da periodontite previne infarto ou AVC.

O tratamento periodontal permanece indicado para controlar inflamação, preservar dentes, reduzir progressão destrutiva e melhorar a saúde oral. Há evidência de que a terapia pode modificar alguns marcadores vasculares e inflamatórios, porém não existem ensaios com poder suficiente para estabelecer redução de eventos cardiovasculares maiores. Assim, o benefício cardiovascular deve ser considerado possível, mas não prometido ao paciente.

Na prática, a associação reforça a necessidade de cuidado integrado. Pacientes com periodontite devem ter seus fatores cardiovasculares tradicionais reconhecidos e controlados, e pessoas com doença cardiovascular devem receber atenção regular à saúde bucal. Essa integração é justificável independentemente de a periodontite ser posteriormente confirmada como fator causal modificável, porque reduz carga de doença oral e cria oportunidades para prevenção global de condições crônicas.

REFERÊNCIAS

ARBILDO-VEGA, H. I. et al. Periodontal disease and cardiovascular disease: umbrella review. BMC Oral Health, 2024. doi:10.1186/s12903-024-04907-1.

DESTEFANO, F.; ANDA, R. F.; KAHN, H. S.; WILLIAMSON, D. F.; RUSSELL, C. M. Dental disease and risk of coronary heart disease and mortality. BMJ, v. 306, p. 688–691, 1993. doi:10.1136/bmj.306.6879.688.

DIETRICH, T.; JIMENEZ, M.; KRALL KAYE, E. A.; VOKONAS, P. S.; GARCIA, R. I. Age-dependent associations between chronic periodontitis/edentulism and risk of coronary heart disease. Circulation, v. 117, p. 1668–1674, 2008. doi:10.1161/CIRCULATIONAHA.107.711507.

FAGUNDES, N. C. F. et al. Periodontitis as a risk factor for stroke: a systematic review and meta-analysis. Vascular Health and Risk Management, v. 15, p. 519–532, 2019.

GAO, S. et al. Periodontitis and number of teeth in the risk of coronary heart disease: an updated meta-analysis. Medical Science Monitor, v. 27, e930112, 2021. doi:10.12659/MSM.930112.

HANSEN, G. M.; EGEBERG, A.; HOLMSTRUP, P.; HANSEN, P. R. Relation of periodontitis to risk of cardiovascular and all-cause mortality (from a Danish nationwide cohort study). American Journal of Cardiology, v. 118, p. 489–493, 2016. doi:10.1016/j.amjcard.2016.05.036.

HERRERA, D. et al. Association between periodontal diseases and cardiovascular diseases, diabetes and respiratory diseases: consensus report of the joint workshop by the European Federation of Periodontology and WONCA Europe. Journal of Clinical Periodontology, v. 50, p. 819–841, 2023. doi:10.1111/jcpe.13807.

HOWELL, T. H.; RIDKER, P. M.; AJANI, U. A.; HENNEKENS, C. H.; CHRISTEN, W. G. Periodontal disease and risk of subsequent cardiovascular

disease in U.S. male physicians. Journal of the American College of Cardiology, v. 37, p. 445–450, 2001. doi:10.1016/S0735-1097(00)01130-X.

HUJOEL, P. P.; DRANGSHOLT, M.; SPIEKERMAN, C.; DEROUEN, T. A. Periodontal disease and coronary heart disease risk. JAMA, v. 284, p. 1406–1410, 2000. doi:10.1001/jama.284.11.1406.

HUMPHREY, L. L.; FU, R.; BUCKLEY, D. I.; FREEMAN, M.; HELFAND, M. Periodontal disease and coronary heart disease incidence: a systematic review and meta-analysis. Journal of General Internal Medicine, v. 23, p. 2079–2086, 2008. doi:10.1007/s11606-008-0787-6.

JIMENEZ, M.; KRALL, E. A.; GARCIA, R. I.; VOKONAS, P. S.; DIETRICH, T. Periodontitis and incidence of cerebrovascular disease in men. Annals of Neurology, v. 66, p. 505–512, 2009. doi:10.1002/ana.21742.

JOSHIPURA, K. J.; HUNG, H. C.; RIMM, E. B.; WILLETT, W. C.; ASCHERIO, A. Periodontal disease, tooth loss, and incidence of ischemic stroke. Stroke, v. 34, p. 47–52, 2003. doi:10.1161/01.STR.0000052974.79428.0C.

LAMONTE, M. J. et al. History of periodontitis diagnosis and edentulism as predictors of cardiovascular disease, stroke, and mortality in postmenopausal women. Journal of the American Heart Association, v. 6, e004518, 2017. doi:10.1161/JAHA.116.004518.

LARVIN, H.; KANG, J.; AGGARWAL, V. R.; PAVITT, S.; WU, J. Risk of incident cardiovascular disease in people with periodontal disease: a systematic review and meta-analysis. Clinical and Experimental Dental Research, v. 7, p. 109–122, 2021. doi:10.1002/cre2.336.

LEE, Y. L.; HU, H. Y.; HUANG, N.; HWANG, D. K.; CHOU, P.; CHU, D. Dental prophylaxis and periodontal treatment are protective factors to ischemic stroke. Stroke, v. 44, p. 1026–1030, 2013. doi:10.1161/STROKEAHA.111.000076.

LENG, Y. et al. Periodontal disease is associated with the risk of cardiovascular disease independent of sex: a meta-analysis. Frontiers in Cardiovascular Medicine, v. 10, 1114927, 2023. doi:10.3389/fcvm.2023.1114927.

LOCKHART, P. B. et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association? A scientific statement from the American Heart Association. Circulation, v. 125, p. 2520–2544, 2012. doi:10.1161/CIR.0b013e31825719f3.

LOBO, M. G. et al. Treating periodontal disease in patients with myocardial infarction: a randomized clinical trial. European Journal of Internal Medicine, v. 71, p. 76–80, 2020.

MENG, X.; CHEN, X. Periodontitis and risk of stroke: a systematic review and meta-analysis of observational studies. Frontiers in Neurology, v. 16, 1700946, 2025. doi:10.3389/fneur.2025.1700946.

MOLINA, A. et al. Effect of periodontal therapy on endothelial function and serum biomarkers in patients with periodontitis and established cardiovascular disease: a pilot study. Frontiers in Oral Health, v. 6, 1488941, 2025. doi:10.3389/froh.2025.1488941.

PAPAPANOU, P. N. et al. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. Journal of Clinical Periodontology, v. 45, Suppl. 20, p. S162–S170, 2018. doi:10.1111/jcpe.12946.

SAFFI, M. A. L. et al. Periodontal therapy and endothelial function in coronary artery disease: a randomized controlled trial. Oral Diseases, v. 24, p. 1349–1357, 2018.

SANGHVI, M. M. et al. The association between periodontal disease and cardiovascular disease: insights from imaging, observational, and genetic data. JACC: Advances, v. 3, n. 10, 101241, 2024. doi:10.1016/j.jacadv.2024.101241.

SANZ, M. et al. Periodontitis and cardiovascular diseases: consensus report. Journal of Clinical Periodontology, v. 47, p. 268–288, 2020. doi:10.1111/jcpe.13189.

SEN, S. et al. Periodontal disease, regular dental care use, and incident ischemic stroke. Stroke, v. 49, p. 355–362, 2018. doi:10.1161/STROKEAHA.117.018990.

TONETTI, M. S. et al. Treatment of periodontitis and endothelial function. New England Journal of Medicine, v. 356, p. 911–920, 2007. doi:10.1056/NEJMoa063186.

TRAN, A. H. et al. Periodontal disease and atherosclerotic cardiovascular disease: a scientific statement from the American Heart Association. Circulation, v. 153, n. 6, p. e73–e88, 2026. doi:10.1161/CIR.0000000000001390.

WAGNER, A. K. et al. Treated periodontitis and recurrent events after first-time myocardial infarction: a Danish nationwide cohort study. Journal of Clinical Periodontology, v. 50, p. 1305–1314, 2023. doi:10.1111/jcpe.13853.

WOOD, S.; LOGUE, L.; MEYER, J. et al. Combined influence of dental caries and periodontal disease on ischemic stroke risk. Neurology Open Access, v. 1, n. 4, e000036, 2025. doi:10.1212/WN9.0000000000000036.

YU, Y. H.; CHASMAN, D. I.; BURING, J. E.; ROSE, L.; RIDKER, P. M. Cardiovascular risks associated with incident and prevalent periodontal

disease. Journal of Clinical Periodontology, v. 42, p. 21–28, 2015. doi:10.1111/jcpe.12335.

CAPÍTULO 11

OBESIDADE MATERNA E COMPLICAÇÕES APÓS CESARIANA

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Nélio Barreto Vieira

RESUMO

A obesidade materna acrescenta complexidade à cesariana por mecanismos anatômicos, metabólicos, inflamatórios, farmacocinéticos e organizacionais. Este capítulo sintetiza evidências sobre complicações da ferida, hemorragia pós-parto, duração operatória, anestesia, tromboembolismo, efetividade das medidas preventivas e morbidade neonatal. Foi realizada busca estruturada em PubMed/MEDLINE e fontes complementares, priorizando coortes de grande porte, ensaios clínicos, revisões sistemáticas e diretrizes com dados diretamente aplicáveis ao perioperatório obstétrico. A evidência observacional mostra gradiente consistente para complicações da ferida: em uma coorte de 2.444 cesarianas, a frequência aumentou de 6,6% no IMC inferior a 30 kg/m² para 22,9% no IMC igual ou superior a 50 kg/m². A hemorragia apresenta arquitetura mais heterogênea, porque perda sanguínea, atonia e transfusão não são desfechos intercambiáveis. O intervalo incisão–nascimento e a morbidade neonatal também aumentam nos extremos do IMC. Entre as intervenções, azitromicina adjunta em cesarianas não eletivas reduziu infecção composta de 12,0% para 6,1%; cefalexina e metronidazol por 48 horas reduziram infecção do sítio cirúrgico de 15,4% para 6,4% em um ensaio anterior, embora um ensaio de 2025, realizado sobre profilaxia pré-operatória ampliada, não tenha confirmado benefício adicional. A pressão negativa profilática não reduziu infecção no maior ensaio e aumentou reações cutâneas. Doses de enoxaparina baseadas no peso alcançaram concentrações anti-Xa profiláticas com maior frequência, mas a tradução para prevenção de eventos clínicos permanece incerta. Conclui-se que

o cuidado deve ser escalonado pelo risco, distinguir cesariana eletiva de não eletiva, integrar anestesia, obstetrícia, enfermagem e neonatologia, e privilegiar pacotes de segurança com benefício demonstrável em vez de tecnologias aplicadas rotineiramente sem seleção clínica.

Palavras-chave: obesidade materna; cesariana; infecção de sítio cirúrgico; hemorragia pós-parto; anestesia obstétrica; tromboprofilaxia; morbidade neonatal.

1 INTRODUÇÃO

A obesidade na gestação deixou de constituir uma condição ocasional e passou a fazer parte da rotina dos serviços de pré-natal, centros obstétricos e unidades de terapia intensiva. Seu impacto começa antes da cirurgia, com maior frequência de diabetes, hipertensão, apneia do sono, macrossomia, indução do parto e falha de progressão, e se prolonga após a alta, quando infecção, abertura da ferida, tromboembolismo e dificuldade de acesso ao seguimento podem surgir. A cesariana, por sua vez, acrescenta incisão abdominal, dissecção de planos, manipulação uterina, imobilidade e exposição a antibióticos e anticoagulantes. O encontro entre ambas não deve ser interpretado apenas como soma de dois fatores de risco, mas como uma situação clínica na qual anatomia, fisiologia, logística e qualidade do sistema interagem [1-4].

O índice de massa corporal é uma medida imperfeita de adiposidade, porém continua sendo a variável mais utilizada para estratificação obstétrica. Na prática, mulheres classificadas na mesma categoria podem diferir quanto à distribuição do tecido adiposo, massa magra, circunferência abdominal, mobilidade, função respiratória e presença de comorbidades. Além disso, o peso pode ter sido registrado antes da gestação, no início do pré-natal ou na admissão para o parto. Essas diferenças metodológicas explicam parte da heterogeneidade entre estudos e devem ser consideradas quando se traduz uma estimativa populacional para uma paciente específica.

A ferida operatória é o domínio em que a relação dose–resposta se mostra mais consistente. Em 2.444 cesarianas, Conner et al. observaram complicações em 6,6% das mulheres com IMC inferior a 30 kg/m², 9,2% no IMC de 30,0 a 39,9 kg/m², 16,8% no intervalo de 40,0 a 49,9 kg/m² e 22,9% no IMC igual ou superior a 50 kg/m² [7]. O gradiente é clinicamente relevante porque combina maior área de tecido subcutâneo, perfusão relativa reduzida, tensão mecânica, umidade em pregas, duração operatória e maior prevalência de diabetes. Contudo, não significa que a infecção seja inevitável: técnica, antibioticoprofilaxia, fechamento, vigilância e acesso ao retorno modificam o risco.

O domínio hemorrágico é mais difícil de resumir. A obesidade pode prolongar a operação, aumentar a área de dissecção, associar-se à macrossomia e à atonia uterina e dificultar a quantificação visual do sangramento. Ao mesmo tempo, maior volume plasmático pode retardar sinais clínicos e alterar a relação entre perda estimada, queda da hemoglobina e indicação de transfusão. Por isso, estudos que utilizam perda igual ou superior a 1.000 mL, transfusão, atonia, uso de uterotônicos ou hemorragia grave não estão medindo exatamente o mesmo fenômeno. A interpretação adequada requer separar o evento fisiopatológico do limiar assistencial empregado para reconhecê-lo e tratá-lo.

O feto também participa do processo perioperatório. O acesso uterino pode ser mais demorado devido à espessura do subcutâneo, aderências de cesarianas anteriores, dificuldade de exposição e necessidade de reposicionamento. Em situações urgentes, minutos adicionais entre a incisão e o nascimento podem coincidir com hipóxia já em curso. Estudos observacionais mostraram aumento do intervalo incisão–nascimento e de desfechos neonatais compostos com o crescimento do IMC [15-17]. Esses achados não autorizam atribuir causalidade ao tempo isoladamente, mas sustentam preparação antecipada da equipe e dos equipamentos.

A anestesia obstétrica constitui outro eixo decisivo. Pontos anatômicos menos palpáveis, maior profundidade do espaço peridural, posicionamento difícil, intolerância ao decúbito, risco de dessaturação rápida e necessidade de

manguitos e mesas compatíveis podem transformar um procedimento habitual em uma intervenção de elevada demanda técnica. A avaliação precoce permite planejar ultrassom, pessoal experiente, estratégia de via aérea e monitorização. O objetivo não é classificar a paciente como "difícil", mas reduzir a improvisação do serviço diante de necessidades previsíveis.

Por fim, parte do risco atribuído biologicamente à obesidade é mediada pelo sistema. Atrasos por falta de equipamentos, estigma, linguagem desrespeitosa, mobiliário inadequado, transferência insegura e ausência de seguimento pós-alta não são consequências inevitáveis do IMC. Uma síntese profissional deve, portanto, distinguir três camadas: susceptibilidade biológica; complexidade técnica; e capacidade organizacional. Este capítulo integra essas dimensões, descreve a magnitude dos principais achados e examina quais intervenções apresentam benefício real, quais permanecem incertas e quais podem produzir dano quando aplicadas indiscriminadamente.

2 OBJETIVOS

Sintetizar e interpretar a evidência sobre complicações maternas, operatórias, anestésicas, trombóticas e neonatais após cesariana em mulheres com obesidade; descrever gradientes de risco segundo classes de IMC; comparar estimativas relativas e absolutas; examinar ensaios de prevenção de infecção e tromboprofilaxia; identificar limitações metodológicas; e traduzir os resultados para um modelo perioperatório escalonado, centrado na segurança e aplicável a serviços com diferentes níveis de recursos.

3 MÉTODOS

3.1 Delineamento e padrão de relato

Foi conduzida revisão sistemática narrativa com síntese multidomínio. O relato foi orientado pelos itens aplicáveis do PRISMA 2020 [5], com adaptação

para um capítulo que reúne estudos de prognóstico, coortes de exposição, ensaios clínicos e recomendações profissionais. A opção por síntese narrativa, em vez de uma única meta-análise, decorreu da heterogeneidade entre definições de obesidade, momento de aferição do peso, urgência da cesariana, desfechos compostos, duração do seguimento e medidas de efeito. Misturar essas estimativas em um valor global produziria precisão aparente, mas pouco significado clínico.

3.2 Pergunta, população e elegibilidade

A pergunta foi organizada pelo modelo PECO: mulheres submetidas à cesariana; exposição a sobrepeso, obesidade ou superobesidade mensurados por IMC; comparação com mulheres sem obesidade, categorias inferiores de IMC ou estratégias preventivas alternativas; e desfechos maternos, operatórios, anestésicos, trombóticos e neonatais. Foram elegíveis coortes, estudos caso-controle, ensaios randomizados, revisões sistemáticas e diretrizes com resultados quantitativos. Foram excluídos relatos isolados, séries sem comparação, estudos exclusivamente sobre parto vaginal e publicações em que o efeito da cesariana não pudesse ser separado de outras vias de parto.

3.3 Fontes de informação e estratégia de busca

A busca principal foi realizada em PubMed/MEDLINE e complementada por PubMed Central, páginas oficiais de periódicos, registros de ensaios e rastreamento das listas de referências. Os termos combinaram obesity, body mass index, morbid obesity, super obesity, cesarean ou caesarean, wound complication, surgical site infection, hemorrhage, transfusion, anesthesia, incision-to-delivery, neonatal morbidity, thromboprophylaxis, enoxaparin, cefazolin, azithromycin e negative pressure wound therapy. A atualização final foi efetuada em 1º de agosto de 2026. A estratégia completa é apresentada no Apêndice A, permitindo reprodução e adaptação para outras bases.

3.4 Seleção, extração e verificação

A seleção privilegiou estudos com denominadores claros, categorias de IMC definidas, seguimento suficiente e estimativas ajustadas. Para cada publicação foram extraídos país, desenho, período, tamanho amostral, critério de exposição, urgência da cirurgia, desfechos, medida de efeito, intervalo de confiança e principais variáveis de ajuste. Dados usados em figuras foram conferidos com os valores apresentados no resumo, texto ou tabelas dos artigos. Quando uma publicação apresentava múltiplos desfechos, manteve-se a distinção entre frequência bruta e associação ajustada. Nenhuma estimativa ausente foi imputada e nenhuma curva foi digitalizada para gerar valores não reportados.

3.5 Medidas de efeito e interpretação

Foram apresentadas frequências, diferenças absolutas, riscos relativos e odds ratios ajustados, conforme a publicação original. A ênfase recaiu sobre risco absoluto quando disponível, porque uma redução relativa expressiva pode corresponder a pequeno benefício quando o risco basal é baixo. Para intervenções, a direção foi classificada como benefício, neutralidade ou possível dano somente quando o desfecho, a população e o comparador eram clinicamente equivalentes. Biomarcadores intermediários, como concentração anti-Xa, não foram tratados como substitutos automáticos de tromboembolismo clínico.

3.6 Avaliação crítica do risco de viés

Coortes foram examinadas segundo os domínios do ROBINS-I [6]: confundimento, seleção, classificação da exposição, desvios, dados faltantes, mensuração e relato seletivo. Ensaios foram avaliados conceitualmente pelos domínios do RoB 2, com atenção à randomização, aderência, mensuração e

perdas. A certeza foi considerada maior quando havia dose–resposta, grandes amostras, estimativas ajustadas e resultados convergentes; e menor quando o desfecho era composto, raro, dependente de prática institucional ou baseado em marcador farmacodinâmico. Essa avaliação foi usada para qualificar a linguagem, não para converter estudos heterogêneos em um escore único.

3.7 Síntese visual e reprodutibilidade

As visualizações foram construídas exclusivamente com dados publicados. O painel de 1.000 cesarianas utiliza as taxas de Conner et al. [7]; a curva de tempo e morbidade neonatal usa dados de Conner et al. [15]; os gráficos de intervenções empregam resultados de Tita et al., Valent et al., Tuuli et al., Overcash et al. e Bruno et al. [21,23-25,29]. Os diâmetros das bolhas representam o logaritmo da amostra, e não qualidade metodológica. As figuras evitam ranking entre estudos que medem fenômenos distintos e indicam explicitamente quando o desfecho é clínico ou intermediário.

4 RESULTADOS

4.1 Características e arquitetura da evidência

A síntese central reuniu grandes registros dos Estados Unidos, Dinamarca, Inglaterra e Nova Zelândia, além de coortes anestésicas da Ásia e ensaios clínicos de prevenção de infecção e tromboprofilaxia. As amostras variaram de 26 participantes em estudo farmacocinético a 239.348 partos em análise populacional. A evidência observacional concentrou-se em ferida, hemorragia e tempo operatório; os ensaios avaliaram antibioticoprofilaxia, curativos de pressão negativa e esquemas de enoxaparina. Poucos estudos mediram simultaneamente desfechos maternos, neonatais, experiência da paciente, custo e eventos após a alta.

Tabela 1. Estudos observacionais centrais incluídos na síntese.

Estudo	Desenho/N	Exposição	Domínio	Achado principal
Conner et al. 2014	Coorte; 2.444	IMC <30 a ≥50	Ferida	6,6% a 22,9%; aOR 3,0 no IMC ≥50
Stamilio; Scifres 2014	Coorte secundária; 585	<30; 30–45; >45	Infecção	18,8%; aOR 2,7 na obesidade extrema
Smid et al. 2015	Registro MFMU; 38.229	<30; 30–45; >45	Ferida/readmissão	aOR 1,65 para qualquer complicação; 5,47 para abertura
Leth et al. 2011	Prospectiva; 2.492	Obesidade vs. não	Infecção em 30 dias	aOR 1,43 geral e 1,86 hospitalar
Fyfe et al. 2012	Coorte; 11.363	Normal, sobrepeso, obesidade	Hemorragia	13,7%; 17,4%; 24,2%; aOR 1,73
Butwick et al. 2018	Populacional; 239.348	Classes de IMC	Hemorragia grave	Direção dependente da via de parto
Whitley et al. 2023	20 hospitais; 27.708	Categorias de IMC	HPP/transfusão	aOR 1,69 para HPP e 0,73 para transfusão
Conner et al. 2013	Coorte; 2.266	IMC <30 a ≥50	Tempo/neonatal	9,4 a 16,0 min; morbidade 23,0% a 32,1%
Girsen et al. 2014	Registro multicêntrico	Categorias de IMC	Tempo operatório	aOR 2,81 para intervalo ≥18 min
Smid et al. 2016	Registro MFMU; 41.262	IMC 18,5–29,9 a ≥50	Neonatal	aOR 1,81 aguda e 2,08 grave
Nivatpumin et al. 2023	Coorte; 494	IMC 40–49,9 vs. ≥50	Anestesia	Mais vasopressor, fluidos e sangramento

Fonte: elaboração própria a partir das publicações citadas [7-18].

4.2 Complicações da ferida: o gradiente mais consistente

A complicação da ferida foi o desfecho com gradiente mais estável. Conner et al. observaram aumento de 6,6% para 22,9% entre os extremos do IMC, com aOR 3,0 no grupo igual ou superior a 50 kg/m² [7]. Traduzido para 1.000 cesarianas, isso corresponde a aproximadamente 66 eventos no grupo sem obesidade e 229 eventos no grupo de superobesidade. A diferença absoluta de 163 eventos por 1.000 ajuda a compreender por que pequenas melhorias em

antibioticoprofilaxia, técnica, vigilância e retorno podem produzir impacto assistencial relevante.

O registro Maternal-Fetal Medicine Units reforçou que desfechos compostos podem esconder componentes de maior magnitude. Na obesidade extrema, as estimativas ajustadas chegaram a 3,77 para infecção, 5,47 para abertura da ferida e 2,97 para readmissão relacionada à ferida [9]. O padrão sugere que a falha mecânica e o diagnóstico após a alta merecem a mesma atenção que a infecção hospitalar. A coorte dinamarquesa de Leth et al. mostrou aumento de infecções diagnosticadas durante a internação e após a alta [10], evidenciando que sistemas que não capturam consultas externas subestimam a carga real.

Os mecanismos são convergentes, mas não exclusivos: menor perfusão relativa do tecido adiposo, maior espessura subcutânea, tensão nas bordas, formação de espaço morto, duração da operação, diabetes, colonização cutânea, umidade e dificuldade de inspeção. A contribuição de cada mecanismo varia entre pacientes. Assim, o IMC deve disparar preparação e vigilância, mas não substituir avaliação de glicemia, urgência, rotura de membranas, tempo de trabalho de parto, cesariana prévia, espessura do subcutâneo e condições sociais para o cuidado domiciliar.

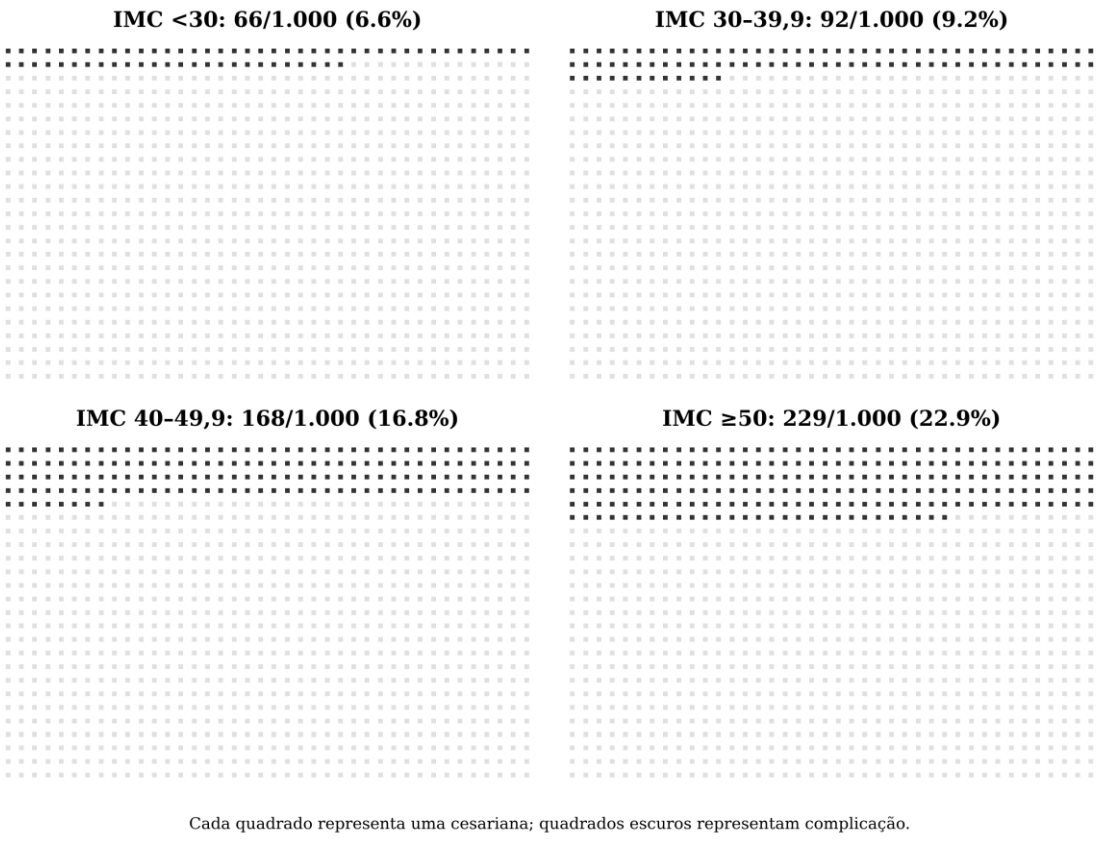


Figura 1. Painel de 1.000 cesarianas: frequência de complicações da ferida por categoria de IMC.

Fonte: dados publicados por Conner et al. [7]. Cada quadrado representa uma cesariana; não se trata de projeção causal.

4.3 Hemorragia: por que perda sanguínea e transfusão divergem

Fyfe et al. encontraram hemorragia pós-parto igual ou superior a 1.000 mL em 13,7% das mulheres com IMC normal, 17,4% das mulheres com sobrepeso e 24,2% das mulheres com obesidade; após cesariana, a associação ajustada foi 1,73 [12]. Whitley et al. também observaram maior chance de hemorragia definida por perda quantitativa, mas menor chance de transfusão no grupo com maior IMC [14]. Essa aparente contradição não invalida os estudos: indica que volume perdido e decisão transfusional são influenciados por reserva circulatória, hemoglobina basal, sintomas, protocolos e limiares institucionais.

Butwick et al., em análise populacional, mostraram que a direção e a magnitude variaram segundo a via de parto e a classe de obesidade [13]. A heterogeneidade é biologicamente plausível. Atonia, macrossomia, indução prolongada, oxitocina, pré-eclâmpsia e cirurgia repetida podem elevar a perda. Por outro lado, a mensuração visual tende a ser imprecisa, especialmente em campos maiores e com irrigação. A quantificação objetiva, a comunicação cumulativa do volume e a avaliação clínica permanecem mais úteis que a expectativa baseada no IMC.

Do ponto de vista operacional, obesidade deve aumentar a prontidão, não levar automaticamente a transfusão preventiva. O planejamento inclui acesso venoso confiável, disponibilidade de uterotônicos, conhecimento do risco de atonia, hemograma, tipagem quando indicada, sistema de resposta à hemorragia e equipamentos compatíveis. A prevenção de atraso é particularmente importante porque posicionamento, obtenção de acesso e transporte podem exigir mais tempo quando não foram preparados antes da urgência.

4.4 Tempo operatório, extração fetal e morbidade neonatal

O intervalo entre a incisão e o nascimento aumentou de 9,4 para 16,0 minutos conforme o IMC, enquanto a morbidade neonatal composta passou de 23,0% para 32,1% [15]. A progressão paralela não demonstra que cada minuto adicional produza dano; o intervalo também funciona como marcador de dificuldade cirúrgica, aderências, experiência, urgência e condição fetal prévia. Ainda assim, o dado é suficiente para justificar antecipação de pessoal, instrumentação, exposição e estratégia de extração.

Girsen et al. encontraram aOR 2,81 para intervalo igual ou superior a 18 minutos na obesidade mórbida [16]. No registro MFMU, a superobesidade associou-se a morbidade neonatal aguda de aproximadamente 10% versus 5% e grave de 6% versus 3% em grupos de referência, com estimativas ajustadas de 1,81 e 2,08 [17]. Parte da associação pode ser mediada por diabetes, hipertensão,

indicação da cesariana, macrossomia e prematuridade, mesmo quando os modelos realizam ajuste.

Na prática, o achado recomenda distinguir cesariana programada de cesariana urgente. Na programada, a equipe dispõe de tempo para posicionar a paciente, preparar exposição, confirmar equipamentos e discutir o plano. Na urgente, a margem temporal é menor e as falhas de sistema tornam-se mais perigosas. Protocolos devem evitar que equipamentos especiais sejam procurados somente depois da decisão de operar.

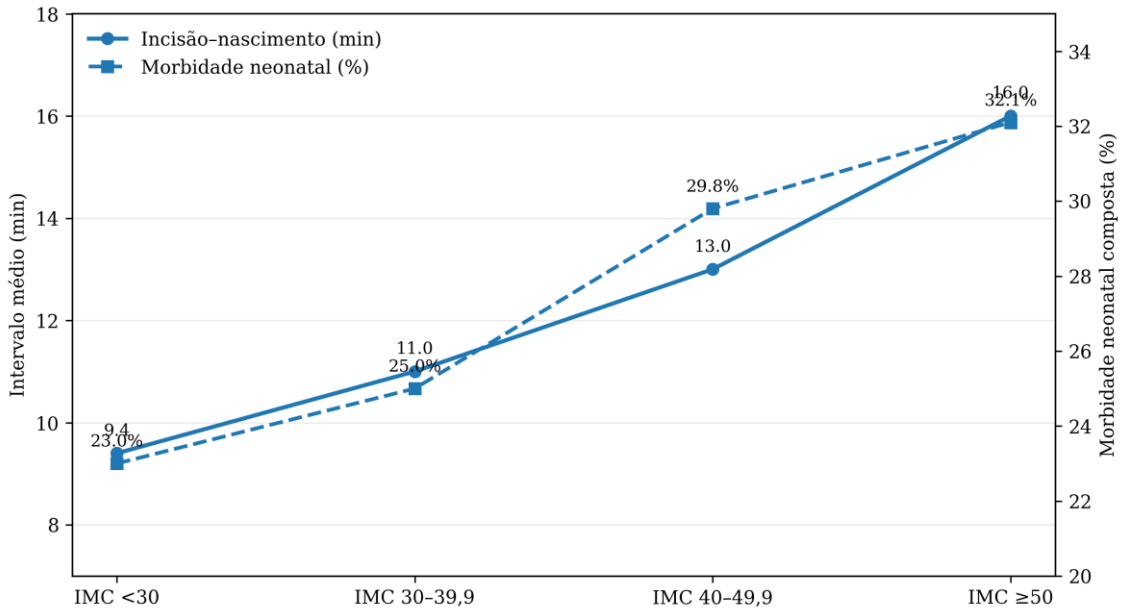


Figura 2. Progressão conjunta do intervalo incisão–nascimento e da morbidade neonatal.

Fonte: dados publicados por Conner et al. [15]. A associação é observacional e não atribui causalidade ao tempo isoladamente.

4.5 Anestesia, ventilação e estabilidade hemodinâmica

A anestesia neuroaxial permanece a estratégia preferida quando viável, por evitar manipulação de via aérea em uma população com dessaturação potencialmente rápida. Entretanto, referências anatômicas menos palpáveis, maior profundidade do espaço, posicionamento limitado e necessidade de múltiplas tentativas podem prolongar a instalação. Avaliação antenatal ou

precoce na admissão permite identificar apneia do sono, limitação funcional, refluxo, acesso venoso difícil e história de bloqueio prévio.

Em 494 mulheres, a comparação entre obesidade mórbida e superobesidade mostrou maior necessidade de vasopressores, fluidos intravenosos e maior perda sanguínea no grupo com IMC igual ou superior a 50 kg/m² [18]. O achado não significa que o IMC prediga sozinho falha anestésica, mas que a reserva e a resposta hemodinâmica devem ser consideradas. Manguitos adequados, posicionamento em rampa, pré-oxigenação, plano de via aérea e disponibilidade de ultrassom são elementos estruturais, não acessórios.

O planejamento anestésico também reduz risco neonatal indiretamente. Tentativas prolongadas, conversão inesperada para anestesia geral e dificuldade de transferência podem atrasar a incisão. A comunicação entre obstetra, anestesiologista, enfermagem e neonatologia deve ocorrer antes que a urgência elimine o tempo de preparação. O cuidado profissional evita linguagem culpabilizante e descreve necessidades concretas: profundidade, mobilidade, equipamento, risco respiratório e estratégia de resgate.

4.6 Prevenção de infecção: farmacologia e ensaios clínicos

4.6.1 Cefazolina e efeito do peso sobre a exposição

A antibioticoprofilaxia deve alcançar concentrações adequadas no momento da incisão e permanecer ativa durante o período de maior inoculação. Estudos farmacocinéticos em gestantes com obesidade demonstraram relação entre IMC, dose e concentrações plasmáticas e adiposas. Swank et al. observaram maior concentração no tecido adiposo com 3 g de cefazolina e sustentaram o uso de dose ampliada em mulheres com obesidade [27]. Young et al., em ensaio randomizado com 26 participantes, mostraram maior área sob a curva plasmática com 3 g, embora ambas as doses tenham alcançado concentrações adiposas acima de alvos microbiológicos no contexto estudado [28]. Esses achados apoiam ajuste

ao peso e redose quando duração e perda sanguínea justificarem, mas não autorizam inferir que 3 g reduz infecção clínica em todos os cenários.

4.6.2 Azitromicina adjunta e antibiótico pós-operatório

No ensaio C/SOAP, 2.013 mulheres submetidas a cesariana não eletiva durante trabalho de parto ou após rotura de membranas receberam profilaxia padrão com azitromicina ou placebo. A infecção materna composta caiu de 12,0% para 6,1%; endometrite, de 6,1% para 3,8%; e infecção da ferida, de 6,6% para 2,4% [23]. O benefício foi observado em uma população de risco infeccioso elevado e não deve ser transferido automaticamente para cesarianas programadas sem trabalho de parto ou membranas rotas.

Valent et al. randomizaram 403 mulheres com obesidade para cefalexina e metronidazol por 48 horas ou placebo após cefalosporina pré-operatória. A infecção do sítio cirúrgico ocorreu em 6,4% versus 15,4%, diferença absoluta de 9,0 pontos percentuais [24]. Entretanto, um ensaio publicado em 2025 avaliou a mesma estratégia após profilaxia pré-operatória de espectro ampliado, incluindo azitromicina, e não demonstrou redução significativa das complicações infecciosas em 30 dias [31]. A discrepância é informativa: quando o cuidado de base melhora, o benefício incremental de antibióticos adicionais pode diminuir, enquanto exposição, eventos adversos e pressão seletiva permanecem.

4.6.3 Pressão negativa profilática

A pressão negativa em incisão fechada tem racional mecânico e resultados conflitantes. Um ensaio pragmático dinamarquês sugeriu redução de infecção que exigia antibiótico [26], mas o maior ensaio multicêntrico, com 1.608 mulheres com obesidade, encontrou infecção em 3,6% com pressão negativa e 3,4% com curativo padrão. Reações cutâneas ocorreram em 7,0% e 0,6%, respectivamente [25]. O risco basal baixo, o tipo de dispositivo, a definição de infecção e a qualidade do cuidado padrão influenciam a estimativa. A evidência não sustenta

uso rotineiro universal; seleção pode ser discutida em pacientes com risco muito elevado, desde que custo, pele e preferências sejam considerados.

Tabela 2. Ensaios e estudos farmacológicos diretamente aplicáveis à prevenção perioperatória.

Estudo	Desenho/população	Comparação	Desfecho	Resultado central
Tita et al. 2016	RCT; 2.013; não eletiva	Azitromicina 500 mg IV + padrão vs. placebo	Infecção composta	6,1% vs. 12,0%; RR 0,51
Valent et al. 2017	RCT; 403; obesidade	Cefalexina + metronidazol 48 h vs. placebo	Infecção de sítio cirúrgico	6,4% vs. 15,4%; RR 0,41
Hyldig et al. 2019	RCT pragmático; 876	Pressão negativa vs. curativo padrão	Infecção com antibiótico	Redução no estudo; efeito dependente do contexto
Tuuli et al. 2020	RCT; 1.608; obesidade	Pressão negativa vs. curativo padrão	Infecção / reação cutânea	3,6% vs. 3,4%; pele 7,0% vs. 0,6%
Young et al. 2015	RCT farmacocinético; 26	Cefazolina 2 g vs. 3 g	Exposição plasmática/adiposa	Maior AUC com 3 g; desfecho intermediário
Overcash et al. 2015	Coorte sequencial; 85; IMC ≥40	Enoxaparina 0,5 mg/kg 12/12 h vs. IMC	Anti-Xa profilático	86% vs. 26%
Bruno et al. 2022	RCT; 146 pós-cesariana	Enoxaparina por peso vs. dose fixa	Anti-Xa profilático	66% vs. 44%; RR 1,49
Saad et al. 2025	RCT; obesidade; profilaxia ampliada	Cefalexina + metronidazol 48 h vs. placebo	Infecção em 30 dias	Sem benefício incremental significativo

Fonte: dados publicados nos estudos [23-31]. Resultados farmacocinéticos e anti-Xa não equivalem a eventos clínicos.

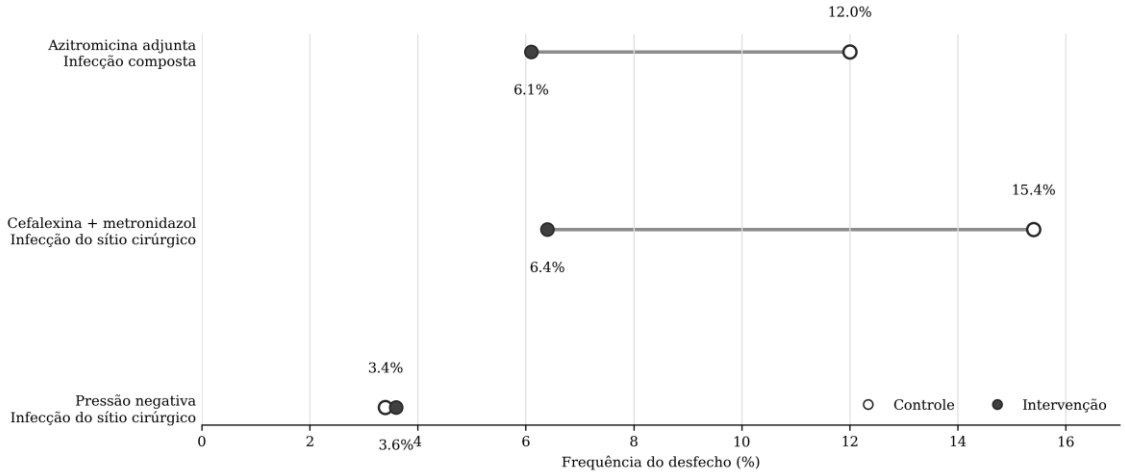


Figura 3. Mudança absoluta nos desfechos infecciosos em ensaios selecionados.

Fonte: Tita et al. [23], Valent et al. [24] e Tuuli et al. [25]. Populações e cuidados de base são diferentes; as linhas não constituem meta-análise.

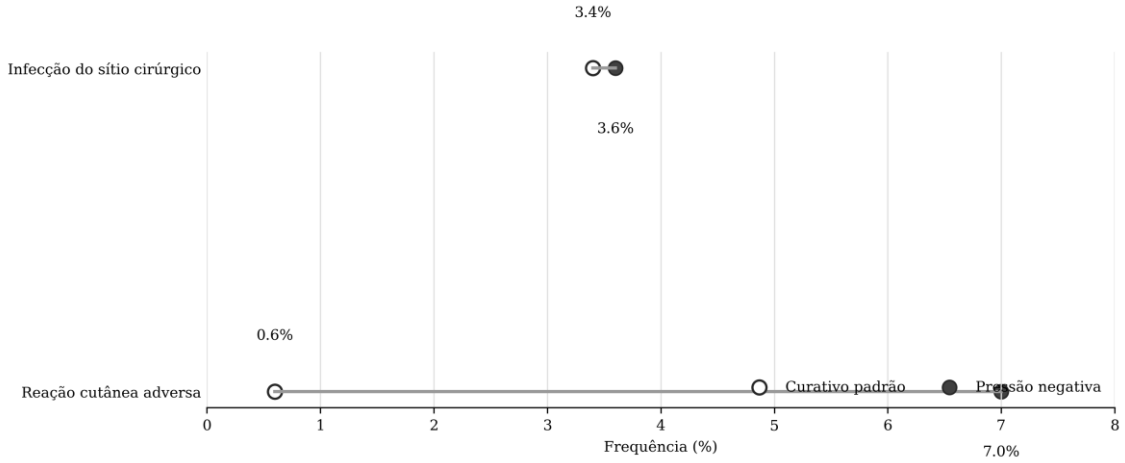


Figura 4. Balanço benefício–dano da pressão negativa no maior ensaio multicêntrico.

Fonte: Tuuli et al. [25]. O desfecho de reação cutânea inclui eventos locais relacionados ao dispositivo.

4.7 Tromboembolismo e adequação da farmacoprofilaxia

A cesariana e a obesidade acumulam fatores trombóticos no período em que o risco de tromboembolismo venoso já está elevado. A prevenção deve começar com avaliação de história de trombose, trombofilia, imobilidade, infecção, hemorragia, pré-eclâmpsia e outras condições. A SMFM recomenda

compressão pneumática para todas as mulheres submetidas à cesariana, iniciada antes da cirurgia e mantida até deambulação, e sugere dose intermediária de enoxaparina quando farmacoprofilaxia é indicada em obesidade classe III [30].

A evidência sobre dose é mais robusta para concentração anti-Xa do que para trombose clínica. Overcash et al. observaram níveis profiláticos em 86% das mulheres com dose de 0,5 mg/kg a cada 12 horas, comparados a 26% com esquema estratificado pelo IMC [21]. Bruno et al. confirmaram maior alcance do alvo com dose por peso: 66% versus 44% na análise por intenção de tratar, sem eventos tromboembólicos em nenhum grupo [29]. O ensaio foi interrompido precocemente por eficácia farmacodinâmica, mas não possuía poder para demonstrar redução de trombose.

O risco de sangramento e complicação da ferida deve ser integrado à decisão. Em Bruno et al., complicações da ferida ocorreram em 7% no grupo por peso e 1% no grupo de dose fixa, com intervalo de confiança muito amplo [29]. O achado é impreciso e não prova dano, mas recorda que alcançar um biomarcador não é o único objetivo. Protocolos devem especificar indicação, momento de início após anestesia neuroaxial, função renal, duração, educação na alta e sinais de sangramento e trombose.

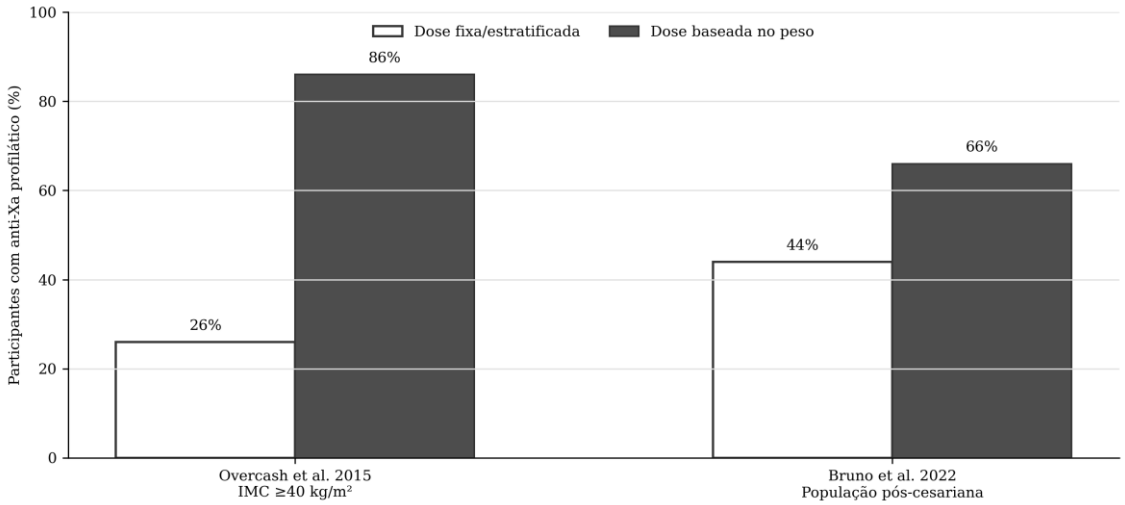


Figura 5. Alcance de concentração anti-Xa profilática com esquemas de enoxaparina após cesariana.

Fonte: Overcash et al. [21] e Bruno et al. [29]. Anti-Xa é marcador intermediário; os estudos não demonstraram superioridade em tromboembolismo clínico.

4.8 Evolução temporal e distribuição dos domínios

A linha do tempo evidencia concentração de grandes registros entre 2013 e 2018, seguida por estudos que aprofundaram anestesia, hemorragia, curativos e farmacoprofilaxia. O tamanho amostral melhora a precisão para eventos raros, mas não elimina confundimento, erros de classificação ou ausência de seguimento pós-alta. Ensaios menores respondem a perguntas específicas, porém sua aplicabilidade depende do cuidado de base, urgência da cesariana e perfil de risco.

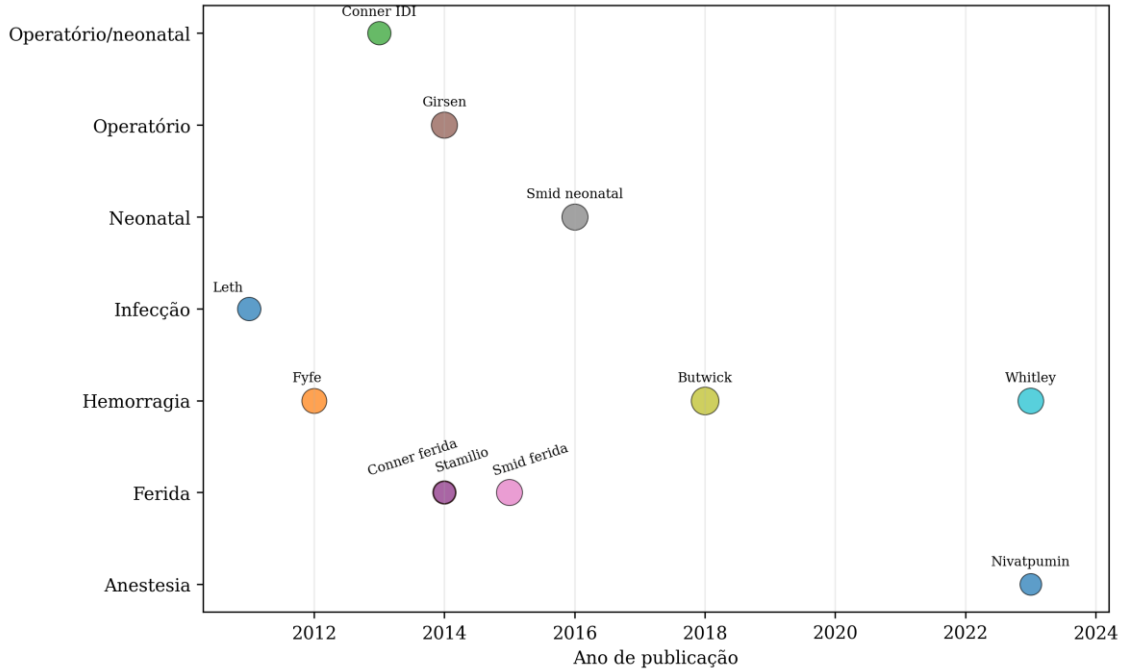


Figura 6. Linha do tempo da evidência por domínio e tamanho amostral.

Fonte: elaboração própria a partir dos estudos [7-18,23-31]. O diâmetro representa o logaritmo da amostra, não qualidade metodológica.

4.9 Magnitude e precisão dos efeitos ajustados

O gráfico de efeitos ajustados mostra maior magnitude em componentes específicos da ferida, sobretudo abertura e infecção, enquanto hemorragia e morbidade neonatal apresentam associações moderadas. A transfusão constitui exceção porque seu valor depende de decisão clínica e pode seguir direção diferente da perda. Em escalas logarítmicas, intervalos amplos indicam menor precisão. A leitura deve considerar risco basal, definição do desfecho e variáveis incluídas no modelo.

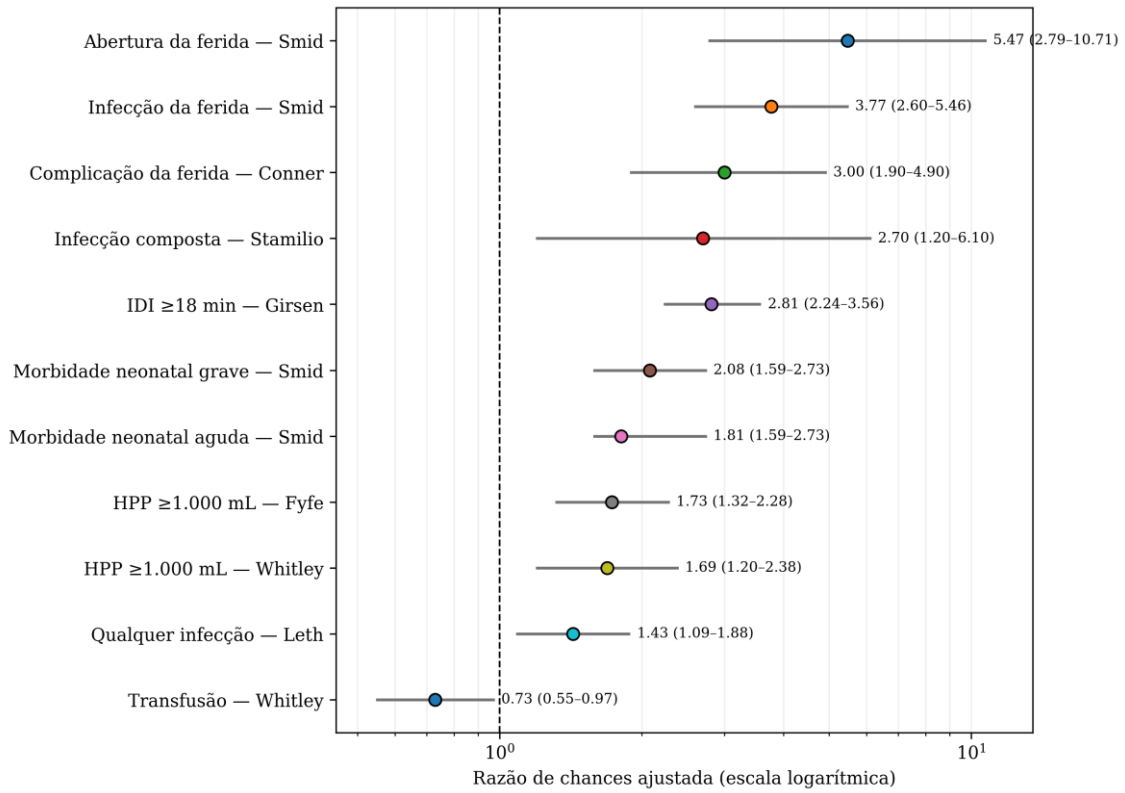


Figura 7. Estimativas ajustadas publicadas para desfechos selecionados.

Fonte: estudos citados [7-17]. Escala logarítmica; a linha vertical marca ausência de associação.

4.10 Cobertura dos domínios e lacunas de mensuração

A matriz de cobertura mostra fragmentação: estudos de ferida raramente mediram desfechos neonatais; estudos de hemorragia nem sempre capturaram infecção após a alta; análises neonatais pouco detalharam anestesia; e ensaios de tromboprofilaxia usaram biomarcadores em lugar de eventos clínicos. A ausência de uma célula não significa resultado negativo, mas que o domínio não foi mensurado ou não foi reportado de forma utilizável. Essa fragmentação limita modelos de decisão que precisam equilibrar simultaneamente infecção, sangramento, trombose, experiência e custo.

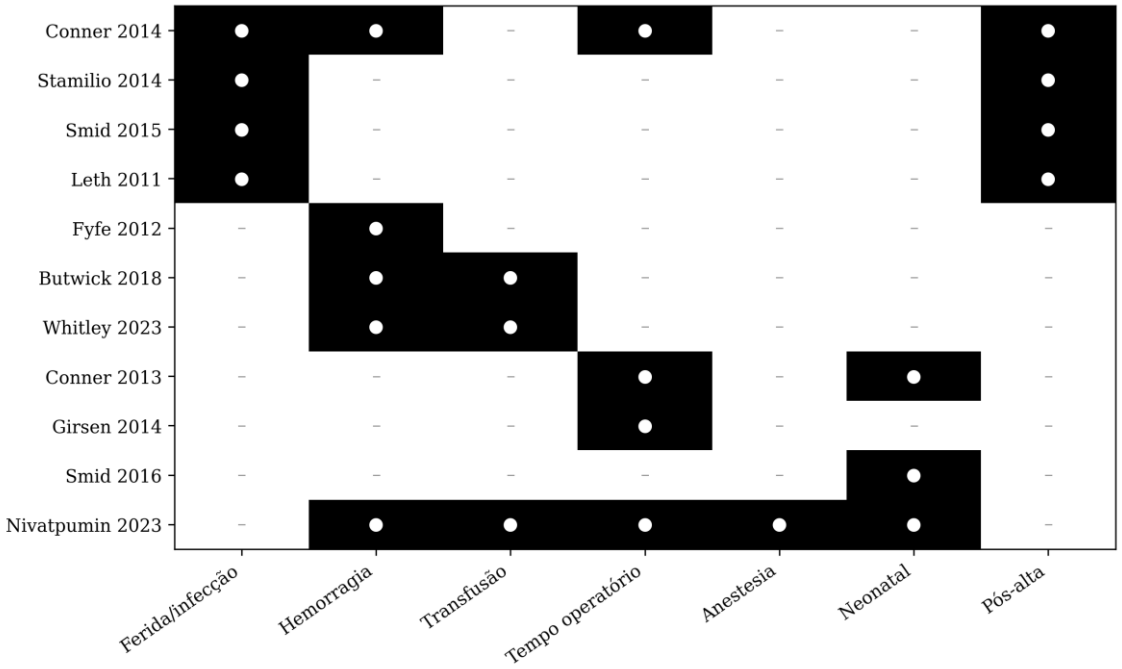


Figura 8. Matriz de cobertura dos domínios avaliados nos estudos centrais.

Fonte: elaboração própria a partir dos métodos e desfechos publicados [7-18,23-31]. Célula preenchida significa que o domínio foi medido, não que houve associação.

4.11 Síntese da qualidade e aplicabilidade

Tabela 3. Julgamento crítico por conjunto de evidências.

Conjunto	Risco de viés	Limitação principal	Força	Aplicabilidade
Ferida por classes de IMC	Moderado	Confundimento residual; definições distintas	Dose–resposta e consistência	Alta para planejamento de vigilância
Hemorragia e transfusão	Moderado a sério	Desfechos não equivalentes; limiares institucionais	Grandes amostras e ajuste	Moderada; interpretar por componente
Tempo e neonatal	Moderado	Confundimento por urgência e indicação	Gradiente e registros multicêntricos	Moderada para organização da equipe
Azitromicina em não eletiva	Baixo no RCT	Aplicabilidade restrita ao cenário estudado	Grande ensaio, desfechos clínicos	Alta para população semelhante

Conjunto	Risco de viés	Limitação principal	Força	Aplicabilidade
Antibiótico oral pós-operatório	Baixo a moderado	Cuidado de base mudou entre ensaios	Resultados clínicos e absolutos	Condicional ao regime pré-operatório
Pressão negativa	Baixo nos grandes RCTs	Dispositivos e risco basal diferentes	Desfecho clínico e dano cutâneo	Não sustenta uso universal
Enoxaparina baseada no peso	Baixo/moderado	Marcador substituto; poucos eventos	Convergência para anti-Xa	Moderada para farmacodinâmica, baixa para VTE

Fonte: avaliação crítica do capítulo com base nos domínios ROBINS-I e RoB 2 [6].

5 DISCUSSÃO

A principal mensagem é que a obesidade materna modifica a cesariana de modo graduado e multidimensional, mas não uniforme. O sinal mais estável aparece na ferida, sustentado por dose–resposta, grandes coortes e efeitos ajustados. Hemorragia, anestesia e neonatal apresentam associações clinicamente importantes, porém mais dependentes de indicação cirúrgica, comorbidades, urgência e práticas institucionais. Essa diferença de consistência deve orientar a linguagem: é apropriado afirmar que complicações da ferida aumentam com o IMC; é mais prudente dizer que a obesidade se associa a maior complexidade hemorrágica, sem presumir que toda mulher terá maior necessidade de transfusão.

A interpretação por risco absoluto é essencial. No gradiente de Conner et al., a diferença entre 6,6% e 22,9% representa grande incremento e justifica vigilância intensificada. No maior ensaio de pressão negativa, a diferença entre 3,4% e 3,6% não sugere benefício e foi acompanhada por aumento de reações cutâneas. A mesma tecnologia pode parecer promissora em estudos com risco basal alto e neutra em cenários com excelente cuidado padrão. Portanto, custo e complexidade não devem ser confundidos com efetividade.

Os ensaios de antibióticos ilustram a importância do contexto. A azitromicina adjunta produziu benefício claro em cesarianas não eletivas durante trabalho de parto ou após rotura de membranas [23]. O ensaio de cefalexina e metronidazol de 2017 mostrou redução absoluta de 9 pontos percentuais [24]. Contudo, o estudo de 2025 não confirmou ganho quando o cuidado pré-operatório já incluía cobertura ampliada [31]. A conclusão madura não é que um ensaio esteja "certo" e outro "errado", mas que o efeito incremental depende do risco residual após as medidas anteriores. Pacotes devem ser avaliados como sequência, evitando duplicação antimicrobiana sem benefício.

A farmacocinética da cefazolina reforça que doses padronizadas podem produzir exposições diferentes, mas também mostra os limites de biomarcadores. Concentrações plasmáticas e adiposas ajudam a escolher dose e redose, porém infecção depende de inoculação, técnica, glicemia, duração, microrganismos e cuidado da ferida. Da mesma forma, anti-Xa informa exposição à enoxaparina, mas não substitui trombose, sangramento ou qualidade de vida. A pesquisa obstétrica frequentemente usa marcadores porque eventos graves são raros; a prática deve reconhecer essa incerteza e não transformar alvo laboratorial em verdade clínica absoluta.

A relação entre tempo incisão–nascimento e desfecho neonatal deve ser tratada com cuidado causal. Um intervalo prolongado pode ser consequência de aderências, dificuldade anatômica, operador menos experiente ou necessidade de manobras adicionais. A condição fetal que motivou a cesariana pode já estar comprometida. Ajustes estatísticos reduzem, mas não eliminam, confundimento. Ainda assim, o achado é operacionalmente útil: se dificuldade é previsível, preparação antecipada é uma intervenção de baixo risco. O serviço pode organizar mesa, exposição, equipe neonatal, instrumentação e plano de extração antes da incisão.

A evidência anestésica indica maior necessidade de preparação e não inevitabilidade de falha. Ultrassom para acesso neuroaxial e vascular, posicionamento em rampa, pré-oxigenação, equipamento de via aérea,

manguitos adequados e pessoal para transferência são respostas a necessidades mensuráveis. Instituições que não dispõem desses recursos podem converter uma diferença anatômica em evento adverso. A qualidade do sistema é, portanto, modificador de efeito: o mesmo IMC pode produzir trajetórias diferentes em ambientes preparados e não preparados.

A tromboprofilaxia exige equilíbrio particular. A recomendação de compressão mecânica para todas as cesarianas é simples e amplamente aplicável [30]. A farmacoprofilaxia deve seguir estratificação, porque obesidade isolada, obesidade associada a história de trombose e obesidade com hemorragia recente representam cenários diferentes. Doses por peso alcançam mais frequentemente o alvo anti-Xa, mas a evidência não demonstrou redução de eventos clínicos e sugere necessidade de monitorar ferida e sangramento. Protocolos devem explicitar quem recebe, quando inicia e por quanto tempo, em vez de depender de decisões improvisadas.

A dimensão da equidade não pode ser separada da segurança. Estigma pode levar a comunicação menos respeitosa, atribuição indevida de sintomas ao peso e atraso na investigação. Equipamentos inadequados expõem paciente e trabalhadores a lesões. Orientações de alta que presumem capacidade de visualizar a incisão ou retornar facilmente ao hospital podem falhar. O planejamento deve incluir forma de inspecionar a ferida, apoio familiar quando desejado, telemonitoramento ou contato precoce e linguagem centrada na pessoa. O objetivo é reduzir risco sem reduzir a paciente ao IMC.

Este capítulo possui fortalezas: integra coortes, ensaios, farmacologia e diretrizes; apresenta riscos absolutos; distingue desfechos clínicos de intermediários; e evita uma meta-análise artificial entre definições incompatíveis. As limitações incluem predominância de estudos de países de alta renda, categorias de IMC distintas, ausência de dados individuais, possível sobreposição de registros, seguimento pós-alta incompleto e menor representação de mulheres com IMC extremamente alto. A busca concentrada em PubMed e fontes complementares pode não capturar toda a literatura

regional. Além disso, várias intervenções foram avaliadas antes da adoção de pacotes modernos, reduzindo transportabilidade.

Apesar das limitações, a convergência permite uma conclusão clínica robusta: o risco deve ser tratado como trajetória, não como rótulo. Avaliação pré-operatória, preparação da sala, profilaxia adequada, técnica cuidadosa, resposta à hemorragia, mobilização, tromboprofilaxia indicada e seguimento da ferida formam um contínuo. Nenhum dispositivo isolado substitui esse sistema. A estratégia mais profissional é selecionar intervenções segundo o risco residual e medir resultados locais, incluindo danos e experiência da paciente.

6 IMPLICAÇÕES PARA A PRÁTICA

A tradução dos resultados requer organização por momento assistencial. A tabela a seguir não constitui prescrição individual e deve ser adaptada às diretrizes nacionais, recursos locais, alergias, função renal, anestesia neuroaxial e perfil hemorrágico. Seu propósito é converter a evidência em perguntas verificáveis para a equipe.

Tabela 4. Matriz perioperatória de risco, ação e indicador de qualidade.

Momento	Avaliação	Ação prioritária	Indicador
Pré-operatório	Comorbidades, via aérea, mobilidade, cesariana prévia, hemograma	Revisão anestésica e obstétrica; plano compartilhado	Plano registrado antes da decisão urgente
Infraestrutura	Peso e dimensões, transferência, manguito, mesa	Confirmar capacidade dos equipamentos e pessoal	Ausência de atraso por equipamento
Antibiotico profilaxia	Peso, urgência, trabalho de parto, membranas, alergia	Dose adequada e espectro conforme cenário; redose indicada	Administração antes da incisão e registro de horário
Hemorragia	Anemia, placenta, macrossomia, indução, cirurgia repetida	Quantificação objetiva, uterotônicos e protocolo de resposta	Tempo até escalonamento e reconciliação da perda

Momento	Avaliação	Ação prioritária	Indicador
Ferida	IMC extremo, diabetes, subcutâneo, higiene, acesso ao retorno	Fechamento cuidadoso e estratégia individualizada	Infecção e abertura em 30 dias, não apenas internação
Trombose	História, trombofilia, imobilidade, sangramento, função renal	Compressão e farmacoprofilaxia quando indicada	Início, duração e adesão documentados
Neonatal	Urgência, traçado fetal, macrossomia, extração difícil	Equipe neonatal e plano de extração disponíveis	Intervalo incisão–nascimento contextualizado
Alta	Visualização da ferida, apoio, distância, letramento	Orientação escrita, contato precoce e sinais de alerta	Retorno em tempo oportuno e readmissão evitável

Fonte: síntese do capítulo e recomendações clínicas [2-4,19-31].

7 AGENDA DE PESQUISA

Pesquisas futuras devem padronizar o momento de aferição do peso e as categorias de IMC, incluir medidas de distribuição adiposa e capacidade funcional, e registrar urgência, trabalho de parto, rotura de membranas, indicação da cesariana, experiência do operador, espessura subcutânea e técnica anestésica. Desfechos de ferida precisam de vigilância ativa por pelo menos 30 dias, com captura de atendimentos fora do hospital de origem.

É prioritário desenvolver coortes multicêntricas no Brasil e em outros países de renda média, conectando prontuário hospitalar, atenção primária, sistemas de mortalidade e readmissões. Diferenças de infraestrutura podem modificar o efeito da obesidade e de tecnologias. Estudos econômicos devem incorporar custo do dispositivo, tempo da equipe, readmissão, antimicrobianos, afastamento e experiência da paciente, em vez de limitar-se ao preço inicial.

Ensaios pragmáticos devem comparar pacotes sequenciais e risco residual. Para infecção, a pergunta não é apenas se um antibiótico ou curativo funciona isoladamente, mas se acrescenta benefício após profilaxia adequada, preparo de pele, controle glicêmico e seguimento. Para tromboprofilaxia, estudos precisam

ser grandes o suficiente para eventos clínicos ou usar desenhos colaborativos, mantendo avaliação de sangramento, ferida e preferência.

Modelos preditivos devem ser externamente validados, calibrados por cenário e avaliados quanto a equidade. Um algoritmo que reproduz práticas de um hospital pode falhar em populações com diferentes prevalências, acesso ou critérios de transfusão. A utilidade deve ser demonstrada por melhora de decisão e desfechos, não apenas por área sob a curva. Participação de pacientes na escolha dos desfechos ajudará a incluir dor, mobilidade, comunicação, autonomia e estigma.

8 CONCLUSÃO

A obesidade materna, especialmente nas classes III e na superobesidade, está associada a maior complexidade da cesariana. A evidência é mais consistente para complicações da ferida e prolongamento do acesso fetal; hemorragia, anestesia, trombose e morbidade neonatal exigem interpretação contextual. Ensaios mostram que algumas intervenções reduzem infecção em populações bem definidas, enquanto outras, como pressão negativa aplicada rotineiramente, podem não oferecer benefício e acrescentar dano. Doses de enoxaparina baseadas no peso melhoram a adequação de anti-Xa, mas ainda não demonstram superioridade em eventos tromboembólicos. O cuidado de maior qualidade combina preparação, profilaxia baseada no cenário, técnica, vigilância pós-alta, respeito e infraestrutura. O IMC deve orientar prevenção proporcional, nunca substituir julgamento clínico nem justificar estigma.

REFERÊNCIAS

1. World Health Organization. Obesity and overweight [Internet]. Geneva: WHO; 2025 [cited 2026 Aug 1].

2. Poston L, Caleyachetty R, Cnattingius S, Corvalán C, Uauy R, Herring S, et al. Preconceptional and maternal obesity: epidemiology and health consequences.

Lancet Diabetes Endocrinol. 2016;4(12):1025-1036. doi:10.1016/S2213-8587(16)30217-0.

3. Catalano PM, Shankar K. Obesity and pregnancy: mechanisms of short term and long term adverse consequences for mother and child. BMJ. 2017;356:j1. doi:10.1136/bmj.j1.

4. American College of Obstetricians and Gynecologists. Obesity in pregnancy: ACOG Practice Bulletin, Number 230. Obstet Gynecol. 2021;137(6):e128-e144. doi:10.1097/AOG.0000000000004395.

5. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

6. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

7. Conner SN, Verticchio JC, Tuuli MG, Odibo AO, Macones GA, Cahill AG. Maternal obesity and risk of postcesarean wound complications. Am J Perinatol. 2014;31(4):299-304. doi:10.1055/s-0033-1348402.

8. Stamilio DM, Scifres CM. Extreme obesity and postcesarean maternal complications. Obstet Gynecol. 2014;124(2 Pt 1):227-232. doi:10.1097/AOG.0000000000000384.

9. Smid MC, Dotters-Katz SK, Silver RM, Kuller JA, Boggess KA, Stamilio DM. Extreme obesity and postcesarean wound complications in the Maternal-Fetal Medicine Unit Cesarean Registry. Am J Perinatol. 2015;32(14):1336-1342.

10. Leth RA, Uldbjerg N, Nørgaard M, Møller JK, Thomsen RW. Obesity, diabetes, and the risk of infections diagnosed in hospital and post-discharge infections after cesarean section: a prospective cohort study. Acta Obstet Gynecol Scand. 2011;90(5):501-509.

11. Wloch C, Wilson J, Lamagni T, Harrington P, Charlett A, Sheridan E. Risk factors for surgical site infection following caesarean section in England: results from a multicentre cohort study. BJOG. 2012;119(11):1324-1333.

12. Fyfe EM, Thompson JMD, Anderson NH, Groom KM, McCowan LME. Maternal obesity and postpartum haemorrhage after vaginal and caesarean delivery among nulliparous women at term: a retrospective cohort study. BMC Pregnancy Childbirth. 2012;12:112.

13. Butwick AJ, Abreo A, Bateman BT, Lee HC, El-Sayed YY, Stephansson O, et al. Effect of maternal body mass index on postpartum hemorrhage. Anesthesiology. 2018;128(4):774-783.

14. Whitley J, Dazelle W, Kripalani S, Ahmadzia H. The association between body mass index and postpartum hemorrhage after cesarean delivery. Sci Rep. 2023;13:11998. doi:10.1038/s41598-023-38526-7.

15. Conner SN, Tuuli MG, Longman RE, Odibo AO, Macones GA, Cahill AG. Impact of obesity on incision-to-delivery interval and neonatal outcomes at cesarean delivery. Am J Obstet Gynecol. 2013;209(4):386.e1-386.e6.

16. Girsen AI, Osmundson SS, Naqvi M, Garabedian MJ, Lyell DJ. Body mass index and operative times at cesarean delivery. Obstet Gynecol. 2014;124(4):684-689. doi:10.1097/AOG.0000000000000462.

17. Smid MC, Vladutiu CJ, Dotters-Katz SK, Manuck TA, Boggess KA, Stamilio DM. Maternal super obesity and neonatal morbidity after term cesarean delivery. Am J Perinatol. 2016;33(12):1198-1204.

18. Nivatpumin P, Lertbunnaphong T, Maneewan S, Vittayaprechapon N. Comparison of perioperative outcomes and anesthetic-related complications of morbidly obese and super-obese parturients delivering by cesarean section. Ann Med. 2023;55(1):2197299.

19. Hussamy DJ, Wortman AC, McIntire DD, Leveno KJ, Casey BM, Roberts SW. Closed incision negative pressure therapy in morbidly obese women undergoing cesarean delivery: a randomized controlled trial. Obstet Gynecol. 2019;134(4):781-789.

20. Zaki MN, Wing DA, McNulty JA. Wound complications in obese women after cesarean: a comparison of staples versus subcuticular suture. J Perinatol. 2016;36(10):819-822.

21. Overcash RT, Somers AT, LaCoursiere DY. Enoxaparin dosing after cesarean delivery in morbidly obese women. Obstet Gynecol. 2015;125(6):1371-1376. doi:10.1097/AOG.0000000000000873.

22. Słabuszewska-Jóźwiak A, Szymański JK, Ciebiera M, Sarecka-Hujar B, Jakiel G. A systematic review and meta-analysis of wound complications after a caesarean section in obese women. J Clin Med. 2021;10(4):675.

23. Tita ATN, Szychowski JM, Boggess K, Saade G, Longo S, Clark E, et al. Adjunctive azithromycin prophylaxis for cesarean delivery. N Engl J Med. 2016;375(13):1231-1241. doi:10.1056/NEJMoa1602044.

24. Valent AM, DeArmond C, Houston JM, Reddy S, Masters HR, Gold A, et al. Effect of post-cesarean delivery oral cephalexin and metronidazole on surgical site infection among obese women: a randomized clinical trial. JAMA. 2017;318(11):1026-1034. doi:10.1001/jama.2017.10567.

25. Tuuli MG, Liu J, Tita ATN, Longo S, Trudell A, Carter EB, et al. Effect of prophylactic negative pressure wound therapy vs standard wound dressing on surgical-site infection in obese women after cesarean delivery: a randomized clinical trial. JAMA. 2020;324(12):1180-1189. doi:10.1001/jama.2020.13361.

26. Hyldig N, Vinter CA, Kruse M, Mogensen O, Bille C, Sorensen JA, et al. Prophylactic incisional negative pressure wound therapy reduces the risk of surgical site infection after caesarean section in obese women: a pragmatic randomised clinical trial. BJOG. 2019;126(5):628-635. doi:10.1111/1471-0528.15413.

27. Swank ML, Wing DA, Nicolau DP, McNulty JA. Increased 3-gram cefazolin dosing for cesarean delivery prophylaxis in obese women. Am J Obstet Gynecol. 2015;213(3):415.e1-415.e8. doi:10.1016/j.ajog.2015.05.030.

28. Young OM, Shaik IH, Twedt R, Binstock A, Althouse AD, Venkataramanan R, et al. Pharmacokinetics of cefazolin prophylaxis in obese gravidae at time of cesarean delivery. Am J Obstet Gynecol. 2015;213(4):541.e1-541.e7. doi:10.1016/j.ajog.2015.06.034.

29. Bruno AM, Allshouse AA, Campbell HM, Branch DW, Lim MY, Silver RM, et al. Weight-based compared with fixed-dose enoxaparin prophylaxis after cesarean delivery: a randomized controlled trial. Obstet Gynecol. 2022;140(4):575-583. doi:10.1097/AOG.0000000000004937.

30. Society for Maternal-Fetal Medicine; Pacheco LD, Saade G, Metz TD. Society for Maternal-Fetal Medicine Consult Series #51: thromboembolism prophylaxis for cesarean delivery. Am J Obstet Gynecol. 2020;223(2):B11-B17. doi:10.1016/j.ajog.2020.04.032.

31. Saad AF, Goldman B, Spencer N, Kuhlmann M, McDonnold M, Saade GR. Prophylactic oral cephalexin and metronidazole compared with placebo after cesarean delivery to reduce infection complications in women with obesity: a randomized controlled trial. Obstet Gynecol. 2025;146(1):113-120. doi:10.1097/AOG.0000000000005936.

APÊNDICE A — ESTRATÉGIA DE BUSCA REPRODUZÍVEL

Exemplo para PubMed/MEDLINE: ((obesity[Title/Abstract] OR obese[Title/Abstract] OR "body mass index"[Title/Abstract] OR "morbid

obesity"[Title/Abstract] OR "super obesity"[Title/Abstract]) AND (pregnancy[Title/Abstract] OR maternal[Title/Abstract] OR parturient*[Title/Abstract]) AND (cesarean[Title/Abstract] OR caesarean[Title/Abstract])) AND (wound[Title/Abstract] OR "surgical site infection"[Title/Abstract] OR hemorrhage[Title/Abstract] OR transfusion[Title/Abstract] OR anesthesia[Title/Abstract] OR neonatal[Title/Abstract] OR thromboprophylaxis[Title/Abstract] OR enoxaparin[Title/Abstract] OR cefazolin[Title/Abstract] OR azithromycin[Title/Abstract] OR "negative pressure"[Title/Abstract]). Aplicar filtros de humanos quando necessário, sem restringir inicialmente por idioma ou desenho.

Buscas complementares: nomes dos autores e títulos de estudos-chave; rastreamento de referências de revisões; consulta a registros de ensaios para confirmar população, comparadores e desfechos. A atualização deve documentar data, base, estratégia e número de registros antes de qualquer submissão como revisão sistemática formal.

APÊNDICE B — DICIONÁRIO MÍNIMO DE EXTRAÇÃO

Tabela B1. Variáveis recomendadas para estudos futuros.

Bloco	Variáveis mínimas
População	País, centro, período, idade, paridade, urgência, trabalho de parto, comorbidades
Exposição	Momento do peso, método de altura, categorias de IMC, IMC ≥50 e ≥60
Cirurgia	Indicação, primária/repetida, incisão, aderências, tempo total e incisão–nascimento
Prevenção	Antibiótico, redose, preparo de pele, fechamento, tromboprofilaxia, pressão negativa
Maternos	Infecção, abertura, seroma, hematoma, HPP, transfusão, trombose, UTI, readmissão
Neonatais	Apgar, pH, reanimação, ventilação, trauma, desconforto respiratório, UTI, morte

Bloco	Variáveis mínimas
Experiência	Dor, mobilidade, compreensão, estigma, satisfação, barreiras de retorno
Análise	Estimativa bruta e ajustada, confundidores, dados faltantes, seguimento, perdas

Fonte: elaboração própria.

CAPÍTULO 12

COMO, PARA QUEM E EM QUAIS CONTEXTOS O CUIDADO INTEGRADO MELHORA A SAÚDE MENTAL DE PESSOAS COM MULTIMORBIDADE? REVISÃO REALISTA COM ANÁLISE COMPARATIVA QUALITATIVA

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Leila Silveira Vieira Bezerra; Saulo de França Oliveira; Uilna Natércia Soares Feitosa; Nélio Barreto Vieira

RESUMO

Objetivo: explicar como, para quem e em quais contextos o cuidado integrado melhora a saúde mental de pessoas com multimorbidade. Métodos: foi realizada revisão realista orientada pelos padrões RAMESES, combinada a análise comparativa qualitativa crisp-set. Foram selecionados 11 ensaios ou programas avaliativos reais que integraram saúde mental e condições crônicas. Os casos foram codificados quanto a seleção de sofrimento clinicamente relevante, psicoterapia direcionada, mensuração sistemática, ajuste terapêutico, coordenação e continuidade. Configurações contexto–mecanismo–resultado foram construídas e casos discordantes foram examinados. Resultados: programas efetivos combinaram, com alta regularidade, sofrimento clinicamente identificado, conteúdo psicológico direcionado, mensuração da resposta e continuidade. A configuração D*P*M*R apresentou consistência elevada e ampla cobertura. Coordenação isolada não foi suficiente: o programa 3D melhorou experiência de cuidado, mas não qualidade de vida ou sintomas, enquanto Step-Dep não preveniu depressão em população subclínica. COMPASS mostrou que equipes complexas não foram necessárias quando psicoterapia digital especializada, monitoramento e suporte longitudinal estiveram disponíveis. Conclusão: o cuidado integrado funciona quando recursos organizacionais ativam acesso, confiança, autoeficácia e resposta terapêutica adaptativa. Não

existe componente único suficiente; o benefício depende da combinação entre seleção adequada, tratamento psicológico, mensuração e continuidade, ajustada à capacidade da atenção primária e às vulnerabilidades do paciente.

Palavras-chave: multimorbidade; saúde mental; cuidado integrado; atenção primária; cuidado colaborativo; revisão realista; análise comparativa qualitativa.

ABSTRACT

Objective: to explain how, for whom, and under which circumstances integrated care improves mental health in people with multimorbidity. Methods: a realist review guided by RAMESES standards was combined with crisp-set qualitative comparative analysis. Eleven real trials or evaluated programs integrating mental health and chronic-condition care were selected. Cases were coded for clinically relevant distress selection, targeted psychotherapy, systematic measurement, treatment adjustment, care coordination, and continuity. Context–mechanism–outcome configurations and deviant cases were examined. Results: effective programs regularly combined clinically identified distress, targeted psychological content, response measurement, and continuity. The D*P*M*R configuration showed high consistency and broad coverage. Coordination alone was insufficient: 3D improved care experience but not quality of life or symptoms, while Step-Dep did not prevent depression in a subclinical population. COMPASS showed that a complex team was not necessary when specialized digital psychotherapy, monitoring, and longitudinal support were available. Conclusion: integrated care works when organizational resources activate access, trust, self-efficacy, and adaptive treatment response. No single component was sufficient; benefit depended on combinations of appropriate selection, psychological treatment, measurement, and continuity adapted to primary-care capacity and patient vulnerability.

Keywords: multimorbidity; mental health; integrated care; primary care; collaborative care; realist review; qualitative comparative analysis.

1 INTRODUÇÃO

Multimorbidade transforma o cuidado em um problema de coordenação. Pessoas com duas ou mais condições crônicas circulam entre consultas, exames, medicamentos e recomendações que frequentemente foram organizados por doença. Quando depressão, ansiedade ou sofrimento persistente se somam a esse percurso, a fragmentação se torna mais prejudicial: sintomas mentais reduzem energia e adesão, enquanto a carga física e terapêutica aumenta desesperança, preocupação e isolamento. Modelos integrados procuram substituir trajetórias paralelas por planos compartilhados, mas seus resultados são heterogêneos.[1-4]

Revisões tradicionais respondem se uma categoria média de intervenção funciona. Para cuidado integrado, essa pergunta é incompleta. O mesmo rótulo pode designar um enfermeiro gestor de caso, uma equipe multiprofissional, psicoterapia na atenção primária, telemonitoramento, reunião de especialistas ou revisão global de medicamentos. Além disso, recursos idênticos podem ativar mecanismos diferentes conforme acesso, vínculo, disponibilidade de profissionais, gravidade mental, complexidade física e apoio familiar. A questão relevante passa a ser: qual combinação funciona, em qual contexto e por meio de qual mecanismo?[5-8]

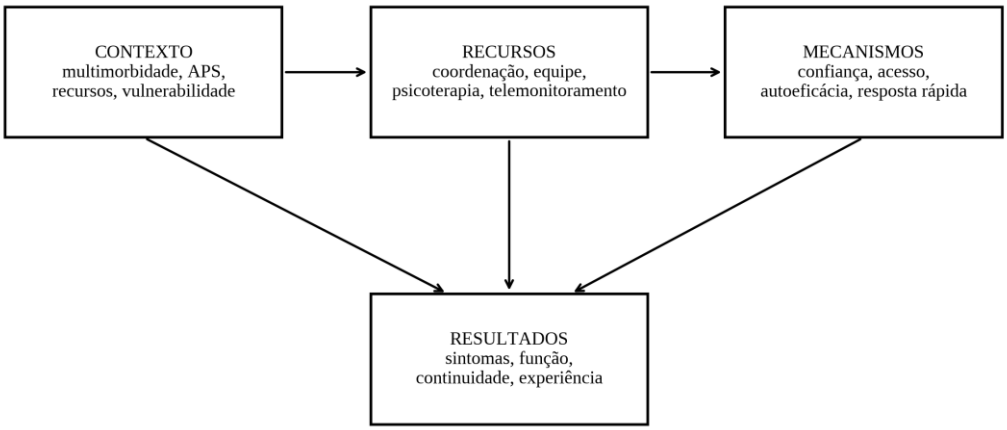
A revisão realista parte da ideia de que programas oferecem recursos e oportunidades, mas os resultados dependem das respostas desencadeadas nas pessoas e nos profissionais. Coordenação pode gerar confiança e reduzir repetição; monitoramento pode produzir sensação de acompanhamento e permitir ajuste precoce; psicoterapia pode aumentar autoeficácia; contato remoto pode reduzir barreiras. Porém, coordenação sem tratamento específico pode apenas organizar melhor um cuidado insuficiente, e mensuração sem capacidade de resposta pode transformar rastreamento em registro burocrático.[5,6,9]

A análise comparativa qualitativa complementa essa lógica ao tratar causalidade como configuração. Em vez de estimar o efeito isolado de cada componente, procura combinações suficientes ou necessárias entre casos. Essa

abordagem é adequada a um conjunto intermediário de programas complexos, mas exige codificação transparente e interpretação junto à teoria. O objetivo deste capítulo foi desenvolver e testar configurações de cuidado integrado para saúde mental em multimorbidade, usando ensaios reais, casos positivos e casos sem benefício.[10-12]

A Figura 1 apresenta a cadeia contexto–recurso–mecanismo–resultado que orientou a revisão. Ela antecede os métodos para explicitar que mecanismos não são componentes visíveis, mas respostas geradas quando recursos são oferecidos em circunstâncias específicas.

Figura 1. Modelo inicial contexto–mecanismo–resultado para cuidado integrado em multimorbidade.



Fonte: elaboração dos autores com base na metodologia realista.[5,6]

A Figura 1 mostra que equipe, coordenação e telemonitoramento foram tratados como recursos, não como mecanismos. Os mecanismos postulados incluíram confiança, percepção de acesso, autoeficácia, redução de fragmentação e possibilidade de resposta rápida. Essa distinção evitou afirmar que uma estrutura organizacional produz benefício automaticamente; o resultado dependeu de o recurso ser utilizado, compreendido e ajustado à necessidade.

2 MÉTODOS

2.1 Delineamento e desenvolvimento da teoria inicial

Foi realizada revisão realista com análise comparativa qualitativa crisp-set. A condução foi orientada pelos padrões RAMESES para sínteses realistas e pelos princípios de transparência do PRISMA 2020. A teoria inicial foi construída a partir de literatura sobre cuidado colaborativo, atenção primária, multimorbidade e implementação. Foram propostas seis famílias de recursos: equipe multiprofissional, coordenação ou gestão de casos, mensuração sistemática, psicoterapia integrada, ajuste de tratamento e continuidade presencial ou remota.[5,6,13]

A pergunta foi estruturada em três eixos: em quais contextos o cuidado integrado foi oferecido; quais mecanismos foram plausivelmente ativados; e quais resultados mentais, funcionais ou experienciais ocorreram. Foram considerados contextos a presença de depressão selecionada, sofrimento subclínico, multimorbidade geral, ruralidade, pós-operatório, câncer, limitação de acesso e integração prévia da atenção primária. Os mecanismos foram inferidos a partir da descrição do programa, processo e resultados, e não apenas de declarações dos autores.

2.2 Busca, seleção e extração

A busca foi realizada em PubMed/MEDLINE até 1º de agosto de 2026, combinando multimorbidity, depression, anxiety, integrated care, collaborative care, primary care, case management, psychological therapy, telecare, randomized trial e realist review. Foram rastreadas revisões e referências. Foram elegíveis ensaios ou avaliações comparativas de adultos com pelo menos uma condição física crônica e depressão, ansiedade ou risco mental, além de programas gerais de multimorbidade que incluíssem componente de saúde mental.

Foram selecionados 11 casos com descrição suficiente da configuração: TEAMcare, COINCIDE, TrueBlue, HOPE, Vera, Step-Dep, COMPASS, 3D, RAINBOW, Bypassing the Blues e INCPAD.[14-24] Foram extraídos população, tamanho amostral, cenário, profissionais, componentes, intensidade, comparador, resultado principal e duração. Resultados foram considerados positivos quando o ensaio demonstrou melhora estatisticamente significativa do desfecho mental primário ou de sofrimento no tempo principal; resultados exclusivamente de experiência foram descritos separadamente.

2.3 Codificação e análise comparativa qualitativa

Cada caso recebeu codificação binária para seis condições: D, seleção de depressão ou sofrimento clinicamente relevante; P, psicoterapia ou tratamento psicológico direcionado; M, mensuração sistemática da resposta; A, possibilidade explícita de ajuste ou intensificação; C, coordenação ou gestão de casos; e R, continuidade longitudinal, incluindo contatos remotos. O resultado Y indicou benefício mental no desfecho principal. A codificação foi baseada nos relatórios publicados e revisada por coerência interna. Ausência de descrição suficiente foi tratada como ausência do componente, decisão conservadora que foi discutida como limitação.

Foram construídas tabela-verdade, medidas de consistência e cobertura para condições isoladas e configurações teoricamente relevantes. Na análise crisp-set, consistência de suficiência representou a proporção de casos com a configuração que apresentou o resultado; cobertura representou a proporção de casos positivos explicada. Para necessidade, a consistência representou a proporção de casos positivos que continha a condição. Não foi aplicado algoritmo de minimização automática como substituto da teoria; as soluções foram escolhidas pela plausibilidade e examinadas contra casos discordantes.[10-12]

2.4 Construção das configurações realistas

Os achados quantitativos da QCA foram integrados à leitura realista. Para cada configuração, foram formuladas proposições do tipo: quando determinado contexto está presente, certos recursos ativam mecanismos que produzem resultado. Casos negativos foram usados para refinar as proposições. O programa 3D testou coordenação ampla sem psicoterapia direcionada; Step-Dep testou intervenção em população subclínica; COMPASS testou psicoterapia digital especializada sem equipe médica complexa. Essa variação permitiu separar componentes frequentes de componentes realmente explicativos.

3 RESULTADOS

3.1 Panorama dos casos

A Tabela 1 descreve os casos antes da codificação. Sua apresentação permite compreender que a comparação não colocou programas idênticos lado a lado: houve diferenças de condição física, cenário, intensidade e alvo mental. A QCA foi usada justamente para explorar essas combinações, e não para produzir um ranking simples.

Tabela 1. Programas de cuidado integrado incluídos na revisão realista.

Caso	n	Contexto	Recursos centrais	Resultado
TEAMcare	214	APS integrada; diabetes/DCV + depressão	Enfermeiro gestor, psiquiatra, treat-to-target	Melhora mental e controle clínico
COINCIDE	387	36 práticas; diabetes/DCV + depressão	Psicoterapia breve + enfermeiro da prática	Redução modesta de depressão
TrueBlue	≈400	APS; diabetes/DCV + depressão	Enfermeiro, plano e seguimento	Benefício sustentado

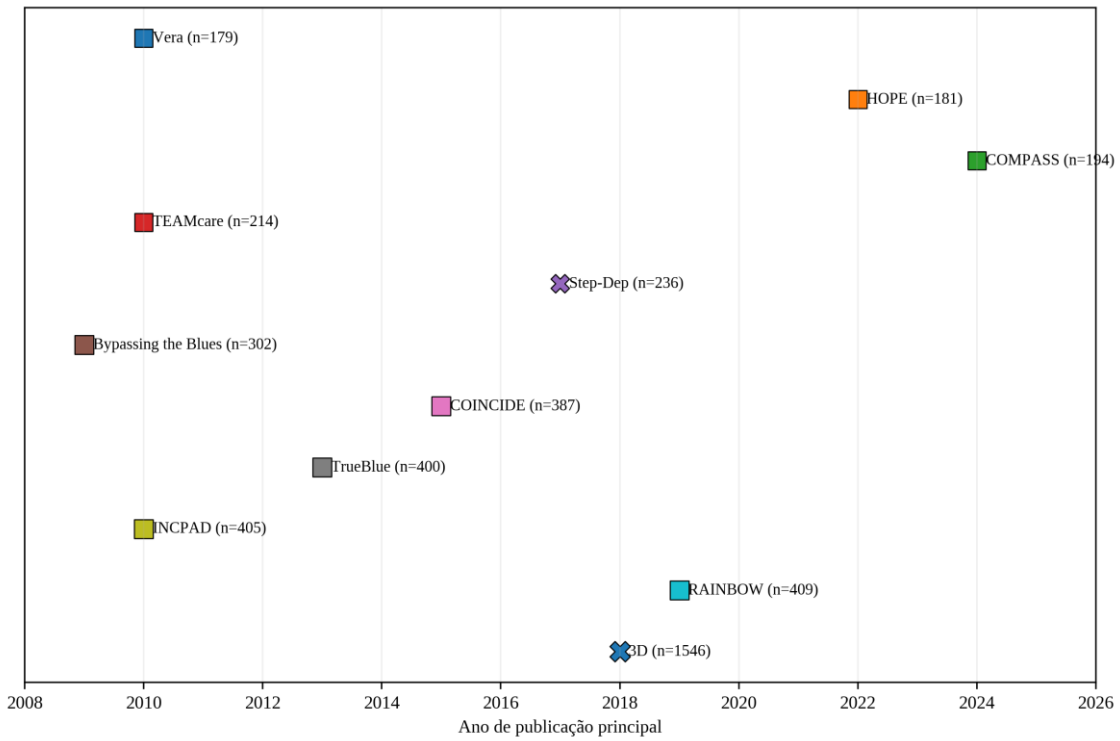
Caso	n	Contexto	Recursos centrais	Resultado
HOPE	181	Índia rural; multimorbidade + depressão	Cuidado colaborativo adaptado	Menor depressão até 12 meses
Vera	179	Porto Rico; doenças crônicas + depressão	Médico, especialista e gestor	Melhora de sintomas e função
Step-Dep	236	Sintomas subclínicos	Cuidado escalonado preventivo	Sem prevenção superior
COMPASS	194	Condições crônicas diversas	TCC digital com terapeuta	DMP 0,71 para sofrimento
3D	1.546	Multimorbidade geral	Revisão multidisciplinar semestral	Melhor experiência; sem QV
RAINBOW	409	Obesidade + depressão	Peso + terapia de solução de problemas	Efeito modesto duplo
Bypassing the Blues	302	Pós-CABG + depressão	Cuidado colaborativo por telefone	Melhora de depressão/QV
INCPAD	405	Câncer + dor/depressão	Telecare e ajuste algorítmico	Melhora de dor/depressão

Fonte: ensaios e relatórios originais.[14-24] APS: atenção primária à saúde; CABG: cirurgia de revascularização; DCV: doença cardiovascular; DMP: diferença média padronizada; QV: qualidade de vida.

A Tabela 1 mostra que os casos positivos variaram de programas intensivos, como TEAMcare, a uma intervenção digital focal, como COMPASS. Dois casos foram particularmente informativos por não produzirem o resultado mental esperado: Step-Dep trabalhou com prevenção em sintomas subclínicos, e 3D reorganizou cuidado geral sem fornecer psicoterapia específica. Essa diversidade sustentou a análise configuracional.

A Figura 2 apresenta distribuição temporal e tamanho dos ensaios. Ela foi construída para mostrar que a base incluiu desde estudos pequenos e focalizados até um ensaio pragmático de mais de 1.500 participantes.

Figura 2. Evolução temporal e dimensão dos casos avaliados.



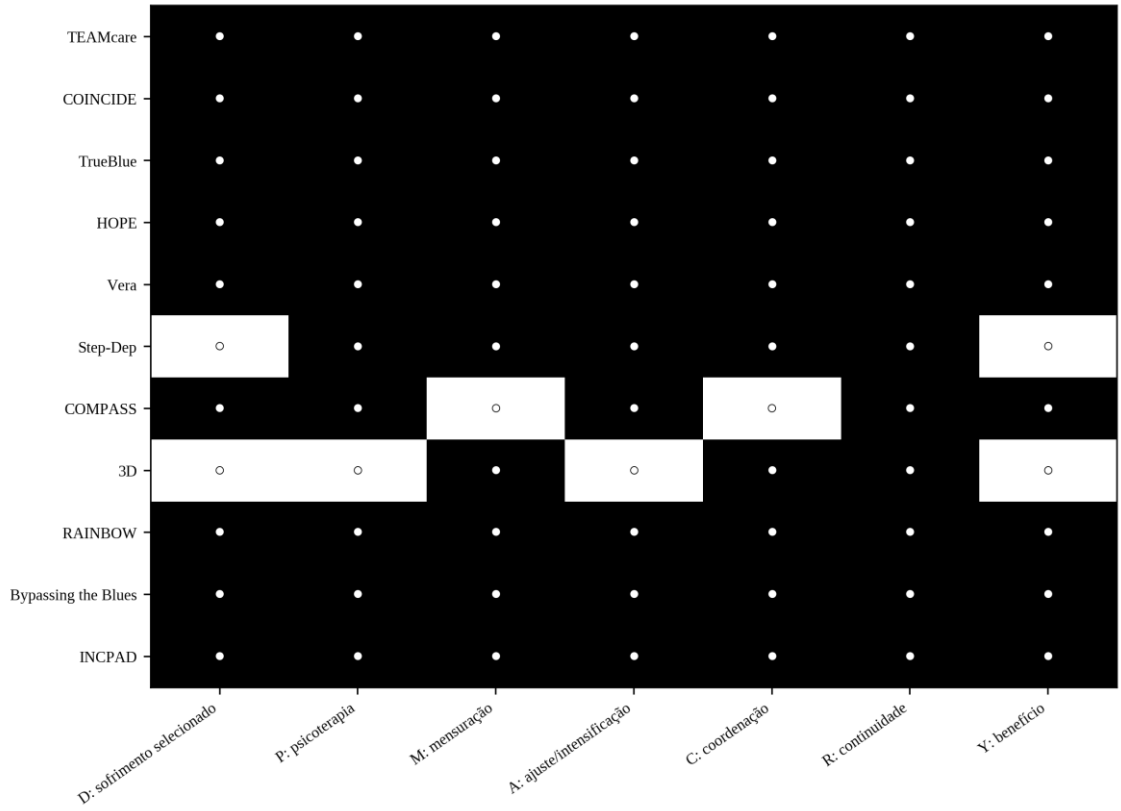
Fonte: elaboração dos autores com dados dos ensaios.[14-24] Quadrados indicam resultado mental favorável; X indica ausência de benefício no desfecho principal.

A Figura 2 evidencia que os resultados positivos não foram restritos a uma época ou tamanho de amostra. O maior ensaio, 3D, foi negativo para qualidade de vida, demonstrando que precisão não compensa desalinhamento entre intervenção e desfecho. COMPASS, publicado mais recentemente, mostrou que suporte digital especializado pode gerar efeito expressivo mesmo sem extensa reorganização clínica.

3.2 Codificação dos componentes e casos discordantes

A Figura 3 apresenta a matriz de condições. A leitura por linhas permite identificar a arquitetura de cada programa; a leitura por colunas mostra quais componentes acompanharam os resultados positivos.

Figura 3. Matriz crisp-set das condições e do resultado nos 11 casos.



Fonte: codificação dos autores com base nos relatórios originais.[14-24] D: sofrimento selecionado; P: psicoterapia; M: mensuração; A: ajuste; C: coordenação; R: continuidade; Y: benefício mental.

A Figura 3 mostra que D, P, M e R estiveram presentes em quase todos os casos positivos. Coordenação e ajuste foram frequentes, mas COMPASS obteve benefício sem gestão clínica complexa, indicando que esses recursos não foram necessários em todos os contextos. Step-Dep continha vários componentes, porém não selecionou depressão clinicamente estabelecida; 3D continha coordenação e continuidade, mas não tratamento psicológico focal.

A Tabela 2 traduz a matriz em proposições realistas preliminares. Cada linha liga contexto, recurso, mecanismo e resultado, evitando tratar a configuração como associação puramente matemática.

Tabela 2. Configurações contexto–mecanismo–resultado refinadas pelos casos.

Contexto	Recursos	Mecanismo inferido	Resultado
Depressão + doença crônica; APS com capacidade de seguimento	Gestor, mensuração, psicoterapia e ajuste	Paciente percebe acompanhamento; equipe responde à não melhora	Redução de depressão e melhor autocuidado
Barreiras geográficas ou funcionais	Telefone/digital + suporte profissional	Acesso e continuidade aumentam; menor custo de participação	Melhora de sintomas e adesão
Multimorbidade geral sem sofrimento selecionado	Revisão ampla e coordenação	Maior organização, mas pouca ativação de mudança psicológica	Melhor experiência sem benefício mental claro
Sintomas subclínicos	Intervenção preventiva escalonada	Baixa necessidade percebida e menor margem de melhora	Ausência de prevenção superior
Condição crônica diversa com sofrimento elevado	TCC digital especializada + mensuração	Conteúdo relevante e feedback aumentam autoeficácia	Grande redução de sofrimento
Alta complexidade e controle clínico inadequado	Treat-to-target e supervisão especializada	Resposta rápida e alinhamento de metas	Benefício mental e físico combinado

Fonte: síntese realista dos autores, sustentada pelos casos incluídos.[14-24]

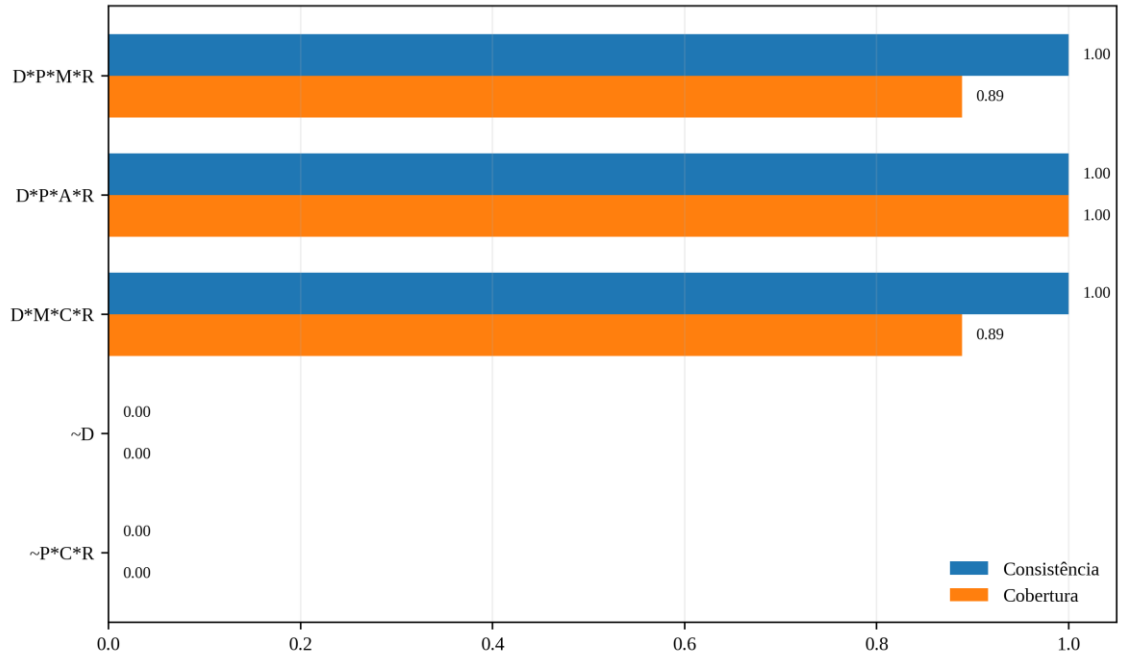
A Tabela 2 mostra que mecanismos semelhantes podem ser ativados por recursos diferentes. Continuidade pode ser obtida por visitas de enfermagem, telefonemas ou plataforma digital; o elemento causal provável é a percepção de que a resposta será observada e de que existe caminho para ajuste. Em contrapartida, uma reunião multidisciplinar que não altera o tratamento psicológico pode melhorar experiência sem reduzir sintomas.

3.3 Configurações suficientes e condições necessárias

A Figura 4 compara consistência e cobertura das configurações escolhidas a priori. A análise privilegiou combinações interpretáveis e evitou minimização

excessiva, que poderia produzir soluções matematicamente simples, porém clinicamente opacas.

Figura 4. Consistência e cobertura das configurações examinadas.

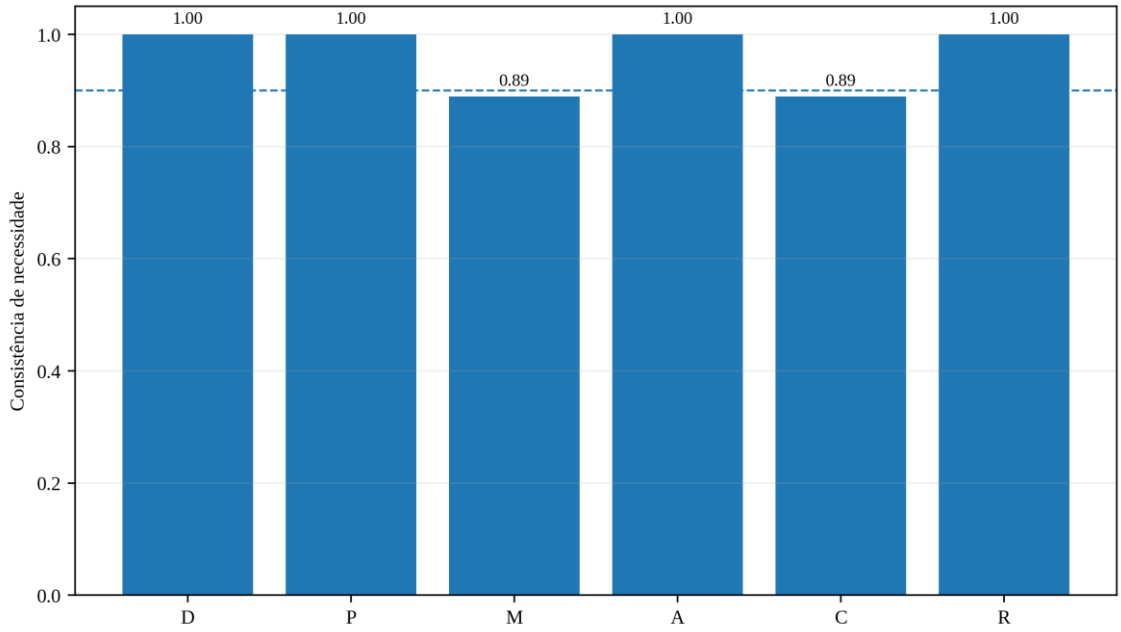


Fonte: cálculos dos autores a partir da matriz crisp-set.

A Figura 4 indica que D*P*M*R e D*P*A*R apresentaram consistência muito alta e ampla cobertura. A combinação D*M*C*R também foi frequente, mas seu poder explicativo dependeu de P: coordenação com mensuração, sem tratamento psicológico, não produziu melhora no 3D. A configuração ~D concentrou os casos negativos, sustentando a importância de selecionar sofrimento clinicamente relevante quando o objetivo principal é reduzir sintomas.

A Figura 5 examina necessidade das condições individuais. Ela foi inserida para mostrar que um componente pode aparecer em todos os casos positivos sem, isoladamente, ser suficiente.

Figura 5. Consistência de necessidade das condições nos casos com benefício mental.



Fonte: cálculos dos autores. A linha tracejada de 0,90 representa limiar convencional de necessidade.

A Figura 5 mostra que seleção de sofrimento, psicoterapia direcionada, mensuração e continuidade atingiram ou se aproximaram do limiar de necessidade. Coordenação e ajuste ficaram abaixo porque COMPASS produziu benefício por via digital focal. A interpretação é configuracional: os componentes de alta necessidade não constituem "ingredientes mágicos", mas criam as condições para que tratamento e resposta adaptativa ocorram.

A Tabela 3 resume a tabela-verdade e os casos que refinaram a teoria. Ela foi apresentada antes do modelo final para tornar visível a evidência que sustentou cada inferência.

Tabela 3. Configurações selecionadas, consistência, cobertura e casos explicativos.

Configuração	Casos	Consistência	Cobertura	Casos informativos
D*P*M*R	8	1,00	0,89	TEAMcare, COINCIDE, HOPE, RAINBOW

Configuração	Casos	Consistência	Cobertura	Casos informativos
D*P*A*R	9	1,00	1,00	TEAMcare, COINCIDE, HOPE, RAINBOW
D*M*C*R	8	1,00	0,89	TEAMcare, COINCIDE, HOPE, RAINBOW
~D	2	0,00	0,00	Step-Dep e 3D
~P*C*R	1	0,00	0,00	3D

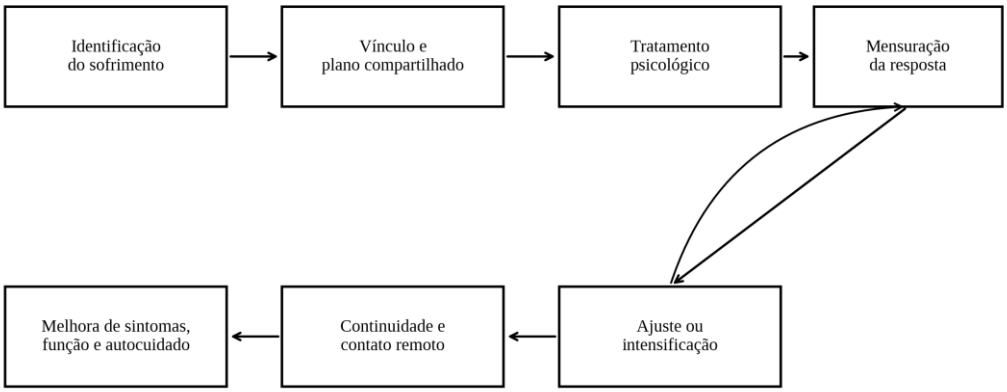
Fonte: cálculos dos autores. * representa conjunção; ~ representa ausência.

A Tabela 3 confirma que as soluções com sofrimento selecionado, psicoterapia, mensuração e continuidade explicaram a maior parte dos programas efetivos. A configuração ~D não foi interpretada como causa suficiente de fracasso, mas como marcador de baixa margem de melhora quando o desfecho exigia redução de depressão. Os casos negativos impediram que coordenação e equipe multiprofissional fossem indevidamente descritas como suficientes.

3.4 Modelo realista refinado

A Figura 6 integra os achados em um ciclo operacional. Ela foi posicionada ao final dos resultados porque representa teoria refinada pela comparação dos casos, e não um modelo definido apenas antes da revisão.

Figura 6. Ciclo de cuidado integrado capaz de sustentar melhora da saúde mental.



Fonte: elaboração dos autores a partir da síntese realista e da QCA.

A Figura 6 mostra que identificação, tratamento, mensuração e ajuste formam um circuito. Continuidade fecha esse circuito ao garantir que a informação produza ação. Quando uma etapa se rompe, o cuidado perde potência: rastreamento sem tratamento gera diagnóstico sem resposta; psicoterapia sem seguimento pode atenuar; coordenação sem conteúdo mental específico organiza o fluxo, mas não necessariamente modifica sintomas.

4 DISCUSSÃO

Esta revisão mostrou que o cuidado integrado não melhora saúde mental apenas porque reúne profissionais. Os casos efetivos compartilharam uma lógica: identificaram sofrimento clinicamente relevante, ofereceram tratamento psicológico direcionado, mediram resposta e mantiveram continuidade suficiente para ajustar o plano. Equipes, gestão de casos e tecnologia foram meios

para ativar mecanismos de acesso, confiança, autoeficácia e resposta adaptativa. A configuração foi mais importante do que qualquer componente isolado.

TEAMcare exemplificou a integração mais completa. Enfermeiros acompanharam metas de depressão, glicemia, pressão e lipídios, com supervisão e intensificação. O efeito não decorreu apenas de mais contatos; o programa converteu medidas em decisões. A pessoa percebeu um plano coerente para problemas físicos e mentais, e a equipe recebeu autorização e estrutura para agir quando metas não foram alcançadas. Esse desenho explica por que houve melhora simultânea de depressão e controle clínico.[14]

COINCIDE demonstrou que uma intervenção breve, realizada em parceria com enfermeiros de práticas gerais, pode produzir benefício modesto em pessoas com diabetes ou doença cardiovascular. O tamanho menor do efeito em comparação ao TEAMcare é plausível diante da intensidade e duração. Ainda assim, a integração no ambiente em que a pessoa já recebia cuidado físico reduziu barreiras e normalizou o tratamento mental. O mecanismo central parece ter sido tornar psicoterapia acessível e conectada ao manejo da doença.[15]

Programas como TrueBlue, HOPE e Vera reforçaram a importância da adaptação local. Em contextos rurais ou com menor disponibilidade de especialistas, gestores e profissionais de atenção primária assumiram funções ampliadas, apoiados por protocolos e supervisão. A fidelidade não exigiu reproduzir a mesma profissão ou plataforma; exigiu preservar função: identificar, acompanhar, tratar e ajustar. Essa flexibilidade é uma força do cuidado colaborativo, mas depende de treinamento, tempo protegido e rotas de referência.[16-18]

COMPASS foi o caso que mais claramente mostrou equifinalidade. A intervenção não precisou de extensa equipe multiprofissional ou gestão médica porque selecionou sofrimento elevado e ofereceu TCC digital específica para viver com uma condição física, acompanhada por terapeuta e mensuração. O efeito padronizado de 0,71 foi maior do que o observado em muitas intervenções organizacionais. Isso não significa que plataformas digitais substituam cuidado

integrado, mas que a função de integração pode ser alcançada por uma trajetória focal quando o problema clínico e a intervenção estão bem alinhados.[19]

Os casos negativos foram igualmente importantes. O 3D melhorou experiência e percepção de cuidado centrado na pessoa, mas não qualidade de vida. O programa reorganizou revisões, medicamentos e prioridades, porém não selecionou sofrimento mental nem forneceu psicoterapia direcionada. Esse resultado mostra que integração geral é valiosa, mas não deve ser avaliada apenas por um desfecho que exige intervenção mais específica. A ausência de melhora mental não invalida o 3D; delimita o que seu mecanismo era capaz de produzir.[20]

Step-Dep tentou prevenir depressão maior em pessoas com sintomas subclínicos e condições crônicas, mas não superou o cuidado usual. A baixa incidência, a menor necessidade percebida e a possibilidade de melhora espontânea reduziram margem de benefício. Na linguagem realista, o recurso foi oferecido em contexto no qual o mecanismo de engajamento e urgência podia não ser ativado. Programas preventivos talvez precisem segmentar risco com maior precisão, usar estratégias mais leves ou escolher desfechos de bem-estar e função, em vez de conversão diagnóstica.[21]

RAINBOW mostrou que depressão e obesidade podem ser tratadas em um programa combinado. A magnitude foi modesta, mas houve mudança em ambos os domínios. O caso sustenta a ideia de que metas físicas e mentais podem se reforçar quando não competem por espaço. Intervenções separadas podem aumentar carga; um plano compartilhado reduz mensagens contraditórias e permite que melhora de energia favoreça atividade física, enquanto mudança comportamental favorece autoeficácia.[22]

Bypassing the Blues e INCPAD ampliaram a teoria para transições e oncologia. O telefone foi importante não por sua tecnologia, mas porque preservou continuidade diante de recuperação cirúrgica, dor, fadiga e deslocamento. Monitoramento remoto permitiu detectar não resposta e ajustar tratamento. Esses casos explicam por que acompanhamento remoto funciona

melhor quando está ligado a uma equipe com capacidade de agir; chamadas de cortesia sem rota terapêutica provavelmente não ativariam o mesmo mecanismo.[23,24]

A síntese recente de Sweeney e colaboradores encontrou benefícios pequenos para intervenções organizacionais e orientadas ao paciente, com evidência de que cuidado colaborativo pode manter redução de depressão por períodos mais longos.[25] Nossos achados explicam essa média: programas organizacionais têm efeitos modestos porque atuam por vários elos e incluem populações complexas; sua vantagem é continuidade. Programas focais podem produzir efeito maior no curto prazo, mas dependem de adesão e manutenção. A combinação entre conteúdo específico e arquitetura longitudinal parece mais promissora.

As implicações para implementação são diretas. Antes de criar uma nova equipe, o serviço deve mapear funções: quem rastreia, quem confirma necessidade, quem oferece intervenção, quem mede resposta, quem ajusta e quem mantém contato. Sem essas respostas, multiprofissionalidade pode se reduzir a encaminhamentos. Liderança, prontuário compartilhado, tempo protegido, supervisão e financiamento são contextos habilitadores. Barreiras incluem rotatividade, agendas incompatíveis, fragmentação tecnológica e ausência de serviços para intensificação.[26-29]

O apoio familiar apareceu menos explicitamente nos ensaios do que outros componentes, mas pode atuar como contexto e recurso. Famílias podem apoiar rotina, transporte, comunicação e segurança, porém também aumentar conflito ou sobrecarga. A participação deve ser consentida e orientada às preferências. Em pessoas com limitações cognitivas, dependência funcional ou regimes complexos, o mecanismo de continuidade pode depender de cuidador; em outras, autonomia e privacidade são centrais.

A análise crisp-set trouxe clareza, mas simplificou programas complexos em presença ou ausência. Componentes podem ter intensidades diferentes, e a documentação incompleta pode ter sido codificada como ausência. O número de

casos foi pequeno e a seleção intencional, não exaustiva. Consistência e cobertura não constituíram testes probabilísticos e não corrigiram risco de viés dos ensaios. A QCA serviu para organizar comparação e refinar teoria, devendo ser replicada com conjunto maior e codificação independente.

A revisão realista também dependeu de inferência de mecanismos. Poucos ensaios mediram confiança, acesso percebido, autoeficácia ou resposta da equipe diretamente. Assim, as proposições foram sustentadas por padrões e descrições, mas permanecem testáveis. Estudos futuros deveriam incorporar avaliação de processo, entrevistas, dados de fidelidade e mediação. Uma mesma redução de depressão pode ocorrer por acesso, relação terapêutica, solução de problemas ou melhora física; distinguir essas vias permitirá adaptar programas.

No conjunto, a evidência favorece um desenho modular. Serviços com recursos limitados podem iniciar por rastreamento validado, psicoterapia breve ou digital apoiada e contato longitudinal. Serviços com maior complexidade podem acrescentar gestor de caso, supervisão psiquiátrica e treat-to-target físico. O núcleo deve permanecer: necessidade real, tratamento específico, mensuração e continuidade. A adaptação pode alterar a forma sem perder a função.

A sustentabilidade financeira modifica a configuração causal. Programas baseados em atividade não remunerada ou em projetos temporários podem demonstrar eficácia e desaparecer ao final do financiamento. O cuidado integrado requer pagamento por coordenação, tempo para discussão de casos, supervisão e sistemas de informação. Quando o modelo remunera apenas consultas isoladas, profissionais tendem a retornar à fragmentação. A institucionalização deve definir responsabilidades, carga de trabalho e indicadores que valorizem desfechos compartilhados, e não apenas volume de atendimentos.[26-29]

A equidade também precisa ser tratada como condição de funcionamento. Telemonitoramento reduz deslocamentos, mas pode excluir pessoas sem internet, dispositivos, privacidade ou alfabetização digital. Psicoterapia breve na atenção primária amplia acesso, porém serviços localizados em territórios

vulneráveis frequentemente têm maior rotatividade e menor disponibilidade de especialistas. Um programa pode apresentar boa eficácia média e, ainda assim, ampliar desigualdades se alcançar principalmente usuários com maior capacidade de navegação. Avaliações devem estratificar participação, abandono e benefício por renda, território, raça, escolaridade e deficiência, sem transformar vulnerabilidade em característica individual desconectada do sistema.

Os desfechos escolhidos influenciam a aparente efetividade. Redução de PHQ-9 ou de uma escala combinada é importante, mas pacientes com multimorbidade podem valorizar energia, sono, participação, capacidade de cumprir prioridades e menor carga de tratamento. O 3D ilustrou que experiência de cuidado pode melhorar mesmo sem mudança em qualidade de vida global. Um conjunto mínimo deveria incluir sintomas, função, metas pessoais, segurança, uso de serviços, experiência e carga do cuidador. Essa amplitude permite reconhecer benefícios organizacionais sem confundi-los com efeito psicoterapêutico.

A governança de dados passou a ser parte do mecanismo de continuidade. Mensuração só produz resposta quando o resultado chega ao profissional certo, em tempo útil, e está ligado a uma regra de ação. Prontuários que não se comunicam, alertas excessivos e registros duplicados podem aumentar carga. Em contraste, painéis simples, listas de pacientes sem resposta e planos compartilhados facilitam priorização. A tecnologia deve apoiar o ciclo clínico e preservar confidencialidade; não deve substituir julgamento nem automatizar encaminhamentos sem considerar contexto e preferência.

No contexto da atenção primária brasileira, a lógica modular é especialmente pertinente. Equipes de Saúde da Família já dispõem de vínculo territorial, agentes comunitários e acompanhamento longitudinal, mas enfrentam diferenças locais de acesso a psicologia e psiquiatria. Matriciamento, teleconsultoria, grupos, intervenções breves e gestão de casos podem assumir formas diferentes sem perder a função central. A implementação precisa respeitar atribuições, evitar sobrecarga e estabelecer critérios de intensificação

para atenção especializada. A coordenação deve se apoiar na rede existente, e não criar uma camada paralela de cuidado.

Esses achados também sugerem uma agenda experimental mais explicativa. Ensaios futuros devem comparar configurações, e não apenas pacotes completos contra cuidado usual. Desenhos fatoriais, SMART e avaliações híbridas de efetividade e implementação podem testar dose, sequência e escalonamento. Análises de mediação devem verificar se acesso, autoeficácia, aliança e resposta rápida explicam os resultados. A QCA pode ser atualizada com maior número de casos e conjuntos fuzzy para representar intensidade, desde que a calibração seja pré-especificada e discutida com especialistas e usuários.

5 CONCLUSÃO

O cuidado integrado melhorou a saúde mental de pessoas com multimorbidade quando converteu organização em tratamento responsivo. A combinação mais consistente reuniu sofrimento clinicamente relevante, intervenção psicológica direcionada, mensuração e continuidade, frequentemente apoiadas por coordenação e capacidade de ajuste. Equipes multiprofissionais e tecnologia não foram suficientes nem sempre necessárias; sua utilidade dependeu dos mecanismos que ativaram. Casos negativos mostraram que coordenação geral pode melhorar experiência sem reduzir sintomas e que prevenção em populações subclínicas exige outra lógica. A implementação deve começar por funções essenciais, adaptar recursos ao contexto e avaliar mecanismos, não apenas componentes.

REFERÊNCIAS

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education. Lancet. 2012;380:37-43. doi:10.1016/S0140-6736(12)60240-2.

2. Coventry PA, Small N, Panagioti M, Adeyemi I, Bee P. Living with complexity: marshalling resources for self-management of multimorbidity. BMC Health Serv Res. 2015;15:98. doi:10.1186/s12913-015-0799-5.

3. Ngo VK, Rubinstein A, Ganju V, et al. Grand challenges: integrating mental health care into the non-communicable disease agenda. PLoS Med. 2013;10:e1001443. doi:10.1371/journal.pmed.1001443.

4. Leijten FRM, Struckmann V, van Ginneken E, et al. The SELFIE framework for integrated care for multi-morbidity. Health Policy. 2018;122:12-22. doi:10.1016/j.healthpol.2017.06.002.

5. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review: a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy. 2005;10 Suppl 1:21-34. doi:10.1258/1355819054308530.

6. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. BMC Med. 2013;11:21. doi:10.1186/1741-7015-11-21.

7. Mewes JC, Steuten LMG, Ijzerman MJ, van der Wilt GJ. How do integrated care programmes work for patients with cardiovascular disease, COPD, depression, diabetes and multi-morbidity? Int J Integr Care. 2018;18:7. doi:10.5334/ijic.3634.

8. Threapleton DE, Chung RY, Wong SYS, et al. Integrated care for older populations and its implementation facilitators and barriers: a rapid scoping review. Int J Qual Health Care. 2017;29:612-624. doi:10.1093/intqhc/mzx095.

9. Archer J, Bower P, Gilbody S, et al. Collaborative care for depression and anxiety problems. Cochrane Database Syst Rev. 2012;(10):CD006525. doi:10.1002/14651858.CD006525.pub2.

10. Ragin CC. The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. Berkeley: University of California Press; 1987.

11. Rihoux B, Ragin CC, editors. Configurational Comparative Methods: Qualitative Comparative Analysis and Related Techniques. Thousand Oaks: Sage; 2009.

12. Schneider CQ, Wagemann C. Set-Theoretic Methods for the Social Sciences. Cambridge: Cambridge University Press; 2012.

13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

14. Katon WJ, Lin EHB, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med. 2010;363:2611-2620. doi:10.1056/NEJMoa1003955.

15. Coventry P, Lovell K, Dickens C, et al. Integrated primary care for patients with mental and physical multimorbidity: cluster randomised controlled trial. BMJ. 2015;350:h638. doi:10.1136/bmj.h638.

16. Morgan MAJ, Coates MJ, Dunbar JA, et al. The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease. BMJ Open. 2013;3:e002171. doi:10.1136/bmjopen-2012-002171.

17. Srinivasan K, Heylen E, Johnson Pradeep R, et al. Collaborative care for depression among patients with multimorbidity in rural South India: the HOPE trial. BMC Psychiatry. 2022;22:394. doi:10.1186/s12888-022-04000-3.

18. Vera M, Perez-Pedrogo C, Huertas SE, et al. Collaborative care for depressed patients with chronic medical conditions: a randomized trial in Puerto Rico. Psychiatr Serv. 2010;61:144-150. doi:10.1176/ps.2010.61.2.144.

19. Picariello F, Hulme K, Seaton N, et al. A randomized controlled trial of a digital cognitive-behavioral therapy program (COMPASS) for managing depression and anxiety related to living with a long-term physical health condition. Psychol Med. 2024;54:1796-1809. doi:10.1017/S0033291723003756.

20. Salisbury C, Man M-S, Bower P, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. Lancet. 2018;392:41-50. doi:10.1016/S0140-6736(18)31308-4.

21. Pols AD, van Dijk SE, Bosmans JE, et al. Effectiveness of a stepped-care intervention to prevent major depression in patients with diabetes mellitus type 2 and/or coronary heart disease and subthreshold depression. PLoS One. 2017;12:e0181023. doi:10.1371/journal.pone.0181023.

22. Ma J, Rosas LG, Lv N, et al. Effect of integrated behavioral weight loss treatment and problem-solving therapy on body mass index and depressive symptoms among patients with obesity and depression: the RAINBOW randomized clinical trial. JAMA. 2019;321:869-879. doi:10.1001/jama.2019.0557.

23. Rollman BL, Belnap BH, LeMenager MS, et al. Telephone-delivered collaborative care for treating post-CABG depression. JAMA. 2009;302:2095-2103. doi:10.1001/jama.2009.1670.

24. Kroenke K, Theobald D, Wu J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. JAMA. 2010;304:163-171. doi:10.1001/jama.2010.944.

25. Sweeney K, Gilarova M, Ng L, et al. Primary care and community interventions for multimorbidity involving depression or anxiety: systematic review with meta-analysis. BMJ Med. 2026;5:e002400. doi:10.1136/bmjmed-2025-002400.

26. Overbeck G, Davidsen AS, Kousgaard MB. Enablers and barriers to implementing collaborative care for anxiety and depression. Implement Sci. 2016;11:165. doi:10.1186/s13012-016-0519-y.

27. Wakida EK, Talib ZM, Akena D, et al. Barriers and facilitators to the integration of mental health services into primary health care: a systematic review. Syst Rev. 2018;7:211. doi:10.1186/s13643-018-0882-7.

28. Ramanuj PP, Pincus HA. Collaborative care: enough of the why; what about the how? Br J Psychiatry. 2019;215:573-576. doi:10.1192/bjp.2019.99.

29. Kappelin C, Hjärthag F, Norman C, et al. Specific content for collaborative care: a systematic review. Fam Pract. 2022;39:725-734. doi:10.1093/fampra/cmab079.

CAPÍTULO 13

COMPLICAÇÕES RESPIRATÓRIAS PERIOPERATÓRIAS EM CRIANÇAS E ADOLESCENTES COM OBESIDADE

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Nélio Barreto Vieira

RESUMO

A obesidade modifica a fisiologia respiratória, a anatomia das vias aéreas, a farmacologia dos anestésicos e a reserva funcional de crianças e adolescentes. O objetivo deste capítulo foi sintetizar a associação entre excesso de adiposidade e eventos respiratórios perioperatórios, examinar a contribuição da gravidade da obesidade, da apneia obstrutiva do sono e das comorbidades respiratórias e integrar estimativas publicadas por meio de uma meta-análise bayesiana exploratória. Foram pesquisados PubMed/MEDLINE e fontes complementares, priorizando estudos observacionais, coortes multicêntricas, estudos de adenotonsilectomia, sedação procedimental e revisões clínicas. Cinco estimativas ajustadas, provenientes de contextos cirúrgicos heterogêneos, foram modeladas em estrutura normal-normal com prior fracamente informativo para o efeito médio e prior half-normal para a heterogeneidade. A mediana posterior da razão de efeito foi 3.27 (intervalo de credibilidade de 95%: 1.67–7.39); a probabilidade de aumento do risco foi de 99.9% e o intervalo preditivo para um novo contexto foi amplo (0.72–17.66), refletindo diferenças entre cirurgia geral, cirurgia de via aérea e obesidade grave. A análise de sensibilidade restrita a cirurgia geral ou não aérea produziu estimativa menor, porém ainda compatível com aumento do risco: 1.89 (1.09–4.23). Estudos de adenotonsilectomia mostraram riscos particularmente elevados, enquanto coortes recentes indicaram que o IMC isolado não substitui a avaliação de hipoxemia noturna, infecção respiratória recente, doença pulmonar, idade e eventos precoces na recuperação anestésica.

Conclui-se que a obesidade é um marcador clínico relevante, sobretudo quando grave ou combinada à apneia do sono, mas o risco é fenotípico e contextual. A prevenção requer triagem estruturada, planejamento da via aérea, pré-oxigenação, ventilação protetora, minimização de opioides, reversão completa do bloqueio neuromuscular e monitorização pós-operatória proporcional ao risco.

Palavras-chave: obesidade pediátrica; anestesia; complicações respiratórias; via aérea; apneia obstrutiva do sono; cirurgia pediátrica; meta-análise bayesiana.

1 INTRODUÇÃO

A obesidade pediátrica representa um desafio perioperatório que não pode ser reduzido ao valor isolado do índice de massa corporal. Durante o crescimento, peso e altura mudam rapidamente, razão pela qual a classificação depende de curvas específicas para idade e sexo. Além disso, duas crianças no mesmo percentil podem apresentar distribuição adiposa, massa muscular, circunferência cervical, mobilidade e perfil cardiometabólico muito diferentes. Para a anestesia, o aspecto central não é apenas a existência de excesso de peso, mas o modo como esse excesso modifica a reserva respiratória, a via aérea superior e a resposta aos fármacos [1-4].

A parede torácica mais pesada, o deslocamento cefálico do diafragma e a redução da capacidade residual funcional tornam a dessaturação mais rápida durante apneia. Em decúbito dorsal e sob anestesia geral, essas alterações podem ser intensificadas. O fechamento de pequenas vias aéreas ocorre mais cedo, a atelectasia se instala com maior facilidade e o recrutamento alveolar pode ser perdido após indução, intubação, mudança de posição ou ventilação com parâmetros inadequados. Em crianças pequenas, cuja demanda metabólica de oxigênio já é elevada, a margem temporal para correção de uma obstrução é especialmente curta [3-5].

A anatomia da via aérea também pode ser menos favorável. Tecido adiposo cervical, macroglossia relativa, hipertrofia adenotonsilar, limitação de

posicionamento e maior prevalência de apneia obstrutiva do sono ampliam a possibilidade de ventilação difícil, obstrução, laringoespasmo e hipoxemia. A dificuldade, contudo, não se manifesta uniformemente: séries de cirurgia ambulatorial registraram poucos casos de laringoscopia difícil, enquanto estudos prospectivos identificaram maior frequência de ventilação sob máscara difícil e obstrução em crianças com obesidade. A diferença decorre do tipo de cirurgia, seleção pré-operatória e definição do evento [6-9].

Os eventos respiratórios adversos perioperatórios incluem dessaturação, obstrução de via aérea, laringoespasmo, broncoespasmo, tosse intensa, necessidade de ventilação com pressão positiva, reintubação, suporte não invasivo e internação não planejada. Alguns estudos utilizam desfechos compostos que reúnem eventos leves e graves; outros registram apenas complicações que exigem intervenção. Uma incidência baixa em banco administrativo, portanto, não contradiz necessariamente uma incidência alta em estudos com observação contínua, porque a sensibilidade para detectar episódios breves é muito distinta [7,10-12].

A cirurgia de via aérea constitui um cenário particular. Crianças encaminhadas para adenotonsilectomia frequentemente têm distúrbio respiratório do sono, hipoxemia noturna e sensibilidade aumentada a opioides. A obesidade pode atuar como causa da apneia, marcador de sua gravidade ou fator que dificulta a recuperação. Em estudos caso-controle, o risco de complicações respiratórias após adenotonsilectomia foi várias vezes maior em crianças com obesidade, mas coortes mais recentes demonstraram que parâmetros de polissonografia e a evolução nas primeiras horas de recuperação podem predizer melhor a necessidade de suporte noturno [13-18].

Outro problema é a dose dos medicamentos. A utilização automática do peso corporal total pode produzir exposição excessiva a agentes lipofílicos ou opioides, enquanto o peso ideal pode subestimar a necessidade de outros fármacos. A literatura recomenda individualização conforme propriedades farmacocinéticas, efeito clínico e monitorização. Técnicas poupadoras de

opioides, anestesia regional quando possível e reversão objetiva do bloqueio neuromuscular são componentes de segurança, não apenas estratégias de recuperação acelerada [3,19,20].

A decisão sobre alta ambulatorial também exige cautela. Obesidade sem comorbidades, cirurgia não aérea de pequena magnitude e recuperação sem eventos podem ser compatíveis com alta no mesmo dia. Em contraste, obesidade grave, apneia moderada ou grave, saturação noturna baixa, necessidade de oxigênio, eventos recorrentes na unidade pós-anestésica e ausência de suporte domiciliar justificam observação prolongada. O desafio é evitar tanto a banalização do risco quanto internações indiscriminadas baseadas apenas no IMC [15-18].

Este capítulo foi elaborado para integrar a evidência sem criar um ranking artificial entre estudos heterogêneos. A análise quantitativa foi tratada como exploratória, com modelo bayesiano capaz de expressar incerteza e heterogeneidade. O propósito central é transformar os achados em um quadro clínico útil: identificar quem apresenta maior risco, em que momento o risco se concentra, quais medidas são sustentadas pela literatura e quais lacunas impedem recomendações mais precisas.

2 OBJETIVOS

O objetivo principal foi sintetizar a evidência sobre complicações respiratórias perioperatórias em crianças e adolescentes com obesidade submetidos a anestesia, sedação ou cirurgia. Os objetivos específicos foram: descrever a magnitude dos eventos em diferentes contextos; examinar o gradiente associado à obesidade grave; analisar a interação com apneia obstrutiva do sono, asma, infecção respiratória recente e idade; estimar um efeito médio exploratório por meta-análise bayesiana; e propor um percurso assistencial baseado em risco, desde a avaliação pré-operatória até a alta.

3 MÉTODO

3.1 Desenho e pergunta clínica

Foi conduzida uma síntese multidomínio de evidências com componente quantitativo bayesiano. A pergunta foi estruturada da seguinte forma: em pacientes de 2 a 18 anos submetidos a anestesia, sedação ou cirurgia, a obesidade, em comparação com peso adequado, associa-se a maior ocorrência de eventos respiratórios perioperatórios? Estudos sem comparador foram mantidos quando contribuíam para estratificação de risco, monitorização ou definição de critérios de internação. O relato foi orientado por princípios do PRISMA 2020, sem caracterizar a busca como revisão sistemática exaustiva registrada [21].

3.2 Fontes, busca e elegibilidade

A busca foi atualizada em 1º de agosto de 2026 no PubMed/MEDLINE e complementada por rastreamento de referências de revisões e estudos-chave. Foram combinados termos relativos a pediatric obesity, child, adolescent, anesthesia, surgery, sedation, airway, respiratory adverse events, oxygen desaturation, laryngospasm, bronchospasm, tonsillectomy e obstructive sleep apnea. Foram elegíveis estudos humanos que apresentassem resultados respiratórios perioperatórios, estimativas ajustadas, frequências de eventos ou modelos de predição. Excluíram-se relatos isolados, séries sem dados úteis e estudos exclusivamente adultos.

3.3 Extração e organização dos desfechos

A extração contemplou desenho, ano, tamanho da amostra, classificação da obesidade, procedimento, desfecho, medida de associação e variáveis de ajuste. Os desfechos foram organizados em cinco domínios: eventos respiratórios compostos; via aérea e ventilação; hipoxemia e necessidade de oxigênio;

recuperação e internação; e marcadores de apneia do sono. Estimativas ajustadas foram preferidas às comparações brutas. Quando um artigo apresentava múltiplos desfechos correlacionados, apenas a medida mais abrangente ou clinicamente representativa foi selecionada para a síntese principal, evitando duplicação do mesmo grupo de participantes.

3.4 Avaliação crítica

A avaliação crítica considerou seleção da amostra, mensuração objetiva do peso e da altura, clareza da definição do evento, controle de confusão, completude do seguimento e aplicabilidade. Estudos observacionais foram interpretados segundo os domínios do ROBINS-I [22]. Não foi calculado escore numérico, pois a soma de itens pode ocultar vieses graves. A certeza foi classificada narrativamente como maior quando havia consistência, temporalidade clara, gradiente de gravidade e ajuste para apneia ou doença respiratória, e menor quando os intervalos eram amplos ou o evento dependia de registro administrativo.

3.5 Meta-análise bayesiana exploratória

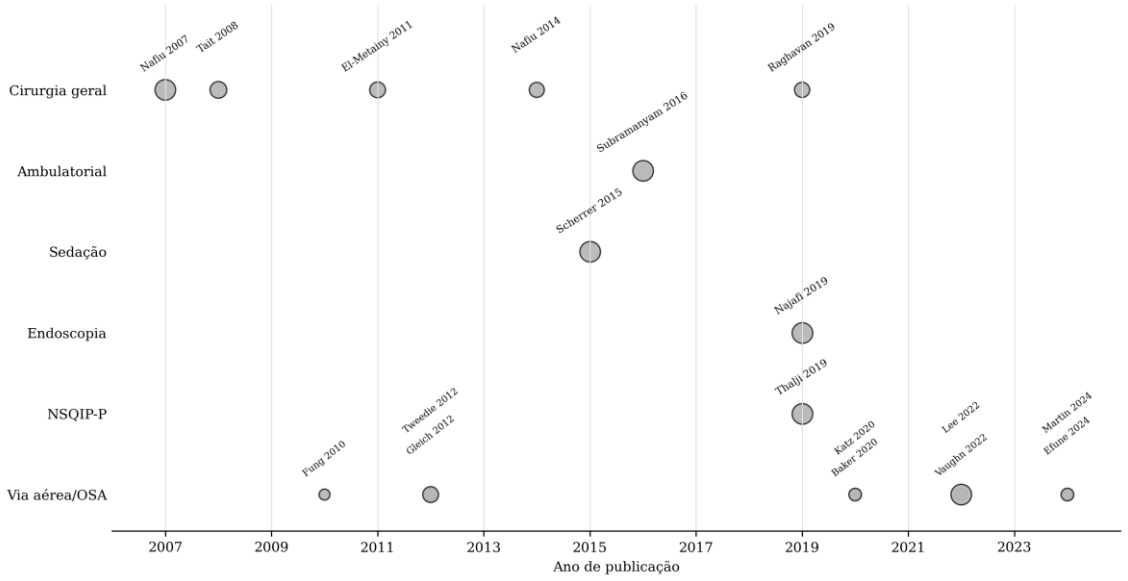
Cinco estudos apresentaram estimativas ajustadas com intervalos de confiança suficientemente completos e desfechos respiratórios relacionados. As razões de risco e odds ratios foram analisadas na escala logarítmica como medidas aproximadas de efeito relativo, decisão aceitável apenas para exploração e explicitamente reconhecida como fonte de heterogeneidade. O modelo normal-normal assumiu $y_i \sim \text{Normal}(\theta_i, SE_i^2)$ e $\theta_i \sim \text{Normal}(\mu, \tau^2)$. Para μ foi adotada prior $\text{Normal}(0, 2^2)$, e para τ prior half-normal com escala 0,5. Foram obtidos intervalo de credibilidade de 95%, probabilidade posterior de efeito superior a 1 e intervalo preditivo para um novo contexto. A análise de sensibilidade excluiu estudos de cirurgia de via aérea, mantendo cirurgia geral ou não aérea.

4 RESULTADOS

4.1 Estrutura do corpo de evidências

Dezoito estudos clínicos centrais foram organizados na síntese, além de revisões e diretrizes. As amostras variaram de 60 crianças em monitorização respiratória intensiva após adenotonsilectomia a mais de 117 mil procedimentos em banco cirúrgico nacional. A linha do tempo mostra três movimentos: estudos iniciais focados em qualidade anestésica e eventos durante cirurgia geral; expansão para grandes bases de sedação e cirurgia ambulatorial; e aprofundamento recente da monitorização após cirurgia de via aérea. Essa evolução ampliou a precisão, mas também aumentou a heterogeneidade conceitual dos desfechos.

Figura 1. Linha do tempo e paisagem dos estudos clínicos centrais, com tamanho das bolhas proporcional à amostra.



Fonte: Elaboração própria a partir dos estudos incluídos; o tamanho das amostras corresponde aos valores publicados.

Tabela 1. Características dos estudos clínicos centrais e contribuição para a síntese.

Estudo	Desenho	N	Contexto	Exposição	Achado principal
Nafiu et al., 2007	Retrospectivo	6.094	Cirurgia geral	Sobrepeso/obesidade	Via aérea difícil, obstrução, PACU prolongada
Tait et al., 2008	Prospectivo	2.025	Cirurgia eletiva	Obesidade	Mais ventilação difícil, obstrução e dessaturação
Fung et al., 2010	Caso-controle	98	Adenotonsilectomia	Obesidade	Complicações 75,5% vs 26,5%; OR 8,54
El-Metainy et al., 2011	Prospectivo	1.465	Cirurgia geral	Obesidade	RR ajustado 1,49 para eventos respiratórios
Gleich et al., 2012	Caso-controle	300	Tonsilectomia	Obesidade grave	Complicações de via aérea 15% vs 2%
Tweedie et al., 2012	Coorte	1.735	Adeno/tonsilectomia	Comorbidades	2,4% de admissão em UTI; obesidade entre fatores
Nafiu & Onyewuche, 2014	Prospectivo	1.102	Não cardíaca	Obesidade abdominal	OR ajustado 2,35 para eventos respiratórios
Scherrer et al., 2015	Registro prospectivo	28.792	Sedação	Obesidade	Obstrução OR 1,94; dessaturação OR 1,99
Subramanyam et al., 2016	Derivação/validação	19.059	Ambulatorial	Obesidade mórbida	Componente independente em escore de risco
Najafi et al., 2019	Coorte	3.435	Endoscopia	Sobrepeso/obesidade	OR 4,25 para obesidade mórbida
Raghavan et al., 2019	Retrospectivo	1.158	Não aérea	Obesidade grave	OR 10,04; IC amplo
Thalji et al., 2019	NSQIP-P	117.272	Cirurgia eletiva	Classes de IMC	Complicações pulmonares graves raras
Katz et al., 2020	Coorte	374	Adenotonsilectomia	OSA/obesidade	Preditores clínicos e polissonográficos
Baker et al., 2020	Retrospectivo	165	Adenotonsilectomia	Alto risco	Necessidade após 2 h indicou maior cuidado
Lee et al., 2022	Retrospectivo	155	Adenotonsilectomia	Obesidade	Nadir e tempo <90% previram complicações

Estudo	Desenho	N	Contexto	Exposição	Achado principal
Vaughn et al., 2022	Coorte	7.072	Adenotonsilectomia	Obesidade	Pequeno efeito isolado do peso
Efune et al., 2024	Prospectivo	60	Adenotonsilectomia	Alto risco	RVM detectou 65% vs 35% clinicamente
Martin et al., 2024	Retrospectivo	401	Tonsilectomia	Faixas de IMC	IMC ≥35: OR 2,30 para oxigênio

Fonte: Dados extraídos das publicações citadas [6-18,23-29].

4.2 Cirurgia geral e procedimentos não aéreos

Os estudos de cirurgia geral mostram associação consistente, porém de magnitude moderada quando comparados às séries de adenotonsilectomia. Em 2.025 crianças, Tait et al. observaram maior incidência de ventilação sob máscara difícil, obstrução, dessaturação maior que 10% do basal e eventos respiratórios críticos entre pacientes com obesidade. O modelo identificou obesidade, idade inferior a 10 anos, cirurgia de via aérea e apneia do sono como fatores relevantes [7].

El-Metainy et al. avaliaram 1.465 crianças de 2 a 16 anos. Após controle por idade, a razão de risco para eventos respiratórios adversos foi 1,49, com aumento de dessaturação, ventilação difícil, obstrução e broncoespasmo. A associação da obesidade com apneia do sono foi forte, OR 4,03, e a frequência de eventos foi maior nas faixas etárias mais jovens [8].

A medida central da adiposidade também acrescentou informação. Em 1.102 crianças, circunferência abdominal no percentil 90 ou superior associou-se a OR ajustado 2,35 para eventos respiratórios e a permanência mais longa na recuperação. Esse resultado sugere que o IMC pode não captar integralmente o efeito mecânico da adiposidade abdominal sobre o diafragma e a capacidade residual funcional [9].

Raghavan et al. procuraram isolar o efeito da obesidade grave em cirurgia ambulatorial não aérea. Eventos respiratórios ocorreram em 3,44% das crianças com obesidade grave e 0,34% das de peso normal, OR 10,04. O intervalo de

confiança, de 1,01 a 99,5, evidencia grande imprecisão decorrente do pequeno número de eventos. O estudo é importante por reduzir a confusão causada por cirurgia de via aérea, mas não permite estimar com precisão a magnitude verdadeira [10].

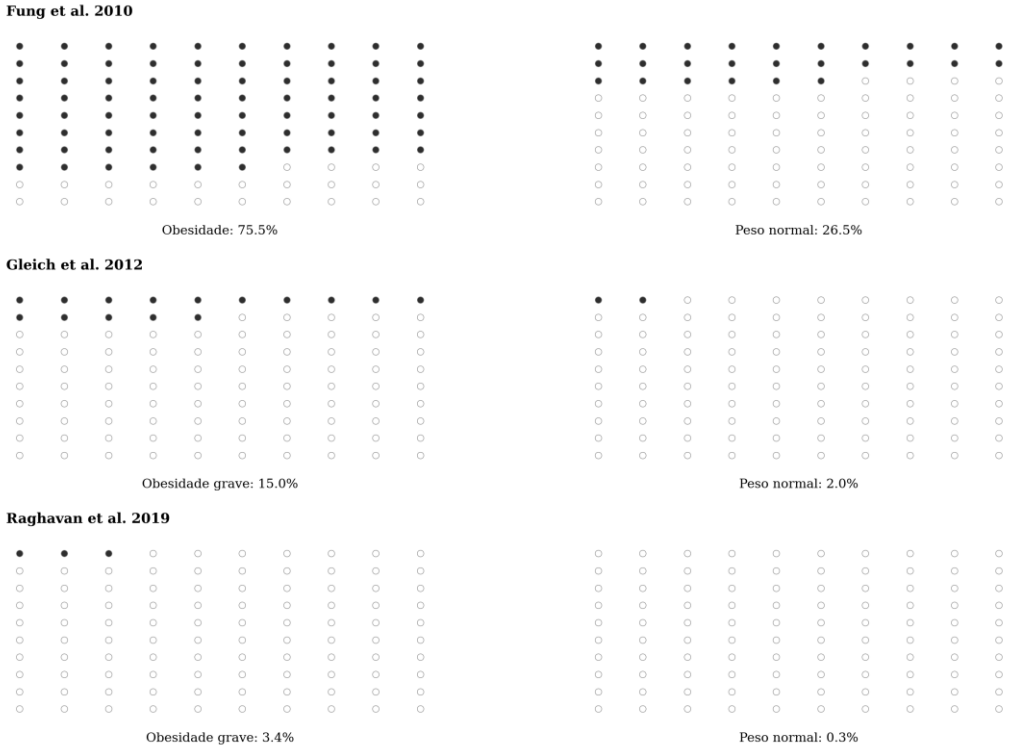
4.3 Cirurgia de via aérea e obesidade grave

Na adenotonsilectomia, a obesidade se combina à apneia do sono e à manipulação direta da via aérea. Fung et al. relataram complicações em 75,5% das crianças com obesidade e em 26,5% dos controles, OR 8,54. Eventos maiores ocorreram em 10 versus 2 crianças; obstrução de via aérea superior foi particularmente frequente na recuperação imediata. O tempo de internação médio foi 18 horas no grupo com obesidade e 8 horas no controle [13].

Gleich et al. compararam 100 crianças com obesidade grave a 200 controles. Complicações de via aérea ocorreram em 15% e 2%, respectivamente. Mesmo após ajuste para doença respiratória preexistente, a obesidade grave permaneceu associada a eventos, OR 7,7. Quando admissões planejadas foram excluídas, a chance de internação não programada permaneceu maior, OR 3,80 [14].

Essas estimativas elevadas não devem ser aplicadas a toda cirurgia pediátrica. Elas representam uma população selecionada, com alta prevalência de distúrbio respiratório do sono e procedimento diretamente relacionado à via aérea. O contraste com cirurgia geral é clinicamente informativo: o risco não decorre apenas do tecido adiposo, mas da interação entre fenótipo respiratório, procedimento e medicamentos administrados.

Figura 2. Risco absoluto de complicações respiratórias em três comparações publicadas; cada matriz representa 100 pacientes.



Fonte: Fung et al. [13], Gleich et al. [14] e Raghavan et al. [10]. Valores arredondados apenas para representação de 100 ícones.

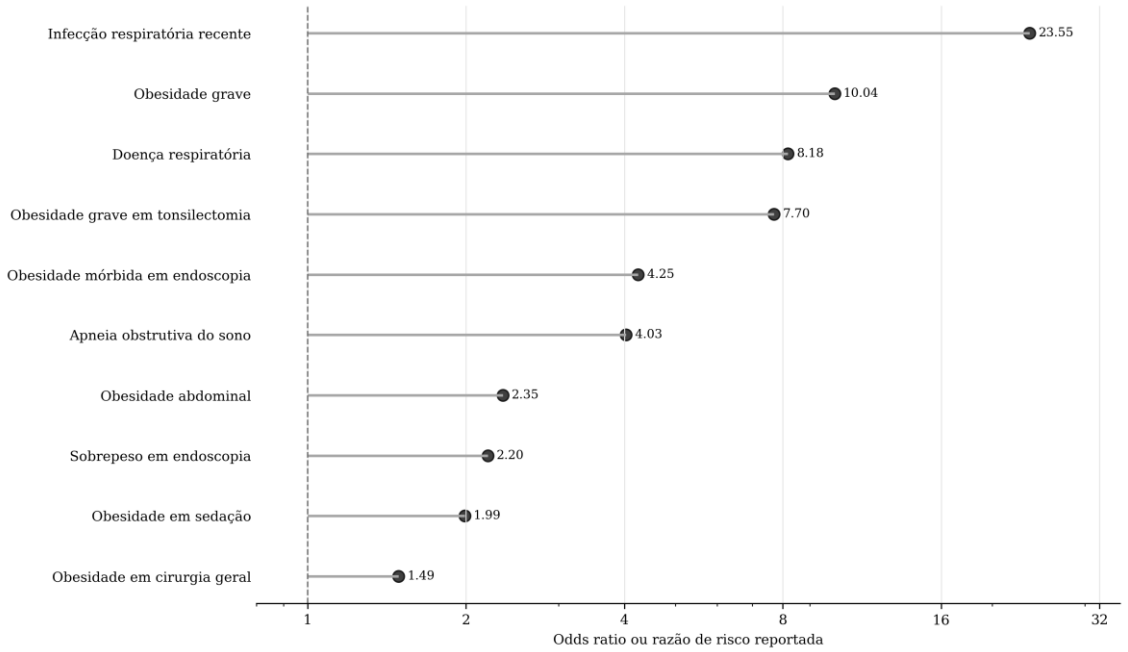
4.4 Sedação, endoscopia e eventos específicos

No registro do Pediatric Sedation Research Consortium, 28.792 sedações foram avaliadas e 5.153 envolveram pacientes com obesidade. O total de eventos adversos foi maior, OR 1,49. Obstrução de via aérea apresentou OR 1,94, dessaturação OR 1,99, laringoespasmo OR 2,30 e recuperação prolongada OR 2,66. A obesidade permaneceu associada a eventos leves e moderados, mas não aos eventos maiores, o que reforça a necessidade de competência para intervenções precoces [23].

Em 3.435 endoscopias pediátricas sob cuidado anestésico monitorado, Najafi et al. observaram incidência respiratória de 3,3%. A obesidade mórbida associou-se a OR 4,25; doença respiratória a OR 8,18; e infecção respiratória recente a OR 23,55. A magnitude do último efeito demonstra que a decisão de

prosseguir ou adiar um procedimento não deve ser guiada apenas pelo peso. Sintomas ativos, secreção, febre, sibilância e urgência do procedimento podem ter impacto maior e mais imediato [24].

Figura 3. Espectro de modificadores respiratórios reportados em diferentes contextos clínicos.

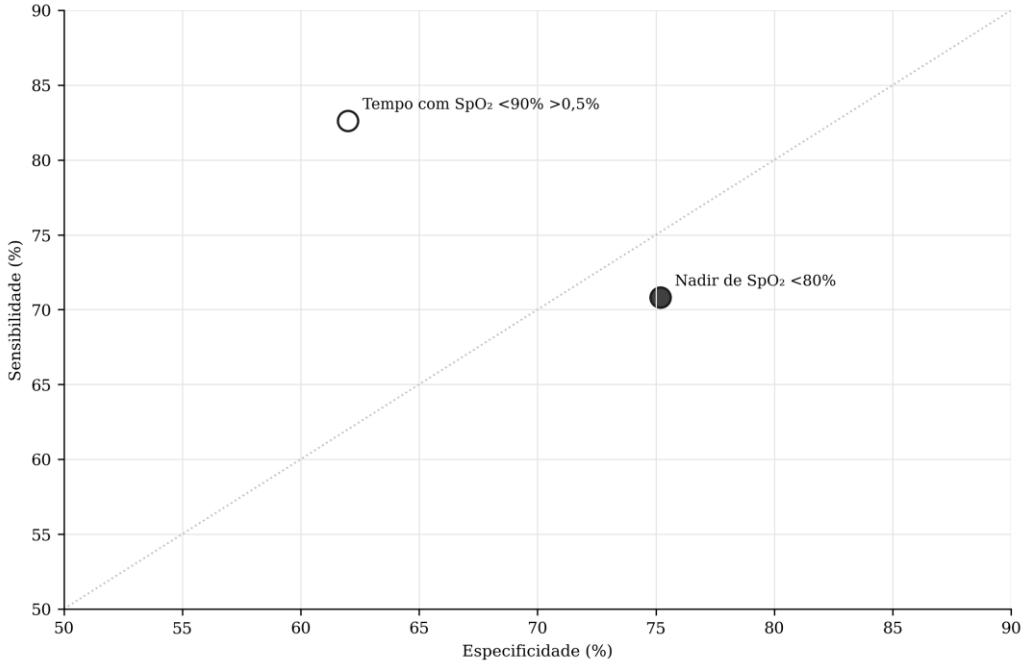


Fonte: Estimativas dos estudos indicados. Os valores não devem ser comparados como ranking porque populações, desfechos e ajustes são distintos.

4.5 Monitorização, polissonografia e necessidade de suporte

A polissonografia ajuda a identificar a reserva noturna e a gravidade da apneia. Em 155 crianças com obesidade submetidas a adenotonsilectomia, menor nadir de saturação e maior tempo de sono abaixo de 90% foram preditores independentes de complicações respiratórias noturnas. Nadir inferior a 80% apresentou sensibilidade de 70,8% e especificidade de 75,2%; mais de 0,5% do sono com saturação inferior a 90% apresentou sensibilidade de 82,6% e especificidade de 62,0%. O modelo que combinou variáveis pré e pós-operatórias alcançou AUC 0,89 [16].

Figura 4. Desempenho de dois critérios polissonográficos para prever complicações respiratórias noturnas.



Fonte: Lee et al. [16]. Sensibilidade e especificidade reproduzidas do artigo.

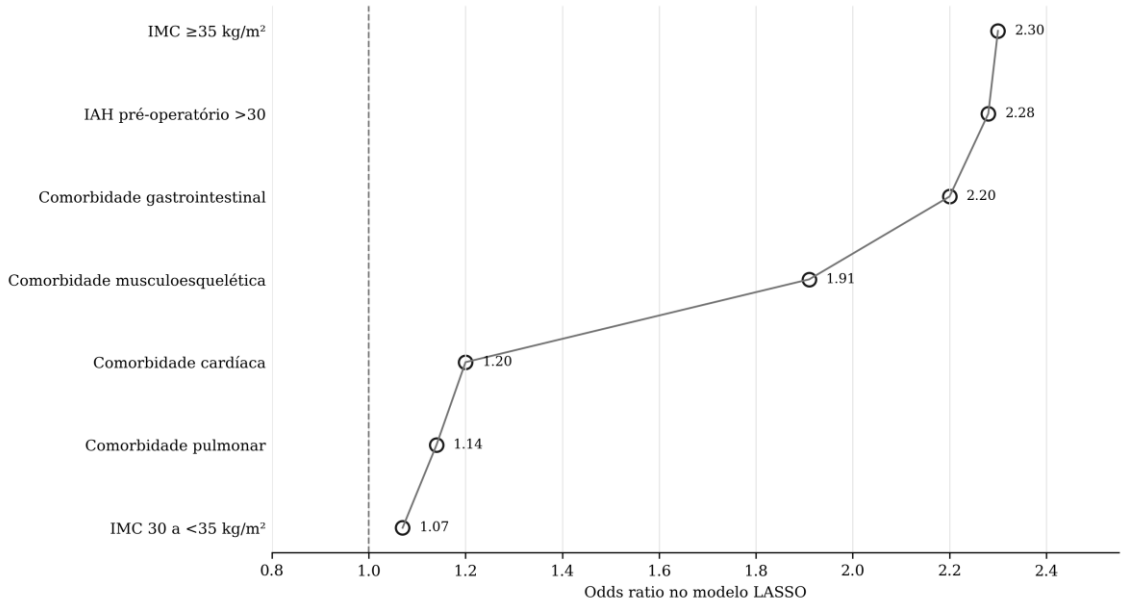
A evolução inicial na recuperação pode ser tão importante quanto os fatores pré-operatórios. Baker et al. observaram que a maioria das crianças admitidas em terapia intensiva não necessitava suporte além de duas horas. A persistência da necessidade respiratória após esse período foi mais útil para identificar quem realmente demandava cuidado intensivo do que obesidade ou polissonografia anormal isoladamente [17]. Esse achado não autoriza alta precoce indiscriminada, mas favorece critérios dinâmicos baseados na resposta clínica.

A monitorização convencional pode subestimar hipoventilação. Em coorte prospectiva de alto risco, monitor de volume respiratório detectou episódios de ventilação inferior a 40% do previsto em 65% das crianças, enquanto eventos clinicamente aparentes ocorreram em 35%. A discrepância demonstra que saturação preservada sob oxigênio suplementar não exclui ventilação inadequada. Capnografia ou tecnologias equivalentes podem ter valor em

pacientes selecionados, embora ainda faltem estudos que demonstrem redução de desfechos graves [18].

Em 401 crianças submetidas a tonsilectomia, 9,07% receberam oxigênio suplementar. O modelo LASSO identificou IMC igual ou superior a 35 kg/m², OR 2,30, e índice de apneia-hipopneia maior que 30, OR 2,28, como fatores de magnitude semelhante. Comorbidades gastrointestinais, musculoesqueléticas, cardíacas e pulmonares também contribuíram. O resultado ilustra por que um modelo multivariável é mais apropriado do que um limiar único de IMC [29].

Figura 5. Fatores associados ao uso de oxigênio suplementar após tonsilectomia.



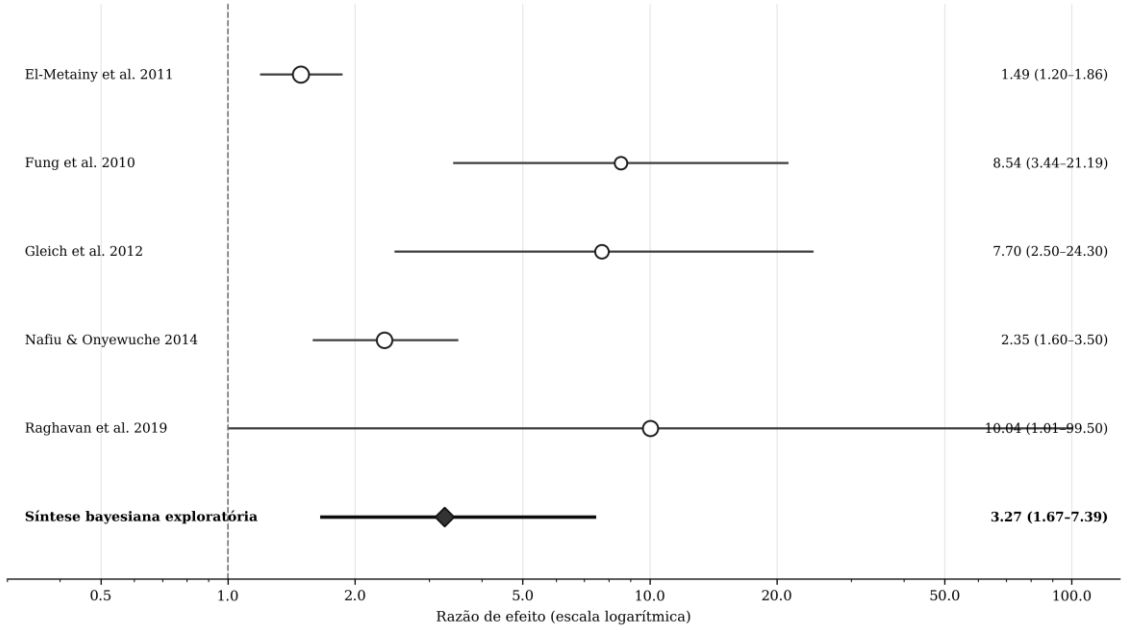
Fonte: Martin et al. [29]. Odds ratios do modelo LASSO; não foram fornecidos intervalos de confiança no resumo.

4.6 Síntese bayesiana

A síntese principal reuniu cinco estimativas ajustadas. A mediana posterior da razão de efeito foi 3,27, com intervalo de credibilidade de 95% de 1,67 a 7.39. A probabilidade posterior de aumento do risco foi 99,9% e a probabilidade de um efeito médio superior a duas vezes foi 92,8%. A heterogeneidade posterior mediana na escala logarítmica foi 0,63, com intervalo

de 0,27 a 1.18. O intervalo preditivo de 0,72 a 17,66 inclui desde associação pequena até aumento muito expressivo e resume a dependência do contexto.

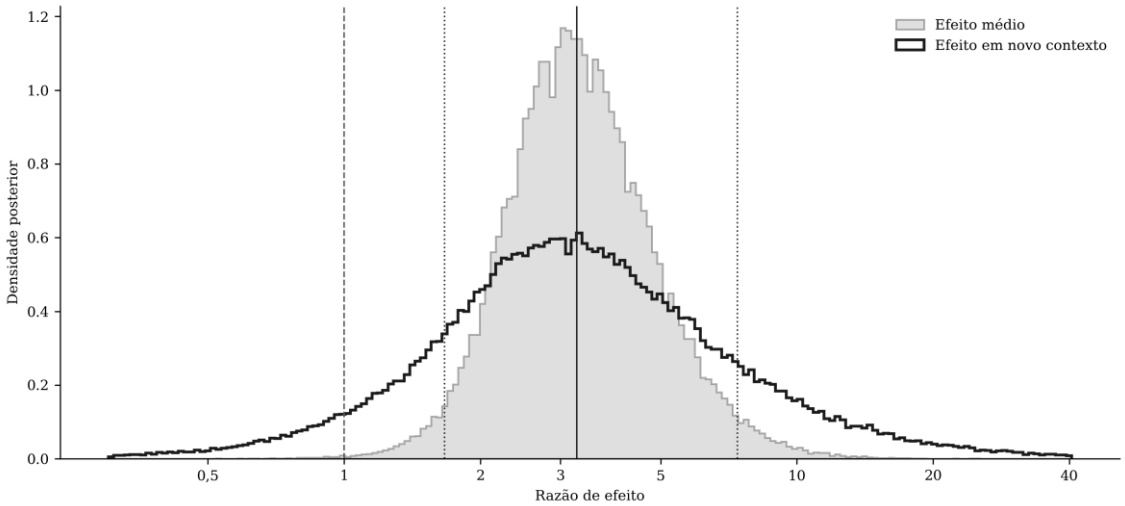
Figura 6. Estimativas ajustadas incluídas e síntese bayesiana exploratória.



Fonte: Cálculos próprios a partir das estimativas publicadas por El-Metainy et al. [8], Fung et al. [13], Gleich et al. [14], Nafiu e Onyewuche [9] e Raghavan et al. [10].

A distribuição posterior do efeito médio permaneceu quase integralmente acima da nulidade, mas a distribuição preditiva foi muito mais larga. Essa diferença é essencial: o modelo pode sustentar que, em média, há aumento do risco, sem garantir que a magnitude seja a mesma em cada serviço, procedimento ou fenótipo. A incerteza preditiva deve ser usada para planejamento e não ocultada por uma única estimativa combinada.

Figura 7. Distribuições posteriores do efeito médio e do efeito esperado em um novo contexto.



Fonte: Meta-análise bayesiana exploratória com prior Normal(0, 2^2) para μ e half-normal(0,5) para τ.

Na análise de sensibilidade restrita a cirurgia geral ou não aérea, a mediana posterior caiu para 1,89 (1,09–4,23), e o intervalo preditivo foi 0,64–7.22. O resultado confirma que as séries de adenotonsilectomia ampliam substancialmente a estimativa global. A análise principal deve, portanto, ser lida como síntese de contextos e não como efeito causal único.

Tabela 2. Estimativas quantitativas selecionadas e contexto de interpretação.

Estudo	Desfecho/exposição	Medida	Estimativa	IC 95%	Observação
El-Metainy 2011	Eventos respiratórios	RR ajustado	1,49	1,20–1,86	Cirurgia geral; controle por idade
Fung 2010	Complicações respiratórias	OR	8,54	3,44–21,19	Adenotonsilectomia; caso-controle
Gleich 2012	Eventos perioperatórios	OR ajustado	7,70	2,50–24,30	Ajuste para doença respiratória
Nafiu 2014	Eventos respiratórios	OR ajustado	2,35	1,60–3,50	Obesidade abdominal
Raghavan 2019	Eventos respiratório-via aérea	OR	10,04	1,01–99,50	Poucos eventos; grande imprecisão

Estudo	Desfecho/exposição	Medida	Estimativa	IC 95%	Observação
Scherrer 2015	Obstrução	OR	1,94	1,54–2,44	Sedação procedimental
Scherrer 2015	Dessaturação	OR	1,99	1,50–2,63	Sedação procedimental
Najafi 2019	Obesidade mórbida	OR	4,25	não informado	Endoscopia sob MAC
Martin 2024	Oxigênio, IMC ≥35	OR LASSO	2,30	não informado	Tonsilectomia

Fonte: Publicações originais [8-10,13,14,23,24,29].

5 DISCUSSÃO

5.1 Interpretação principal

A síntese indica que a obesidade pediátrica está associada a maior risco respiratório perioperatório, mas a magnitude depende do fenótipo e do procedimento. A estimativa bayesiana média foi superior a três vezes, porém o intervalo preditivo foi amplo e incluiu cenários de efeito discreto. O achado não é uma fragilidade do modelo; é uma representação honesta de uma literatura em que adenotonsilectomia, cirurgia geral, sedação e obesidade abdominal foram estudadas com definições distintas.

Os estudos de cirurgia de via aérea produziram os maiores efeitos. Nesses casos, obesidade, hipertrofia adenotonsilar, apneia do sono, exposição a opioides e edema pós-operatório atuam simultaneamente. Em cirurgia não aérea, o efeito médio foi menor e mais próximo de duas vezes. Essa diferença sustenta uma abordagem estratificada: não é adequado atribuir a uma adolescente com obesidade moderada submetida a procedimento periférico o mesmo risco de uma criança pequena com obesidade grave, apneia e tonsilectomia.

A gravidade da obesidade importa. Estudos que separaram obesidade grave encontraram eventos mais frequentes, enquanto alguns trabalhos que agruparam todas as crianças acima do percentil 95 observaram efeitos pequenos.

O agrupamento dilui o risco das classes extremas e dificulta identificar relações dose-resposta. Medidas adicionais, como circunferência abdominal, cervical e parâmetros de polissonografia, podem melhorar a caracterização, embora ainda não exista modelo universal validado para todos os procedimentos.

5.2 Mecanismos fisiopatológicos e farmacológicos

A redução da capacidade residual funcional aproxima o volume pulmonar do volume de fechamento, favorecendo colapso de pequenas vias aéreas. A indução anestésica e o bloqueio neuromuscular reduzem tônus faríngeo e diafragmático; a posição supina aumenta compressão abdominal; e a ventilação com volume corrente excessivo ou pressão expiratória insuficiente pode ampliar atelectasia. Em consequência, a criança pode apresentar reserva reduzida antes que qualquer evento clínico se torne visível.

A apneia obstrutiva do sono acrescenta intermitência hipóxica, fragmentação do sono e alterações de controle ventilatório. Crianças com hipoxemia noturna podem ter maior sensibilidade a opioides, especialmente após remoção da obstrução adenotonsilar. A analgesia deve ser multimodal, com paracetamol, anti-inflamatórios quando apropriados, dexmedetomidina ou técnicas regionais segundo o procedimento, e opioides titulados pelo efeito, nunca por uma fórmula baseada exclusivamente no peso total [3,19,20].

A farmacocinética em obesidade pediátrica é heterogênea. O volume de distribuição de fármacos lipofílicos pode aumentar, enquanto depuração e maturação orgânica variam com idade e comorbidades. Peso corporal total, peso ideal e massa magra são escalas úteis para medicamentos diferentes. Protocolos devem especificar qual peso utilizar e incorporar monitorização objetiva, sobretudo para bloqueadores neuromusculares e sua reversão.

5.3 Interação com infecção respiratória, asma e idade

A infecção respiratória recente foi um dos fatores de maior magnitude em endoscopia pediátrica. O epitélio hiper-reativo, a produção de secreção e a redução do calibre das vias aéreas podem persistir após melhora clínica. Obesidade e infecção não devem ser tratadas como fatores independentes simples: uma criança com reserva reduzida tolera pior laringoespasmo ou broncoespasmo. A decisão de adiar depende de gravidade dos sintomas, febre, secreção, ausculta, urgência do procedimento e possibilidade de otimização.

A asma deve estar controlada, com revisão do uso de broncodilatadores, exacerbações e internações anteriores. Em crianças pequenas, o risco aumenta porque o diâmetro da via aérea é menor, a capacidade de oxigênio é limitada e a cooperação para pré-oxigenação pode ser difícil. O achado de maior risco em idades jovens aparece em diferentes coortes e reforça a necessidade de equipe experiente e equipamentos adequados [7,8,24].

5.4 Avaliação pré-operatória

A avaliação deve registrar peso e altura medidos, percentil ou escore-z, gravidade da obesidade, circunferência cervical quando útil, sintomas de ronco, pausas respiratórias, sonolência, cefaleia matinal, tolerância ao exercício, asma, refluxo, infecção recente e uso de pressão positiva. A disponibilidade de polissonografia varia; sua ausência não deve impedir a identificação de alto risco. Questionários podem apoiar triagem, mas não substituem julgamento clínico.

O exame da via aérea inclui abertura oral, mobilidade cervical, tamanho de língua e tonsilas e distribuição adiposa. A preparação precisa contemplar máscara de tamanho correto, cânulas, dispositivos supraglóticos, videolaringoscópio e plano de resgate. Equipamentos de transferência, mesa e manguito de pressão adequados também fazem parte da segurança. Essas medidas evitam que uma característica corporal previsível se transforme em falha organizacional.

5.5 Indução, ventilação e emergência

Pré-oxigenação eficaz é prioritária. Posição cefálica elevada e alinhamento do meato auditivo com o esterno podem melhorar a patência e prolongar o tempo seguro de apneia. A escolha entre indução inalatória e intravenosa depende de acesso, cooperação e risco de via aérea. Após controle da via aérea, volume corrente ajustado ao peso ideal, pressão expiratória positiva individualizada e manobras de recrutamento seletivas podem reduzir colapso, evitando pressões excessivas.

A emergência é um período de risco concentrado. Bloqueio neuromuscular residual, secreções, dor, agitação e opioides favorecem obstrução. Monitorização quantitativa da transmissão neuromuscular e reversão completa devem preceder extubação. A literatura de revisão favorece extubação desperta em crianças com obesidade e alto risco respiratório, mas técnica, procedimento e experiência da equipe devem ser considerados. O ponto decisivo é uma estratégia planejada, com capacidade imediata de ventilação e suporte [3,20].

5.6 Recuperação, internação e alta

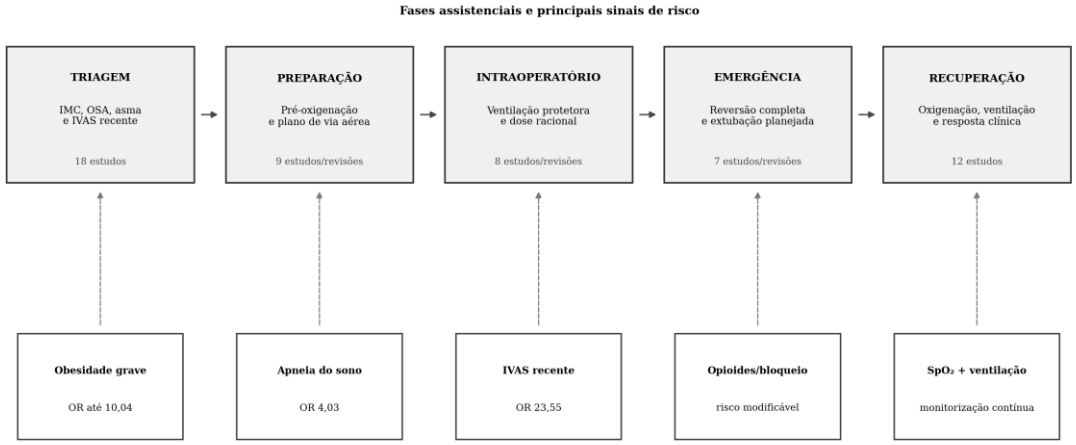
A monitorização deve avaliar oxigenação e ventilação. O oxigênio suplementar pode normalizar a saturação sem corrigir hipoventilação; em pacientes de maior risco, capnografia ou monitorização de volume podem acrescentar informação. Eventos repetidos, necessidade de manobras, oxigênio persistente, sedação excessiva, dor que exige opioides ou incapacidade de manter saturação durante sono contraindicam alta imediata.

A internação baseada apenas no percentil de IMC pode consumir recursos sem melhorar desfechos. Os dados favorecem combinação de gravidade da obesidade, apneia, nadir de saturação, tempo abaixo de 90%, comorbidades e curso na recuperação. A criança que permanece estável, desperta, ventilando adequadamente, com analgesia não opioide eficaz e suporte familiar pode ser

elegível para alta conforme protocolo local. Em contraste, eventos após duas horas, oxigênio contínuo e hipoxemia noturna importante justificam observação.

5.7 Modelo assistencial integrado

Figura 8. Percurso perioperatório e pontos de intervenção sustentados pela literatura.



Fonte: Síntese própria dos estudos clínicos e revisões [3,7-20,23-29]. O número de estudos corresponde aos domínios cobertos, não a uma medida de eficácia.

Tabela 3. Pacote assistencial proposto para redução do risco respiratório.

Fase	Ação	Justificativa
Pré-operatório	Medir IMC por idade/sexo; identificar obesidade grave	Evita classificação imprecisa e reconhece gradiente
Pré-operatório	Triar OSA, asma e infecção respiratória recente	Comorbidades podem superar o efeito isolado do IMC
Planejamento	Equipe e equipamento para via aérea difícil	Reduz atraso diante de obstrução ou ventilação difícil
Indução	Posição elevada e pré-oxigenação adequada	Reserva funcional reduzida e dessaturação rápida
Manutenção	Ventilação protetora e dose baseada em escala apropriada	Menos atelectasia e exposição farmacológica excessiva
Analgesia	Estratégia multimodal poupadora de opioides	OSA aumenta vulnerabilidade à depressão respiratória

Fase	Ação	Justificativa
Emergência	Reversão quantitativa e extubação planejada	Bloqueio residual e obstrução concentram risco
Recuperação	Monitorizar oxigenação e ventilação	SpO_2 isolada pode não detectar hipoventilação
Alta	Usar critérios dinâmicos e contexto domiciliar	Evita alta insegura e internação baseada apenas no IMC

Fonte: Síntese clínica das evidências [3,15-20,29].

5.8 Equidade, comunicação e estigma

O cuidado deve evitar linguagem estigmatizante. A criança e a família não devem ser responsabilizadas por limitações de equipamento ou por atrasos do serviço. Comentários depreciativos podem reduzir confiança e adesão. A comunicação deve explicar que o peso é apenas um dos fatores utilizados para planejar oxigenação, medicamentos e monitorização, enfatizando que as adaptações visam segurança e conforto.

A equidade também envolve disponibilidade de polissonografia, anestesista pediátrico, videolaringoscopia e observação noturna. Serviços com menor infraestrutura podem precisar de critérios mais conservadores de encaminhamento. A ausência de tecnologia avançada não elimina a possibilidade de cuidado seguro, desde que haja triagem, planejamento, protocolos de transferência e reconhecimento precoce de deterioração.

5.9 Limitações da evidência e da síntese

A maior parte dos estudos é observacional e suscetível a confusão por apneia, asma, gravidade clínica e seleção de pacientes. Definições de obesidade e eventos respiratórios variam. Alguns estudos agrupam sobrepeso e obesidade; outros usam percentil 98, IMC absoluto ou circunferência abdominal. Bancos administrativos capturam eventos graves, enquanto observação direta identifica episódios leves e transitórios.

A meta-análise reuniu medidas diferentes, incluindo razão de risco e odds ratio, e contextos distintos. Por isso foi classificada como exploratória. O modelo bayesiano quantificou a heterogeneidade, mas não corrige falta de comparabilidade. A estimativa combinada não deve ser utilizada para consentimento individual sem considerar procedimento e fenótipo. O intervalo preditivo e a análise de sensibilidade são mais informativos do que o número central isolado.

Poucos estudos avaliam intervenções especificamente em crianças com obesidade. Recomendações sobre ventilação, analgesia e extubação derivam de fisiologia, estudos gerais de anestesia pediátrica e revisões. Há escassez de dados em países de baixa e média renda, cirurgia de emergência, crianças com deficiência e adolescentes submetidos à cirurgia bariátrica.

Tabela 4. Certeza narrativa e principais limitações por domínio.

Domínio	Certeza	Fortalezas	Limitações
Obesidade e eventos respiratórios	Moderada	Consistência; estudos prospectivos	Heterogeneidade de desfechos
Obesidade grave	Moderada	Gradiente e efeitos grandes	Poucos eventos em algumas coortes
Adenotonsilectomia	Moderada	Caso-controle e coortes concordantes	Confusão por OSA e indicação cirúrgica
Sedação/endoscopia	Moderada	Registros grandes e ajuste	Eventos majoritariamente leves/moderados
Polissonografia para risco noturno	Baixa a moderada	Capacidade preditiva em coortes	Validação externa limitada
Monitorização ventilatória avançada	Baixa	Detecta eventos ocultos	Impacto clínico ainda incerto
Pacote anestésico preventivo	Baixa a moderada	Plausibilidade e consenso	Poucos ensaios específicos

5.10 Modelos de predição e utilidade clínica

Modelos de risco podem organizar variáveis que, isoladamente, têm baixo valor discriminativo. Subramanyam et al. desenvolveram um escore com 19.059 procedimentos ambulatoriais, dividido em coortes de derivação e validação. Idade igual ou inferior a três anos, estado físico ASA II ou III, obesidade mórbida, doença pulmonar preexistente e realização de cirurgia, em vez de radiologia, foram os cinco preditores. O ponto de corte de quatro identificou uma categoria de maior risco com sensibilidade de 77,6% na derivação e 76,2% na validação. Entretanto, as estatísticas C de 0,64 e 0,63 mostram discriminação apenas moderada [25].

O desempenho limitado é esperado porque eventos respiratórios dependem de fatores não totalmente capturados antes da anestesia: dificuldade de pré-oxigenação, secreções, resposta à instrumentação, profundidade anestésica, dose de opioide, bloqueio residual e qualidade da recuperação. Um escore pré-operatório deve ser utilizado para preparar recursos, não para criar falsa segurança. Crianças inicialmente classificadas como baixo risco precisam ser reavaliadas quando ocorrem obstrução, dessaturação ou necessidade de manobras na unidade pós-anestésica.

Modelos específicos de adenotonsilectomia apresentam melhor desempenho quando incorporam parâmetros polissonográficos e dados da recuperação. No estudo de Lee et al., a combinação de variáveis pré e pós-operatórias alcançou AUC 0,89, substancialmente superior ao desempenho de modelos gerais. Essa vantagem decorre de maior homogeneidade do procedimento e de variáveis diretamente ligadas à fisiopatologia da apneia. Ainda assim, o estudo foi unicêntrico e requer validação externa antes de ser transformado em regra de alta [16].

A análise de Martin et al. utilizou LASSO para selecionar fatores associados ao oxigênio suplementar. O método é útil quando há muitos candidatos e risco de sobreajuste, mas os coeficientes dependem da população e

da frequência dos eventos. A ausência de intervalos de confiança no resumo e de validação externa impede classificar o modelo como ferramenta pronta para decisão. Sua principal contribuição é demonstrar que IMC extremo, apneia grave e comorbidades têm magnitudes semelhantes e devem ser avaliados em conjunto [29].

Tabela 5. Modelos e critérios de predição respiratória relevantes para crianças com obesidade.

Estudo	N	Variáveis	Desempenho	Aplicação
Subramanyam 2016	19.059	Idade ≤3; ASA II/III; obesidade mórbida; doença pulmonar; cirurgia	C 0,64/0,63; sensibilidade 77,6%/76,2%	Triagem geral; discriminação moderada
Lee 2022	155	Nadir de SpO_2; tempo <90%; variáveis da PACU	AUC 0,89	Promissor para adenotonsilectomia; validar externamente
Baker 2020	165	Necessidade de suporte após 2 horas	Critério dinâmico; sem AUC	Apoia reclassificação durante recuperação
Martin 2024	401	IMC, IAH e comorbidades selecionados por LASSO	ORs selecionados; sem AUC no resumo	Gera hipóteses; não é escore validado

Fonte: Estudos originais [16,17,25,29].

5.11 Implicações multiprofissionais e segurança do sistema

A prevenção de complicações não é responsabilidade exclusiva do anestesiologista. A equipe de enfermagem participa da mensuração antropométrica, triagem de sintomas, preparo de dispositivos, posicionamento, vigilância na recuperação e educação para alta. Registros incompletos de altura, peso, ronco, uso de CPAP ou infecção recente reduzem a capacidade de estratificação. Protocolos devem definir campos obrigatórios e critérios de comunicação imediata ao profissional responsável.

A fisioterapia respiratória pode contribuir em crianças com doença neuromuscular, limitação ventilatória ou uso domiciliar de suporte,

especialmente quando o procedimento exige internação. Não há evidência para aplicação rotineira de intervenções fisioterapêuticas em toda criança com obesidade, mas avaliação seletiva pode auxiliar na adaptação de interfaces, expansão pulmonar, tosse assistida e continuidade do suporte habitual. A indicação deve ser baseada em função, não apenas no IMC.

O cirurgião precisa informar duração esperada, posição, possibilidade de sangramento, intensidade de dor e necessidade de imobilização. Procedimentos que limitam ventilação ou demandam opioides aumentam a importância da observação. Em cirurgia de via aérea, o plano deve integrar anestesia e otorrinolaringologia para definir momento de extubação, hemostasia, uso de antieméticos e critérios de transferência.

A farmácia clínica pode apoiar protocolos de dose que indiquem explicitamente quando utilizar peso total, ideal ou massa magra. Erros de escala são particularmente perigosos em opioides, sedativos e bloqueadores neuromusculares. Sistemas eletrônicos podem emitir alerta quando a dose excede faixa recomendada, mas alertas excessivos tendem a ser ignorados. A implementação deve priorizar medicamentos de maior risco e permitir revisão por especialista.

A segurança do sistema exige equipamento compatível com o peso e o tamanho corporal, inclusive mesas, macas, dispositivos de transferência, manguitos e interfaces de ventilação. A falta desses recursos produz atraso, constrangimento e risco ocupacional. Auditorias devem monitorar disponibilidade e tempo até obtenção do equipamento, porque esses desfechos organizacionais não aparecem nos estudos clínicos tradicionais, embora influenciem diretamente o cuidado.

Na alta, responsáveis devem receber instruções claras sobre sonolência excessiva, pausas respiratórias, esforço, cianose, dificuldade para despertar e necessidade de procurar serviço de emergência. A orientação deve considerar distância do hospital, transporte, alfabetização em saúde e possibilidade de

observar a criança durante o sono. Uma alta tecnicamente correta pode tornar-se insegura quando o contexto domiciliar não permite vigilância adequada.

6 MATRIZ PARA DECISÃO CLÍNICA

Tabela 6. Estratificação clínica pragmática proposta.

Categoria	Características	Conduta sugerida
Baixo	Obesidade não grave; cirurgia não aérea; sem OSA/asma ativa; recuperação estável	Protocolo padrão reforçado; alta conforme critérios
Intermediário	Obesidade grave ou ronco/OSA leve; procedimento moderado; necessidade limitada de opioide	Planejamento avançado; observação prolongada; considerar monitorização ventilatória
Alto	OSA moderada/grave; nadir <80%; infecção recente; doença pulmonar; cirurgia de via aérea	Equipe experiente; estratégia poupadora de opioides; internação/monitorização contínua
Muito alto	Eventos recorrentes na PACU; oxigênio persistente; suporte ventilatório; múltiplas comorbidades	Unidade de alta dependência ou terapia intensiva conforme resposta

Fonte: Proposta interpretativa; não constitui escore validado.

A matriz não substitui protocolos institucionais nem avaliação especializada. Seu valor é tornar explícitos os elementos que devem ser considerados conjuntamente. A passagem de uma categoria para outra pode ocorrer durante o perioperatório: uma criança inicialmente intermediária que apresenta obstruções repetidas na recuperação deve ser reclassificada e observada em ambiente compatível com o novo risco.

7 AGENDA DE PESQUISA

Estudos futuros devem padronizar eventos respiratórios por gravidade, momento e intervenção necessária. A publicação separada de dessaturação breve, laringoespasmo, ventilação com máscara, reintubação e internação permitiria comparações mais úteis. Também é necessário reportar peso, altura, percentil, classe de obesidade, circunferência abdominal ou cervical e presença de apneia.

Coortes multicêntricas devem desenvolver modelos que combinem gravidade da obesidade, idade, sintomas respiratórios, polissonografia, tipo de cirurgia, dispositivo de via aérea, dose de opioide e eventos na recuperação. A validação externa é indispensável, pois modelos construídos em centros terciários podem não se aplicar a hospitais gerais.

Ensaios pragmáticos podem comparar pacotes de pré-oxigenação, posição elevada, pressão expiratória individualizada, analgesia poupadora de opioides e monitorização ventilatória. Desfechos devem incluir eventos clínicos, tempo de recuperação, internação não planejada, conforto da família e custos. A simples detecção de mais episódios por monitor não garante benefício se não houver protocolo de resposta.

Por fim, estudos qualitativos devem examinar experiência de crianças e responsáveis, estigma relacionado ao peso e barreiras estruturais. A segurança perioperatória depende tanto de decisões farmacológicas quanto de comunicação respeitosa, equipamento adequado e acesso equitativo a serviços especializados.

8 CONCLUSÃO

A obesidade pediátrica associa-se a aumento do risco de eventos respiratórios perioperatórios, sobretudo quando grave, acompanhada de apneia obstrutiva do sono ou submetida a cirurgia de via aérea. A meta-análise bayesiana exploratória indicou aumento médio do risco, mas o amplo intervalo preditivo confirma que a magnitude depende do contexto. IMC isolado não é suficiente para decidir internação ou alta. A avaliação deve integrar idade, gravidade da adiposidade, infecção respiratória recente, asma, polissonografia, tipo de procedimento, necessidade de opioides e evolução na recuperação. O cuidado profissional combina pré-oxigenação, plano de via aérea, ventilação protetora, farmacologia racional, reversão completa e monitorização proporcional ao risco. A criança deve ser tratada com segurança e respeito, sem

transformar obesidade em rótulo nem em justificativa para falhas organizacionais.

DECLARAÇÕES

Financiamento: não houve financiamento específico. Aspectos éticos: foram utilizados exclusivamente dados agregados publicados. Integridade dos dados: nenhuma amostra, frequência ou estimativa clínica foi inventada; a análise bayesiana foi reproduzida a partir das medidas e intervalos informados nas publicações selecionadas.

REFERÊNCIAS

1. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC growth charts for the United States: methods and development. Vital Health Stat 11. 2002;(246):1-190.

2. Hampl SE, Hassink SG, Skinner AC, Armstrong SC, Barlow SE, Bolling CF, et al. Clinical practice guideline for the evaluation and treatment of children and adolescents with obesity. Pediatrics. 2023;151(2):e2022060640. doi:10.1542/peds.2022-060640.

3. Marjanovic V, Budic I, Golubovic M, Breschan C. Perioperative respiratory adverse events during ambulatory anesthesia in obese children. Ir J Med Sci. 2022;191(3):1305-1313. doi:10.1007/s11845-021-02659-3.

4. Chidambaran V, Tewari A, Mahmoud M. Anesthetic and pharmacologic considerations in perioperative care of obese children. J Clin Anesth. 2018;45:39-50.

5. von Ungern-Sternberg BS, Trachsel D, Erb TO, Hammer J. Forced expiratory flows and volumes in intubated and paralyzed infants and children: normative data up to 5 years of age. J Appl Physiol. 2009;107(1):105-111.

6. Nafiu OO, Reynolds PI, Bamgbade OA, Tremper KK, Welch K, Kasa-Vubu JZ. Childhood body mass index and perioperative complications. Paediatr Anaesth. 2007;17(5):426-430.

7. Tait AR, Voepel-Lewis T, Burke C, Kostrzewa A, Lewis I. Incidence and risk factors for perioperative adverse respiratory events in children who are obese. Anesthesiology. 2008;108(3):375-380.

8. El-Metainy S, Ghoneim T, Aridae E, Abdel Wahab M. Incidence of perioperative adverse events in obese children undergoing elective general surgery. Br J Anaesth. 2011;106(3):359-363. doi:10.1093/bja/aeq368.

9. Nafiu OO, Onyewuche V. Association of abdominal obesity in children with perioperative respiratory adverse events. J Perianesth Nurs. 2014;29(2):84-93. doi:10.1016/j.jopan.2013.03.010.

10. Raghavan KC, Moo DXY, Tan ZH. Severe obesity in children as an independent risk factor for perioperative respiratory adverse events during anaesthesia for minor non-airway surgery: a retrospective observational study. Proc Singapore Healthc. 2019;28(2):83-89. doi:10.1177/2010105818802994.

11. Thalji L, Shi Y, Hanson KT, Hyder JA. Characterizing the spectrum of body mass index associated with severe postoperative pulmonary complications in children. J Anesth. 2019;33(3):372-380.

12. Setzer N, Saade E. Childhood obesity and anesthetic morbidity. Paediatr Anaesth. 2007;17(4):321-326.

13. Fung E, Cave D, Witmans M, Gan K, El-Hakim H. Postoperative respiratory complications and recovery in obese children following adenotonsillectomy for sleep-disordered breathing: a case-control study. Otolaryngol Head Neck Surg. 2010;142(6):898-905. doi:10.1016/j.otohns.2010.02.012.

14. Gleich SJ, Olson MD, Sprung J, Weingarten TN, Schroeder DR, Warner DO, et al. Perioperative outcomes of severely obese children undergoing tonsillectomy. Paediatr Anaesth. 2012;22(12):1171-1178. doi:10.1111/j.1460-9592.2012.03905.x.

15. Tweedie DJ, Bajaj Y, Ifeacho SN, Jonas NE, Jephson CG, Cochrane LA, et al. Peri-operative complications after adenotonsillectomy in a UK pediatric tertiary referral centre. Int J Pediatr Otorhinolaryngol. 2012;76(6):809-815.

16. Lee TC, Walia H, Manickam PV, et al. Predictors of overnight postoperative respiratory complications in obese children undergoing adenotonsillectomy for obstructive sleep apnea. Int J Pediatr Otorhinolaryngol. 2022;162:111306.

17. Baker AK, et al. Assessing frequency of respiratory complications in children undergoing adenotonsillectomy. Pediatr Crit Care Med. 2020;21(7):e409-e415.

18. Efune PN, Pinales P, Park J, Poppino KF, Mitchell RB, Szmuk P. Pediatric obstructive sleep apnea: a prospective observational study of respiratory events in the immediate recovery period after adenotonsillectomy. Anaesth Crit Care Pain Med. 2024;43(4):101385. doi:10.1016/j.accpm.2024.101385.

19. Lejus C, et al. Peri-operative management of overweight and obese children and adolescents. Lancet Child Adolesc Health. 2018;2(6):453-464.

20. Templeton TW, Sommerfield D, Hii J, Sommerfield A, Matava CT, von Ungern-Sternberg BS. Risk assessment and optimization strategies to reduce perioperative respiratory adverse events in pediatric anesthesia—part 2: anesthesia-related risk and treatment options. Paediatr Anaesth. 2022;32(2):217-227. doi:10.1111/pan.14376.

21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

22. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

23. Scherrer PD, Mallory MD, Cravero JP, Lowrie L, Hertzog JH, Berkenbosch JW; Pediatric Sedation Research Consortium. The impact of obesity on pediatric procedural sedation-related outcomes. Paediatr Anaesth. 2015;25(7):689-697. doi:10.1111/pan.12627.

24. Najafi N, Veyckemans F, Vanhonacker D, et al. Incidence and risk factors for adverse events during monitored anaesthesia care for paediatric gastrointestinal endoscopy. Eur J Anaesthesiol. 2019;36(6):390-399. doi:10.1097/EJA.0000000000000995.

25. Subramanyam R, Yeramaneni S, Hossain MM, Anneken AM, Varughese AM. Perioperative respiratory adverse events in pediatric ambulatory anesthesia: development and validation of a risk prediction tool. Anesth Analg. 2016;122(5):1578-1585.

26. Kiekkas P, Stefanopoulos N, Bakalis N, Kefaliakos A, Konstantinou E. Perioperative adverse respiratory events in overweight/obese children: systematic review. J Perianesth Nurs. 2016;31(1):11-22.

27. Katz SL, Monsour A, Barrowman N, Hoey L, Bromwich M, Momoli F, et al. Predictors of postoperative respiratory complications in children undergoing adenotonsillectomy. J Clin Sleep Med. 2020;16(1):41-48.

28. Vaughn WL, et al. Evaluating obesity as a risk factor for complications after pediatric adenotonsillectomy. Int J Pediatr Otorhinolaryngol. 2022;162:111303.

29. Martin E, et al. Supplemental oxygen requirement after tonsillectomy in children older than 3 years of age. Int J Pediatr Otorhinolaryngol. 2024;178:111887.

30. Mitchell RB, Archer SM, Ishman SL, Rosenfeld RM, Coles S, Finestone SA, et al. Clinical practice guideline: tonsillectomy in children (update). Otolaryngol Head Neck Surg. 2019;160(1 Suppl):S1-S42.

31. Schwengel DA, Dalesio NM, Stierer TL. Pediatric obstructive sleep apnea: perioperative management and outcome. Curr Opin Anaesthesiol. 2014;27(3):286-291.

32. Oofuvong M, Geater AF, Chongsuvivatwong V, et al. Excess costs and length of hospital stay attributable to perioperative respiratory events in children. Anesth Analg. 2015;120(2):411-419. doi:10.1213/ANE.0000000000000557.

33. Röver C. Bayesian random-effects meta-analysis using the bayesmeta R package. J Stat Softw. 2020;93(6):1-51. doi:10.18637/jss.v093.i06.

APÊNDICE A — ESTRATÉGIA DE BUSCA

Estratégia-base para PubMed/MEDLINE: ((obesity[Title/Abstract] OR obese[Title/Abstract] OR overweight[Title/Abstract] OR "body mass index"[Title/Abstract] OR "severe obesity"[Title/Abstract]) AND (child*[Title/Abstract] OR pediatric[Title/Abstract] OR adolescent*[Title/Abstract]) AND (anesthesia[Title/Abstract] OR anaesthesia[Title/Abstract] OR surgery[Title/Abstract] OR sedation[Title/Abstract]) AND (respiratory[Title/Abstract] OR airway[Title/Abstract] OR desaturation[Title/Abstract] OR hypoxemia[Title/Abstract] OR laryngospasm[Title/Abstract] OR bronchospasm[Title/Abstract] OR "sleep apnea"[Title/Abstract])). A busca foi complementada pelos termos tonsillectomy, adenotonsillectomy, postoperative oxygen e perioperative respiratory adverse events.

Para atualização formal, recomenda-se registrar a data, a plataforma, o número de registros, os motivos de exclusão e a estratégia completa de cada base.

A presente síntese priorizou reprodutibilidade das estimativas incluídas e não apresentou contagens PRISMA que não pudessem ser auditadas.

APÊNDICE B — DICIONÁRIO MÍNIMO PARA NOVOS ESTUDOS

Tabela B1. Variáveis mínimas para padronizar pesquisa e auditoria clínica.

Domínio	Variáveis
Antropometria	peso e altura medidos; percentil; escore-z; classe de obesidade; circunferência cervical/abdominal
Fenótipo respiratório	ronco, pausas, AHI, nadir de SpO_2, tempo <90%, asma, IVAS recente
Procedimento	tipo, duração, via aérea ou não aérea, urgência, posição
Anestesia	indução, dispositivo, tentativas, ventilação, PEEP, opioides, bloqueio e reversão
Eventos	obstrução, dessaturação, laringoespasmo, broncoespasmo, ventilação, reintubação
Recuperação	tempo, oxigênio, capnografia, suporte após 2 h, internação não planejada
Contexto	nível do hospital, experiência da equipe, disponibilidade de UTI e equipamentos

APÊNDICE C — ESPECIFICAÇÃO DA META-ANÁLISE BAYESIANA

A síntese utilizou o logaritmo da razão de efeito e o erro-padrão derivado do intervalo de confiança: SE = [ln(limite superior) − ln(limite inferior)]/(2 × 1,96). Para cada estudo, y_i foi modelado por distribuição normal com média θ_i e variância conhecida SE_i^2. Os efeitos verdadeiros θ_i foram assumidos normalmente distribuídos em torno de μ, com variância τ^2. A prior de μ, Normal(0, 2^2), permite desde efeitos protetores expressivos até aumentos muito grandes; a prior half-normal de τ, escala 0,5, favorece heterogeneidade pequena ou moderada sem excluir valores altos.

Resultados numéricos: efeito médio posterior mediano = 3,265; CrI 95% = 1,667–7,389; probabilidade de razão de efeito >1 = 0,9985; probabilidade >2 = 0,9276; τ mediano = 0,630; intervalo preditivo = 0,722–17,656. Na sensibilidade

sem cirurgia de via aérea, a mediana foi 1,894, CrI 95% 1,087–4,228. Esses resultados são exploratórios e dependem da aproximação entre OR e RR.

A análise pode ser reproduzida em R com os pacotes bayesmeta, brms ou rstanarm. Recomenda-se verificar priors alternativas para τ, análise leave-one-out, separação por procedimento e metarregressão apenas quando houver número suficiente de estudos. A publicação formal de uma meta-análise deve incluir arquivo de dados, código, versão dos pacotes e registro das decisões de inclusão.

APÊNDICE D — INDICADORES PARA AUDITORIA INSTITUCIONAL

Tabela D1. Indicadores propostos para auditoria de segurança respiratória.

Indicador	Definição	Meta sugerida
Documentação antropométrica	% com peso, altura e percentil registrados	≥95%
Triagem de OSA	% com ronco/apneia/CPAP documentados	≥90%
Infecção respiratória	% com sintomas recentes registrados	≥90%
Plano de via aérea	% de alto risco com plano e dispositivo de resgate	100%
Bloqueio neuromuscular	% com monitorização quantitativa quando utilizado	≥95%
Recuperação	Taxa de obstrução, dessaturação e oxigênio persistente	Monitorar tendência
Alta	% com critérios e orientação documentados	100%
Internação não planejada	Eventos por 100 procedimentos e motivo	Monitorar tendência

Fonte: Proposta para melhoria da qualidade; metas devem ser adaptadas à estrutura local.

CAPÍTULO 14

MUDANÇAS DE DESFECHOS, RELATO SELETIVO E INTERPRETAÇÃO EXAGERADA EM ENSAIOS CLÍNICOS DE INTERVENÇÕES CIRÚRGICAS: ESTUDO METAEPIDEMIOLÓGICO DE PROTOCOLOS, REGISTROS E PUBLICAÇÕES

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Nélio Barreto Vieira

RESUMO

Objetivo: examinar a frequência, a forma e as implicações das discrepâncias entre protocolos, registros e publicações de ensaios clínicos cirúrgicos, com ênfase em mudanças do desfecho primário, relato seletivo e spin. Método: foi conduzido um estudo metaepidemiológico de segunda ordem, apoiado em investigações metodológicas que compararam documentos de planejamento e relatórios finais ou avaliaram linguagem de interpretação em cirurgia geral, ortopedia, cirurgia bariátrica, cirurgia plástica e estudos piloto cirúrgicos. As fontes foram pesquisadas em PubMed/MEDLINE, Web of Science, Scopus, ClinicalTrials.gov e referências dos estudos elegíveis. Foram extraídos desenho, especialidade, tamanho da amostra metodológica, tipo de documento confrontado, definição de discrepância, frequência de spin e fatores associados. Resultados: a literatura mostrou que o problema não se restringiu a uma especialidade. Spin foi identificado em 112 de 250 resumos de ensaios ortopédicos com desfecho primário não significativo e em 59 de 72 estudos piloto ou de viabilidade cirúrgicos. Em cirurgia plástica, 85% dos ensaios não significativos empregaram alguma estratégia de spin; na cirurgia bariátrica, a frequência variou conforme a seção, alcançando 72% na discussão e 76% na conclusão. Estudos transversais também registraram inconsistências entre resumo e texto completo, subnotificação de desfechos relevantes ao paciente e

baixa disponibilidade de protocolos anteriores à publicação. Conclusão: o registro prospectivo isolado não assegurou transparência. Auditorias precisaram reconstruir a cronologia do ensaio, comparar cada definição do desfecho e avaliar se a redação final preservou a incerteza estatística e clínica.

Palavras-chave: ensaios clínicos; cirurgia; relato seletivo; desfechos; spin; integridade científica.

ABSTRACT

Objective: to examine the frequency, forms, and implications of discrepancies between protocols, registries, and publications of surgical clinical trials, focusing on primary outcome changes, selective reporting, and spin. Methods: a second-order meta-epidemiological study was conducted using methodological investigations that compared planning documents with final reports or assessed interpretive language in general surgery, orthopaedics, bariatric surgery, plastic surgery, and surgical pilot trials. PubMed/MEDLINE, Web of Science, Scopus, ClinicalTrials.gov, and reference lists were searched. Study design, surgical field, methodological sample size, documents compared, discrepancy definitions, spin frequency, and associated factors were extracted. Results: spin was identified in 112 of 250 abstracts of orthopaedic trials with non-significant primary outcomes and in 59 of 72 surgical pilot or feasibility trials. In plastic surgery, 85% of non-significant trials used spin; in bariatric surgery, prevalence differed by manuscript section and reached 72% in discussions and 76% in conclusions. Other methodological surveys documented abstract–full-text inconsistencies, incomplete reporting of patient-important outcomes, and limited access to protocols made available before publication. Conclusion: prospective registration alone did not guarantee transparent reporting. A credible audit must reconstruct trial chronology, compare every outcome definition, and determine whether the final wording preserves statistical and clinical uncertainty.

Keywords: clinical trials; surgery; selective reporting; outcomes; spin; research integrity.

1 INTRODUÇÃO

Ensaios randomizados sustentam decisões sobre indicação, técnica, momento operatório e segurança, mas a autoridade atribuída ao delineamento não elimina distorções produzidas durante a passagem do protocolo ao artigo. Um mesmo estudo pode ter um desfecho principal no cálculo amostral, outro no registro, uma terceira formulação no manuscrito e uma conclusão que enfatiza resultados secundários. A literatura clássica demonstrou que resultados estatisticamente significativos apresentaram maior probabilidade de relato completo e que mudanças de desfechos nem sempre foram acompanhadas de justificativa explícita.[1-4] Em cirurgia, o problema ganhou importância adicional porque cegamento, padronização técnica, curva de aprendizagem, preferência do cirurgião e dependência do centro tornam os ensaios mais vulneráveis a decisões analíticas posteriores.

Relato seletivo e spin não foram tratados como sinônimos. O primeiro correspondeu à escolha de quais desfechos, tempos, escalas, populações ou análises apareceram no relatório; o segundo correspondeu ao modo de apresentar e interpretar resultados, sobretudo quando o desfecho primário não sustentou superioridade. A distorção pôde ocorrer pela ênfase em comparações intragrupo, subgrupos, análises per protocol, tendências sem significância ou afirmações de equivalência sem margem previamente definida.[5,6] Essas estratégias não exigiram fraude deliberada para produzir dano: bastou que deslocassem a atenção do leitor e aumentassem a impressão de benefício.

A investigação metaepidemiológica foi particularmente útil porque examinou o comportamento da própria literatura. Em vez de perguntar se uma operação funcionou, perguntou-se se o caminho entre planejamento e publicação permaneceu verificável. Esse tipo de pesquisa permitiu quantificar discrepâncias, localizar em que seção o spin se concentrou e testar associações com financiamento, impacto do periódico, registro prospectivo ou tamanho amostral. Os levantamentos disponíveis em ortopedia, cirurgia bariátrica, cirurgia plástica

e estudos piloto cirúrgicos mostraram frequências altas e heterogêneas, sugerindo que o problema não poderia ser atribuído apenas a uma área ou período.[7-11]

O objetivo deste capítulo foi realizar uma síntese metaepidemiológica estruturada de estudos que confrontaram protocolos, registros e publicações ou avaliaram spin em ensaios de intervenções cirúrgicas. Pretendeu-se descrever a frequência observada, comparar definições, identificar padrões recorrentes e propor um modelo de auditoria aplicável a novos estudos. Não se buscou combinar indiscriminadamente proporções provenientes de universos diferentes; a prioridade foi preservar denominadores, contexto e seção do artigo em que cada desvio foi medido.

2 MÉTODO

Foi conduzido um estudo metaepidemiológico de segunda ordem, definido como a análise de investigações metodológicas que avaliaram ensaios clínicos individuais. O protocolo de trabalho estabeleceu previamente a pergunta, os domínios de extração e a regra de não combinar estimativas quando as definições de spin, os critérios de não significância ou as unidades de análise fossem incompatíveis. A redação seguiu recomendações para estudos metodológicos e incorporou itens do PRISMA 2020 aplicáveis à identificação e seleção das fontes.[12,13]

A busca foi realizada em PubMed/MEDLINE, Scopus e Web of Science, complementada por busca dirigida no ClinicalTrials.gov e rastreamento de referências. Foram combinados termos relacionados a randomized trial, surgery, surgical, outcome switching, selective outcome reporting, protocol discrepancy, registration, spin, overinterpretation e pilot trial. Foram aceitos estudos publicados até 31 de julho de 2026 que incluíram ensaios cirúrgicos ou apresentaram resultados separáveis para intervenções cirúrgicas. Comentários sem dados originais foram utilizados apenas para contextualização.

Dois níveis de evidência foram considerados. O núcleo cirúrgico reuniu levantamentos de ortopedia, cirurgia bariátrica, cirurgia plástica, estudos piloto/viabilidade e amostras multiespecialidade que incluíram cirurgia. O núcleo contextual reuniu estudos clássicos de comparação protocolo-publicação, pesquisas sobre concordância de desfechos primários e avaliações de disponibilidade de protocolos. Essa separação evitou que estimativas oriundas de reabilitação, farmacologia ou periódicos gerais fossem indevidamente apresentadas como prevalência cirúrgica.

Foram extraídos autores, ano, especialidade, desenho da pesquisa metodológica, janela de publicação, número de ensaios avaliados, condição de elegibilidade do desfecho primário, documentos confrontados, unidade de análise, definição de spin, frequência geral e frequência por seção. Quando numerador e denominador estavam disponíveis, foram calculados intervalos de confiança de 95% pela aproximação de Wilson para fins gráficos. Percentuais sem denominador verificável foram mantidos como publicados e não receberam intervalos recalculados.

As discrepâncias foram classificadas em omissão, introdução, promoção de secundário a primário, rebaixamento de primário, mudança de tempo, mudança de métrica, mudança de população analítica e alteração de direção interpretativa. Spin foi classificado conforme foco em resultados secundários, comparações intragrupo, subgrupos, tendências, equivalência indevida, minimização de eventos adversos ou recomendação clínica não sustentada pelo desfecho primário.[5,6] O financiamento e os conflitos de interesse foram registrados quando os estudos metodológicos os haviam analisado.

A síntese foi predominantemente narrativa e comparativa. Não foi realizada metanálise global porque os estudos diferiram na especialidade, no tipo de documento examinado, na seleção exclusiva de ensaios com resultado primário não significativo e na seção do manuscrito avaliada. Foram produzidos painéis gráficos com prevalências, matrizes de cobertura e fluxos de auditoria.

Os cálculos derivados foram limitados a proporções com numeradores e denominadores explicitamente publicados.

A Tabela 1 foi organizada para preservar a unidade de análise de cada investigação. Algumas pesquisas selecionaram exclusivamente ensaios com desfecho primário não significativo, enquanto outras avaliaram estudos piloto ou confrontaram protocolos e artigos. Essa distinção foi essencial para evitar comparação direta entre percentuais que respondiam a perguntas diferentes.

Tabela 1. Estudos metaepidemiológicos e dados reais utilizados na síntese.

Estudo	Campo/unidade	Amostra metodológica	Achado principal	Documento comparado
Boutron et al., 2010	RCTs com primário não significativo	72	Spin na conclusão do resumo: 58,3%; no texto: 50,0%	Artigo completo
Fleming et al., 2016	Literatura clínica com componente cirúrgico	110	Spin em 27% das conclusões de resumos; 28% eram intervenções cirúrgicas	Artigos
Jain et al., 2020	Ortopedia	250	112 resumos com spin (44,8%)	Resumo e texto
Rassy et al., 2021	Cirurgia bariátrica	RCTs não significativos	37% nos resultados, 72% na discussão e 76% na conclusão	Artigo completo
Yuan et al., 2023	Cirurgia plástica	RCTs não significativos	85% apresentaram spin	Artigo completo
McKechnie et al., 2025	Pilotos/viabilidade em cirurgia	72	59 estudos com spin (81,9%)	Resumo, resultados e discussão
Mathieu et al., 2025	RCTs com protocolo disponível	215	3,7% divergiram no primário entre protocolo e publicação	Protocolo, registro e artigo

O conjunto mostrou duas regularidades. Primeiro, a distorção apareceu em várias especialidades e formatos de estudo. Segundo, a frequência aumentou

nas seções interpretativas: em cirurgia bariátrica, por exemplo, o spin foi menos frequente na apresentação objetiva dos resultados do que na discussão e na conclusão. Esse deslocamento indicou que a leitura apenas das tabelas não bastou para avaliar fidelidade científica.

A Figura 1 apresenta os estudos em ordem temporal. O tamanho dos círculos representa aproximadamente o número de ensaios avaliados quando esse denominador foi informado; a posição vertical apenas separa as investigações. O objetivo não foi sugerir tendência secular uniforme, mas mostrar que estimativas elevadas continuaram sendo observadas depois da consolidação do registro prospectivo.

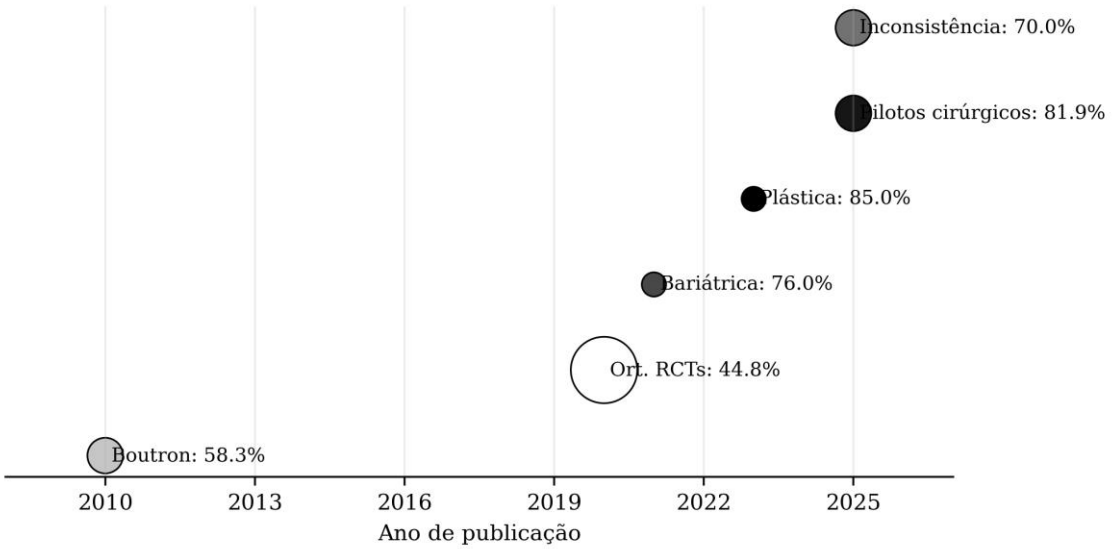


Figura 1. Evolução temporal e magnitude das estimativas de spin e inconsistência em estudos metodológicos.

A persistência de percentuais altos em 2025 enfraqueceu a hipótese de que o registro, por si só, teria resolvido o problema. A expansão dos registros aumentou a possibilidade de auditoria, porém não impediu redações que privilegiaram eficácia em estudos de viabilidade nem inconsistências entre resumo e texto completo. A governança editorial precisou, portanto, verificar concordância, e não apenas a existência de um número de registro.

Para permitir comparação estatística transparente, a Figura 2 foi restrita a quatro estimativas com contagens publicadas. Os intervalos de confiança foram

recalculados pelo método de Wilson e a linha tracejada marcou 50%, valor que corresponde à presença do problema em metade dos artigos avaliados.

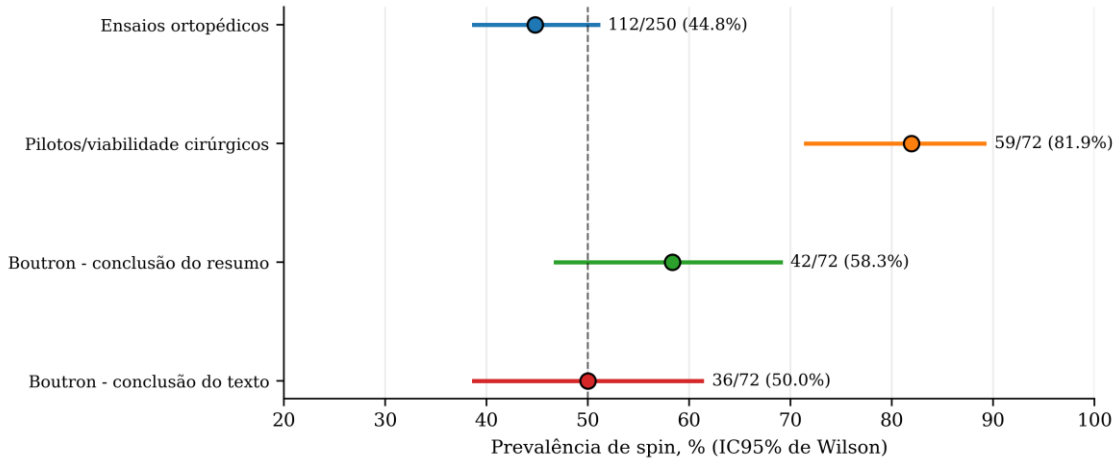


Figura 2. Prevalência de spin em investigações com numerador e denominador disponíveis.

Os ensaios ortopédicos apresentaram estimativa abaixo de 50%, mas o grande denominador tornou o intervalo relativamente estreito. Nos estudos piloto cirúrgicos, mais de quatro quintos dos relatos empregaram spin. A diferença refletiu parcialmente os critérios: em pilotos, qualquer foco deslocado da viabilidade para eficácia foi considerado distorção, uma definição apropriada ao propósito específico desses estudos.

3 RESULTADOS

Foram incluídas sete investigações centrais e estudos contextuais que fundamentaram definições e mecanismos de viés. O levantamento ortopédico avaliou 250 ensaios com desfecho primário estatisticamente não significativo e identificou spin em 112 resumos, correspondendo a 44,8%.[8] A magnitude foi suficiente para tornar plausível que um leitor limitado ao resumo recebesse uma interpretação mais favorável do que aquela sustentada pelo resultado primário.

Nos estudos piloto ou de viabilidade cirúrgicos, 59 de 72 artigos utilizaram spin, prevalência de 81,9% e intervalo de confiança publicado de 71,4% a 89,3%.[11] O padrão dominante foi tratar o estudo como teste preliminar de eficácia, embora

o propósito de um piloto fosse estimar recrutamento, adesão, aceitabilidade, completude dos dados e viabilidade da intervenção ou do ensaio definitivo. A distorção foi especialmente relevante porque resultados de eficácia em amostras pequenas apresentaram grande incerteza.

Na cirurgia bariátrica, o perfil variou dentro do próprio manuscrito. Spin foi reconhecido em 37% das seções de resultados, 72% das discussões e 76% das conclusões; aproximadamente 55% dos resumos utilizaram linguagem distorcida.[9] A escalada entre resultados e conclusão indicou que parte importante do problema foi introduzida durante a interpretação. Foco em objetivos secundários, análises modificadas e benefícios metabólicos não definidos como primários foram estratégias recorrentes.

Na cirurgia plástica, 85% dos ensaios com desfecho primário não significativo apresentaram alguma forma de spin.[10] As estratégias mais frequentes incluíram afirmar equivalência diante de ausência de superioridade e concentrar a discussão em outro objetivo. Como não inferioridade e equivalência exigem desenho, margem, cálculo amostral e análise apropriados, a simples falta de significância não autorizou concluir que procedimentos eram semelhantes.

A análise clássica de Boutron e colaboradores ofereceu um referencial para interpretar os estudos cirúrgicos. Entre 72 ensaios, spin apareceu em 42 conclusões de resumos e em 36 conclusões do texto completo. Comparações intragrupo, resultados secundários, subgrupos e linguagem de equivalência foram usados para deslocar a atenção do primário não significativo.[5] Embora a amostra não fosse exclusivamente cirúrgica, suas categorias foram reproduzidas por investigações posteriores em especialidades operatórias.

O confronto entre protocolo, registro e publicação revelou outra dimensão. Em uma amostra de 215 ensaios com protocolos em inglês, oito apresentaram discrepância no desfecho primário entre protocolo e publicação; divergências entre protocolo e registro também ocorreram.[14] Em uma revisão de protocolos de reabilitação, apenas 727 de 5.597 registros foram associados a publicações, e a concordância do desfecho primário foi de 48,7% no resumo e

72,6% no texto.[15] Esses números foram utilizados como referência contextual, não como prevalência cirúrgica.

A Figura 3 traduz a extração em uma trilha de auditoria. O desfecho foi acompanhado desde o protocolo até a conclusão, e cada mudança recebeu classificação própria. O desenho evitou o erro de registrar apenas "houve discrepância", sem dizer se o problema foi omissão, promoção, rebaixamento, introdução ou mudança interpretativa.

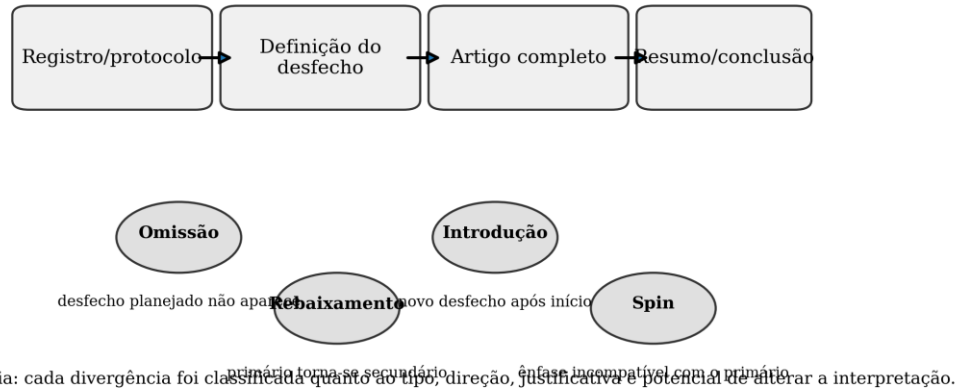


Figura 3. Fluxo operacional para auditoria de mudanças de desfechos e spin.

A principal vantagem desse fluxo foi separar modificações justificadas de alterações que poderiam favorecer resultados. Mudanças ocorreram legitimamente por novos conhecimentos, impossibilidade de mensuração ou decisão do comitê independente; contudo, precisaram ser datadas, documentadas antes do acesso aos resultados e explicadas no artigo. Ausência de justificativa não provou intenção, mas reduziu a confiabilidade.

A Tabela 2 apresenta os domínios mínimos que foram aplicados ao confrontar registros e publicações. A matriz foi desenhada para uso por revisores, editores e grupos de revisão sistemática, com categorias suficientemente específicas para permitir concordância entre avaliadores.

Tabela 2. Matriz de classificação para auditoria de ensaios cirúrgicos.

Domínio	Pergunta de auditoria	Classificação sugerida	Efeito potencial
Primário	A definição, métrica e tempo permaneceram iguais?	concordante / modificado / substituído	altera hipótese confirmatória
Secundários	Resultados não significativos foram omitidos?	completo / parcial / ausente	superestima benefício
População	ITT foi trocada por per protocol?	planejada / pós-hoc / não informada	quebra randomização
Tempo	O ponto temporal foi escolhido após o resultado?	pré-especificado / alterado	seleção do melhor momento
Segurança	Eventos adversos receberam igual visibilidade?	adequado / minimizado / omitido	distorce balanço benefício-risco
Conclusão	A linguagem refletiu o primário?	fiel / tendência / equivalência / recomendação	amplifica certeza

A classificação mostrou que discrepância de desfecho não foi o único mecanismo capaz de alterar a mensagem. Mesmo com o primário preservado, mudança de população, seleção de tempo, omissão de eventos adversos ou linguagem de equivalência puderam transformar um resultado inconclusivo em recomendação aparentemente positiva.

A Figura 4 demonstra que os estudos metodológicos não examinaram os mesmos componentes. As pesquisas de spin concentraram-se nos resultados, discussão e conclusão, enquanto estudos de concordância incorporaram registro e protocolo. A célula escura indica que o domínio foi avaliado; “não” significa que não integrou o objetivo central do estudo.

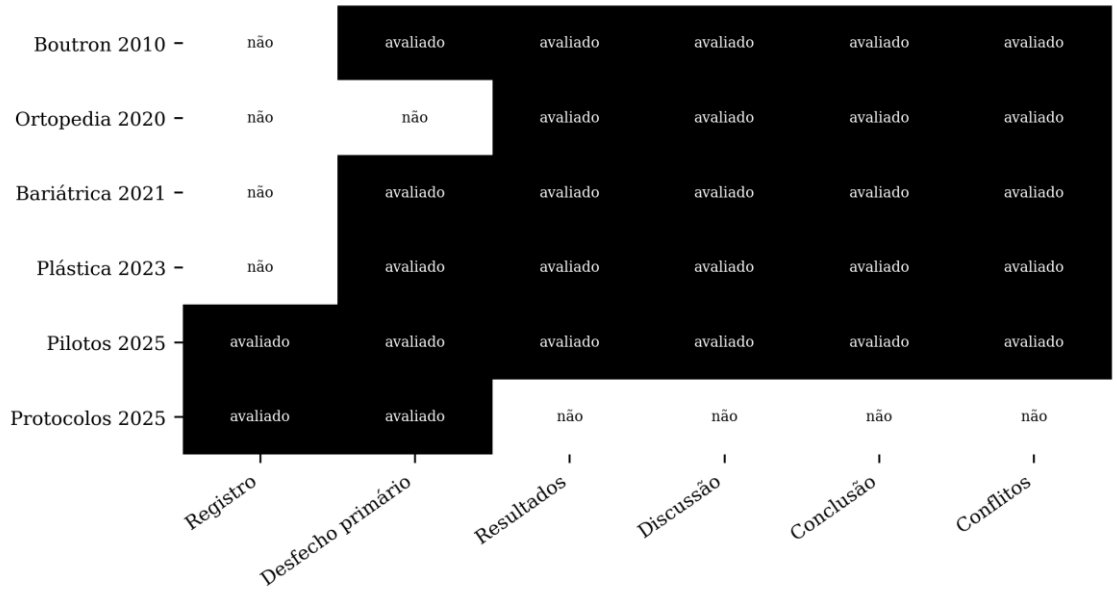


Figura 4. Matriz de cobertura dos domínios avaliados pelas principais investigações.

Essa heterogeneidade explicou por que uma metanálise simples de prevalências seria enganosa. Um estudo que examinou apenas resumos teve menor oportunidade de detectar distorções no texto completo, e uma pesquisa de protocolos não avaliou necessariamente linguagem de conclusão. A síntese precisou ser estratificada pelo ponto da cadeia documental.

A Figura 5 utiliza os percentuais publicados no estudo de cirurgia bariátrica e os dispõe em coordenadas polares. A representação evidencia a diferença entre descrição e interpretação, mantendo cada seção no mesmo raio percentual.

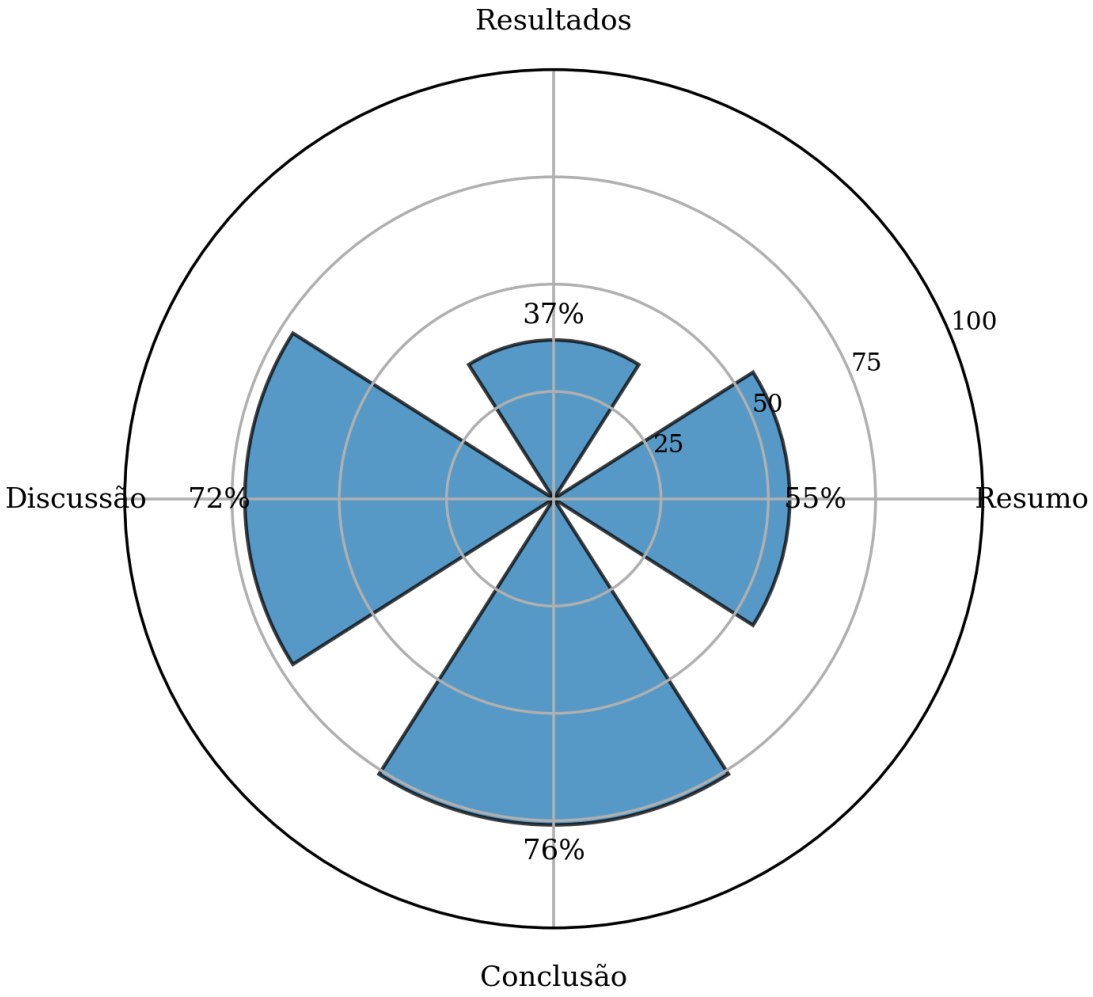


Figura 5. Distribuição do spin por seção em ensaios randomizados de cirurgia bariátrica.

O perfil não sustentou a ideia de que o spin se originou apenas de resultados mal relatados. A maior concentração na discussão e na conclusão mostrou que a distorção foi construída depois que as análises já estavam apresentadas. Editores e leitores precisaram examinar se a conclusão retomava o desfecho primário e sua incerteza, em vez de aceitar a narrativa final como resumo neutro.

A Figura 6 utiliza uma investigação externa à cirurgia para ilustrar a magnitude da perda documental que pode ocorrer antes mesmo da avaliação de spin. Dos 5.597 protocolos registrados, 727 foram associados a publicações. Entre

esses artigos, a concordância do desfecho primário foi menor no resumo do que no texto completo.

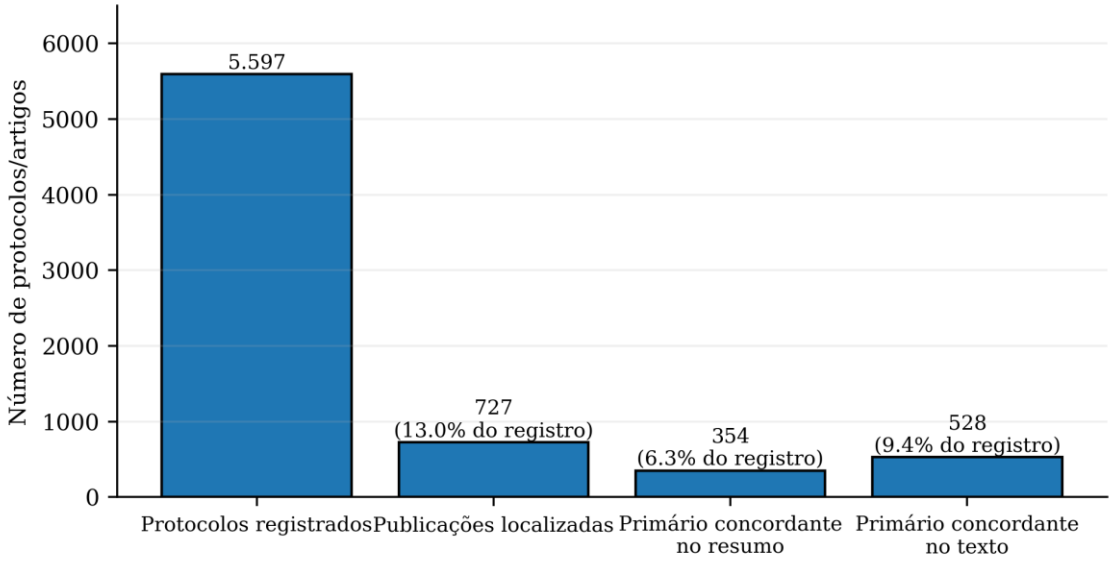


Figura 6. Perda de transparência entre protocolos registrados e publicações em estudo contextual de reabilitação.

A figura não foi apresentada como estimativa cirúrgica. Sua função foi demonstrar que a cadeia de transparência depende de quatro passos: publicação, vinculação correta ao registro, manutenção do desfecho e fidelidade do resumo. A falha em qualquer etapa dificulta auditorias e pode introduzir viés em revisões sistemáticas que dependem do resumo para triagem.

4 DISCUSSÃO

Os resultados mostraram que a integridade de um ensaio cirúrgico não pôde ser inferida apenas pelo selo “randomizado” ou pela presença de registro. Em várias especialidades, a linguagem final ultrapassou o que o desfecho primário permitia concluir. O padrão foi coerente com a literatura geral: quando o resultado confirmatório não foi significativo, autores frequentemente deslocaram a ênfase para análises secundárias ou para a ausência de diferença, tratada como equivalência.[5,16] Na prática, essa mudança alterou a pergunta

respondida. Um ensaio planejado para testar superioridade terminou apresentado como evidência de semelhança, segurança ou promessa clínica.

A cirurgia bariátrica ofereceu um exemplo especialmente claro porque a frequência aumentou das seções objetivas para as interpretativas. A discussão científica exige contextualização, mas contextualizar não significa substituir o resultado principal por uma seleção de achados favoráveis. Em intervenções que afetam peso, diabetes, qualidade de vida, eventos adversos e reoperações, o grande número de desfechos cria oportunidades para destacar o melhor resultado. A solução não foi reduzir a riqueza clínica, e sim estabelecer hierarquia explícita e relatar todos os resultados com medidas de efeito e incerteza.

Nos pilotos cirúrgicos, o problema adquiriu outra forma. Ensaios de viabilidade não foram dimensionados para eficácia, e uma diferença nominalmente significativa pôde surgir por acaso ou instabilidade. Quando 81,9% dos artigos deslocaram a narrativa para benefício, o leitor recebeu uma impressão prematura de efetividade. A consequência foi mais ampla do que comunicação: estimativas infladas puderam orientar o desenho do ensaio definitivo, induzir cálculo amostral inadequado ou favorecer adoção precoce de tecnologia.[11,17]

A alta prevalência observada em cirurgia plástica e ortopedia teve relevância clínica porque muitas decisões envolveram procedimentos com preferência do paciente, desfechos autorrelatados e impossibilidade de cegamento. Esses elementos não invalidaram os estudos, mas aumentaram a necessidade de pré-especificação e interpretação disciplinada. A afirmação de equivalência mereceu atenção particular. Ausência de significância em um teste de superioridade significou apenas que os dados não forneceram evidência suficiente de diferença; não demonstrou que a diferença clinicamente importante estivesse excluída.

Os achados também mostraram que o resumo funcionou como ponto crítico. Profissionais frequentemente leem título, resumo e conclusão antes do texto completo, e revisões usam o resumo durante a triagem. Spin no resumo,

portanto, tem maior capacidade de propagação. Estudos experimentais demonstraram que resumos distorcidos influenciam a interpretação dos leitores, ainda que o conteúdo dos resultados permaneça disponível.[18] A revisão editorial deveria comparar de forma explícita a conclusão do resumo com a estimativa do desfecho primário e seu intervalo de confiança.

Financiamento e conflitos de interesse foram hipóteses plausíveis, mas os estudos nem sempre encontraram associação consistente. Isso impediu reduzir o fenômeno a patrocínio comercial. Pressão por novidade, preferência por resultados positivos, cultura da especialidade, limitações do periódico e desconhecimento metodológico também contribuíram. A resposta mais efetiva precisou ser sistêmica: protocolos acessíveis, versões datadas do registro, planos de análise, tabelas de concordância e revisão por avaliadores treinados.

A adoção do CONSORT, do SPIRIT e de extensões específicas melhorou a estrutura de relato, porém checklists somente funcionaram quando verificados.[12,19] A presença de um item marcado não assegurou que mudanças fossem explicadas. Uma prática editorial de alto rendimento seria exigir uma tabela com desfecho original, versão do registro, versão analisada, motivo e data da alteração. Essa tabela tornaria a discrepância visível sem obrigar o leitor a reconstruir múltiplos documentos.

Para revisões sistemáticas, a implicação foi direta. O risco de viés por seleção do resultado relatado não deveria ser julgado apenas por impressão. Revisores precisaram localizar registros, protocolos e planos de análise, além de considerar que o registro pode ter sido atualizado depois do início do recrutamento. Ferramentas como RoB 2 incorporaram o domínio de seleção do resultado, mas a qualidade da decisão dependeu da cronologia documental.[20] Quando o protocolo não foi encontrado, a ausência deveria aumentar a incerteza, e não ser tratada como evidência de concordância.

Este estudo teve limitações. A síntese reuniu investigações com definições e universos distintos, algumas com denominadores pequenos. Percentuais por seção não equivaleram à prevalência por artigo, e as pesquisas focadas em

desfechos não significativos naturalmente apresentaram risco maior de spin do que amostras de todos os ensaios. Além disso, o campo evoluiu, e novos registros, suplementos ou cartas poderiam esclarecer mudanças. Essas limitações justificaram a opção por comparação estruturada, sem uma prevalência global artificial.

Apesar disso, a convergência foi forte. Independentemente da especialidade, a interpretação exagerada apareceu com frequência suficiente para afetar confiança, síntese e prática. A transparência precisou ser entendida como propriedade verificável do percurso completo do ensaio. Um artigo fiel não é aquele que evita qualquer mudança, mas aquele que registra quando a mudança ocorreu, por que ocorreu e como alterou a interpretação.

A Tabela 3 converte os achados em responsabilidades distribuídas. O controle não foi atribuído apenas aos autores, porque várias decisões ocorreram durante o acompanhamento, a análise e a revisão editorial.

Tabela 3. Recomendações operacionais para reduzir relato seletivo e spin.

Ator	Ação recomendada	Momento	Produto verificável
Pesquisadores	registrar antes do primeiro participante e anexar plano estatístico	antes do recrutamento	versão datada e protocolo
Comitê independente	aprovar mudanças sem acesso indevido aos resultados	durante o ensaio	ata e justificativa
Autores	apresentar tabela de concordância	submissão	mapa protocolo-registro-artigo
Revisores	conferir primário, tempo e população	peer review	parecer com discrepâncias
Editores	exigir correção da conclusão e dados de segurança	decisão editorial	resumo fiel
Revisores sistemáticos	buscar documentos e classificar viés	síntese	julgamento RoB 2 fundamentado

A proposta central foi produzir rastreabilidade. Quando cada ator gerou um documento verificável, mudanças legítimas deixaram de ser confundidas

com seleção pós-hoc. O custo operacional foi pequeno em comparação com o custo científico de ensaios cuja mensagem não pôde ser reconstruída.

5 CONCLUSÃO

Mudanças de desfechos, relato seletivo e spin permaneceram frequentes em diferentes segmentos da pesquisa cirúrgica. Estimativas reais variaram de 44,8% nos resumos de ensaios ortopédicos a 81,9% em pilotos cirúrgicos, e alcançaram 85% em ensaios não significativos de cirurgia plástica. A interpretação exagerada concentrou-se nas seções de discussão e conclusão, revelando que a distorção não dependeu apenas da análise, mas também da narrativa. Protocolos acessíveis, registros prospectivos datados, planos estatísticos e tabelas de concordância deveriam integrar a submissão. Para leitores e revisores, a regra mais segura foi acompanhar o desfecho desde sua definição inicial até a última frase da conclusão.

REFERÊNCIAS

1. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291(20):2457-2465. doi:10.1001/jama.291.20.2457.

2. Chan AW, Krleža-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ. 2004;171(7):735-740. doi:10.1503/cmaj.1041086.

3. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008;3(8):e3081. doi:10.1371/journal.pone.0003081.

4. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias: an updated review. PLoS One. 2013;8(7):e66844. doi:10.1371/journal.pone.0066844.

5. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010;303(20):2058-2064. doi:10.1001/jama.2010.651.

6. Boutron I, Ravaud P. Misrepresentation and distortion of research in biomedical literature. Proc Natl Acad Sci U S A. 2018;115(11):2613-2619. doi:10.1073/pnas.1710755115.

7. Fleming PS. Evidence of spin in clinical trials in the surgical literature. Ann Transl Med. 2016;4(19):385. doi:10.21037/atm.2016.08.21.

8. Jain V, et al. Analyzing spin in abstracts of orthopaedic randomized controlled trials with statistically insignificant primary endpoints. Arthroscopy. 2020;36(5):1443-1449. doi:10.1016/j.arthro.2020.01.019.

9. Rassy N, Rives-Lange C, Carette C, Barsamian C, Moszkowicz D, Thereaux J, et al. Spin occurs in bariatric surgery randomized controlled trials with a statistically nonsignificant primary outcome: a systematic review. J Clin Epidemiol. 2021;139:185-193. doi:10.1016/j.jclinepi.2021.04.013.

10. Yuan M, Wu J, Li A, Gallo L, Chin B, Murphy J, et al. Spin in plastic surgery randomized controlled trials with statistically nonsignificant primary outcomes: a systematic review. Plast Reconstr Surg. 2023;151(4):773e-782e. PMID:36442055.

11. McKechnie T, et al. Spin reporting is common in pilot randomized controlled trials in surgery: a methodological survey. J Surg Res. 2025;305:32-41. PMID:39742654.

12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

13. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017;22(4):139-142.

14. Liang F, et al. Comparison of primary endpoints between publications, registries and protocols in randomized controlled trials. J Clin Epidemiol. 2017;91:75-84. PMCID:PMC5722592.

15. Komukai K, et al. Publication bias and selective outcome reporting in randomized controlled trials related to rehabilitation. Arch Phys Med Rehabil. 2024;105(2):322-330. PMID:37364686.

16. Mathieu S, et al. Protocol publication rate and comparison between article, registry and protocol in randomized controlled trials. BMC Med Res Methodol. 2025;25:30. PMCID:PMC11786558.

17. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ. 2016;355:i5239. doi:10.1136/bmj.i5239.

18. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in cancer: the SPIIN randomized controlled trial. J Clin Oncol. 2014;32(36):4120-4126.

19. Chan AW, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration. BMJ. 2013;346:e7586. doi:10.1136/bmj.e7586.

20. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Cochrane Methods. 2019.

21. Wang X, et al. Impact of selective reporting bias on stroke trials: a cross-sectional meta-research study. BMC Med Res Methodol. 2024;24:252. PMCID:PMC11514957.

22. Stoll M, et al. Discrepancies from registered protocols and spin occurred in more than half of randomized controlled trials of nonpharmacological interventions. J Clin Epidemiol. 2020;128:97-104. PMID:32828837.

23. Menon R, et al. Has the reporting of patient-important outcomes improved in surgical trials? Br J Surg. 2021;108(8):e282-e284. PMID:33982387.

24. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials. 2017;18:122. doi:10.1186/s13063-017-1870-2.

CAPÍTULO 15

ESTIMULAÇÃO CEREBRAL PROFUNDA PRECOCE VERSUS TARDIA NA DOENÇA DE PARKINSON: PROTOCOLO DE EMULAÇÃO DE ENSAIO CLÍNICO-ALVO E ESTUDO DE VIABILIDADE COM DADOS REAIS

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: especificar uma emulação de ensaio clínico-alvo para comparar estimulação cerebral profunda (DBS) precoce versus tardia em adultos com doença de Parkinson e avaliar a viabilidade da estratégia com evidências reais de ensaios e coortes. Método: o protocolo do ensaio-alvo foi definido antes da análise, incluindo elegibilidade, estratégias, tempo zero, janela de graça, desfechos, estimandos e plano de controle de confusão. A viabilidade foi examinada em estudos reais, entre eles o EARLYSTIM, o ensaio piloto de DBS em estágio inicial, coortes longitudinais, uma análise de decisão e uma coorte multicêntrica chinesa de 1.717 pacientes. Resultados: o EARLYSTIM incluiu 251 participantes com duração média da doença de 7,5 anos e mostrou melhora de qualidade de vida e controle motor em comparação ao tratamento clínico. A coorte chinesa agrupou pacientes com duração inferior a cinco anos (n=141), entre cinco e dez anos (n=978) e igual ou superior a dez anos (n=598); todos melhoraram em 24 meses, mas o grupo intermediário apresentou maior benefício motor, emocional e de qualidade de vida. Esses dados indicaram que "precoce" não deveria ser definido apenas pelo tempo desde o diagnóstico. A emulação proposta utilizará como tempo zero a primeira data em que o paciente satisfez simultaneamente critérios de levodopa-responsividade, complicações motoras clinicamente relevantes e elegibilidade neurocirúrgica. Estratégias não distinguíveis no tempo zero serão tratadas por clonagem, censura ao desvio e

ponderação. Conclusão: a pergunta sobre timing da DBS pôde ser formulada de modo causal, desde que indicação, início do seguimento e atribuição da estratégia fossem alinhados. A literatura real sustentou benefício em fases não avançadas, mas também sugeriu uma janela intermediária e reforçou a necessidade de ajustar prognóstico, cognição, sintomas axiais e resposta à levodopa.

Palavras-chave: doença de Parkinson; estimulação cerebral profunda; emulação de ensaio-alvo; inferência causal; tempo imortal; prontuários eletrônicos.

ABSTRACT

Objective: to specify a target trial emulation comparing early versus delayed deep brain stimulation (DBS) in adults with Parkinson disease and to assess feasibility using real trial and cohort evidence. Methods: the target trial protocol was specified before analysis, including eligibility, treatment strategies, time zero, grace period, outcomes, estimands, and confounding control. Feasibility was examined using the EARLYSTIM trial, the early-stage DBS pilot trial, longitudinal cohorts, a decision analysis, and a Chinese multicentre cohort of 1,717 patients. Results: EARLYSTIM enrolled 251 participants with a mean disease duration of 7.5 years and demonstrated better quality of life and motor control than medical therapy. The Chinese cohort classified patients into disease duration below five years (n=141), five to ten years (n=978), and ten years or longer (n=598); all groups improved at 24 months, but the intermediate-duration group achieved the greatest motor, mood, and quality-of-life gains. Therefore, early DBS should not be defined solely by years since diagnosis. The proposed emulation defines time zero as the first date on which levodopa responsiveness, clinically relevant motor complications, and neurosurgical eligibility coexisted. Strategies indistinguishable at time zero will be handled through cloning, censoring, and weighting. Conclusion: the timing question can be addressed causally only when eligibility, treatment assignment, and follow-up initiation are aligned. Real evidence supports DBS before very advanced disease, while also

suggesting an intermediate window and the need to account for prognosis, cognition, axial symptoms, and levodopa response.

Keywords: Parkinson disease; deep brain stimulation; target trial emulation; causal inference; immortal time; electronic health records.

1 INTRODUÇÃO

A estimulação cerebral profunda tornou-se tratamento consolidado para flutuações motoras, discinesias e tremor refratário na doença de Parkinson. Tradicionalmente, a cirurgia foi adiada até fases avançadas, quando o paciente já apresentava longa duração de doença, perda funcional e exposição prolongada a complicações da levodopa.[1-4] Essa prática foi influenciada pelo caráter invasivo do procedimento, pela necessidade de seleção criteriosa e pelo receio de efeitos cognitivos, psiquiátricos e cirúrgicos. Entretanto, esperar demais também pode reduzir a oportunidade de preservar trabalho, autonomia, participação social e qualidade de vida.

O ensaio EARLYSTIM modificou o debate ao demonstrar superioridade da estimulação subtalâmica associada ao tratamento clínico em pacientes com complicações motoras relativamente recentes. Os participantes tinham idade média próxima de 52 anos, duração média da doença de 7,5 anos e até três anos de complicações motoras.[5,6] O estudo não avaliou pessoas recém-diagnosticadas, mas mostrou que a DBS poderia ser considerada antes do estágio clássico de doença avançada. Estudos piloto e análises subsequentes levantaram a possibilidade de modificar a progressão do tremor ou preservar funcionamento social, embora sem prova de efeito neuroprotetor global.[7-9]

A pergunta "precoce ou tardia" é difícil de responder em prontuários porque o paciente precisa sobreviver, permanecer elegível e chegar à cirurgia. Se todo o período entre indicação e operação for atribuído retrospectivamente ao grupo operado, cria-se tempo imortal: durante esse intervalo o paciente, por definição, não poderia ter sofrido determinados desfechos antes de receber a exposição que o classificou.[10-12] Além disso, pacientes operados cedo diferem

daqueles operados tarde quanto a idade, gravidade, cognição, resposta à levodopa, sintomas axiais, acesso e experiência do centro.

A emulação de ensaio-alvo obriga o pesquisador a descrever o ensaio hipotético que responderia à pergunta e, em seguida, mapear cada componente para os dados observacionais.[10,13,14] O método não cria randomização nem elimina confundimento não medido, mas previne erros de desenho, sobretudo desalinhamento do tempo zero, usuários prevalentes e classificação por informação futura. O objetivo deste capítulo foi especificar uma emulação de ensaio-alvo para DBS precoce versus tardia e testar sua plausibilidade com dados publicados reais.

2 MÉTODO

Foi elaborado um protocolo de emulação de ensaio clínico-alvo conforme o framework de Hernán e Robins e as recomendações TARGET para relato transparente.[10,13-15] A especificação foi concluída antes da construção dos gráficos e da síntese de viabilidade. Como não houve acesso a prontuários individuais nesta etapa, não foram produzidas estimativas causais próprias; o produto foi um protocolo publicável acompanhado de análise de viabilidade baseada em estudos reais.

A população-alvo será composta por adultos com doença de Parkinson idiopática, diagnóstico clínico confirmado, resposta documentada à levodopa e complicações motoras clinicamente relevantes. Serão excluídos parkinsonismos atípicos, demência, psicose não controlada, contraindicação cirúrgica maior e ausência de acompanhamento longitudinal. A elegibilidade será considerada em avaliações sequenciais, permitindo que uma pessoa entrasse no primeiro momento em que todos os critérios estivessem presentes.

A estratégia precoce foi definida como implantação de DBS em até seis meses após o tempo zero, mantendo tratamento clínico otimizado durante a janela. A estratégia tardia correspondeu ao adiamento por pelo menos 24 meses,

salvo indicação de resgate por deterioração grave. A janela de graça reconheceu que cirurgia não ocorre no mesmo dia da decisão. Como as duas estratégias eram compatíveis no início da janela, cada participante seria clonado em ambas, censurado quando se desviasse e ponderado pela probabilidade inversa de permanecer aderente.[12,16]

O tempo zero foi a primeira data em que coexistiram levodopa-responsividade, complicações motoras que afetavam função ou qualidade de vida e avaliação de elegibilidade neurocirúrgica. Essa definição impediu usar a data do diagnóstico para o grupo tardio e a data da cirurgia para o grupo precoce, comparação que criaria grupos em momentos prognósticos diferentes. A data será reconstruída a partir de escalas, notas clínicas, ajustes de medicação, encaminhamento e reunião multidisciplinar.

Os desfechos primários foram incapacidade funcional e qualidade de vida em cinco anos. Desfechos secundários incluíram MDS-UPDRS, tempo em estado on sem discinesia problemática, dose equivalente de levodopa, quedas, hospitalizações, declínio cognitivo, depressão, ansiedade, institucionalização, eventos adversos cirúrgicos e mortalidade. Foram definidos pontos em 24, 60 e 120 meses, com risco cumulativo, diferença de risco, razão de riscos e tempo médio restrito de sobrevivência.

O conjunto mínimo de confundidores incluiu idade, sexo, duração da doença, idade de início, MDS-UPDRS, resposta aguda à levodopa, tempo off, discinesia, LEDD, sintomas axiais, quedas, MoCA/MMSE, depressão, ansiedade, comorbidades, alvo planejado, volume do centro e situação socioeconômica. Pesos de tratamento e censura seriam estimados por modelos flexíveis com verificação de positividade, truncamento e balanço. Análises de sensibilidade incluiriam g-formula paramétrica, controle negativo e E-value para confundimento não medido.

A viabilidade foi examinada por busca de ensaios, coortes e análises sobre timing. Foram extraídos tamanho amostral, duração da doença, idade, desenho, desfecho e magnitude de benefício. Os dados da coorte chinesa de 2025 foram

utilizados para descrever grupos de duração e melhora em 24 meses; os valores foram transcritos do artigo aberto. Resultados de escalas distintas não foram combinados em metanálise.

A Tabela 1 apresenta lado a lado o ensaio que idealmente seria randomizado e sua tradução para registros. Essa apresentação tornou visível onde a emulação preservou o protocolo e onde precisou de aproximações.

Tabela 1. Especificação do ensaio-alvo e mapeamento para dados observacionais.

Componente	Ensaio-alvo hipotético	Emulação com registros/prontuários
Elegibilidade	DP idiopática, levodopa-responsiva, complicação funcional, sem demência	algoritmo com diagnóstico, escalas, notas e avaliação multidisciplinar
Estratégia 1	DBS em até 6 meses	procedimento registrado na janela de graça
Estratégia 2	adiar DBS por ≥24 meses	ausência de implante; resgate modelado como desvio
Tempo zero	randomização após elegibilidade	primeira data em que todos os critérios coexistiram
Seguimento	até 10 anos	eventos, perda, morte ou fim da base
Estimando	ITT e per-protocolo	clonagem-censura-ponderação e g-formula
Desfechos	função, QV, quedas, cognição, morte	escalas, códigos, medicação e registros de óbito

O item mais crítico foi o tempo zero. Quando a elegibilidade não pôde ser reconstruída antes da cirurgia, o indivíduo não deveria ser incluído na análise principal, porque qualquer classificação dependeria de informação futura. Essa regra reduziu amostra, mas protegeu validade.

A Figura 1 resume a ordem causal do estudo. Elegibilidade, atribuição da estratégia e início do seguimento foram tratados como eventos simultâneos, seguidos por uma janela operacional de implementação e por desfechos definidos previamente.

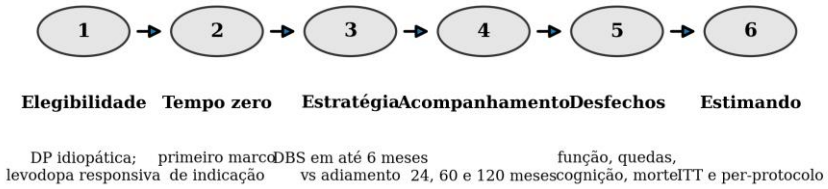


Figura 1. Relógio do ensaio-alvo para DBS precoce versus tardia.

A sequência evitou que a cirurgia fosse usada como ponto de partida apenas para os expostos. O mesmo relógio foi aplicado a ambas as estratégias, permitindo que riscos em dois, cinco e dez anos fossem comparados a partir de uma condição clínica equivalente.

A Figura 2 representa os principais caminhos de confusão. Idade, duração, cognição, resposta à levodopa, sintomas axiais e características do centro influenciaram tanto a probabilidade de cirurgia precoce quanto os desfechos.

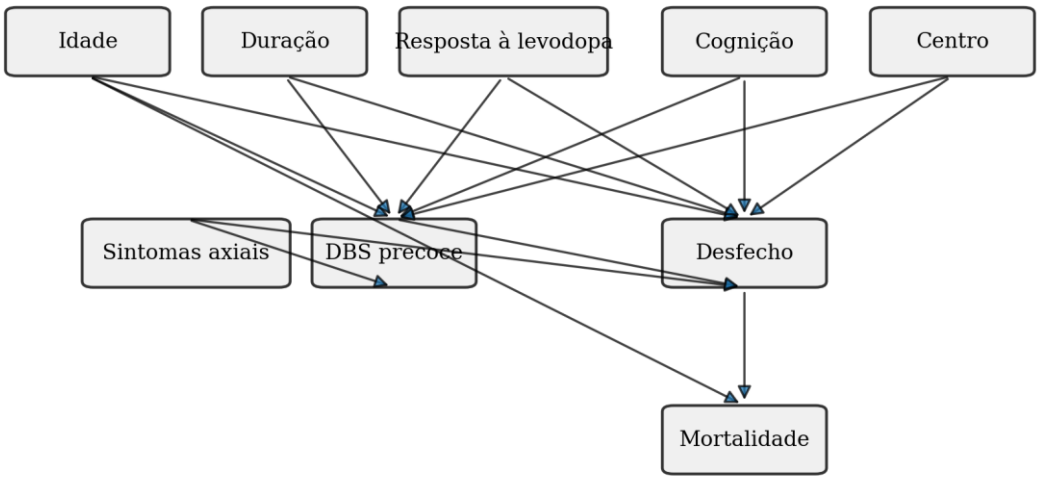


Figura 2. Grafo causal direcionado para o efeito do timing da DBS.

O grafo mostrou por que ajustar somente idade e duração seria insuficiente. Resposta à levodopa e gravidade tiveram relações diferentes conforme o estágio, e o centro afetou seleção, técnica, programação e seguimento. Variáveis pós-operatórias não foram utilizadas como confundidores basais.

3 RESULTADOS DA ANÁLISE DE VIABILIDADE

O EARLYSTIM randomizou 251 pacientes e demonstrou que estimulação subtalâmica associada ao tratamento clínico foi superior ao tratamento clínico isolado em pessoas com complicações motoras precoces. A qualidade de vida melhorou aproximadamente 26% no grupo cirúrgico e apresentou pequena piora no grupo clínico, além de benefícios em função motora, atividades diárias e complicações relacionadas à levodopa.[5] O estudo forneceu prova de conceito para antecipar a DBS, porém sua população continuou altamente selecionada e relativamente jovem.

O estudo piloto norte-americano de DBS em estágio inicial incluiu 30 participantes e avaliou segurança e sinais de efeito em fase ainda mais precoce.[7]

A análise posterior sugeriu menor progressão do tremor de repouso após washout prolongado em pacientes estimulados, resultado classificado como evidência de classe II para esse desfecho específico.[8] O achado não demonstrou modificação global da doença e precisou ser interpretado como hipótese, mas confirmou que bancos longitudinais continham marcadores úteis para emulação.

A coorte multicêntrica chinesa examinou 1.717 pacientes submetidos a STN-DBS em sete centros. Foram incluídos 141 com duração inferior a cinco anos, 978 com cinco a dez anos e 598 com dez anos ou mais. Aos 24 meses, MDS-UPDRS-III off, ansiedade, depressão e PDQ-39 melhoraram, respectivamente, 46,7%, 54,4%, 43,4% e 47,9% no conjunto.[17] O grupo de cinco a dez anos obteve maior melhora motora do que os grupos curto e longo e também superou o grupo longo em humor e qualidade de vida.

A mesma coorte revelou interação prognóstica importante. Resposta à levodopa foi fator positivo em todos os grupos, enquanto maior gravidade motora basal predisse melhor resposta nos grupos intermediário e longo, mas pior resposta no grupo curto.[17] Isso sugeriu que uma duração muito curta associada a progressão rápida poderia representar fenótipo diferente ou diagnóstico menos estável. Portanto, o timing não deveria ser modelado apenas como ponto de corte temporal.

A análise de decisão de Espay e colaboradores estimou 22,3 anos de vida ajustados por qualidade para DBS precoce e 19,8 para estratégia tardia, ganho de 2,5 QALYs.[18] Como se tratou de modelo, o resultado dependeu das probabilidades e utilidades utilizadas. Ainda assim, demonstrou que benefícios acumulados de anos funcionais poderiam compensar riscos iniciais e custos do procedimento.

Ensaios e coortes de doença avançada mostraram benefício motor robusto, mas também redução gradual de ganhos em qualidade de vida e progressão de sintomas não responsivos à dopamina. No seguimento de 36 meses do estudo VA, o tempo diário de boa função permaneceu melhor que no basal, enquanto alguns domínios cognitivos e sociais perderam parte do ganho inicial.[19] Essa

trajetória reforçou a hipótese de que operar apenas no estágio avançado pode ocorrer depois que componentes axiais e cognitivos passaram a limitar o efeito global.

A Figura 3 reúne estudos de desenhos diferentes e, por isso, não representa metanálise. O eixo horizontal mostra a duração aproximada da doença no momento da cirurgia; o eixo vertical utiliza a melhora percentual no domínio principal relatado, e o tamanho do círculo reflete a amostra.

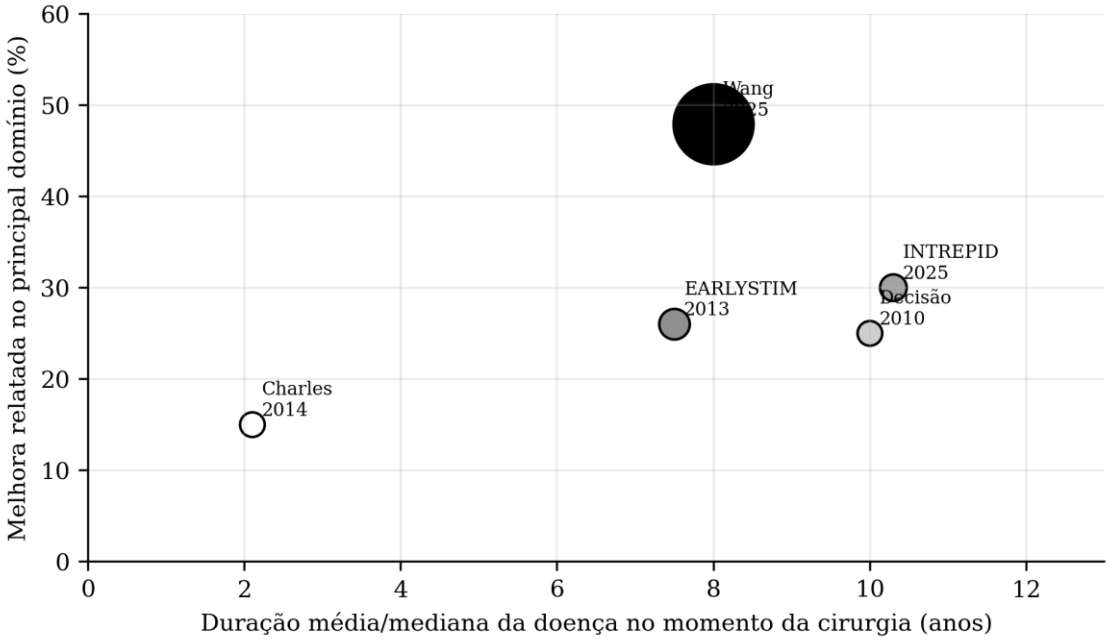


Figura 3. Relação entre duração da doença, tamanho amostral e benefício principal em estudos reais.

A distribuição ilustrou que benefício não aumentou linearmente com cirurgia cada vez mais precoce. Ensaios em torno de sete a dez anos mostraram ganhos relevantes, enquanto estudos muito precoces foram menores e exploratórios. A coorte de 2025 forneceu o maior suporte empírico para uma janela intermediária.

A Figura 4 combina número de pacientes e idade mediana nos três grupos da coorte chinesa. O grupo intermediário concentrou 57% da amostra, enquanto os pacientes com duração longa foram operados em idade mais avançada.

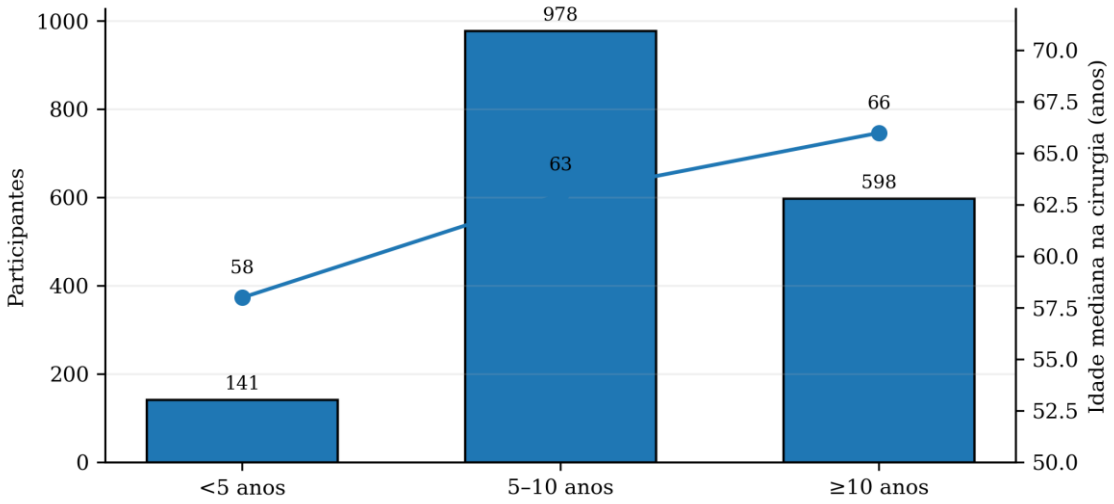


Figura 4. Composição da coorte multicêntrica chinesa por duração e idade cirúrgica.

As diferenças de idade e gravidade mostraram que uma comparação bruta entre grupos produziria confusão. A emulação precisou balancear não apenas duração, mas o conjunto de variáveis que determinou elegibilidade e prognóstico. A grande amostra intermediária também ofereceu melhor precisão para estimativas nessa janela.

A Tabela 2 sintetiza as fontes que informaram elegibilidade, definição de timing, desfechos e pressupostos. Ensaios forneceram comparação causal com terapia clínica; coortes ofereceram heterogeneidade e duração; o modelo econômico explorou consequências de longo prazo.

Tabela 2. Evidências reais utilizadas para avaliar a viabilidade do ensaio-alvo.

Fonte real	N	Timing/duração	Resultado relevante para a emulação	Limitação
EARLYSTIM, 2013	251	média 7,5 anos; complicações ≤3 anos	melhora de qualidade de vida e função vs terapia clínica	seleção jovem; aberto
Charles et al., 2014	30	estágio inicial	viabilidade e segurança	piloto pequeno

Fonte real	N	Timing/duração	Resultado relevante para a emulação	Limitação
Hacker et al., 2018	30	estágio inicial	sinal de menor progressão do tremor	análise pós-hoc
Wang et al., 2025	1.717	<5, 5–10 e ≥10 anos	maior benefício no grupo 5–10 anos	observacional
Weaver et al., 2012	299	doença avançada	benefício motor com atenuação de alguns domínios em 36 meses	não compara timing
Espay et al., 2010	modelo	precoce vs tardia	ganho estimado de 2,5 QALYs	dependente de pressupostos

Nenhuma fonte respondeu sozinha à pergunta proposta. A emulação foi necessária justamente para comparar estratégias de timing em população mais ampla, mas deveria ser calibrada contra ensaios e testar se reproduzia efeitos conhecidos em subgrupos semelhantes ao EARLYSTIM.

A Figura 5 contrasta uma classificação retrospectiva incorreta com a emulação. Na abordagem inadequada, o intervalo entre indicação e cirurgia foi atribuído ao grupo operado, embora a pessoa ainda não tivesse recebido DBS.

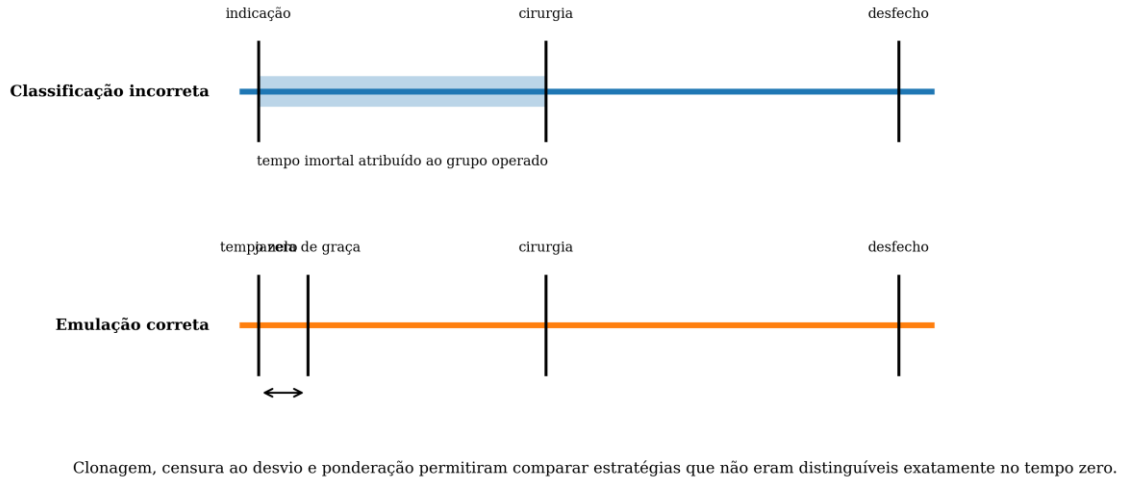


Figura 5. Representação do viés de tempo imortal e sua correção.

Na abordagem correta, o seguimento começou no mesmo marco para todos. A janela de graça foi tratada por clonagem e censura ao desvio. Assim, o paciente que não realizou cirurgia dentro do prazo deixou de contribuir para a estratégia precoce, sem que o período anterior fosse classificado com base no futuro.

A Tabela 3 especifica desfechos que puderam ser medidos de modo consistente em dados clínicos e administrativos. A escolha combinou escalas sensíveis à DBS com eventos objetivos de longo prazo.

Tabela 3. Desfechos e definições para a emulação.

Desfecho	Definição operacional	Horizonte	Fonte preferencial	Análise
Incapacidade	MDS-UPDRS II/III e atividades diárias	24, 60, 120 meses	escalas e consulta	diferença média e trajetória
Qualidade de vida	PDQ-39 summary index	24 e 60 meses	avaliação estruturada	diferença padronizada
Quedas	queda com atendimento ou registro clínico	anual	EHR e urgência	risco cumulativo
Cognição	MoCA/MMSE, diagnóstico de demência	anual	escalas e códigos	RMST sem demência

Desfecho	Definição operacional	Horizonte	Fonte preferencial	Análise
Hospitalização	qualquer internação e causa	5 anos	administrativo	taxa e tempo até evento
Mortalidade	óbito por qualquer causa	10 anos	registro de óbito	risco, HR e RMST
Segurança	hemorragia, infecção, revisão de hardware	90 dias e 2 anos	cirurgia e dispositivos	incidência cumulativa

A mortalidade foi mantida como desfecho, mas não como única medida de benefício. DBS é primariamente sintomática; portanto, função, tempo de boa mobilidade e qualidade de vida foram mais próximos da decisão do paciente. Cognição e quedas funcionaram simultaneamente como resultados e marcadores de progressão não dopaminérgica.

A Figura 6 mostra que a emulação não dependeria de um único modelo. O efeito semelhante à intenção de tratar estimaria a consequência de adotar a estratégia precoce, enquanto o per-protocolo estimaria aderência sustentada às regras de timing.

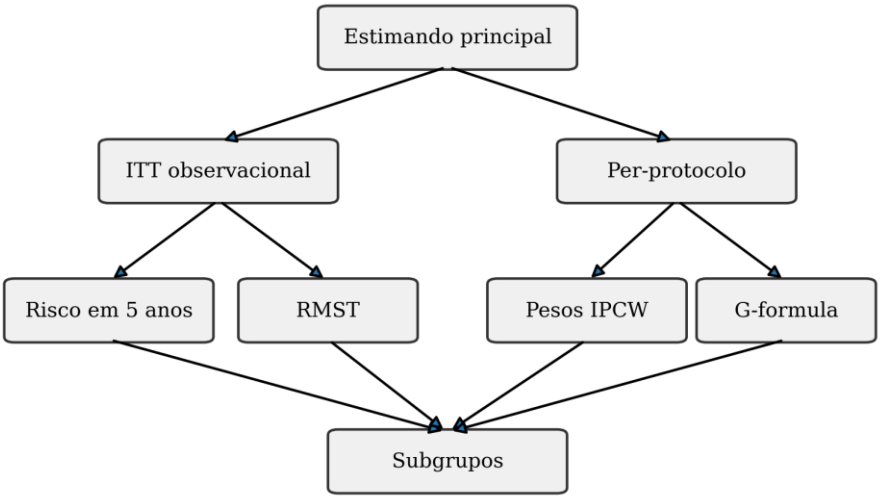


Figura 6. Árvore de estimandos e análises planejadas.

Risco em cinco anos e tempo médio restrito foram priorizados por serem interpretáveis. Pesos de censura e g-formula ofereceram estimadores complementares. Resultados só seriam considerados robustos quando direção e magnitude permanecessem coerentes entre abordagens e após avaliação de balanço e positividade.

4 DISCUSSÃO

A literatura real apoiou abandonar a dicotomia simplista entre operar muito cedo e operar apenas no estágio avançado. O EARLYSTIM mostrou benefício antes da deterioração tardia, mas a coorte chinesa sugeriu maior ganho entre cinco e dez anos. Essa aparente tensão desapareceu quando “precoce” foi entendido como fase de complicações motoras relevantes, ainda com resposta dopaminérgica e reserva funcional, e não como poucos anos desde o diagnóstico.[5,17]

A resposta à levodopa emergiu como variável central. Ela indicou que circuitos responsivos à dopamina permaneciam passíveis de modulação e foi fator positivo em todos os grupos da coorte multicêntrica. Em contraste, sintomas

axiais pouco responsivos, quedas e comprometimento cognitivo tendem a progredir apesar da DBS. Uma política de adiamento até que esses problemas dominem o quadro pode reduzir o benefício percebido mesmo quando a cirurgia melhora tremor e flutuações.

A emulação proposta tratou timing como estratégia dinâmica. Na prática, a indicação não ocorreu em um único instante: pacientes foram avaliados repetidamente, modificaram medicação, desenvolveram discinesia e foram encaminhados a centros. Ensaios sequenciais mensais ou trimestrais poderiam aumentar eficiência, permitindo entrada quando critérios fossem preenchidos. Essa abordagem exigiria cuidado para que uma mesma pessoa contribuísse para múltiplos ensaios e para que erros-padrão fossem ajustados.

Clonagem, censura e ponderação foram necessárias porque, durante a janela de seis meses, o paciente ainda poderia seguir qualquer estratégia. Censurar sem ponderar criaria seleção, pois quem desviou poderia ter piorado ou enfrentado barreiras. Os pesos modelariam a probabilidade de permanecer aderente com base na história clínica. Extremos sinalizariam baixa positividade, por exemplo quando centros nunca operassem idosos ou pacientes com menor renda.

Confundimento por indicação continuou sendo a principal ameaça. Pacientes operados cedo podem ser mais jovens, mais informados, ter acesso mais rápido e ser avaliados em centros de alto volume. Por outro lado, doença agressiva pode antecipar cirurgia e produzir pior prognóstico, gerando viés contra a estratégia precoce. O padrão observado no grupo com menos de cinco anos, em que maior gravidade basal se associou a pior resposta, ilustrou essa possibilidade.[17]

Validação negativa foi recomendada. A emulação deveria reproduzir resultados de ensaios em subamostras com critérios semelhantes ao EARLYSTIM. Se não reproduzisse direção ou magnitude aproximada, diferenças de mensuração, confusão ou definição do tempo zero precisariam ser revistas.

Essa calibração não provaria validade, mas forneceria teste empírico mais forte do que confiar apenas em ajuste estatístico.

O alvo cirúrgico e a tecnologia também precisaram ser considerados. STN e GPi diferem em redução de medicação, discinesia e perfil neuropsiquiátrico, e dispositivos adaptativos podem alterar benefício e eventos adversos. Uma análise principal poderia restringir-se a STN-DBS para comparabilidade, seguida por estudo de heterogeneidade por alvo. Misturar todos os sistemas sem modelagem poderia confundir timing com evolução tecnológica.

A carga terapêutica também precisou integrar a interpretação causal. Redução de levodopa não representa benefício isolado quando obtida à custa de apatia, piora da fala ou perda de estabilidade axial; da mesma forma, manutenção de dose elevada pode ser aceitável se houver ganho funcional sustentado. Por isso, a emulação deveria analisar conjuntamente equivalentes de levodopa, tempo em estado OFF, discinesias, atividades diárias e qualidade de vida. Essa abordagem evitaria transformar uma variável intermediária em substituto automático do que realmente importa ao paciente e permitiria reconhecer perfis de benefício distintos entre STN e GPi.

Sintomas neuropsiquiátricos exigiram atenção semelhante. Depressão, ansiedade, impulsividade e apatia influenciam indicação, adesão e percepção de resultado, além de poderem mudar depois da cirurgia. Ajustar por medidas obtidas após a estratégia produziria viés por controle de mediadores; ignorar medidas basais, por outro lado, deixaria confusão relevante. O protocolo, portanto, separou covariáveis prévias, mediadores e desfechos, e previu análises específicas para deterioração cognitiva, hospitalização psiquiátrica e qualidade de vida. A distinção foi essencial para não resumir sucesso a melhora motora.

A heterogeneidade do efeito teria implicação direta na decisão compartilhada. Idade cronológica, duração da doença e escore motor não deveriam ser usados como barreiras isoladas. A estimativa média precisaria ser acompanhada por efeitos em grupos definidos previamente, com especial cuidado para positividade e multiplicidade. Em vez de prometer uma data

universal para DBS, o estudo buscaria identificar combinações de resposta à levodopa, complicações motoras, cognição, sintomas axiais e fragilidade nas quais antecipar a cirurgia produzisse ganho absoluto clinicamente relevante.

O protocolo teve limitações. Não foram analisados prontuários individuais, e algumas variáveis importantes, como apoio familiar, preferência, fragilidade e avaliação detalhada da equipe, podem não estar estruturadas. A definição de complicação motora clinicamente relevante exigiu algoritmo validado. Mortalidade e hospitalizações foram mais fáceis de capturar do que qualidade de vida. Além disso, confundimento residual permaneceu possível mesmo com desenho correto.

Apesar dessas restrições, a emulação ofereceu avanço sobre coortes convencionais. Ao tornar explícitos o ensaio, o tempo zero, a estratégia e o estimando, impediu interpretações baseadas em grupos construídos retrospectivamente. O estudo também produziu uma pergunta clínica reconhecível: entre pacientes que já eram candidatos à DBS, qual foi o efeito de operar em até seis meses em vez de adiar por pelo menos dois anos?

A resposta final exigirá colaboração entre centros com grande volume, registros de dispositivos, bases de mortalidade e medidas repetidas. O desenho deverá ser registrado antes da análise e seguir o TARGET Statement. Até que resultados estejam disponíveis, a evidência favorece considerar DBS antes de perda funcional avançada, mas não sustenta indicação indiscriminada em doença muito precoce ou sem complicações levodopa-responsivas.

5 CONCLUSÃO

A comparação causal entre DBS precoce e tardia dependeu mais da definição do tempo zero do que de modelos sofisticados. A literatura demonstrou benefício em pacientes com complicações motoras relativamente recentes e identificou maior resposta, em uma grande coorte, entre cinco e dez anos de doença. O protocolo proposto alinhou elegibilidade, estratégia e seguimento,

utilizou janela de graça e planejou clonagem-censura-ponderação. A emulação deverá ajustar resposta à levodopa, cognição, sintomas axiais, gravidade, idade e centro. Seu resultado poderá informar uma janela clínica de oportunidade, sem confundir antecipação apropriada com cirurgia indiscriminadamente precoce.

REFERÊNCIAS

1. Deuschl G, Schade-Brittinger C, Krack P, Volkmann J, Schäfer H, Bötzel K, et al. A randomized trial of deep-brain stimulation for Parkinson disease. N Engl J Med. 2006;355(9):896-908. doi:10.1056/NEJMoa060281.

2. Weaver FM, Follett K, Stern M, Hur K, Harris C, Marks WJ Jr, et al. Bilateral deep brain stimulation vs best medical therapy for patients with advanced Parkinson disease. JAMA. 2009;301(1):63-73.

3. Williams A, Gill S, Varma T, Jenkinson C, Quinn N, Mitchell R, et al. Deep brain stimulation plus best medical therapy versus best medical therapy alone for advanced Parkinson disease: PD SURG trial. Lancet Neurol. 2010;9(6):581-591.

4. Hariz M, Blomstedt P. Deep brain stimulation for Parkinson disease. J Intern Med. 2022;292(5):764-778.

5. Schuepbach WMM, Rau J, Knudsen K, Volkmann J, Krack P, Timmermann L, et al. Neurostimulation for Parkinson disease with early motor complications. N Engl J Med. 2013;368(7):610-622. doi:10.1056/NEJMoa1205158.

6. Deuschl G, Schüpbach M, Knudsen K, Pinsker MO, Cornu P, Rau J, et al. Stimulation of the subthalamic nucleus at an earlier disease stage: concept and standards of the EARLYSTIM study. Parkinsonism Relat Disord. 2013;19(1):56-61.

7. Charles D, Konrad PE, Neimat JS, Molinari AL, Tramontana MG, Finder SG, et al. Subthalamic nucleus deep brain stimulation in early stage Parkinson disease. Parkinsonism Relat Disord. 2014;20(7):731-737.

8. Hacker ML, Turchan M, Heusinkveld LE, Currie AD, Millan SH, Molinari AL, et al. Effects of deep brain stimulation on rest tremor progression in early stage Parkinson disease. Neurology. 2018;91(5):e463-e471.

9. Stoker V, Krack P, Tonder L, Barnett G, Durand-Zaleski I, Schnitzler A, et al. Deep brain stimulation impact on social and occupational functioning in

Parkinson disease with early motor complications. Mov Disord Clin Pract. 2020;7(6):672-680.

10. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758-764.

11. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries. J Clin Epidemiol. 2016;79:70-75.

12. Maringe C, Benitez Majano S, Exarchakou A, Smith M, Rachet B, Belot A, et al. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Int J Epidemiol. 2020;49(5):1719-1726.

13. Matthews AA, Danaei G, Islam N, Kurth T. Target trial emulation: applying principles of randomised trials to observational studies. BMJ. 2022;378:e071108.

14. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25:1601-1606.

15. TARGET Working Group. Transparent reporting of observational studies emulating a target trial: the TARGET Statement. BMJ. 2025;390:e087179.

16. Cotter D, et al. Starting right: aligning eligibility and treatment assignment at time zero when emulating a target trial. BMJ. 2026;392:e084909.

17. Wang S, Hu W, Gao Y, Wang A, Chen L, Liang Z, et al. Deep brain stimulation surgical timing, outcomes, and prognostic factors in patients with Parkinson disease: a Chinese retrospective multicenter cohort study. PLoS Med. 2025;22(8):e1004670. doi:10.1371/journal.pmed.1004670.

18. Espay AJ, Vaughan JE, Marras C, Fowler R, Eckman MH. Early versus delayed bilateral subthalamic deep brain stimulation for Parkinson disease: a decision analysis. Mov Disord. 2010;25(10):1456-1463.

19. Weaver FM, Follett KA, Stern M, Luo P, Harris CL, Hur K, et al. Randomized trial of deep brain stimulation for Parkinson disease: thirty-six-month outcomes. Neurology. 2012;79(1):55-65.

20. Hitti FL, Ramayya AG, McShane BJ, Yang AI, Vaughan KA, Baltuch GH. Long-term outcomes following deep brain stimulation for Parkinson disease. J Neurosurg. 2019;132(1):205-210.

21. Mestre TA, Espay AJ, Marras C, Eckman MH, Pollak P, Lang AE. Subthalamic nucleus-deep brain stimulation for early motor complications in Parkinson disease: the EARLYSTIM trial. Mov Disord. 2014;29(14):1753-1760.

22. Hartmann CJ, Fliegen S, Groiss SJ, Wojtecki L, Schnitzler A. An update on best practice of deep brain stimulation in Parkinson disease. Ther Adv Neurol Disord. 2019;12:1756286419838096.

23. Rughani A, Schwalb JM, Sidiropoulos C, Pilitsis J, Ramirez-Zamora A, Sweet JA, et al. CNS systematic review and evidence-based guideline on STN and GPi DBS for Parkinson disease. Neurosurgery. 2018;82(6):753-756.

24. Zoon TJC, van Rooijen G, Balm GMFC, Bergfeld IO, Daams JG, Krack P, et al. Apathy induced by subthalamic nucleus deep brain stimulation in Parkinson disease: a meta-analysis. Mov Disord. 2021;36(2):317-326.

25. Limousin P, Foltynie T. Long-term outcomes of deep brain stimulation in Parkinson disease. Nat Rev Neurol. 2019;15(4):234-242.

26. Deuschl G, Follett KA, Luo P, Rau J, Weaver FM, Paschen S, et al. Comparing two randomized deep brain stimulation trials for Parkinson disease. J Neurosurg. 2020;132(5):1376-1384.

CAPÍTULO 16

OBESIDADE SARCOPÊNICA NO PERIOPERATÓRIO DE PESSOAS IDOSAS: SÍNTESE CRÍTICA, MAPA DE EVIDÊNCIAS E PROPOSTA DE ESTRATIFICAÇÃO CLÍNICA

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Maria Isabel Pires Teixeira; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

A obesidade sarcopênica combina adiposidade excessiva com redução de força, quantidade ou qualidade muscular e constitui um fenótipo de risco frequentemente subdiagnosticado no perioperatório. Em pessoas idosas, o índice de massa corporal pode ocultar baixa reserva funcional, inflamação crônica, infiltração gordurosa do músculo e menor capacidade de responder ao estresse cirúrgico. Este capítulo apresenta uma síntese crítica e sistematizada das evidências sobre definição, diagnóstico, mecanismos e desfechos pós-operatórios, com ênfase em cirurgias abdominais e oncológicas. Foram priorizados consensos internacionais, revisões sistemáticas, meta-análises, coortes e ensaios de pré-habilitação publicados em bases indexadas até agosto de 2026. Não foi realizada nova meta-análise; as estimativas quantitativas foram extraídas de estudos publicados e organizadas em mapa comparativo. A literatura mostra associação consistente entre obesidade sarcopênica e maior ocorrência de complicações totais e graves, infecções, fístulas, mortalidade e pior sobrevida, embora a magnitude varie conforme definição, população e método de mensuração. A avaliação clinicamente útil deve integrar triagem funcional, força de preensão ou teste de levantar da cadeira e mensuração de composição corporal por absorciometria, bioimpedância ou tomografia oportunística. Propõe-se um fluxo perioperatório baseado em identificação precoce, classificação de gravidade, otimização nutricional, treinamento resistido e

aeróbico, controle de comorbidades e monitoramento pós-operatório. A principal implicação é substituir a avaliação centrada exclusivamente no peso por uma análise da capacidade biológica para tolerar a cirurgia.

Palavras-chave: obesidade sarcopênica; sarcopenia; pessoa idosa; complicações pós-operatórias; composição corporal; pré-habilitação.

1 INTRODUÇÃO

A avaliação pré-operatória de pessoas idosas ainda é frequentemente organizada ao redor de diagnósticos isolados, idade cronológica, índice de massa corporal e classificação anestésica. Esses marcadores são úteis, mas não descrevem integralmente a reserva fisiológica que sustentará a resposta à anestesia, à incisão, à inflamação, ao jejum, à imobilidade e ao catabolismo pós-operatório. Duas pessoas com a mesma idade e o mesmo índice de massa corporal podem apresentar capacidades funcionais radicalmente diferentes. Uma pode conservar força, massa muscular e autonomia; outra pode ter musculatura reduzida, infiltração gordurosa e baixa tolerância ao esforço, embora o excesso de tecido adiposo impeça que a perda muscular seja percebida ao exame visual. É nesse intervalo entre aparência corporal e capacidade biológica que a obesidade sarcopênica se torna particularmente relevante para a cirurgia.

A sarcopenia é atualmente compreendida como doença muscular caracterizada principalmente por baixa força, com confirmação por redução da quantidade ou qualidade muscular e classificação de gravidade pela piora do desempenho físico. O consenso EWGSOP2 deslocou o foco exclusivo da massa muscular para a força e a função, reconhecendo que o músculo é um órgão metabólico, locomotor e imunológico. A obesidade sarcopênica, por sua vez, corresponde à coexistência de excesso de adiposidade com baixa massa e função muscular. O consenso conjunto da European Society for Clinical Nutrition and Metabolism e da European Association for the Study of Obesity propôs um processo diagnóstico em duas etapas: primeiro avaliar função muscular e, em seguida, confirmar simultaneamente adiposidade elevada e baixa quantidade

muscular. Essa abordagem evita que o diagnóstico seja reduzido a uma simples razão entre gordura e músculo (Cruz-Jentoft *et al.*, 2019; Donini *et al.*, 2022).

No perioperatório, a combinação entre adiposidade visceral e baixa reserva muscular produz um risco que não é meramente aditivo. O tecido adiposo disfuncional favorece resistência à insulina, inflamação de baixo grau, disfunção endotelial, alteração ventilatória e maior dificuldade técnica. A baixa massa e a baixa qualidade muscular reduzem a reserva de aminoácidos, a capacidade de mobilização, a tosse eficaz, a estabilidade postural e a resposta ao estresse. Em uma cirurgia de grande porte, esses mecanismos convergem para maior vulnerabilidade a infecção, pneumonia, fístula, deiscência, delirium, perda funcional, permanência prolongada e necessidade de cuidados após a alta. A pessoa pode atravessar o procedimento sem complicação imediata e, ainda assim, perder autonomia de maneira clinicamente significativa.

Apesar disso, a obesidade sarcopênica permanece pouco incorporada à rotina. Há barreiras conceituais, porque diferentes estudos utilizam pontos de corte diversos; barreiras operacionais, porque força de preensão, velocidade de marcha e análise de tomografia nem sempre fazem parte do fluxo; e barreiras terapêuticas, porque reduzir peso de forma rápida antes da cirurgia pode agravar a perda muscular. A meta não é simplesmente emagrecer o paciente, mas melhorar a relação entre capacidade funcional, composição corporal e demanda cirúrgica. Essa distinção é central: uma intervenção que reduz gordura e preserva ou aumenta músculo pode ser benéfica; uma restrição energética agressiva que reduz ambos pode piorar a tolerância ao procedimento.

Este capítulo tem como objetivo analisar criticamente o papel da obesidade sarcopênica no risco perioperatório de pessoas idosas, sintetizar os principais resultados quantitativos publicados, discutir limitações diagnósticas e propor um modelo de estratificação e cuidado aplicável à prática. O texto também examina a tomografia computadorizada oportunística, a automação por inteligência artificial, a pré-habilitação multimodal e os indicadores necessários para transformar o diagnóstico em ação clínica. A intenção não é apresentar uma

fórmula universal, mas construir uma ponte entre evidência, decisão cirúrgica e organização do cuidado.

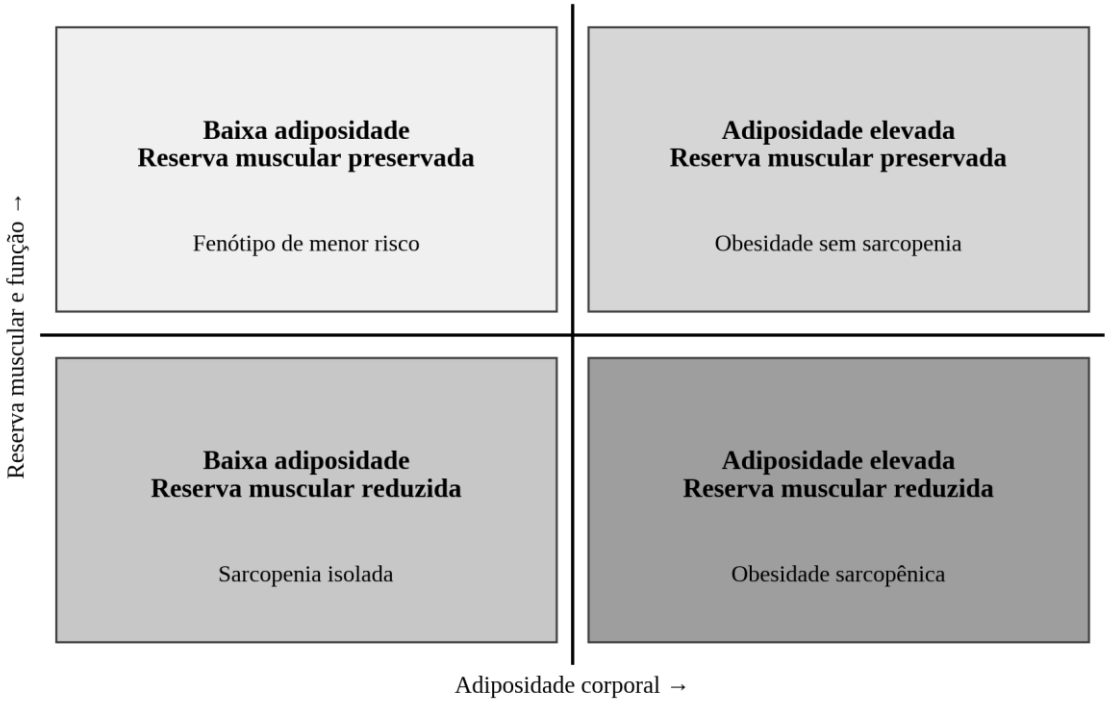


Figura 1 – Matriz fenotípica baseada na interação entre adiposidade e reserva muscular

Fonte: Elaboração própria, com base em Donini *et al.* (2022) e Cruz-Jentoft *et al.* (2019).

2 PERCURSO METODOLÓGICO DA SÍNTESE

Foi realizada uma síntese crítica de natureza sistematizada, orientada por uma pergunta de prognóstico e implementação: em adultos mais velhos submetidos a cirurgia, como a coexistência de adiposidade elevada e baixa reserva muscular se relaciona com complicações, mortalidade, recuperação funcional e organização do cuidado perioperatório? Foram priorizadas publicações indexadas em PubMed/MEDLINE, complementadas por rastreamento de referências e consulta a documentos de sociedades científicas. A busca foi atualizada em 1º de agosto de 2026 e combinou termos relacionados a sarcopenic obesity, sarcopenia, body composition, older adults, surgery,

postoperative complications, mortality, prehabilitation, nutrition e resistance exercise.

Foram valorizados consensos internacionais de definição, revisões sistemáticas, meta-análises, estudos de coorte com avaliação de composição corporal e ensaios clínicos de intervenções perioperatórias. As evidências foram organizadas em cinco domínios: definição e diagnóstico; mecanismos; associação com desfechos; intervenções; e implementação. Estudos centrados apenas em índice de massa corporal, sem avaliação muscular, foram usados apenas para contextualização. Publicações que utilizaram baixa massa muscular como substituto de sarcopenia foram interpretadas com cautela, porque os consensos atuais exigem componente funcional para caracterização completa.

Não foi conduzida nova meta-análise. As razões de chances, razões de risco, hazard ratios e intervalos de confiança apresentados em figuras e tabelas foram transcritos de meta-análises e estudos originais, preservando o tipo de medida de cada publicação. Assim, o mapa de efeitos serve para comparar direção, magnitude e precisão, mas não constitui estimativa combinada entre revisões. Essa escolha evita duplicação de participantes, mistura de definições e falsa precisão. Também foi feita distinção entre evidência de sarcopenia isolada e evidência específica de obesidade sarcopênica.

3 CONCEITO, DIAGNÓSTICO E ARMADILHAS DE MENSURAÇÃO

3.1 Da massa muscular à função

A primeira armadilha é assumir que toda baixa massa muscular equivale a sarcopenia. Em estudos retrospectivos, especialmente oncológicos, a tomografia pré-operatória é frequentemente o único recurso disponível. Nesses casos, o índice de músculo esquelético na terceira vértebra lombar é usado como marcador de depleção muscular. Esse método tem grande valor prognóstico, mas não mede força nem desempenho. Pelo EWGSOP2, a baixa força torna a sarcopenia provável; a baixa quantidade ou qualidade muscular confirma; e o

baixo desempenho físico caracteriza forma grave. Portanto, um laudo de baixa área muscular deve ser interpretado como sinal de alto risco e não necessariamente como diagnóstico completo.

A força de preensão manual é um dos testes mais acessíveis, rápidos e reprodutíveis. O teste de levantar da cadeira pode substituí-la quando há limitação de membros superiores. Velocidade de marcha, Short Physical Performance Battery e Timed Up and Go ajudam a quantificar gravidade e risco funcional, mas podem ser inviáveis em dor intensa, urgência ou incapacidade de deambular. A seleção do teste deve respeitar o contexto. O importante é que ao menos uma medida de função esteja presente, porque a força traduz de maneira mais direta a capacidade do sistema neuromuscular de responder a demandas.

3.2 Instrumentos e pontos de corte

A inexistência de um único ponto de corte universal é a principal fonte de heterogeneidade. Força de preensão varia por sexo, idade e equipamento; massa muscular varia conforme método e ajuste; adiposidade pode ser definida por índice de massa corporal, circunferência da cintura, massa gorda, área de gordura visceral ou combinação. Em população asiática, limiares de índice de massa corporal e gordura visceral diferem daqueles usados em populações ocidentais. Em pacientes com câncer, edema, ascite, desidratação e inflamação alteram bioimpedância e peso. Por isso, o diagnóstico deve ser interpretado como construção multicomponente e não como resultado isolado de um número.

Tabela 1 – Componentes recomendados para identificação de obesidade sarcopênica no perioperatório

Fonte: Elaboração própria com base nos consensos EWGSOP2 e ESPEN/EASO.

Componente	Exemplos de medida	Papel clínico	Limitações principais
Risco inicial	SARC-F, perda funcional, quedas, redução de atividade, doença crônica	Selecionar quem precisa de avaliação completa	Baixa sensibilidade isolada; depende do relato

Componente	Exemplos de medida	Papel clínico	Limitações principais
Função muscular	Força de preensão; teste de levantar da cadeira	Identificar baixa força e sarcopenia provável	Dor, artrite, déficit neurológico e técnica influenciam
Quantidade muscular	DXA, BIA, área muscular em L3 por TC	Confirmar baixa massa muscular	Pontos de corte e ajustes variam
Qualidade muscular	Atenuação em unidades Hounsfield; infiltração gordurosa	Detectar mioesteatose e risco não explicado pela área	Ausência de padronização universal
Adiposidade	Cintura, percentual de gordura, massa gorda, VFA	Confirmar excesso adiposo	IMC pode ocultar distribuição e composição
Desempenho físico	Velocidade de marcha, SPPB, Timed Up and Go	Classificar gravidade e planejar recuperação	Pode ser inviável em urgência ou dor intensa

Nota: BIA = bioimpedância elétrica; DXA = absorciometria por dupla energia; SPPB = Short Physical Performance Battery; TC = tomografia computadorizada; VFA = área de gordura visceral.

3.3 Tomografia oportunística e inteligência artificial

A tomografia abdominal realizada para diagnóstico ou estadiamento oferece uma oportunidade singular: medir músculo, gordura visceral, gordura subcutânea e atenuação muscular sem novo exame. Na terceira vértebra lombar, a área muscular é segmentada e ajustada pela altura para formar o índice de músculo esquelético. A atenuação média em unidades Hounsfield informa infiltração lipídica e qualidade. A gordura visceral acrescenta informação que o índice de massa corporal não fornece. Em pancreatoduodenectomia, por exemplo, uma coorte de 630 pacientes classificou obesidade por área de gordura visceral superior a 100 cm² e sarcopenia por índice muscular específico por sexo, identificando 18,6% com obesidade sarcopênica (Jeong *et al.*, 2024).

O obstáculo histórico foi o tempo necessário para segmentação manual. Ferramentas de inteligência artificial e aprendizado profundo conseguem automatizar a delimitação de músculo e gordura, aproximando a composição

corporal do laudo rotineiro. Revisões sobre tomografia e inteligência artificial mostram desempenho tecnicamente promissor, mas a implementação clínica exige validação externa, controle de qualidade, transparência do algoritmo e integração ao prontuário. Um valor automatizado não elimina a necessidade de contexto: o sistema pode medir a área com precisão e ainda assim classificar incorretamente se o ponto de corte não for adequado à população.

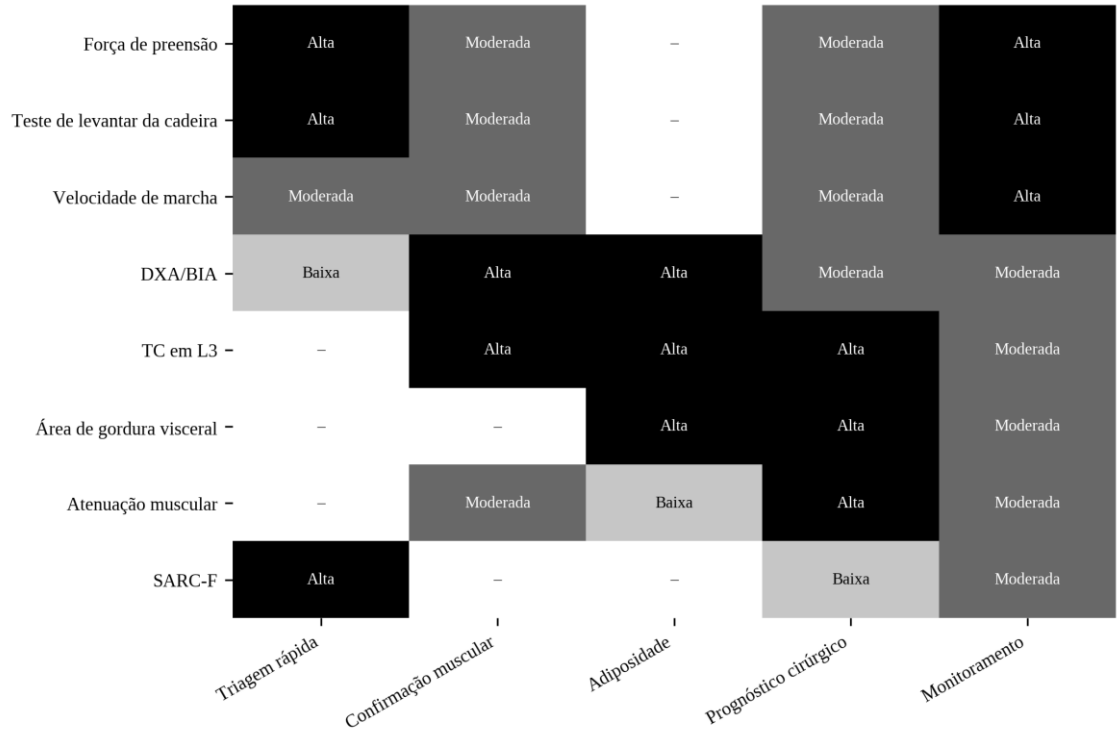


Figura 2 – Matriz de utilidade clínica dos principais instrumentos de avaliação

Fonte: Elaboração própria. A intensidade representa utilidade relativa, não acurácia comparativa formal.

4 MECANISMOS QUE CONECTAM COMPOSIÇÃO CORPORAL E DESFECHOS

A obesidade sarcopênica deve ser entendida como interação entre sistemas, não como soma de dois diagnósticos. A adiposidade visceral secreta adipocinas e mediadores inflamatórios, favorece resistência à insulina e altera a resposta imune. O músculo reduzido ou infiltrado por gordura produz menos

mioquinas e oferece menor reserva de aminoácidos para síntese de proteínas de fase aguda, cicatrização e resposta imunológica. Quando a cirurgia inicia um estado hipercatabólico, a pessoa com pouca reserva muscular atravessa mais rapidamente o limiar de incapacidade funcional.

A gordura visceral também aumenta a complexidade técnica. Campos operatórios profundos, menor mobilidade dos tecidos e dificuldade de visualização podem prolongar operação e aumentar sangramento. Na pancreatoduodenectomia, pacientes com obesidade sarcopênica apresentaram tempo operatório e perda sanguínea maiores que o grupo de referência. A obesidade isolada contribuiu para fístula pancreática clinicamente relevante, enquanto a sarcopenia isolada se associou a complicações maiores; a combinação captou as duas dimensões de risco (Jeong *et al.*, 2024).

Esses mecanismos explicam por que duas pessoas com obesidade podem ter prognósticos diferentes. A presença de força e massa preservadas fornece reserva; sua ausência transforma adiposidade em fardo proporcionalmente maior. Também explica por que intervenções exclusivamente voltadas à perda de peso são insuficientes. A meta perioperatória deve ser reduzir risco metabólico sem aprofundar déficit muscular.

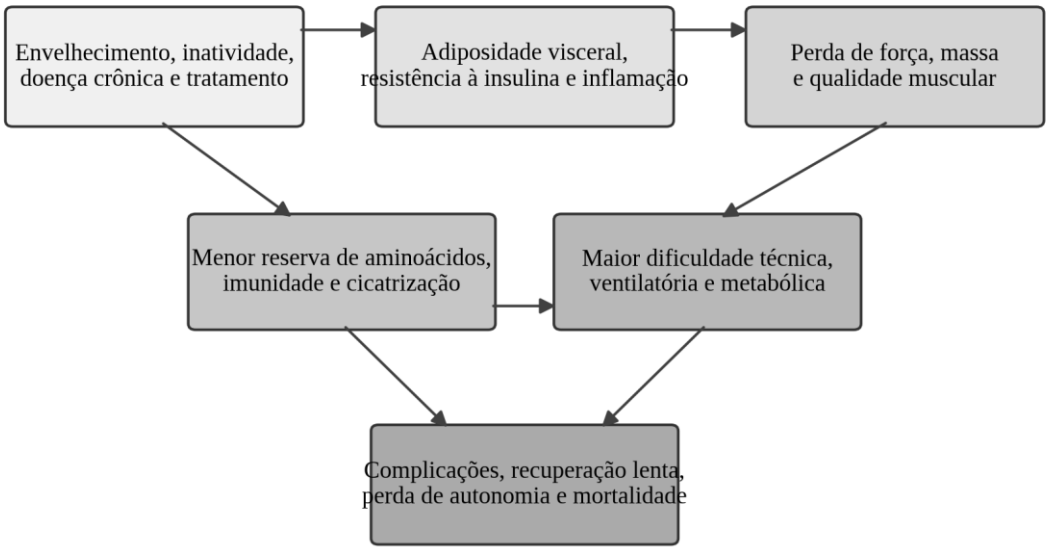


Figura 3 – Cascata fisiopatológica da obesidade sarcopênica no contexto cirúrgico

Fonte: Elaboração própria a partir da síntese das evidências.

5 MAPA DE EVIDÊNCIAS SOBRE DESFECHOS CIRÚRGICOS

5.1 Sínteses amplas e cirurgia abdominal

As evidências mais abrangentes confirmam que a baixa reserva muscular tem valor prognóstico em múltiplas especialidades. A meta-análise de Knoedler *et al.* (2023) reuniu 294 estudos e 97.643 pacientes, dos quais 33.070 apresentavam sarcopenia. A condição se associou a maior mortalidade, mais complicações, maior permanência, menor alta para o domicílio e pior sobrevida em um, três e cinco anos. A consistência foi mantida em análises de oncologia, cirurgia cardiovascular, emergência e transplante. Essa base não é específica para obesidade sarcopênica, mas demonstra que o componente muscular possui relevância transversal.

Quando são exigidos critérios consensuais de sarcopenia, a associação permanece. Park *et al.* (2023) incluíram 20 estudos e 5.421 pacientes submetidos a cirurgia abdominal eletiva. A prevalência combinada foi de aproximadamente 20%, e a sarcopenia aumentou o risco ajustado de complicações pós-operatórias em 56%. Também houve aumento expressivo de mortalidade hospitalar e em 30 e 90 dias. O dado é importante porque reduz parte da heterogeneidade causada por definições exclusivamente radiológicas.

5.2 Obesidade sarcopênica em oncologia gastrointestinal

A evidência específica de obesidade sarcopênica é mais concentrada em oncologia gastrointestinal. Wang *et al.* (2022) reuniram 26 estudos e 8.729 participantes, com idade média agrupada de 65,6 anos. A obesidade sarcopênica aumentou complicações totais em 30% e complicações maiores em mais de duas vezes. Também se associou a pior sobrevida global e livre de doença. A heterogeneidade foi baixa para complicações totais e sobrevida global, mas alta para complicações maiores, refletindo diferenças de definição, cirurgia e perfil tumoral.

Atualizações por sítio anatômico mantiveram a direção. Em câncer colorretal, meta-análise de 20 estudos e 11.264 pacientes encontrou prevalência de 14,5% e aumento de complicações totais e maiores, embora a associação com sobrevida global não tenha alcançado significância no conjunto principal (Yıldız Kopuz *et al.*, 2025). Em tumores do trato gastrointestinal superior, a obesidade sarcopênica duplicou aproximadamente o risco de complicações graves. Em câncer esofagogástrico, revisão de 2025 encontrou maior risco de complicações e pior sobrevida, porém com heterogeneidade substancial.

No câncer gástrico, uma meta-análise estimou prevalência de 16,3% e associações com complicações significativas, complicações graves, infecção de sítio cirúrgico, mortalidade e pior sobrevida. A estimativa para mortalidade apresentou intervalo amplo, indicando menor precisão. Em uma coorte de 155

pacientes submetidos a gastrectomia, 16,8% tinham obesidade sarcopênica; complicações graves ocorreram em 30,8% desse grupo e em 10,9% dos demais, com odds ratio ajustada de 3,95 (Yamagishi *et al.*, 2023).

Em pancreatoduodenectomia, Jeong *et al.* (2024) compararam quatro fenótipos corporais. A obesidade sarcopênica foi independente para complicações maiores, com odds ratio de 1,97, e para fístula pancreática clinicamente relevante, com odds ratio de 2,31. A taxa de complicações maiores foi 23,1% no fenótipo combinado e 11,4% no grupo sem sarcopenia e sem obesidade. Esses resultados sugerem que a composição corporal não apenas prediz se uma complicação ocorrerá, mas também a capacidade de evitar sua progressão para forma clinicamente relevante.

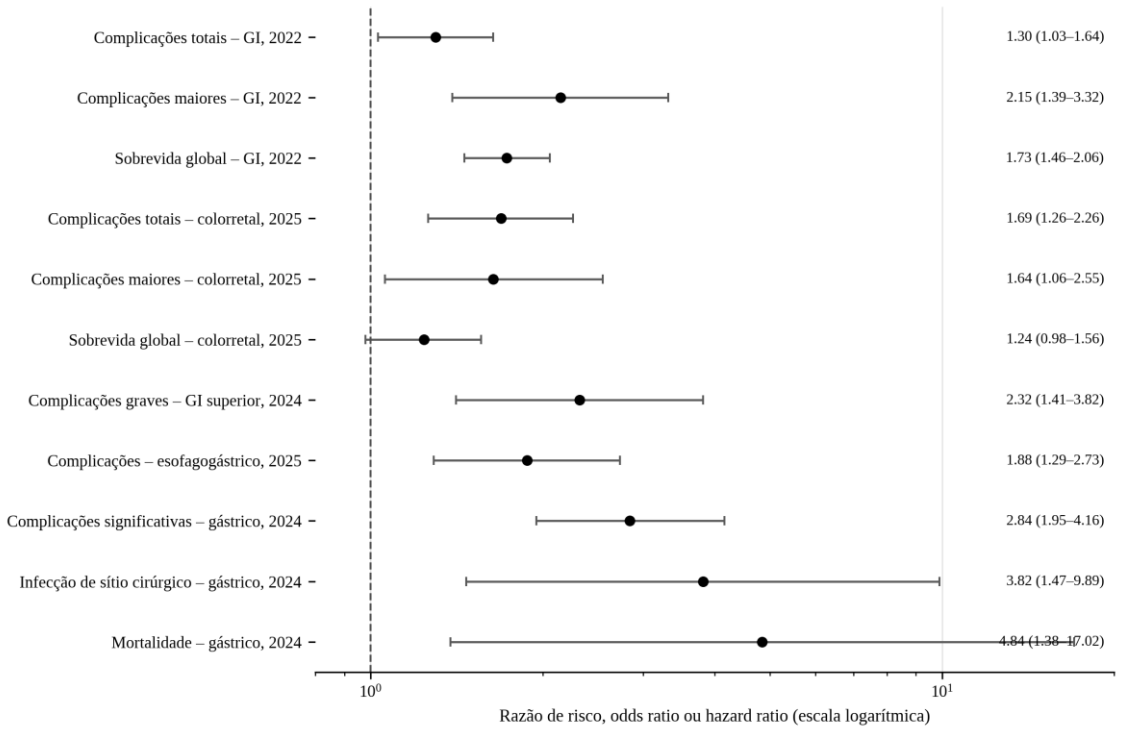


Figura 4 – Mapa comparativo de estimativas publicadas sobre obesidade sarcopênica e desfechos

Fonte: Elaboração própria com dados de Wang *et al.* (2022), Yıldız Kopuz *et al.* (2025), Matsui *et al.* (2025), revisões de tumores gastrointestinais superiores e meta-análise em câncer gástrico. Não constitui nova meta-análise.

Tabela 2 – Síntese quantitativa das principais evidências

Fonte: Dados transcritos das publicações originais; medidas não devem ser combinadas diretamente.

Fonte/população	Amostra	Desfecho	Estimativa publicada	Leitura clínica
Wang et al. (2022), câncer GI	26 estudos; 8.729	Complicações totais	OR 1,30 (1,03–1,64)	Aumento modesto e consistente
Wang et al. (2022), câncer GI	15 estudos; 3.952	Complicações maiores	OR 2,15 (1,39–3,32)	Risco aproximadamente duplicado
Wang et al. (2022), câncer GI	11 estudos; 4.197	Sobrevida global	HR 1,73 (1,46–2,06)	Pior prognóstico de longo prazo
Yıldız Kopuz et al. (2025), colorretal	20 estudos; 11.264	Complicações totais	OR 1,69 (1,26–2,26)	Associação clinicamente relevante
Yıldız Kopuz et al. (2025), colorretal	20 estudos; 11.264	Sobrevida global	HR 1,24 (0,98–1,56)	Incerteza; intervalo inclui nulidade
Matsui et al. (2025), esofagogástrico	Revisão sistemática	Complicações	RR 1,88 (1,29–2,73)	Maior risco, com heterogeneidade
Yamagishi et al. (2023), gastrectomia	155 pacientes	Complicações graves	OR ajustada 3,95 (1,39–11,20)	Efeito grande, porém impreciso
Jeong et al. (2024), pancreatoduodenectomia	630 pacientes	Complicações maiores	OR ajustada 1,97 (1,19–3,26)	Preditor independente
Jeong et al. (2024), pancreatoduodenectomia	630 pacientes	Fístula clinicamente relevante	OR ajustada 2,31 (1,34–4,00)	Aproximadamente duplica o risco

Nota: GI = gastrointestinal; HR = hazard ratio; OR = odds ratio; RR = razão de risco. Intervalos de confiança de 95%.

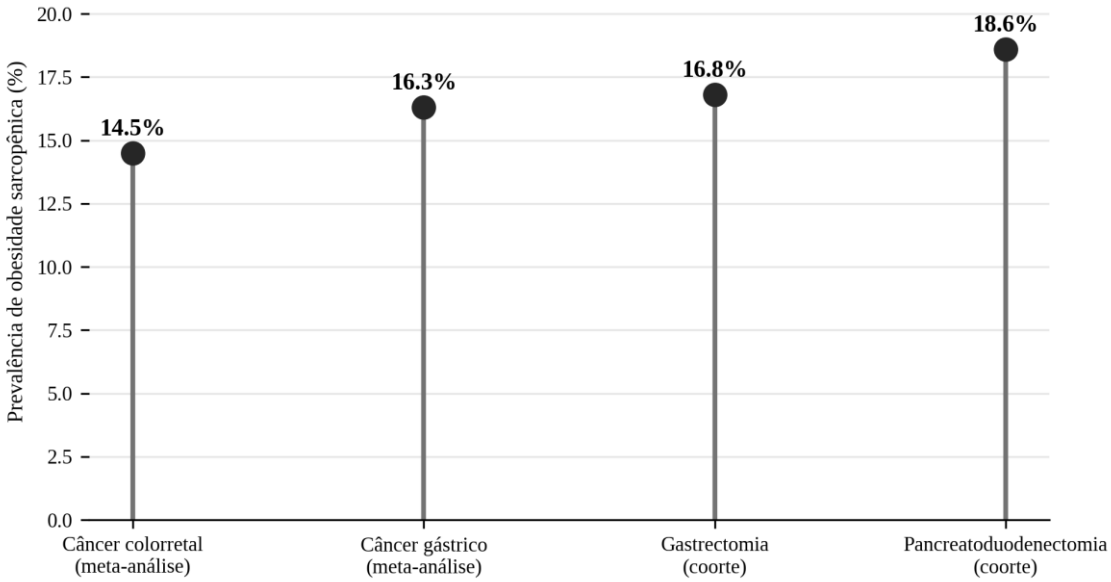


Figura 5 – Prevalência de obesidade sarcopênica em contextos cirúrgicos selecionados

Fonte: Elaboração própria com dados publicados. As populações e definições diferem entre estudos.

5.3 Desfechos funcionais e centrados no paciente

A literatura cirúrgica tradicional privilegia mortalidade e complicações classificadas por Clavien-Dindo, mas a pessoa idosa pode considerar igualmente importante retornar para casa, caminhar, alimentar-se, realizar autocuidado e retomar papéis sociais. A meta-análise ampla de sarcopenia mostrou menor probabilidade de alta domiciliar. Outros estudos relacionam baixa reserva muscular a recuperação mais lenta, maior necessidade de reabilitação e institucionalização. Esses desfechos devem ser incluídos em estudos futuros de obesidade sarcopênica.

A composição corporal pode ainda modificar tolerância a tratamentos oncológicos posteriores. Menor reserva muscular aumenta toxicidade de quimioterapia e pode atrasar tratamento adjuvante. A pior sobrevida observada em algumas meta-análises pode refletir tanto biologia tumoral e inflamação

quanto menor capacidade de completar tratamento. O fenótipo, portanto, conecta cirurgia, oncologia e reabilitação.

6 ESTRATIFICAÇÃO PRÉ-OPERATÓRIA

A triagem deve ser simples o suficiente para alcançar todos os pacientes de risco. Pessoas com 65 anos ou mais, obesidade central, doença oncológica, perda funcional, quedas, baixa atividade, perda de peso não intencional, insuficiência cardíaca, doença renal, diabetes, doença pulmonar ou internação recente merecem investigação. O SARC-F pode iniciar o processo, mas não deve ser a única ferramenta por sua sensibilidade limitada. A observação de dificuldade para levantar da cadeira ou caminhar até o consultório já é um sinal clínico.

A segunda etapa é medir força. A dinamometria deve seguir posicionamento padronizado, melhor de tentativas e registro do lado avaliado. Quando indisponível, o teste de cinco levantamentos da cadeira fornece informação funcional e pode ser repetido. A terceira etapa é quantificar composição corporal. DXA é referência para massa apendicular; bioimpedância é acessível, mas sensível a hidratação; tomografia é especialmente útil quando já existe por indicação clínica. O método deve ser registrado junto com o ponto de corte.

A estratificação proposta considera três eixos: vulnerabilidade biológica, magnitude da cirurgia e tempo disponível. Uma pessoa com baixa força, tomografia mostrando baixa área muscular e gordura visceral elevada, programada para operação de grande porte em quatro semanas, possui alta prioridade para pré-habilitação. Em emergência, a janela prévia não existe, mas o diagnóstico ainda orienta anestesia, vigilância, mobilização, nutrição precoce e planejamento de alta.

Tabela 3 – Matriz prática de estratificação perioperatória
Fonte: Proposta de aplicação clínica baseada na síntese das evidências.

Nível	Achados	Risco esperado	Conduta mínima
Baixo	Força preservada; adiposidade sem baixa massa; boa mobilidade	Risco explicado sobretudo por comorbidades e procedimento	ERAS padrão, orientação proteica e mobilização
Intermediário	Um componente alterado; limitação leve; incerteza diagnóstica	Maior chance de declínio funcional	Confirmar composição, avaliar nutrição e prescrever exercício
Alto	Baixa força + baixa massa + adiposidade elevada	Complicações e recuperação lenta mais prováveis	Pré-habilitação multimodal, avaliação geriátrica e plano de alta
Muito alto	Fenótipo combinado + fragilidade, desnutrição ou cirurgia de grande porte	Risco de complicação grave, perda de autonomia e mortalidade	Decisão compartilhada, otimização intensiva e discussão de local de cuidado

7 INTERVENÇÕES: DA IDENTIFICAÇÃO À MODIFICAÇÃO DO RISCO

7.1 Exercício e pré-habilitação

O treinamento resistido é a intervenção com racional mais direto. Em pessoas com obesidade sarcopênica em idade próxima à aposentadoria, revisão sistemática e meta-análise mostrou redução de percentual de gordura, aumento de massa muscular, ganho de força e pequena melhora de velocidade de marcha com treinamento resistido. A combinação de proteína e exercício também reduziu massa gorda. Os estudos não foram realizados exclusivamente em pacientes cirúrgicos, mas demonstram modificabilidade do fenótipo (Eglseer *et al.*, 2023).

O ensaio PREHEP avaliou seis semanas de exercício e suporte nutricional antes de hepatectomia maior em pacientes com sarcopenia e sugeriu redução de morbidade e segurança do programa. A evidência é promissora, mas não autoriza concluir que todo programa reduzirá complicações em qualquer cirurgia. Adesão, supervisão, intensidade, risco basal e tempo disponível influenciam efeito. Em pessoas idosas frágeis, programas domiciliares apoiados

por profissionais podem ampliar acesso, embora resultados recentes mostrem que nem todos os desfechos respondem igualmente.

7.2 Nutrição e preservação de massa magra

A estratégia nutricional deve evitar dois extremos: subtratar a desnutrição por causa da obesidade e impor restrição energética agressiva que acelere perda muscular. A avaliação deve estimar ingestão, risco nutricional, perda de peso, sintomas gastrointestinais e capacidade de atingir proteína. Diretrizes ESPEN para cirurgia enfatizam alimentação oral ou enteral precoce, redução do jejum e terapia nutricional em pacientes de risco. Em pessoas idosas, necessidades proteicas frequentemente superam o mínimo populacional e devem ser individualizadas conforme função renal, doença aguda e tolerância.

Em câncer e cirurgias complexas, a inflamação pode limitar resposta anabólica. Por isso, nutrição isolada tende a ter efeito menor que a combinação com exercício. O cuidado também inclui corrigir sintomas que impedem alimentação, como dor, náusea, constipação, alterações de paladar e problemas dentários. A meta deve ser mensurável: atingir ingestão prescrita, estabilizar peso quando perda é involuntária, preservar força e iniciar alimentação pós-operatória sem atraso desnecessário.

7.3 Controle metabólico, anestesia e recuperação

Otimização metabólica inclui controle de glicemia, tratamento de anemia, cessação do tabagismo, revisão de medicamentos e manejo de apneia do sono. A anestesia deve antecipar maior dificuldade ventilatória e posicionamento, mas também reconhecer menor reserva muscular. Bloqueio neuromuscular, analgesia e sedação precisam ser cuidadosamente titulados. Estratégias ERAS, como redução de jejum, analgesia multimodal, prevenção de náusea, mobilização precoce e retirada oportuna de dispositivos, são particularmente importantes.

No pós-operatório, cada dia de repouso acelera perda funcional. Mobilização deve começar de forma segura, com metas diárias: sentar fora do leito, caminhar, realizar exercícios respiratórios e progredir atividades. Fisioterapia e enfermagem precisam compartilhar objetivos. O registro de distância percorrida, número de transferências e tolerância alimentar transforma recomendações genéricas em cuidado verificável.

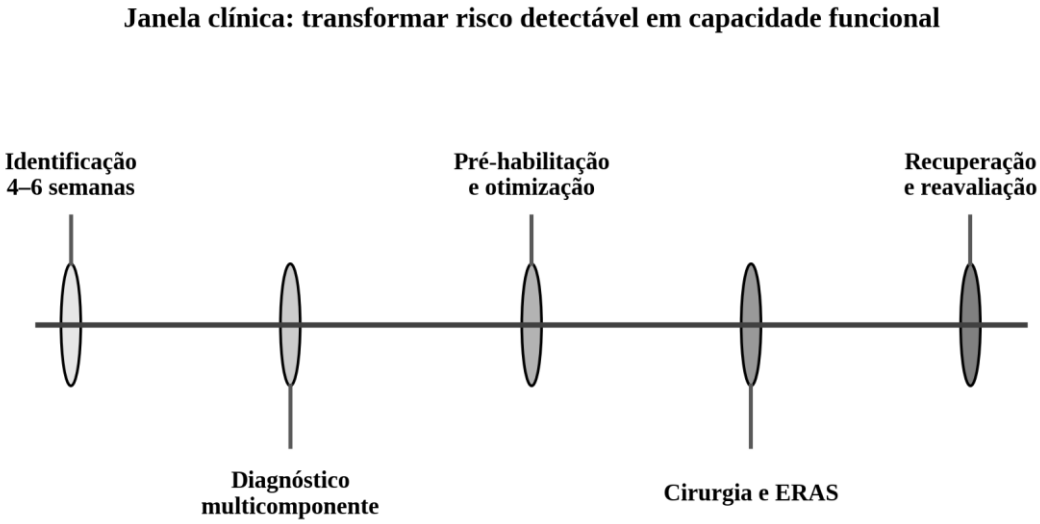


Figura 6 – Linha do tempo para integração da avaliação e intervenção perioperatória

Fonte: Elaboração própria. A duração deve ser adaptada à urgência e à magnitude do procedimento.

Tabela 4 – Componentes de um pacote perioperatório para obesidade sarcopênica

Fonte: Elaboração própria a partir de diretrizes de nutrição cirúrgica e estudos de pré-habilitação.

Momento	Componente	Ação operacional	Indicador
Pré-operatório	Triagem	Aplicar risco funcional e medir força	Percentual avaliado antes da cirurgia

Momento	Componente	Ação operacional	Indicador
Pré-operatório	Composição corporal	Usar DXA/BIA ou TC oportunística	Percentual de alto risco com medida confirmatória
Pré-operatório	Nutrição	Definir meta proteico-energética e tratar barreiras	Adesão e ingestão alcançada
Pré-operatório	Exercício	Combinar resistência e aeróbico com progressão	Sessões concluídas e mudança de força
Intraoperatório	Proteção fisiológica	Planejar ventilação, posicionamento e analgesia	Complicações anestésicas e tempo de ventilação
Pós-operatório	ERAS	Alimentação e mobilização precoces	Tempo até dieta e primeira deambulação
Alta	Continuidade	Reabilitação, suporte social e retorno precoce	Alta domiciliar, readmissão e função em 30 dias

8 ALGORITMO CLÍNICO PROPOSTO

O algoritmo começa pela pergunta “há risco de baixa reserva muscular?”, e não “o paciente parece magro?”. Se houver idade avançada, obesidade central, doença catabólica, perda funcional ou cirurgia de grande porte, realiza-se força de preensão ou teste de levantar da cadeira. Resultado alterado aciona composição corporal. Se a adiposidade e a baixa massa forem confirmadas, classifica-se obesidade sarcopênica e sua gravidade funcional.

O segundo passo é estimar modificabilidade. Tempo até a cirurgia, urgência, estabilidade da doença, sintomas e capacidade de aderir definem intensidade da intervenção. Em operações eletivas, quatro a seis semanas permitem programa estruturado. Em janela menor, priorizam-se nutrição, exercícios simples de resistência, treinamento respiratório e educação. Em emergência, o foco se desloca para proteção intraoperatória, mobilização, nutrição precoce e prevenção de delirium.

O terceiro passo é decisão compartilhada. A composição corporal não deve ser usada para negar cirurgia automaticamente. Ela deve qualificar a conversa sobre risco, alternativas, possibilidade de adiar para otimização e metas do paciente. Um biomarcador é eticamente útil quando melhora o cuidado; torna-se problemático quando cria barreira sem oferta de intervenção.

O quarto passo é reavaliar. A ausência de mudança de massa em poucas semanas não significa fracasso. Força, tolerância ao esforço, ingestão, controle glicêmico e confiança podem melhorar antes que a tomografia mostre alteração. Após a cirurgia, reavaliar função é essencial para detectar perda acumulada e ajustar reabilitação.

9 IMPLEMENTAÇÃO, AUDITORIA E EQUIDADE

A implementação exige definir responsabilidades. A equipe cirúrgica identifica risco e magnitude do procedimento; anestesia integra vulnerabilidade ao plano; nutrição avalia ingestão e prescreve; fisioterapia ou educação física estruturam exercício; enfermagem acompanha metas e mobilização; geriatria aborda fragilidade, cognição e polifarmácia; radiologia ou sistemas automatizados fornecem medidas corporais. Sem coordenação, o diagnóstico gera encaminhamentos fragmentados e perde tempo.

Indicadores precisam acompanhar processo e resultado. Cobertura de triagem, proporção com força medida, tempo até avaliação nutricional, adesão ao exercício, início de alimentação, mobilização no primeiro dia, complicações maiores, alta domiciliar, readmissão e recuperação funcional são exemplos. Metas de cobertura devem ser progressivas e realistas. O objetivo não é produzir um novo campo obrigatório no prontuário, mas mudar trajetórias.

A equidade merece atenção. Pessoas com menor renda podem ter dificuldade para adquirir suplementos, transportar-se para sessões ou acessar ambiente seguro para exercício. Programas exclusivamente presenciais podem selecionar pacientes mais favorecidos. Alternativas domiciliares, materiais

simples, contato remoto e articulação com atenção primária reduzem barreiras. Também é necessário validar pontos de corte em diferentes sexos, etnias e biotipos para evitar classificação desigual.

10 DISCUSSÃO CRÍTICA

A síntese indica que a obesidade sarcopênica representa um fenótipo clinicamente relevante e não uma curiosidade de imagem. A direção dos resultados é predominantemente consistente: mais complicações, maior gravidade, pior sobrevida e recuperação mais difícil. Entretanto, a força da conclusão varia. Para complicações maiores em oncologia gastrointestinal, há associação robusta, mas heterogeneidade. Para sobrevida em câncer colorretal, o efeito combinado não foi estatisticamente conclusivo. Para desfechos pulmonares, o resultado dependeu do modelo estatístico. A interpretação responsável deve reconhecer essa gradação.

A heterogeneidade diagnóstica é o principal problema metodológico. Alguns estudos definem obesidade por índice de massa corporal, outros por gordura visceral; alguns usam área do psoas, outros área muscular total; poucos incluem força. Essa variação pode alterar prevalência e magnitude do efeito. A padronização proposta pelo consenso ESPEN/EASO é avanço importante, mas ainda há distância entre consenso e estudos cirúrgicos. Coortes futuras devem medir função e composição, registrar sexo e etnia, usar desfechos centrados no paciente e repetir avaliação após a cirurgia.

A pré-habilitação é biologicamente plausível e apoiada por resultados promissores, mas não deve ser apresentada como solução universal. Programas variam em duração, intensidade, supervisão e componentes. Pessoas com janela curta ou doença instável podem não alcançar mudanças grandes. Mesmo assim, o processo de avaliar, educar e planejar recuperação pode trazer benefício organizacional. O valor não está apenas em aumentar músculo antes da cirurgia,

mas em alinhar equipe, reduzir jejum e inatividade, garantir proteína e preparar alta.

A tomografia oportunística e a inteligência artificial podem ampliar identificação, mas não devem criar tecnocentrismo. Uma imagem não mede objetivos, suporte social, cognição ou vontade. O melhor modelo combina biomarcadores corporais com avaliação funcional e geriátrica. Essa integração desloca a pergunta de "qual é o risco estatístico?" para "o que pode ser feito para melhorar a trajetória desta pessoa?".

Por fim, a aplicação prática exige evitar estigma. Pessoas com obesidade frequentemente vivenciam julgamento e atribuição moral do risco. O diagnóstico de obesidade sarcopênica não deve reforçar culpabilização. Deve ser comunicado como condição multifatorial, influenciada por envelhecimento, doença, tratamento, ambiente e acesso. A linguagem clínica deve enfatizar força, função e preparação, não inadequação corporal.

11 LIMITAÇÕES DA EVIDÊNCIA E DESTE CAPÍTULO

A maioria dos estudos é observacional e oncológica, o que limita generalização para cirurgias benignas, ortopédicas e vasculares. Definições variam e muitos trabalhos usam baixa massa muscular como equivalente de sarcopenia. Há risco de sobreposição entre meta-análises, porque as mesmas coortes podem aparecer em mais de uma síntese. As estimativas comparadas neste capítulo não foram recalculadas nem combinadas, e os gráficos devem ser interpretados como mapa, não como nova meta-análise.

As propostas de algoritmo e indicadores são derivadas da literatura e da lógica de implementação, mas precisam de validação prospectiva. Pontos de corte e metas institucionais devem ser adaptados à população e aos recursos locais.

12 CONCLUSÃO

A obesidade sarcopênica revela uma vulnerabilidade que o peso corporal isolado não mostra. Em pessoas idosas submetidas a cirurgia, a combinação de adiposidade elevada, baixa força e baixa reserva muscular está associada a maior risco de complicações graves, infecção, fístula, mortalidade, pior sobrevida e perda de autonomia. A magnitude varia entre contextos, mas a direção é suficientemente consistente para justificar avaliação pré-operatória.

A prática deve integrar triagem funcional, medida de força e composição corporal. Tomografias já realizadas podem ser aproveitadas, especialmente com automação validada, mas precisam ser combinadas com avaliação clínica. O diagnóstico deve acionar intervenção: proteína e energia adequadas, exercício resistido e aeróbico, otimização metabólica, ERAS, mobilização precoce e planejamento de alta.

O avanço mais importante é conceitual. O risco perioperatório não deve ser estimado apenas por idade, doenças e índice de massa corporal, mas pela capacidade do organismo de enfrentar e recuperar-se do estresse cirúrgico. Identificar obesidade sarcopênica é transformar um risco invisível em oportunidade de cuidado.

REFERÊNCIAS

BATES, D. D.; PICKHARDT, P. J. CT-derived body composition assessment as a prognostic tool in oncologic patients: from opportunistic research to artificial intelligence-based clinical implementation. AJR American Journal of Roentgenology, v. 219, p. 671–680, 2022.

BEDRIKOVETSKI, S. *et al.* Artificial intelligence for body composition and sarcopenia evaluation on computed tomography: a systematic review and meta-analysis. European Journal of Radiology, v. 149, 110218, 2022.

BERARDI, G. *et al.* Prehabilitation with exercise and nutrition to reduce morbidity of major hepatectomy in patients with sarcopenia: the PREHEP randomized clinical trial. JAMA Surgery, v. 160, n. 10, p. 1068–1075, 2025. DOI: 10.1001/jamasurg.2025.3102.

CRUZ-JENTOFT, A. J. *et al.* Sarcopenia: revised European consensus on definition and diagnosis. Age and Ageing, v. 48, n. 1, p. 16–31, 2019. DOI: 10.1093/ageing/afy169.

DINDO, D.; DEMARTINES, N.; CLAVIEN, P. A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Annals of Surgery, v. 240, n. 2, p. 205–213, 2004.

DONINI, L. M. *et al.* Definition and diagnostic criteria for sarcopenic obesity: ESPEN and EASO consensus statement. Obesity Facts, v. 15, n. 3, p. 321–335, 2022. DOI: 10.1159/000521241.

EGLSEER, D. *et al.* Nutritional and exercise interventions in individuals with sarcopenic obesity around retirement age: a systematic review and meta-analysis. Nutrition Reviews, v. 81, n. 9, p. 1077–1090, 2023. DOI: 10.1093/nutrit/nuad007.

JEONG, J. H. *et al.* The effects of sarcopenic obesity on immediate postoperative outcomes after pancreatoduodenectomy: a retrospective cohort study. Annals of Surgical Treatment and Research, v. 107, n. 4, p. 203–212, 2024. DOI: 10.4174/astr.2024.107.4.203.

KNOEDLER, S. *et al.* Impact of sarcopenia on outcomes in surgical patients: a systematic review and meta-analysis. International Journal of Surgery, v. 109, n. 12, p. 4238–4262, 2023. DOI: 10.1097/JS9.0000000000000688.

LEE, C. M. *et al.* Prognostic impact of myosteatosis in patients with colorectal cancer: a systematic review and meta-analysis. Journal of Cachexia, Sarcopenia and Muscle, v. 11, n. 5, p. 1270–1282, 2020.

MATSUI, R. *et al.* Impact of sarcopenic obesity on postoperative outcomes in patients with oesophago-gastric cancer: a systematic review and meta-analysis. World Journal of Surgery, v. 49, 2025. DOI: 10.1002/wjs.12451.

MEI, K. L.; BATSIS, J. A.; MILLS, J. B.; HOLUBAR, S. D. Sarcopenia and sarcopenic obesity: do they predict inferior oncologic outcomes after gastrointestinal cancer surgery? Perioperative Medicine, v. 5, 30, 2016.

MOURTZAKIS, M. *et al.* A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Applied Physiology, Nutrition, and Metabolism, v. 33, n. 5, p. 997–1006, 2008.

PARK, B. *et al.* Consensus-defined sarcopenia predicts adverse outcomes after elective abdominal surgery: meta-analysis. BJS Open, v. 7, n. 4, zrad065, 2023. DOI: 10.1093/bjsopen/zrad065.

PECORELLI, N. *et al.* Impact of sarcopenic obesity on failure to rescue from major complications following pancreaticoduodenectomy for cancer: results from a multicenter study. Annals of Surgical Oncology, v. 25, p. 308–317, 2018.

PRADO, C. M. *et al.* Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The Lancet Oncology, v. 9, n. 7, p. 629–635, 2008.

SORIA-UTRILLA, V. *et al.* AI-assisted body composition assessment using CT imaging in colorectal cancer patients: predictive capacity for sarcopenia and malnutrition diagnosis. Nutrients, v. 16, 2024.

TREJO-AVILA, M. *et al.* Sarcopenia predicts worse postoperative outcomes and decreased survival rates in patients with colorectal cancer: a systematic review and meta-analysis. International Journal of Colorectal Disease, v. 36, p. 1077–1096, 2021.

VOLKERT, D. *et al.* ESPEN practical guideline: clinical nutrition and hydration in geriatrics. Clinical Nutrition, v. 41, n. 4, p. 958–989, 2022.

WANG, P. *et al.* Sarcopenic obesity and therapeutic outcomes in gastrointestinal surgical oncology: a meta-analysis. Frontiers in Nutrition, v. 9, 921817, 2022. DOI: 10.3389/fnut.2022.921817.

WEIMANN, A. *et al.* ESPEN practical guideline: clinical nutrition in surgery. Clinical Nutrition, v. 40, n. 7, p. 4745–4761, 2021. DOI: 10.1016/j.clnu.2021.03.031.

WEIMANN, A. *et al.* ESPEN guideline on clinical nutrition in surgery: update 2025. Clinical Nutrition, 2025.

YAMAGISHI, S. *et al.* Impact of sarcopenic obesity on severe postoperative complications in patients with gastric cancer undergoing gastrectomy. Digestive Surgery, v. 40, n. 5, p. 143–152, 2023. DOI: 10.1159/000531797.

YANG, Y. *et al.* Adverse postoperative outcomes in elderly patients with sarcopenia. BMC Geriatrics, v. 24, 2024.

YILDIZ KOPUZ, T. N. *et al.* Prognostic impact of sarcopenic obesity on postoperative outcomes in colorectal cancer patients undergoing surgery: a systematic review and meta-analysis. Nutrition and Cancer, v. 77, n. 3, p. 360–371, 2025. DOI: 10.1080/01635581.2025.2450843.

ZHUANG, C. L. *et al.* Sarcopenia is an independent predictor of severe postoperative complications and long-term survival after radical gastrectomy for gastric cancer. Medicine, v. 95, n. 13, e3164, 2016.

CAPÍTULO 17

PREDIÇÃO DE COMPLICAÇÕES PÓS-OPERATÓRIAS EM PESSOAS COM DIABETES POR APRENDIZADO FEDERADO E DADOS SINTÉTICOS MULTICÊNTRICOS: PROTOCOLO DE DESENVOLVIMENTO, VALIDAÇÃO E BENCHMARK COM DADOS REAIS

Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Nélio Barreto Vieira

RESUMO

Objetivo: propor um estudo multicêntrico de desenvolvimento e validação de modelos para complicações pós-operatórias em adultos com diabetes, combinando aprendizado federado e dados sintéticos, e examinar sua viabilidade com resultados reais publicados. Método: foi especificada uma arquitetura federada na qual cada hospital manterá os prontuários locais e compartilhará apenas atualizações protegidas do modelo. Os desfechos previstos são infecção do sítio cirúrgico, deiscência, readmissão, tromboembolismo, eventos cardiovasculares, internação prolongada e mortalidade em 30 dias. A estratégia proposta inclui harmonização, divisão temporal, validação externa, calibração, avaliação de equidade e geração sintética com teste train-on-synthetic/test-on-real. Foram analisados estudos reais de previsão cirúrgica, aprendizado federado e síntese de prontuários. Resultados: um modelo multicêntrico para mortalidade após cirurgia não cardíaca analisou 454.404 pacientes de quatro instituições e obteve AUROC de 0,9376 no desenvolvimento e 0,941 em validação entre hospitais. Um modelo federado de recorrência pós-operatória de câncer gástrico apresentou AUCs de 0,710, 0,798, 0,809 e 0,869 em quatro centros. Em 2026, empilhamento federado para lesão renal aguda após cirurgia cardíaca foi treinado em 43.926 casos de 23 hospitais e superou modelos

locais, sobretudo em centros menores. Dados sintéticos preservadores de privacidade produziram Brier score de 0,00494 versus 0,00489 para dados reais em previsão de câncer de pulmão, embora a discriminação fosse menor. Conclusão: o desenho híbrido foi tecnicamente plausível e ofereceu vantagens para hospitais com menor volume, mas exigiu governança, avaliação de mudança de distribuição, validação em dados reais e transparência sobre a perda de utilidade dos dados sintéticos.

Palavras-chave: diabetes mellitus; complicações pós-operatórias; aprendizado federado; dados sintéticos; inteligência artificial; validação externa.

ABSTRACT

Objective: to propose a multicentre development and validation study for postoperative complication prediction in adults with diabetes by combining federated learning and synthetic data, and to assess feasibility using real published results. Methods: a federated architecture was specified in which each hospital will retain local records and share only protected model updates. Planned outcomes include surgical-site infection, wound dehiscence, readmission, thromboembolism, cardiovascular events, prolonged length of stay, and 30-day mortality. The proposed strategy includes data harmonisation, temporal splitting, external validation, calibration, fairness assessment, and synthetic data generation with train-on-synthetic/test-on-real evaluation. Results: a multicentre model for 30-day mortality after non-cardiac surgery analysed 454,404 patients from four hospitals and achieved an AUROC of 0.9376 during development and 0.941 in cross-hospital validation. A federated model for postoperative gastric cancer recurrence achieved AUCs of 0.710, 0.798, 0.809, and 0.869 across four centres. In 2026, federated stacking for acute kidney injury after cardiac surgery was trained on 43,926 cases from 23 hospitals and outperformed local models, particularly at lower-volume centres. Privacy-preserving synthetic data achieved a Brier score of 0.00494 versus 0.00489 for real-data lung cancer prediction, although discrimination was lower. Conclusion: a

hybrid design was technically plausible and may especially benefit smaller hospitals, but requires governance, distribution-shift assessment, real-data validation, and transparent reporting of synthetic-data utility loss.

Keywords: diabetes mellitus; postoperative complications; federated learning; synthetic data; artificial intelligence; external validation.

1 INTRODUÇÃO

Pessoas com diabetes enfrentam risco pós-operatório heterogêneo. Controle glicêmico, duração da doença, doença renal, obesidade, neuropatia, doença vascular, uso de insulina e urgência interagem com o tipo e a magnitude da cirurgia. Revisões identificaram diabetes como fator de risco para complicações de ferida, vazamento anastomótico e outros eventos, embora a magnitude varie entre especialidades.[1-3] Modelos convencionais frequentemente tratam diabetes como variável binária e perdem informação sobre gravidade metabólica e vulnerabilidade de órgãos.

Aprendizado de máquina pode combinar grande número de variáveis, relações não lineares e dados longitudinais. Entretanto, muitos modelos cirúrgicos foram desenvolvidos em um único centro, com validação interna e pouca avaliação de calibração ou utilidade clínica. Um modelo excelente no hospital de origem pode falhar em outro por diferenças de população, equipamentos, fluxo, definição de desfechos e prevalência.[4-6] A centralização de dados multicêntricos, por sua vez, enfrenta restrições legais, institucionais e de segurança.

Aprendizado federado permite treinar modelos sem transferir prontuários brutos. Cada instituição calcula atualizações locais, e um servidor agrega parâmetros ou previsões. Estudos recentes mostraram desempenho semelhante ou superior ao de modelos locais em mortalidade, recorrência oncológica e lesão renal aguda pós-cirurgia.[7-9] O benefício foi particularmente relevante para centros menores, que possuíam dados insuficientes para treinar modelos estáveis isoladamente.

Dados sintéticos oferecem caminho complementar. Um gerador aprende distribuições do conjunto real e produz registros artificiais para desenvolvimento, teste de código, planejamento e, em alguns cenários, treinamento. Esses registros não são automaticamente anônimos, e sua utilidade deve ser examinada por fidelidade, privacidade, calibração e desempenho train-on-synthetic/test-on-real.[10-13] O objetivo deste capítulo foi especificar um protocolo híbrido para pessoas com diabetes e analisar sua viabilidade com métricas publicadas reais.

2 MÉTODO

Foi elaborado um protocolo multicêntrico de desenvolvimento e validação de modelos prognósticos. A população-alvo incluirá adultos com diabetes tipo 1 ou tipo 2 submetidos a cirurgia de grande ou médio porte. Serão excluídos procedimentos exclusivamente diagnósticos, ausência de confirmação de diabetes e prontuários sem período mínimo de observação. O protocolo seguiu princípios do TRIPOD+AI, PROBAST+AI e recomendações para sistemas de inteligência artificial clínica, com ênfase em separação entre desenvolvimento, validação e avaliação de impacto.[14-16]

Os desfechos foram definidos antes da modelagem: infecção do sítio cirúrgico em 30 dias, deiscência, readmissão não planejada, tromboembolismo venoso, evento cardiovascular maior, internação prolongada, mortalidade em 30 dias e composto de complicação grave. Cada hospital mapeará códigos, microbiologia, procedimentos, readmissões e revisão clínica. Um comitê de adjudicação avaliaria amostra de casos e não casos para estimar validade das definições.

As variáveis candidatas incluíram idade, sexo, tipo de diabetes, duração, HbA1c, glicemia, uso de insulina, complicações micro e macrovasculares, TFG, albumina, hemoglobina, IMC, ASA, tabagismo, mobilidade, história de tromboembolismo, urgência, porte, técnica, duração e transfusão. Variáveis

intraoperatórias seriam usadas em modelos dinâmicos separados para não contaminar o modelo pré-operatório.

A federação adotaria arquitetura horizontal, porque os centros possuiriam variáveis semelhantes em pacientes diferentes. Cada instituição aplicaria o mesmo dicionário, pré-processamento e versão do código. Modelos candidatos incluiriam regressão penalizada, gradient boosting e rede neural tabular. FedAvg funcionaria como referência; FedProx ou métodos de reponderação seriam utilizados quando heterogeneidade e covariate shift produzissem divergência.[17,18]

O conjunto de cada centro seria dividido temporalmente, preservando os anos mais recentes para validação. Além da validação temporal local, seria executada validação leave-one-centre-out: o modelo seria treinado nos demais hospitais e avaliado no centro excluído. Discriminação seria medida por AUROC e AUPRC; calibração por intercepto, slope, Brier score e curvas; utilidade por decision curve analysis. Resultados seriam estratificados por sexo, idade, raça/cor quando disponível, tipo de diabetes, TFG e porte do centro.

Dados sintéticos seriam gerados apenas a partir do conjunto de treinamento local, usando geradores tabulares ou longitudinais com proteção diferencial quando possível. Três usos foram previstos: desenvolvimento de código sem acesso ao real, aumento de classes raras e criação de um conjunto de demonstração. A utilidade seria avaliada por train-on-synthetic/test-on-real, concordância de seleção de variáveis e preservação de calibração. Privacidade seria testada por membership inference, nearest-neighbour e risco de reprodução de registros.[10-13]

A governança proibiria transferência de identificadores e estabeleceria logs, assinatura de versões, cartão do modelo, cartão dos dados sintéticos e plano de resposta a incidente. Atualizações seriam protegidas por secure aggregation; differential privacy seria considerada após análise da perda de utilidade. Nenhum modelo seria implementado apenas por AUROC. Critérios de avanço

incluiriam calibração aceitável, benefício líquido, ausência de dano desproporcional e avaliação prospectiva silenciosa.

A Figura 1 mostra quatro centros exemplificativos conectados a um agregador seguro. Os rótulos representam especialidades distintas, mas o protocolo admite qualquer hospital que mapeie o dicionário comum e cumpra os critérios de qualidade.



Figura 1. Arquitetura federada proposta para predição de complicações pós-operatórias.

O elemento central foi que os dados permanecerão locais. Parâmetros ou predições serão agregados e devolvidos para calibração institucional. Essa arquitetura reduzirá a necessidade de transferir prontuários, mas não eliminará riscos de inferência sobre atualizações, motivo pelo qual foram previstas agregação segura e auditoria.

A Tabela 1 apresenta o núcleo de dados que todos os hospitais precisariam disponibilizar localmente. A escolha privilegiou variáveis clínicas interpretáveis e rotineiramente coletadas, reduzindo dependência de campos proprietários.

Tabela 1. Dicionário mínimo harmonizado para a federação.

Domínio	Variáveis mínimas	Formato comum	Momento
Diabetes	tipo, duração, HbA1c, insulina, hipoglicemia	categorias e unidades padronizadas	pré-operatório

Domínio	Variáveis mínimas	Formato comum	Momento
Rim	creatinina, TFG, diálise, albuminúria	KDIGO/CKD-EPI	pré-operatório
Risco cirúrgico	ASA, urgência, porte, especialidade	vocabulário comum	pré/intra
Inflamação/nutrição	albumina, hemoglobina, leucócitos	unidade SI	pré-operatório
Procedimento	aberto/laparoscópico /robótico, duração	código harmonizado	intraoperatório
Desfechos	SSI, deiscência, TEV, MACE, readmissão, morte	definição adjudicada	0–30 dias

A harmonização foi considerada parte do estudo, não tarefa administrativa. Unidades, tempo de coleta e codificação deveriam ser iguais antes do treinamento; caso contrário, o algoritmo poderia aprender diferenças de documentação em vez de risco biológico.

3 RESULTADOS DO BENCHMARK DE VIABILIDADE

O estudo multicêntrico de mortalidade após cirurgia não cardíaca analisou 454.404 pacientes de quatro instituições e utilizou apenas 12 a 18 variáveis. O XGBoost alcançou AUROC de 0,9376 no centro de desenvolvimento e 0,941 quando validado em outro hospital. Albumina, tempo de protrombina e idade foram variáveis importantes.[6] O resultado demonstrou que modelos relativamente leves podiam transferir-se quando treinados em grande amostra e com entradas objetivas.

Em recorrência pós-operatória de câncer gástrico, um modelo federado robusto utilizou imagens de quatro centros e obteve AUCs de 0,710, 0,798, 0,809 e 0,869. O método superou modelos clínicos e algoritmos federados de referência e apresentou ganho de acurácia de 32,36% em relação ao modelo clínico agregado.[8] A variação entre centros confirmou que desempenho global não substitui avaliação local.

Em 2026, uma rede nacional avaliou lesão renal aguda após cirurgia cardíaca em 43.926 casos de 23 hospitais. Modelos multicêntricos por pooling e empilhamento federado superaram modelos treinados localmente para diferentes graus de lesão renal, e hospitais de menor volume apresentaram os maiores ganhos.[9] Esse resultado foi diretamente relevante ao protocolo, pois complicações em pessoas com diabetes podem ser raras em centros pequenos.

Modelos cirúrgicos automatizados também demonstraram viabilidade de atualização com prontuários. A plataforma MySurgeryRisk foi construída para prever oito complicações a partir de dados extraídos automaticamente, com previsões dinâmicas pré e pós-operatórias.[19,20] Embora não fosse federada nem restrita a diabetes, estabeleceu que variáveis longitudinais e múltiplos desfechos poderiam ser integrados em fluxo clínico.

Para cirurgia bariátrica, um modelo baseado no MBSAQIP identificou diabetes como preditor de evento adverso grave (odds ratio ajustada 1,15) e readmissão (1,08), além de IMC, procedimento, doença cardiovascular, hipertensão e mobilidade limitada.[21] A evidência justificou incluir severidade do diabetes e comorbidades, evitando tratá-lo como marcador isolado.

No campo dos dados sintéticos, Qian e colaboradores geraram coortes preservadoras de privacidade a partir do UK Biobank. Modelos treinados em dados ADSGAN e PATEGAN apresentaram Brier scores próximos ao modelo real; para ADSGAN, 0,00494 versus 0,00489 no real, embora a discriminação fosse inferior.[10] O estudo mostrou que dados sintéticos poderiam apoiar seleção de variáveis, ajuste de hiperparâmetros e preparação do pipeline, mas a validação final deveria ocorrer em dados reais.

EHR-Safe e geradores longitudinais demonstraram fidelidade e resistência a ataques em bases complexas, mas revisões recentes encontraram ausência de consenso sobre métricas de privacidade e utilidade.[11-13] Consequentemente, o protocolo não aceitou a expressão “dados anônimos” sem teste. Cada conjunto sintético deveria informar gerador, parâmetros, proteção, ataques avaliados e tarefas em que preservou desempenho.

A Figura 2 reúne métricas de estudos com tarefas diferentes; por isso, os valores não foram comparados como ranking único. As seis primeiras linhas representam AUROC, a sétima a AUROC calibrada em readmissão de pessoas com diabetes e a última um índice derivado de calibração indicado na nota.

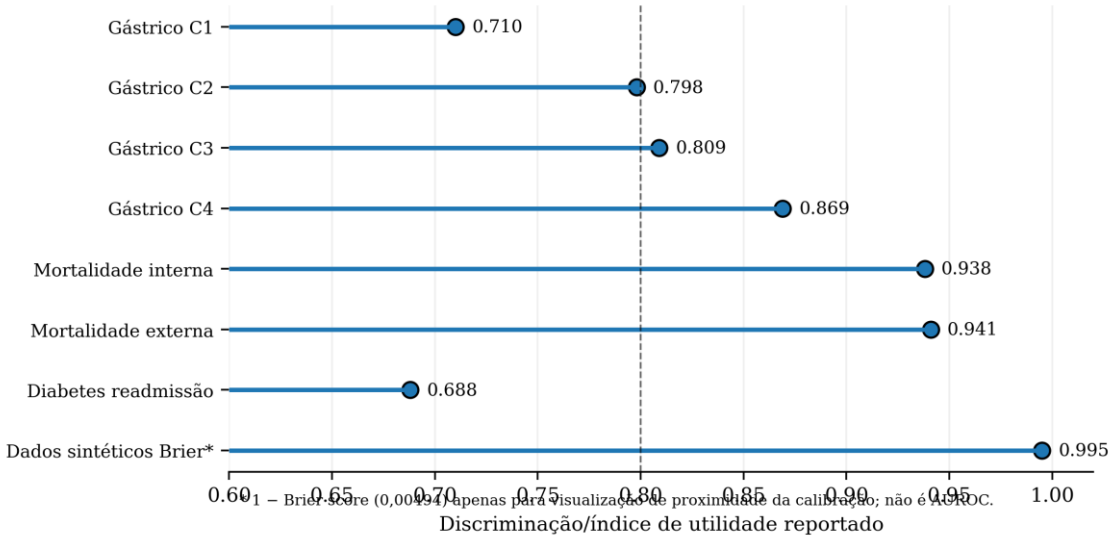


Figura 2. Métricas publicadas em estudos reais de predição multicêntrica, federada e sintética.

O painel mostrou que desempenho elevado foi possível em mortalidade, enquanto tarefas oncológicas federadas apresentaram heterogeneidade entre centros. A proximidade do Brier score sintético ao real indicou boa calibração, mas não autorizou concluir equivalência de discriminação. Cada métrica precisou ser interpretada no contexto do desfecho e da prevalência.

A Tabela 2 separa as contribuições de cada estudo. Alguns demonstraram transferência multicêntrica, outros mostraram robustez federada ou utilidade sintética, e o estudo bariátrico forneceu variáveis prognósticas específicas.

Tabela 2. Estudos reais e métricas que fundamentaram o benchmark.

Estudo real	Dados/centros	Tarefa	Métrica principal	Contribuição ao protocolo
Lee et al., 2022	454.404; 4 centros	mortalidade 30 dias	AUROC 0,9376; externa 0,941	validação entre hospitais

Estudo real	Dados/centros	Tarefa	Métrica principal	Contribuição ao protocolo
Feng et al., 2024	4 centros	recorrência gástrica	AUC 0,710–0,869	FL robusto sob heterogeneidade
Cao et al., 2026	43.926; 23 hospitais	AKI pós-cardíaca	multicêntrico > local	benefício para centros pequenos
Maurer et al., 2025	rede multicêntrica	mortalidade 90 dias	FL semelhante ao centralizado	privacidade sem perda relevante
Qian et al., 2024	UK Biobank	risco de câncer	Brier 0,00494 vs 0,00489	train-synthetic/test-real
El Chaar et al., 2019	MBSAQIP 2015/2016	evento/readmissão bariátrica	diabetes AOR 1,15/1,08	variáveis específicas de diabetes

O benchmark não substituiu treinamento na população-alvo. Ele estabeleceu faixas plausíveis, problemas esperados e critérios de avaliação. O desempenho final poderia ser menor porque a amostra seria restrita a pessoas com diabetes e incluiria desfechos raros e heterogêneos.

A Figura 3 resume, em escala relativa baseada na literatura, os compromissos entre centralização, federação e síntese. Os pontos não representam estimativas clínicas; funcionam como mapa de decisão para o desenho do estudo.

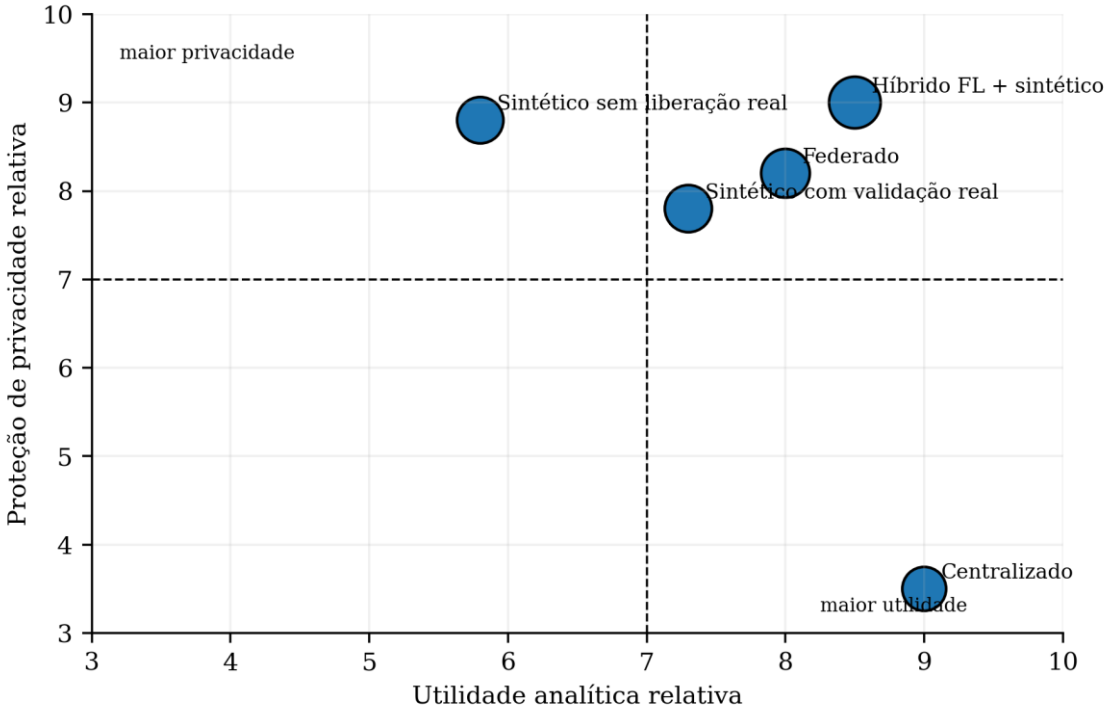


Figura 3. Relação conceitual entre utilidade analítica e proteção de privacidade.

A centralização ofereceu alta flexibilidade, mas exigiu transferência de dados. A federação aumentou privacidade operacional, enquanto dados sintéticos permitiram uso downstream convencional com perda variável de utilidade. O desenho híbrido ocupou o quadrante desejável, mas somente se agregação e geradores fossem auditados.

A Figura 4 apresenta uma sequência diferente de um fluxograma clínico convencional. Cada seta correspondeu a um artefato que poderia ser versionado e inspecionado: dicionário, modelo local, atualização agregada, relatório de validação e conjunto sintético.



Cada etapa deixou um artefato auditável: dicionário, versão do modelo, pesos agregados, relatório de calibração e cartão do conjunto sintético.

Nenhum arquivo de paciente cruzou a fronteira institucional.

Figura 4. Ciclo de vida auditável dos dados e modelos.

A rastreabilidade permitiu investigar queda de desempenho e reproduzir versões. Também evitou que dados sintéticos fossem distribuídos sem documentação de origem. O princípio foi simples: privacidade não deveria reduzir a capacidade de auditoria.

A Tabela 3 amplia a avaliação além da AUROC. Desfechos raros requereram AUPRC, e decisões clínicas dependeram de calibração e benefício líquido. Dados sintéticos exigiram teste em registros reais e ataques de privacidade.

Tabela 3. Métricas de desempenho, utilidade, privacidade e equidade.

Métrica	Pergunta	Aplicação	Critério de alerta
AUROC/AUPRC	o modelo discrimina casos?	todos os desfechos	queda externa >0,05
Calibração	probabilidades correspondem ao risco?	por centro e subgrupo	slope <0,8 ou >1,2
Brier score	erro probabilístico global	comparação real/sintético	piora clinicamente relevante
Decision curve	há benefício líquido?	limiares clínicos	inferior a treat-all/none
TSTR	modelo sintético funciona no real?	dados sintéticos	perda de AUROC e calibração

Métrica	Pergunta	Aplicação	Critério de alerta
Membership inference	há risco de reidentificação?	geradores	ataque acima do acaso
Equidade	erro difere por grupo?	sexo, idade, raça, TFG	diferença persistente não explicada

Os critérios de alerta foram propostos como gatilhos para investigação, não como normas universais. Um modelo poderia ter discriminação estável e calibração inadequada, ou preservar média global enquanto piorava em pacientes com doença renal. A decisão de implementação precisaria integrar todos os domínios.

A Figura 5 representa validação como camadas concêntricas. O modelo somente alcançaria avaliação de impacto depois de atravessar validação interna, temporal, externa e prospectiva.



Figura 5. Camadas sucessivas de validação antes do impacto clínico.

Esse desenho impediu que um resultado de cross-validation fosse apresentado como evidência de utilidade. A etapa prospectiva silenciosa observaria o modelo sem alterar cuidado; apenas depois seriam testados alertas, fluxos e efeitos sobre decisões e resultados.

A Figura 6 mostra quais variáveis poderiam contribuir para cada desfecho. O ponto indica inclusão candidata e o traço ausência de relação prioritária no protocolo. A matriz foi construída a partir de fatores conhecidos e de variáveis presentes em modelos cirúrgicos reais.

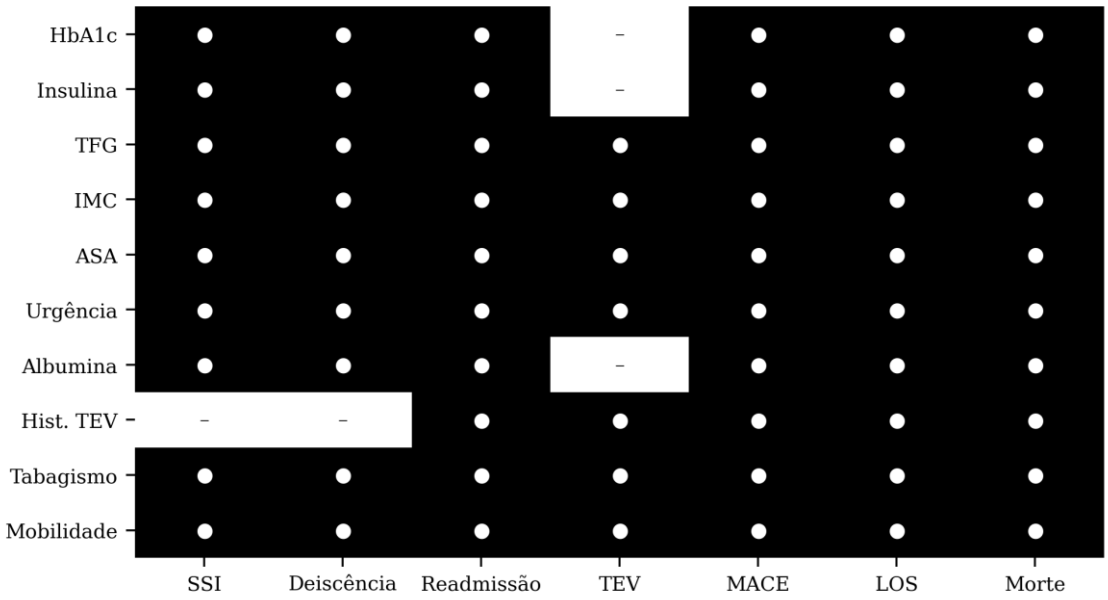


Figura 6. Matriz de ligação entre preditores e desfechos planejados.

A sobreposição foi alta porque fragilidade metabólica, renal e funcional afeta múltiplas complicações. Ainda assim, modelos específicos por desfecho foram preferidos a um composto único, pois infecção, tromboembolismo e mortalidade possuem mecanismos e limiares de decisão distintos.

4 DISCUSSÃO

Os estudos reais demonstraram que colaboração multicêntrica melhorou generalização, mas também expôs heterogeneidade. O intervalo de AUC entre centros no modelo de câncer gástrico não foi falha secundária; foi informação sobre diferenças de equipamento, população e distribuição. Um modelo federado clínico precisou relatar desempenho por hospital e não somente média ponderada.[8]

O achado de 2026 em cirurgia cardíaca teve implicação de equidade institucional. Hospitais pequenos foram os que mais ganharam ao participar da rede, enquanto centros grandes contribuíram com volume e diversidade.[9] Essa assimetria justificou acordos que reconhecessem contribuição dos centros de alto

volume sem restringir acesso ao modelo. A federação deve ser vista como infraestrutura cooperativa, não apenas técnica de privacidade.

Na população com diabetes, definição de caso exigiu maior detalhe. HbA1c isolada não capturou variabilidade glicêmica, hipoglicemia, insulinização e complicações. Doença renal e albumina foram especialmente relevantes, pois influenciaram infecção, permanência e mortalidade. Modelos deveriam incorporar a temporalidade da coleta e evitar usar exames obtidos após deterioração como preditores pré-operatórios.

A transferência entre especialidades também mereceu cautela. Um modelo treinado predominantemente em cirurgia geral poderia superestimar ou subestimar risco em ortopedia, vascular ou bariátrica. A federação permitiu personalização local e modelos hierárquicos, mas personalização excessiva poderia recriar isolamento. Estratégias de camadas globais e locais, reponderação e empilhamento ofereceram equilíbrio entre conhecimento compartilhado e contexto.

Dados sintéticos foram úteis para acelerar desenvolvimento e reduzir acesso precoce a prontuários, porém não deveriam ser apresentados como substituto irrestrito. O estudo de risco pulmonar mostrou calibração próxima em determinadas abordagens e menor discriminação.[10] Em desfechos raros, um gerador pode suavizar caudas, perder interações e criar confiança artificial. Por isso, o protocolo exigiu teste no real e proibiu validação exclusivamente sintética.

Privacidade também não foi binária. Modelos federados podem vazar informação por gradientes, e dados sintéticos podem reproduzir indivíduos raros. Secure aggregation, clipping, ruído diferencial e limitação de consultas reduzem risco, mas podem diminuir utilidade. A documentação precisa informar orçamento de privacidade, ataques avaliados e grupos potencialmente expostos, especialmente combinações raras de procedimento e comorbidade.

A avaliação de equidade deveria contemplar pacientes e instituições. Subgrupos com menor representação podem apresentar calibração pior, enquanto hospitais pequenos podem receber modelo menos ajustado.

Reponderação federada, como FedWeight, foi proposta para corrigir covariate shift sem compartilhar dados individuais.[18] Entretanto, correção algorítmica não substitui coleta adequada nem validação clínica.

A implementação exigiu decisão sobre limiar. Um alerta para SSI pode justificar intervenção de baixo risco; um alerta para mortalidade pode alterar encaminhamento e consentimento. Decision curve analysis ajudou a comparar benefício líquido, mas o limiar deveria ser definido com profissionais e pacientes. Exibir probabilidades sem plano de ação pode aumentar carga e automação enviesada.

A definição dos rótulos foi outro ponto crítico. Infecção do sítio cirúrgico pode ser registrada por vigilância, cultura, antibiótico ou código administrativo; readmissão pode ocorrer em outro hospital; deiscência e tromboembolismo podem aparecer apenas em texto livre. Treinar centros com definições diferentes criaria heterogeneidade artificial e faria o algoritmo aprender práticas de documentação. O protocolo exigiu adjudicação amostral, janelas temporais comuns e análise de sensibilidade por fonte do rótulo, além de estimar subnotificação por vinculação com bases de urgência e mortalidade.

Vazamento temporal também poderia inflar o desempenho. Exames solicitados em resposta a deterioração, códigos lançados após a complicação e informações do resumo de alta não poderiam entrar em um modelo apresentado como pré-operatório. Cada variável deveria carregar um timestamp verificável e ser congelada no momento da previsão. Para modelos atualizados no pós-operatório, novos marcos temporais seriam definidos explicitamente. Essa disciplina permitiria distinguir previsão antecipada de simples reconhecimento automatizado de um evento já em curso.

Governança e responsabilidade clínica não foram tratadas como anexos. Cada centro precisaria manter comitê local, registro de versões, canal para contestação e plano de resposta a drift. O modelo global teria um cartão com população, exclusões, desempenho por centro, limitações e condições de suspensão. Pacientes e profissionais deveriam participar da escolha dos limiares

e das ações decorrentes, porque o valor de um alerta depende do custo da intervenção, da possibilidade de prevenção e do risco de sobrecarga. A federação distribuiria o treinamento, mas não diluiria a responsabilidade pelo uso.

O protocolo teve limitações. A análise foi um benchmark, não treinamento federado realizado na população-alvo. Métricas vieram de tarefas e bases diferentes. Alguns estudos utilizaram imagens, outros dados tabulares, e a prevalência dos desfechos variou. As pontuações de privacidade-utilidade da figura conceitual não foram medidas empíricas. Além disso, legislação e infraestrutura diferem entre países.

Mesmo assim, o desenho ofereceu caminho concreto. A combinação de federação, validação leave-one-centre-out, dados sintéticos para desenvolvimento seguro e avaliação prospectiva poderia produzir modelos mais transportáveis. O critério de sucesso não seria superar um escore em décimos de AUROC, mas fornecer probabilidades calibradas, úteis e auditáveis, sem concentrar dados sensíveis ou ampliar desigualdades.

A Tabela 4 organiza o projeto como sequência com pontos de parada. Essa estrutura evitou avançar para implementação apenas porque um algoritmo conseguiu ser treinado.

Tabela 4. Etapas e critérios de progressão do estudo multicêntrico.

Fase	Produto	Critério de progressão	Motivo para interromper
Harmonização	dicionário e qualidade	≥95% campos essenciais mapeados	definição incompatível
Desenvolvimento local	baseline por centro	calibração e dados suficientes	evento raro sem suporte
Federação	modelo global/personalizado	ganho externo ou estabilidade	degradação em centros
Síntese	conjunto sintético e cartão	TSTR aceitável e ataques negativos	risco de reprodução

Fase	Produto	Critério de progressão	Motivo para interromper
Validação silenciosa	relatório prospectivo	benefício líquido e equidade	drift ou dano potencial
Impacto	ensaio de implementação	melhora de processo/resultado	fadiga de alerta ou inequidade

Interrupções foram tratadas como resultado científico. Falha de harmonização, baixa frequência de eventos, degradação externa ou risco de privacidade indicariam que a tarefa ou a base precisavam ser redefinidas antes de expor pacientes a decisões automatizadas.

5 CONCLUSÃO

Aprendizado federado e dados sintéticos formaram uma estratégia plausível para desenvolver modelos pós-operatórios em pessoas com diabetes sem centralizar prontuários. Evidências reais mostraram alta discriminação em mortalidade multicêntrica, desempenho federado variável entre centros e benefício especial para hospitais de menor volume. Dados sintéticos puderam preservar calibração e apoiar o pipeline, mas não substituíram validação em dados reais. O estudo proposto deverá harmonizar definições, avaliar calibração, AUPRC, benefício líquido, equidade e privacidade, e avançar por etapas até avaliação prospectiva. A inovação relevante não será apenas o algoritmo, mas a capacidade de compartilhar aprendizado mantendo responsabilidade institucional e segurança do paciente.

REFERÊNCIAS

1. Martin ET, Kaye KS, Knott C, Nguyen H, Santarossa M, Evans R, et al. Diabetes and risk of surgical site infection: a systematic review and meta-analysis. Infect Control Hosp Epidemiol. 2016;37(1):88-99.

2. Tan DJH, Yaow CYL, Mok HT, Ng CH, Tai CH, Tham HY, et al. The influence of diabetes on postoperative complications following colorectal surgery. Tech Coloproctol. 2021;25:513-531. PMID:33386511.

3. Slump J, et al. Risk factors for postoperative wound complications after extremity soft-tissue sarcoma resection: a systematic review and meta-analysis. J Plast Reconstr Aesthet Surg. 2019;72(9):1449-1464. PMID:31302071.

4. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: machine learning prediction of postoperative complications. Ann Surg. 2018;268(4):574-583.

5. Shickel B, et al. Dynamic predictions of postoperative complications from explainable, uncertainty-aware and multi-task deep neural networks. Sci Rep. 2023;13:1224. PMCID:PMC9867692.

6. Lee SW, Lee HC, Suh J, Lee KH, Lee H, Seo S, et al. Multi-center validation of machine learning model for preoperative prediction of postoperative mortality. NPJ Digit Med. 2022;5:91. doi:10.1038/s41746-022-00625-6.

7. Maurer MM, et al. Privacy preserving federated learning for 90-day mortality prediction after surgery. NPJ Digit Med. 2025. PMCID:PMC12695362.

8. Feng B, Shi J, Huang L, Yang Z, Feng ST, Li J, et al. Robustly federated learning model for identifying high-risk patients with postoperative gastric cancer recurrence. Nat Commun. 2024;15:742. doi:10.1038/s41467-024-44946-4.

9. Cao J, et al. Exploring the limits of localization: federated model stacking improves hospital-level prediction in a national research network. NPJ Digit Med. 2026;9. doi:10.1038/s41746-026-02634-1.

10. Qian Z, et al. Synthetic data for privacy-preserving clinical risk prediction. Sci Rep. 2024;14:24720. doi:10.1038/s41598-024-72894-y.

11. Yoon J, Drumright LN, van der Schaar M. EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records. NPJ Digit Med. 2023;6:141. doi:10.1038/s41746-023-00888-7.

12. Li J, Cairns BJ, Li J, Zhu T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. NPJ Digit Med. 2023;6:98. doi:10.1038/s41746-023-00834-7.

13. Kaabachi B, Despraz J, Meurers T, Otte A, Brügger U, et al. A scoping review of privacy and utility metrics in medical synthetic data. NPJ Digit Med. 2025;8:44. doi:10.1038/s41746-024-01359-3.

14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning. BMJ. 2024;385:e078378.

15. Moons KGM, et al. PROBAST+AI: an updated quality, risk of bias and applicability assessment tool for prediction models using regression or artificial intelligence. BMJ. 2025.

16. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: CONSORT-AI. Nat Med. 2020;26:1364-1374.

17. McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. Proc AISTATS. 2017;54:1273-1282.

18. Zhu H, et al. FedWeight: mitigating covariate shift of federated learning on electronic health records through patient re-weighting. NPJ Digit Med. 2025;8. doi:10.1038/s41746-025-01661-8.

19. Balch JA, et al. Building an automated machine learning-enabled platform for postoperative complication risk prediction. Ann Surg. 2023;277(5):e1101-e1109. PMCID:PMC9910093.

20. Ren Y, et al. Performance of a machine learning algorithm using electronic health record data to predict postoperative complications and provide clinical decision support. JAMA Netw Open. 2022;5(5):e2213570. PMID:35576007.

21. El Chaar M, Stoltzfus J, Gersin K, Thompson K. A novel risk prediction model for 30-day severe adverse events and readmissions following bariatric surgery based on the MBSAQIP database. Surg Obes Relat Dis. 2019;15(7):1138-1145. doi:10.1016/j.soard.2019.03.005.

22. Nopour R, et al. Comparison of machine learning models to predict complications of bariatric surgery: a systematic review. Obes Surg. 2024;34:4011-4025. PMID:39282871.

23. Guan Z, et al. Artificial intelligence in diabetes management: current status and future directions. Diabetes Metab Syndr Obes. 2023;16:3697-3712. PMCID:PMC10591058.

24. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. NPJ Digit Med. 2023;6:186. doi:10.1038/s41746-023-00927-3.

25. Greenberg JK, et al. Leveraging artificial intelligence and synthetic data to expand electronic health record research in spine surgery. Global Spine J. 2023;13(8):2146-2154. PMCID:PMC10538345.

26. Sella N, et al. Preserving information while respecting privacy through synthetic data generation from electronic health records. NPJ Digit Med. 2025;8. doi:10.1038/s41746-025-01431-6.

27. Salim SS, Ibrahim AA. A machine learning approach for predicting 30-day hospital readmission in patients with diabetes. Healthcare. 2026;14(9):1185. doi:10.3390/healthcare14091185.

28. Aminian A, et al. Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery. Ann Surg. 2020;272(4):617-625. PMID:32029638.

CAPÍTULO 18

ATENÇÃO PRIMÁRIA MULTIPROFISSIONAL E CUIDADO INTEGRADO DE PESSOAS COM DOENÇAS CRÔNICAS: REVISÃO SISTEMÁTICA DE EVIDÊNCIAS, SÍNTESE META-ANALÍTICA POR DOMÍNIOS E INTEGRAÇÃO DA SAÚDE BUCAL

Aline Pinto Monteiro Costa Sousa; Danilo Ferreira de Sousa; Rafaela Nonato de Menezes; Maria Eveline Nascimento Pereira Vieira; Saulo de França Oliveira; Nélio Barreto Vieira

RESUMO

Objetivo: analisar a efetividade do cuidado multiprofissional e integrado na Atenção Primária à Saúde (APS) para pessoas com doenças crônicas e multimorbidade, identificar os componentes organizacionais associados a melhores resultados e examinar, de forma transversal, como a saúde bucal e a odontologia podem ser incorporadas às linhas de cuidado. Métodos: foi conduzida revisão sistematizada de revisões sistemáticas, meta-análises e estudos-chave, com busca principal em PubMed/MEDLINE, verificação em textos completos de periódicos, PubMed Central e documentos institucionais, além de rastreamento de referências. A atualização foi encerrada em 14 de agosto de 2026. Estimativas quantitativas foram extraídas de meta-análises contemporâneas sobre diabetes, hipertensão, depressão associada ao diabetes e funções da APS. Revisões potencialmente sobrepostas foram exibidas lado a lado, sem novo agrupamento estatístico. A integração da saúde bucal foi sintetizada separadamente devido à heterogeneidade de estratégias, desenhos e desfechos. Resultados: no diabetes, o cuidado em equipe reduziu a HbA1c em cerca de meio ponto percentual em duas sínteses independentes; em hipertensão, meta-análise de 61 ensaios e 64.857 participantes estimou redução de 6,31 mmHg na pressão sistólica e menor risco de pressão não controlada. O cuidado colaborativo para depressão e diabetes aumentou a resposta depressiva, com benefício pequeno em

qualidade de vida e efeito glicêmico inconsistente. Em revisão de 119 estudos de cuidado integrado, as intervenções com participação da APS melhoraram acesso e continuidade, mas não produziram benefício consistente para hospitalização, mortalidade ou qualidade de vida. Para saúde bucal, revisões recentes mostraram melhora de acesso, triagem, encaminhamento, higiene oral e capacidade dos profissionais não odontológicos; a certeza para desfechos cardiovasculares permaneceu baixa. Conclusão: equipes multiprofissionais podem melhorar desfechos de doenças crônicas, mas o benefício depende menos do número de profissões e mais da integração funcional entre elas. A odontologia deve participar de fluxos bidirecionais, com triagem de risco oral em consultas gerais e identificação de fatores cardiometabólicos no atendimento odontológico. Na APS, o modelo mais defensável combina plano compartilhado, registro comum, definição de responsabilidades, referência com contrarreferência e acompanhamento longitudinal.

Palavras-chave: Atenção Primária à Saúde; Equipe Multiprofissional; Doença Crônica; Cuidado Integrado; Saúde Bucal; Odontologia; Diabetes Mellitus; Hipertensão; Meta-análise.

ABSTRACT

Objective: to examine the effectiveness of multiprofessional and integrated primary care for people with chronic diseases and multimorbidity, identify organizational components associated with better outcomes, and assess how oral health and dentistry can be incorporated across care pathways. Methods: a systematicized review of systematic reviews, meta-analyses, and key studies was conducted, with a main PubMed/MEDLINE search, full-text verification on journal platforms and PubMed Central, institutional documents, and reference tracking. The update closed on August 14, 2026. Quantitative estimates were extracted from contemporary meta-analyses on diabetes, hypertension, depression with diabetes, and primary care functions. Potentially overlapping reviews were displayed side by side without second-order statistical pooling.

Oral health integration was synthesized separately because of heterogeneity in strategies, designs, and outcomes. Results: team-based diabetes care reduced HbA1c by approximately half a percentage point in two independent syntheses. In hypertension, a meta-analysis of 61 trials and 64,857 participants estimated a 6.31-mmHg reduction in systolic blood pressure and a lower risk of uncontrolled blood pressure. Collaborative care for depression and diabetes improved depression response, produced a small quality-of-life benefit, and had inconsistent glycemic effects. In a review of 119 integrated-care studies, interventions involving primary care improved access and continuity but did not consistently improve hospitalization, mortality, or quality of life. Recent oral-health reviews showed improved access, screening, referral, oral hygiene, and non-dental provider capability, while certainty for cardiovascular outcomes remained low. Conclusion: multiprofessional teams can improve chronic-disease outcomes, but benefit depends less on the number of professions than on functional integration. Dentistry should participate in bidirectional workflows, with oral-risk screening in general care and cardiometabolic risk identification in dental care. The most defensible primary-care model combines a shared plan, common documentation, role clarity, closed-loop referral, and longitudinal follow-up.

Keywords: Primary Health Care; Patient Care Team; Chronic Disease; Integrated Care; Oral Health; Dentistry; Diabetes Mellitus; Hypertension; Meta-analysis.

1 INTRODUÇÃO

A Atenção Primária à Saúde é o nível em que a maior parte das doenças crônicas precisa ser acompanhada por tempo suficiente para que pequenas decisões produzam consequências relevantes. Hipertensão, diabetes, doença cardiovascular, doença pulmonar crônica, transtornos mentais comuns e multimorbidade não são manejados apenas em consultas episódicas: exigem vigilância de sinais, ajuste de medicamentos, educação, mudança de

comportamento, prevenção de complicações, coordenação com especialistas e resposta rápida quando o risco se modifica. A complexidade aumenta quando essas tarefas são distribuídas entre diferentes serviços sem que exista um plano comum. O problema, portanto, não é simplesmente ampliar o número de profissionais, mas construir um sistema em que informações e responsabilidades acompanhem a pessoa ao longo do tempo.1-4

O cuidado multiprofissional emergiu como resposta à insuficiência do modelo centrado em um único profissional para condições que combinam fatores biológicos, comportamentais e sociais. Equipes podem incluir medicina, enfermagem, farmácia, nutrição, psicologia, fisioterapia, serviço social, agentes comunitários e odontologia. A presença de diferentes categorias permite distribuir tarefas, aumentar capacidade de acompanhamento e aproximar decisões do cotidiano do paciente. Entretanto, "multidisciplinar", "interprofissional", "colaborativo" e "integrado" não são termos equivalentes. Uma equipe pode ser multiprofissional e continuar fragmentada, com prontuários, agendas e metas independentes. A integração exige algum grau de trabalho conjunto: decisão compartilhada, comunicação clínica, referência com retorno, definição explícita de papéis e possibilidade de ajustar a conduta a partir da informação produzida por outro profissional.4,5

A evidência para doenças cardiometabólicas é particularmente informativa. Meta-análises de cuidado em equipe no diabetes mostram reduções clinicamente relevantes da hemoglobina glicada e melhora de pressão arterial e lipídios. Em hipertensão, revisões de ensaios randomizados apontam reduções sustentadas de pressão quando enfermeiros, farmacêuticos, médicos e outros profissionais atuam como uma unidade de cuidado.6-9 Ao mesmo tempo, estudos de integração mais ampla mostram que melhorar coordenação e experiência do paciente não garante, por si só, redução de internações ou mortalidade.10 Essa discrepância é útil: sugere que uma intervenção precisa alterar decisões clínicas, autocuidado ou acesso oportuno para que a reorganização do serviço produza efeito biológico mensurável.

A saúde mental ilustra outro componente da integralidade. Depressão e ansiedade podem reduzir adesão, atividade física, comparecimento a consultas e capacidade de organizar regimes terapêuticos. Em pessoas com diabetes e depressão, o cuidado colaborativo produz melhor resposta depressiva e pequeno ganho em qualidade de vida, enquanto o efeito sobre HbA1c é muito mais variável.11,12 Isso mostra que desfechos de diferentes domínios não caminham necessariamente em paralelo. A pessoa pode melhorar emocionalmente sem apresentar mudança glicêmica imediata, e uma redução de HbA1c não implica melhora do sofrimento associado ao tratamento. Modelos multiprofissionais precisam definir quais resultados pretendem modificar e em que horizonte temporal.

A odontologia ocupa posição peculiar nesse debate. Embora a boca seja parte do organismo e compartilhe fatores de risco com doenças crônicas, os sistemas de saúde frequentemente organizam o atendimento odontológico como circuito separado. Diabetes e periodontite apresentam relação bidirecional; tabagismo, dieta, privação social e inflamação atravessam doenças cardiovasculares e bucais; idosos com perda dentária podem apresentar pior capacidade mastigatória e dificuldades nutricionais; xerostomia pode ser agravada por polifarmácia.13-17 A separação administrativa entre medicina e odontologia, portanto, não corresponde à biologia nem às necessidades de pessoas com multimorbidade.

Revisões recentes sobre integração de saúde bucal mostram resultados positivos para triagem, acesso, educação e organização de encaminhamentos, mas ainda pouca evidência de que essas estratégias reduzam eventos cardiovasculares duros.14-18 Essa distinção deve ser preservada. Não é necessário alegar que uma consulta odontológica previne infarto para justificar sua integração à APS. Detectar sangramento gengival, dor, infecção, boca seca ou dificuldade mastigatória em pessoas com diabetes, e identificar hipertensão, risco metabólico ou tabagismo em pessoas atendidas pela odontologia, já amplia a capacidade de cuidado quando existe um fluxo real de resposta.

No Brasil, a discussão tem relevância operacional imediata. As equipes de Saúde da Família e de Atenção Primária podem atuar com equipes de Saúde Bucal e equipes multiprofissionais (eMulti). Em 2026, o Ministério da Saúde passou a monitorar de forma explícita ações interprofissionais compartilhadas da eMulti, reforçando que a multiprofissionalidade precisa ser observável no processo de trabalho e não apenas no cadastro da equipe.19-22 A estrutura institucional, entretanto, não elimina barreiras de agenda, interoperabilidade, definição de responsabilidade e contrarreferência. A questão prática é como transformar coexistência de equipes em continuidade assistencial.

Este capítulo analisou as evidências sobre cuidado multiprofissional e integrado em pessoas com doenças crônicas, utilizou estimativas meta-analíticas reais para os domínios com maior base quantitativa e incorporou a saúde bucal ao longo da arquitetura do cuidado. A intenção não foi produzir um ranking de profissões. O foco foi identificar mecanismos organizacionais capazes de explicar por que algumas equipes melhoram resultados e outras apenas redistribuem tarefas sem modificar a trajetória clínica.

Antes de detalhar os métodos, a Figura 1 apresenta a lógica adotada para o cuidado multiprofissional. A pessoa e a família foram colocadas no centro porque as profissões não funcionam como fins em si mesmas; elas oferecem competências que precisam convergir para problemas clínicos compartilhados. A odontologia foi posicionada no mesmo nível das demais áreas, evitando a ideia de um "anexo" de saúde bucal.

Figura 1. Rede multiprofissional centrada na pessoa na Atenção Primária à Saúde.



Fonte: elaboração do capítulo a partir das evidências citadas.

A Figura 1 evidencia que a integração acontece nas conexões, e não nos círculos isolados. Plano compartilhado, comunicação, seguimento e referência com contrarreferência são mecanismos transversais. Quando cada profissional avalia a pessoa, registra seus achados e encerra o atendimento sem modificar o plano comum, existe diversidade profissional, mas não necessariamente cuidado integrado.

2 OBJETIVOS

O objetivo principal foi analisar a efetividade de modelos multiprofissionais e integrados na APS para pessoas com doenças crônicas e multimorbidade, com atenção a desfechos clínicos, funções da APS, qualidade de vida, saúde mental e utilização de serviços. Os objetivos secundários foram comparar a magnitude de efeitos publicados para diabetes e hipertensão;

examinar o cuidado colaborativo em pessoas com depressão e diabetes; identificar componentes organizacionais associados a melhor desempenho; sintetizar estratégias de integração da saúde bucal; e propor um modelo de trabalho aplicável à APS brasileira, com participação explícita da odontologia.

3 MÉTODOS

3.1 Delineamento e princípio de síntese

Foi conduzida uma revisão sistematizada de evidências com síntese meta-analítica por domínios. O trabalho priorizou revisões sistemáticas e meta-análises capazes de representar perguntas clinicamente distintas: cuidado em equipe no diabetes; cuidado multiprofissional na hipertensão; cuidado colaborativo para depressão associada ao diabetes; integração de serviços com participação da APS; e integração da saúde bucal em ambientes não odontológicos. O relato foi orientado pelos princípios do PRISMA 2020, adaptados a uma síntese de revisões e evidências de implementação.23

A busca principal foi executada no PubMed/MEDLINE e complementada por textos completos em PubMed Central, páginas oficiais dos periódicos, documentos da Organização Mundial da Saúde e do Ministério da Saúde e rastreamento das listas de referências das sínteses elegíveis. A atualização final foi realizada em 14 de agosto de 2026. Foram utilizados termos relacionados a primary health care, multidisciplinary care, interprofessional care, team-based care, integrated care, chronic disease, diabetes, hypertension, multimorbidity, depression, oral health, dentistry, medical-dental integration e bidirectional referral. Publicações de referência anteriores foram mantidas quando continuavam necessárias para explicar conceitos ou comparar a evolução do campo.

3.2 Elegibilidade e unidade de análise

Foram elegíveis revisões sistemáticas, meta-análises e ensaios ou estudos de implementação considerados fundamentais para interpretar mecanismos de cuidado. A população de interesse foi composta por adultos com doenças crônicas acompanhados em APS, atenção ambulatorial ou modelos integrados com participação explícita da atenção primária. Estudos de hipertensão e diabetes foram priorizados para a síntese quantitativa porque dispunham de meta-análises recentes, com medidas diretamente interpretáveis. Para saúde bucal, foram incluídas revisões de integração oral–primária e oral–cardiovascular, além de estudos de implementação citados nessas sínteses.

A unidade quantitativa foi a estimativa meta-analítica publicada, não o participante individual nem o ensaio original. Essa decisão foi necessária porque as revisões de diabetes, hipertensão e cuidado integrado apresentam escopos e períodos parcialmente sobrepostos. Somar estimativas de revisões como se fossem independentes produziria dupla contagem e falsa precisão. Assim, os forest plots deste capítulo funcionaram como painéis de concordância entre sínteses, e não como uma nova meta-meta-análise.

A Tabela 1 explicita as definições operacionais utilizadas. Essa etapa foi importante porque parte da literatura chama de cuidado integrado intervenções que apenas adicionam um profissional, enquanto outras exigem coordenação, compartilhamento de informação e seguimento conjunto. Sem essa distinção, resultados de intervenções estruturalmente diferentes seriam interpretados como pertencentes à mesma categoria.

Tabela 1. Definições operacionais utilizadas para classificar modelos de cuidado.

Termo	Definição operacional no capítulo	Implicação para a análise
Multiprofissional	Dois ou mais grupos profissionais participam do cuidado da mesma população.	Não pressupõe, isoladamente, trabalho conjunto ou plano comum.

Termo	Definição operacional no capítulo	Implicação para a análise
Interprofissional	Profissionais de áreas distintas interagem deliberadamente para planejar, executar ou rever o cuidado.	Exige comunicação clínica e reconhecimento de papéis complementares.
Cuidado colaborativo	Modelo estruturado com responsabilidades compartilhadas, acompanhamento e ajuste orientado por metas; frequentemente empregado em saúde mental e doenças crônicas.	Foi analisado separadamente quando a intervenção possuía componentes próprios, como gestão de caso e revisão sistemática.
Cuidado integrado	Articulação entre setores, níveis ou serviços para oferecer continuidade centrada na pessoa.	O efeito foi examinado tanto nas funções da APS quanto em desfechos amplos.
Co-localização	Serviços diferentes funcionam no mesmo espaço físico.	Facilita comunicação, mas não foi tratada como sinônimo de integração.
Referência em circuito fechado	Encaminhamento inclui motivo, destino, confirmação de acesso e retorno de informação ao profissional/equipe de origem.	Foi considerado componente central para integração da saúde bucal.

Fonte: elaboração do capítulo com base na literatura de cuidado integrado e interprofissional.4,5,10,14

A Tabela 1 mostra que a diferença entre "ter várias profissões" e "trabalhar de forma integrada" é mensurável. Co-localização pode reduzir barreiras logísticas, mas, se o encaminhamento não retorna e as condutas não se comunicam, o paciente continua responsável por integrar sozinho o próprio cuidado. Essa distinção orientou toda a interpretação dos resultados.

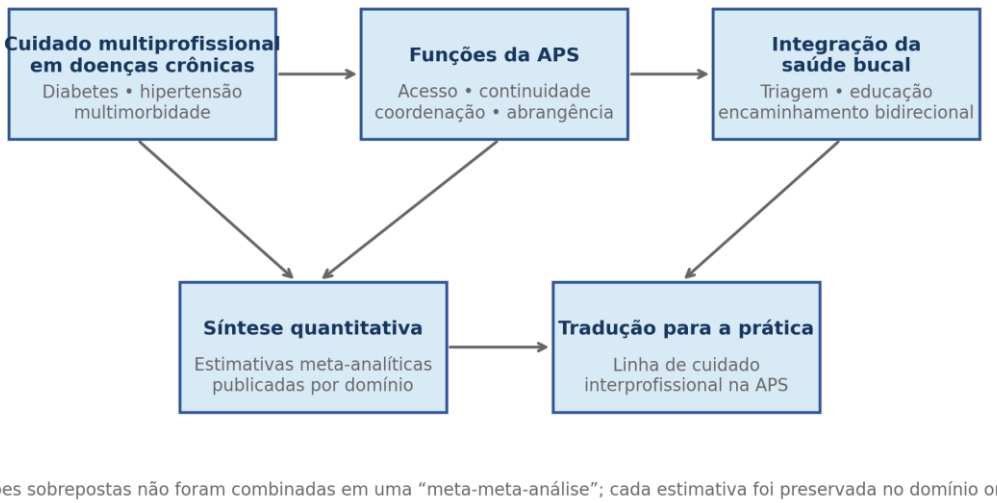
3.3 Extração, comparação e avaliação da certeza

Foram extraídos, quando disponíveis, número de estudos, tamanho amostral, população, composição da equipe, comparador, desfecho, medida de efeito, intervalo de confiança e heterogeneidade. Foram priorizadas estimativas ajustadas ou meta-analíticas conforme publicadas. Para diabetes e pressão arterial, diferenças médias foram mantidas nas unidades originais. Para depressão, razão de risco foi preservada. Para funções da APS e qualidade de vida, foram utilizadas medidas padronizadas conforme os artigos originais.

A confiança foi interpretada a partir do desenho das fontes, amplitude dos intervalos, consistência, risco de viés descrito pelos autores e aplicabilidade ao cenário de APS. Não foi criado um escore numérico próprio. A evidência sobre saúde bucal foi intencionalmente separada entre desfechos orais e de serviço, nos quais a consistência foi maior, e desfechos cardiovasculares, para os quais a revisão mais recente classificou a confiança como baixa.18

A Figura 2 resume a arquitetura de evidência utilizada. Ela substitui um fluxograma PRISMA artificial com contagens que não corresponderiam a uma revisão de novo de todos os ensaios e deixa explícito que o capítulo combinou três correntes: resultados clínicos de equipes, funções da APS e integração da saúde bucal.

Figura 2. Arquitetura da síntese de evidências utilizada no capítulo.



Fonte: elaboração do capítulo a partir das evidências citadas.

A organização da Figura 2 protegeu contra um erro comum em sínteses de segunda ordem: combinar estatisticamente revisões que reutilizam parte dos mesmos estudos. Em vez disso, os efeitos foram preservados em seus domínios e comparados por direção, magnitude e coerência. A saúde bucal foi integrada à

interpretação clínica, mas não forçada a uma metanálise quando os desenhos não sustentavam essa operação.

4 RESULTADOS

4.1 Panorama das evidências centrais

As fontes selecionadas mostraram dois padrões. Primeiro, intervenções orientadas a um problema clínico definido, como diabetes ou hipertensão, apresentaram efeitos mais nítidos sobre biomarcadores. Segundo, intervenções de integração ampla mostraram ganhos mais consistentes em acesso, continuidade e experiência do cuidado do que em mortalidade ou hospitalização. A saúde bucal seguiu padrão semelhante: triagem, educação, encaminhamento e desfechos orais melhoraram com maior frequência do que eventos cardiovasculares de longo prazo.6-11,14-18

A Tabela 2 organiza as principais sínteses que sustentaram o núcleo quantitativo e a integração odontológica. Ela foi construída para tornar visível que cada revisão respondeu a uma pergunta própria; por isso, o número de estudos e participantes não foi somado entre linhas.

Tabela 2. Principais fontes de evidência utilizadas na síntese.

Fonte	Escopo	Base quantitativa	Resultado central
Levengood et al., 2019	Cuidado em equipe no diabetes	35 estudos	HbA1c: −0,50 ponto percentual (IC95% −0,70 a −0,30).
Siaw & Lee, 2019	Cuidado colaborativo em diabetes não controlado	16 estudos	HbA1c: −0,55% (IC95% −0,65 a −0,45); PAS: −4,89 mmHg (IC95% −6,64 a −3,13).
Akasaki et al., 2025	Equipes no manejo da hipertensão	61 RCTs; 64.857 participantes	PAS: −6,31 mmHg (IC95% −7,71 a −4,90); RR de PA não controlada 0,73 (IC95% 0,68–0,79).

Fonte	Escopo	Base quantitativa	Resultado central
Zhang et al., 2025	Cuidado integrado e participação da APS	119 estudos; 69 intervenções com APS	Acesso: 0,17 (IC95% 0,05–0,29); continuidade: 0,32 (0,14–0,50); efeitos amplos inconsistentes.
Wang et al., 2022	Cuidado colaborativo em depressão + diabetes	12 RCTs; 3.109 participantes	Resposta depressiva RR 1,31 (1,23–1,39); QoL SMD 0,12 (0,03–0,21); HbA1c sem efeito consistente.
Harada et al., 2022	Promoção de saúde bucal por profissionais não odontológicos em DM2	5 estudos, 4 países	Melhora de desfechos orais em todos os estudos; heterogeneidade impediu meta-análise.
Usmani et al., 2026	Integração oral em APS/cuidado cardiovascular	17 estudos	Alta confiança para acesso e desfechos orais; baixa confiança para resultados cardiovasculares.

Fonte: Levengood et al.6; Siaw e Lee7; Akasaki et al.8; Zhang et al.10; Wang et al.11; Harada et al.15; Usmani et al.18.

A Tabela 2 indica uma diferença importante entre resultados proximais e distais. Quando a intervenção atua diretamente sobre titulação, adesão, educação e monitoramento, HbA1c e pressão arterial respondem. Quando o objetivo é reorganizar relações entre serviços, as mudanças aparecem primeiro em acesso e continuidade. Para saúde bucal, a evidência é mais madura para processo e condição oral do que para prevenção de eventos cardiovasculares.

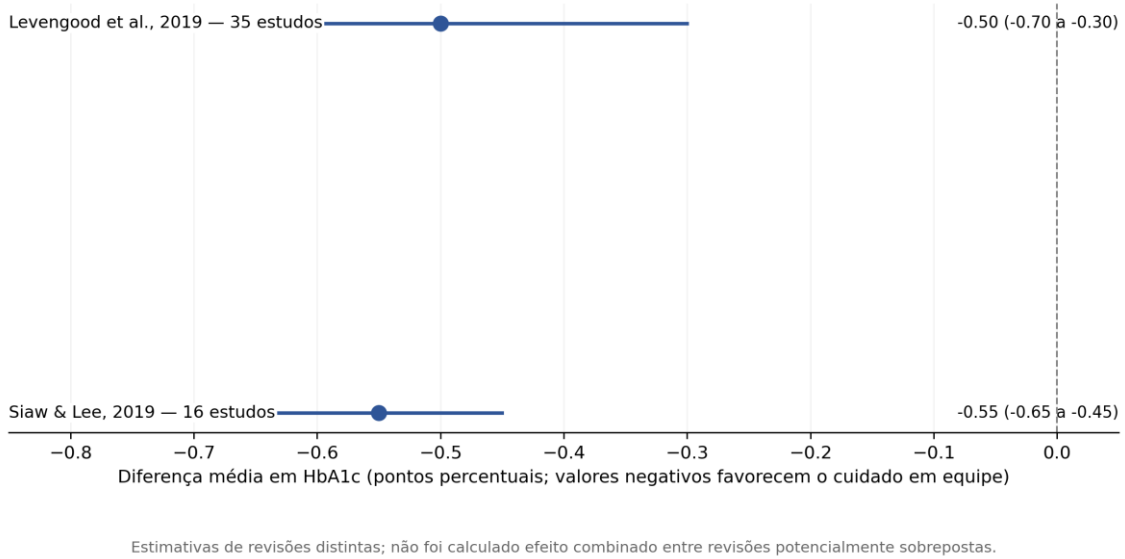
4.2 Diabetes: efeito do cuidado em equipe sobre o controle glicêmico

O diabetes foi o domínio com evidência histórica mais consistente para trabalho em equipe. Levengood e colaboradores incluíram 35 estudos e encontraram redução média de 0,50 ponto percentual na HbA1c em comparação com cuidado usual. A revisão de Siaw e Lee, restrita a pacientes com diabetes não controlado, encontrou estimativa muito semelhante, de 0,55 ponto percentual.6,7 Embora as revisões não devam ser combinadas por possível sobreposição e

diferenças de elegibilidade, a concordância de magnitude reforça a estabilidade do sinal.

A Figura 3 apresenta as duas estimativas em um painel comum. O objetivo não foi criar um novo efeito combinado, mas mostrar visualmente que sínteses com escopos distintos convergiram para uma redução próxima de meio ponto percentual.

Figura 3. Estimativas meta-analíticas publicadas para HbA1c em modelos de cuidado em equipe.



Fonte: estimativas publicadas por Levengood et al.6 e Siaw & Lee.7

Na Figura 3, os dois intervalos de confiança permanecem inteiramente abaixo da nulidade. A interpretação clínica não deve ser que qualquer composição de equipe produzirá exatamente esse efeito. As intervenções variaram em profissionais, intensidade, população e capacidade de modificar tratamento. O achado mais defensável é que redistribuir tarefas e criar acompanhamento estruturado pode produzir melhora glicêmica adicional quando o cuidado em equipe possui autoridade e mecanismos para agir sobre os dados do paciente.

A composição foi frequente com enfermagem e farmácia, profissões capazes de ampliar seguimento, educação, reconciliação medicamentosa e

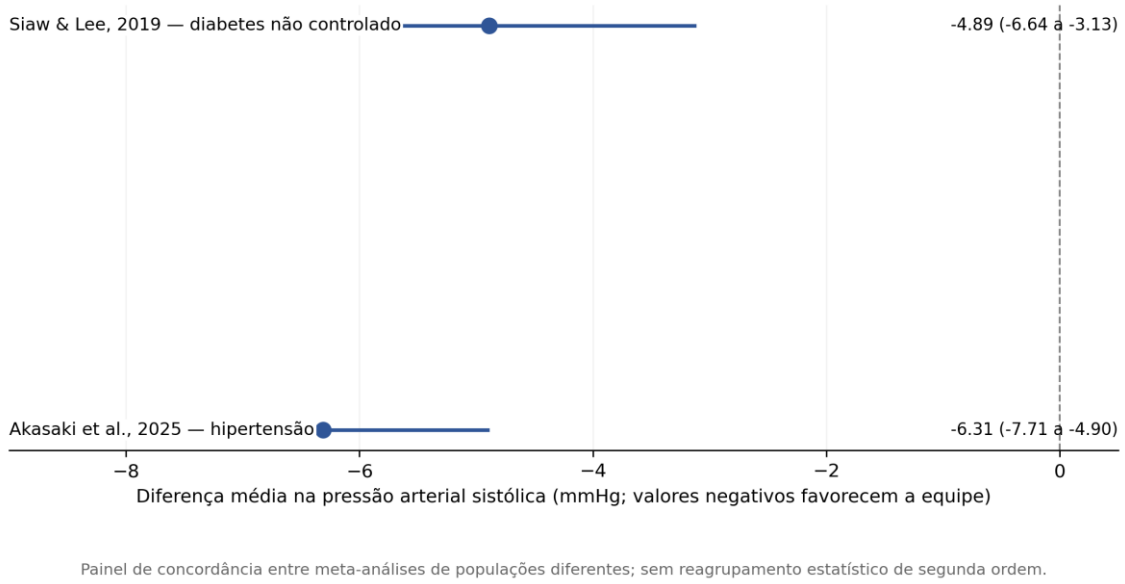
contato entre consultas.6 A presença da odontologia ainda foi rara nas meta-análises de controle glicêmico, apesar da relevância da doença periodontal para pessoas com diabetes. Essa lacuna não significa ausência de importância clínica, mas evidencia que a integração oral ainda foi avaliada mais como promoção, triagem e encaminhamento do que como componente formal dos ensaios de manejo metabólico.

4.3 Hipertensão: equipes e redução da pressão arterial

A meta-análise de Akasaki e colaboradores ampliou substancialmente a base de evidência para hipertensão. Sessenta e um ensaios randomizados, com 64.857 participantes, mostraram redução média de 6,31 mmHg na pressão sistólica e risco 27% menor de pressão não controlada.8 Os benefícios foram observados em diferentes arranjos de liderança, incluindo médicos, enfermeiros e farmacêuticos. A revisão de Siaw e Lee, em população com diabetes não controlado, encontrou redução de 4,89 mmHg, mostrando que o benefício pressórico também aparece quando a hipertensão é tratada dentro de uma condição metabólica mais complexa.7

A Figura 4 coloca essas duas estimativas em paralelo. Como as populações não são equivalentes, o gráfico não apresenta diamante combinado. Ele serve para comparar direção e ordem de grandeza do efeito.

Figura 4. Redução da pressão arterial sistólica em sínteses de cuidado multiprofissional.



Fonte: estimativas publicadas por Siaw & Lee7 e Akasaki et al.8

A Figura 4 mostra reduções consistentes de pressão sistólica, com intervalos que não atravessam zero. Em termos de organização do trabalho, esse resultado é compatível com intervenções em que a equipe melhora adesão, detecta necessidade de intensificação, reduz inércia terapêutica e mantém contato longitudinal.8,9 A presença de profissionais adicionais é mais útil quando eles conseguem fechar esse ciclo; apenas aferir pressão sem capacidade de revisar o plano tem valor limitado.

A odontologia pode participar desse ciclo em sentido inverso. Consultórios odontológicos recebem pessoas que podem não utilizar regularmente outros serviços e oferecem oportunidade para aferição de pressão e reconhecimento de fatores como tabagismo, obesidade e sinais de risco metabólico. A revisão de Usmani et al. descreveu experiências em que o ambiente odontológico ampliou a triagem de hipertensão e diabetes e encaminhou pacientes para a atenção primária.18 A contribuição do dentista, nesse caso, não é substituir o diagnóstico médico, mas funcionar como ponto adicional de identificação e entrada em um fluxo coordenado.

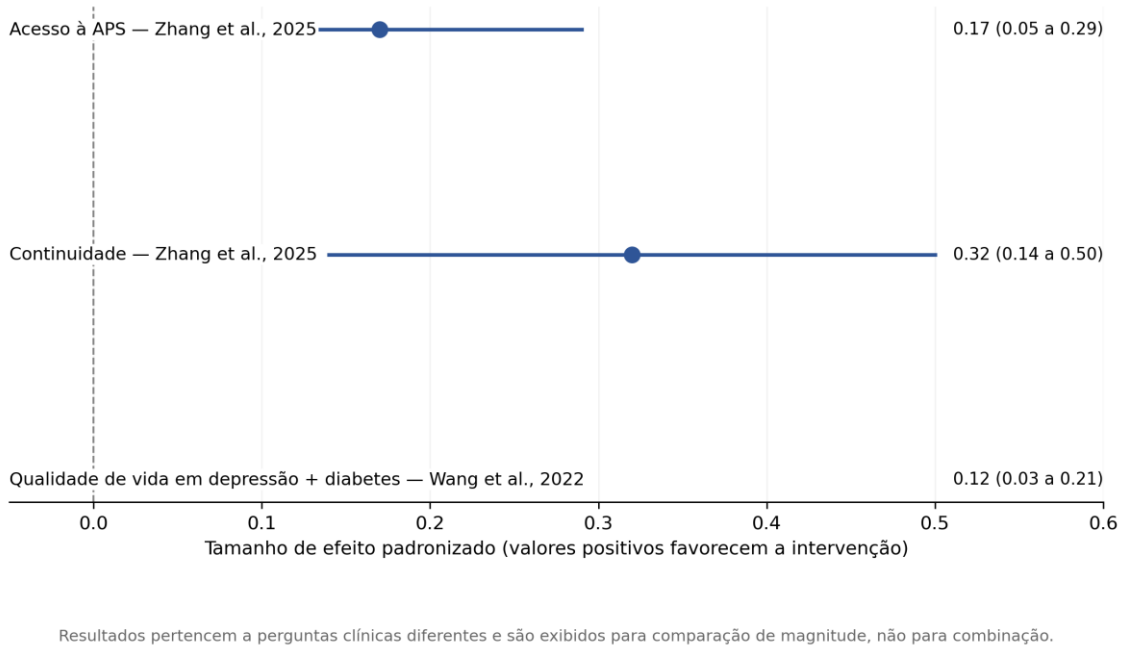
4.4 Saúde mental, multimorbidade e funções da Atenção Primária

A integração de saúde mental com cuidado de doenças crônicas mostrou efeitos mais seletivos. Wang e colaboradores reuniram 12 ensaios, com 3.109 participantes, e encontraram maior resposta da depressão no cuidado colaborativo (RR 1,31; IC95% 1,23–1,39). A qualidade de vida melhorou de forma pequena, enquanto a HbA1c não apresentou diferença consistente e mostrou heterogeneidade elevada.11 O ensaio TEAMcare demonstrou que intervenções estruturadas podem atuar simultaneamente sobre depressão e fatores cardiometabólicos, mas os resultados dependem de gestão ativa, metas e revisão de tratamento.12

Em uma escala organizacional mais ampla, Zhang et al. identificaram 119 estudos de cuidado integrado, dos quais 69 intervenções envolviam a APS. Os efeitos mais claros foram encontrados para acesso e continuidade; hospitalização, mortalidade e qualidade de vida não melhoraram de forma consistente.10 Isso reforça que integração é uma cadeia causal: primeiro o serviço precisa ser mais acessível e contínuo; depois, essa organização deve efetivamente alterar decisões clínicas ou comportamentos para modificar desfechos distais.

A Figura 5 compara três efeitos padronizados publicados: acesso, continuidade e qualidade de vida no cuidado colaborativo de depressão e diabetes. Apesar de pertencerem a perguntas diferentes, todos utilizam escalas padronizadas e ajudam a mostrar a magnitude relativamente pequena a moderada dos benefícios organizacionais e centrados na pessoa.

Figura 5. Magnitude de efeitos padronizados em funções da APS e qualidade de vida.



Fonte: Zhang et al.10 e Wang et al.11. Não houve combinação estatística entre os resultados.

A Figura 5 sugere que a continuidade apresentou a maior estimativa entre os três desfechos exibidos. Isso é coerente com a natureza do cuidado integrado: seu primeiro efeito provável é tornar o acompanhamento menos fragmentado. A ausência de grandes efeitos em qualidade de vida ou eventos clínicos não invalida essa reorganização, mas exige que sistemas de saúde não confundam melhoria de processo com benefício clínico já demonstrado.

O ensaio 3D de multimorbidade oferece exemplo importante. O modelo reorganizou consultas para aumentar foco na pessoa e integração entre problemas, melhorou experiência do cuidado, mas não mostrou melhora da qualidade de vida relacionada à saúde.24 A lição é aplicável à odontologia: inserir uma pergunta de saúde bucal no formulário de enfermagem é insuficiente se não houver acesso ao dentista, retorno de informação e seguimento. Um processo só se torna clinicamente relevante quando existe uma ação possível após a identificação do problema.

5 O QUE DIFERENCIA UMA EQUIPE DE UM CONJUNTO DE PROFISSIONAIS?

A literatura quantitativa não sustenta um ranking universal de profissões. Enfermeiros e farmacêuticos aparecem com frequência em ensaios de diabetes e hipertensão porque conseguem assumir tarefas estruturadas de monitoramento, educação, adesão e ajuste protocolizado. Psicologia é central quando depressão e ansiedade participam do problema clínico. Nutrição atua sobre dieta, peso e risco metabólico; fisioterapia sobre capacidade funcional; serviço social e agentes comunitários sobre acesso, contexto e continuidade. Odontologia identifica e trata doenças bucais e pode atuar como ponto de triagem para risco sistêmico. O valor de cada área depende da necessidade concreta e de sua conexão com o plano comum.

A Tabela 3 apresenta uma matriz funcional. Ela não define reserva de mercado nem pretende reduzir profissões a uma única tarefa. Seu objetivo é mostrar onde as competências se sobrepõem e onde a coordenação evita duplicidade ou lacunas.

Tabela 3. Funções possíveis das profissões no cuidado integrado de doenças crônicas na APS.

Profissão/área	Contribuições frequentes	Ponto de integração com a equipe
Medicina	Diagnóstico, estratificação, prescrição, investigação e coordenação de problemas complexos.	Revisão conjunta de metas, critérios de encaminhamento e escalonamento terapêutico.
Enfermagem	Acompanhamento longitudinal, educação, monitoramento, plano de autocuidado, busca ativa.	Pode funcionar como elo entre dados, pessoa e demais profissionais.
Farmácia	Reconciliação, adesão, interações, simplificação de regime e acompanhamento farmacoterapêutico.	Reduz falhas de uso e oferece retorno estruturado sobre efetividade e segurança.
Nutrição	Avaliação alimentar, metas nutricionais, obesidade, fragilidade e adequação cultural.	Integra dieta às metas de diabetes, pressão, saúde oral e funcionalidade.

Profissão/área	Contribuições frequentes	Ponto de integração com a equipe
Psicologia	Depressão, ansiedade, estresse, enfrentamento, mudança de comportamento.	Ajuda a separar dificuldade de adesão de barreiras emocionais tratáveis.
Fisioterapia/atividade física	Mobilidade, capacidade funcional, exercício terapêutico e reabilitação.	Conecta risco cardiometabólico a função, dor e autonomia.
Serviço social/ACS	Determinantes sociais, acesso, vulnerabilidade, território, vínculo e busca ativa.	Identifica por que um plano tecnicamente adequado pode ser inviável na vida cotidiana.
Odontologia/equipe de saúde bucal	Prevenção e tratamento de doenças bucais; periodonto; dor; infecção; mastigação; xerostomia; educação.	Recebe referência de risco oral e devolve conduta; pode identificar PA elevada, tabagismo e risco metabólico e referir à equipe.

Fonte: elaboração do capítulo a partir das meta-análises de equipe e das revisões de integração oral.6-11,14-18

A Tabela 3 reforça que integração se torna mais necessária justamente onde há sobreposição. Educação em saúde pode ser realizada por várias profissões; se cada uma entrega recomendações discordantes, a equipe aumenta carga cognitiva do paciente. Um plano compartilhado permite que mensagens se reforcem. Na saúde bucal, por exemplo, redução do tabagismo, melhora de dieta e controle glicêmico não são objetivos exclusivamente odontológicos ou médicos; são metas comuns com repercussões em múltiplos desfechos.

6 SAÚDE BUCAL COMO PARTE DA LINHA DE CUIDADO

6.1 Por que integrar odontologia e doenças crônicas?

A integração da saúde bucal pode ser justificada por três caminhos independentes. O primeiro é biológico: diabetes e periodontite interagem, e fatores como tabagismo, dieta e inflamação atravessam doenças bucais e cardiovasculares.13,16,17 O segundo é funcional: dor, perda dentária, doença periodontal, próteses inadequadas e xerostomia podem alterar alimentação, sono, comunicação, adesão e qualidade de vida. O terceiro é organizacional: pessoas podem circular pela atenção médica e odontológica em momentos

diferentes; cada contato é oportunidade de identificar necessidades que o outro serviço deverá manejar.

É importante evitar causalidade excessiva. A associação entre periodontite e doença cardiovascular é consistente, mas diretrizes científicas historicamente distinguem associação epidemiológica de prova de que o tratamento periodontal previna eventos cardiovasculares.17 Em contraste, a relação entre diabetes e doença periodontal possui base bidirecional mais estabelecida, e a integração do cuidado pode melhorar detecção, educação e manejo do risco oral.13,15 O capítulo adota essa gradação de certeza para não transformar integração de serviços em alegação terapêutica não demonstrada.

6.2 O que a evidência de integração mostrou

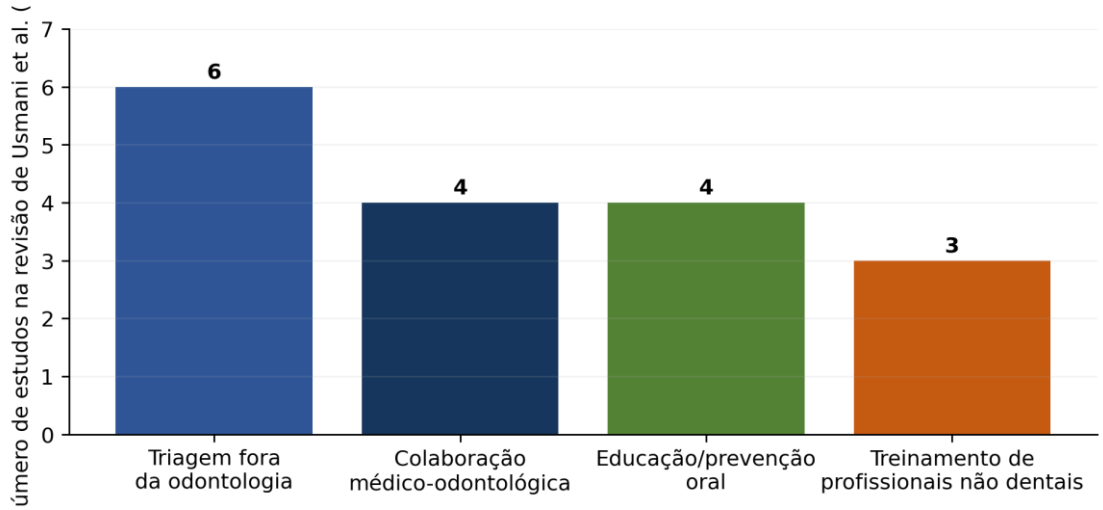
Harada et al. identificaram cinco estudos, em quatro países, nos quais profissionais não odontológicos — principalmente enfermeiros — promoveram saúde bucal em pessoas com diabetes tipo 2. As estratégias incluíram educação, lembretes para cuidado odontológico e protocolos em prática geral. Todos os estudos relataram melhora de algum desfecho oral, mas a heterogeneidade foi suficiente para impedir uma meta-análise responsável.15 Esse resultado é relevante porque mostra que tarefas simples de promoção e identificação podem ser distribuídas, sem substituir a avaliação odontológica quando há doença.

Christian et al. sintetizaram estratégias de integração da saúde bucal à atenção primária e encontraram, na maior parte dos estudos, resultados favoráveis em encaminhamento, documentação, eficiência e acesso.14 A atualização de Usmani et al., publicada em 2026 e direcionada à interface oral–cardiovascular, incluiu 17 estudos e organizou as intervenções em quatro grupos: colaboração médico-odontológica, triagem oral fora do ambiente odontológico, educação/prevenção e treinamento de profissionais não dentais. A revisão atribuiu alta confiança a melhorias de saúde bucal e acesso, confiança moderada

ao ganho de conhecimentos e habilidades dos profissionais e baixa confiança a desfechos cardiovasculares.18

A Figura 6 mostra a distribuição desses 17 estudos segundo as quatro estratégias principais. A maior categoria foi triagem fora da odontologia, seguida por colaboração, educação/prevenção e treinamento.

Figura 6. Estratégias de integração da saúde bucal identificadas por Usmani et al. em 2026.



Fonte: Usmani et al.18

A Figura 6 demonstra que integração não significa necessariamente colocar dentista dentro de toda consulta de doença crônica. Parte relevante da evidência utilizou ferramentas de triagem e educação por profissionais não dentais, combinadas a encaminhamento. O dentista permanece essencial para diagnóstico e tratamento odontológico; o ganho sistêmico ocorre quando outros profissionais conseguem reconhecer necessidade oral e quando a odontologia reconhece riscos que exigem resposta médica ou multiprofissional.

A Tabela 4 detalha resultados exemplares da literatura de integração oral. Ela separa desfechos de processo e saúde bucal de desfechos cardiovasculares para evitar que melhorias de acesso sejam apresentadas como prova de prevenção de eventos.

Tabela 4. Evidências selecionadas sobre integração da saúde bucal em modelos de atenção primária e cardiovascular.

Fonte/estratégia	Achado observado	Interpretação clínica
Harada et al., 2022 — promoção oral por não dentistas em DM2	5 estudos; todos relataram melhora de desfechos orais; 3 relataram melhora em algum desfecho do diabetes; sem meta-análise por heterogeneidade.	Profissionais não odontológicos podem promover e reconhecer risco, mas a base ainda é pequena.
Christian et al., 2023 — integração oral na APS	A maioria das estratégias favoreceu encaminhamento, documentação, eficiência ou acesso.	Integração deve ser avaliada por processos verificáveis, não apenas co-localização.
Usmani et al., 2026 — oral + cardiovascular	17 estudos; alta confiança para saúde bucal e acesso; baixa confiança para desfechos cardiovasculares.	Justifica integração por cuidado integral, mas não autoriza prometer prevenção de eventos cardíacos.
TISH — colaboração bidirecional	Sites odontológicos ampliaram triagem de hipertensão e encaminhamentos; triagem de gengivite na atenção primária foi mais difícil de incorporar.	A direção odontologia→APS pode amadurecer antes da direção APS→odontologia; ambos os fluxos precisam ser monitorados.
Programa Colorado de integração oral-cardiometabólica	Mais de 9.500 pacientes odontológicos foram avaliados; 60% receberam triagem de hipertensão e 16% avaliação de risco para diabetes.	O consultório odontológico pode funcionar como ponto adicional de detecção de risco cardiometabólico.

Fonte: Harada et al.15; Christian et al.14; Usmani et al.18 e estudos de implementação sintetizados nessa revisão.

A Tabela 4 indica que o benefício mais consistente da integração oral ocorre antes do evento cardiovascular: identificação, educação, acesso, documentação e referência. Esses resultados são clinicamente relevantes porque pessoas com doença crônica frequentemente apresentam necessidades bucais não tratadas. Ao mesmo tempo, a tabela deixa explícito que a cadeia causal até infarto, AVC ou mortalidade permanece insuficientemente demonstrada.

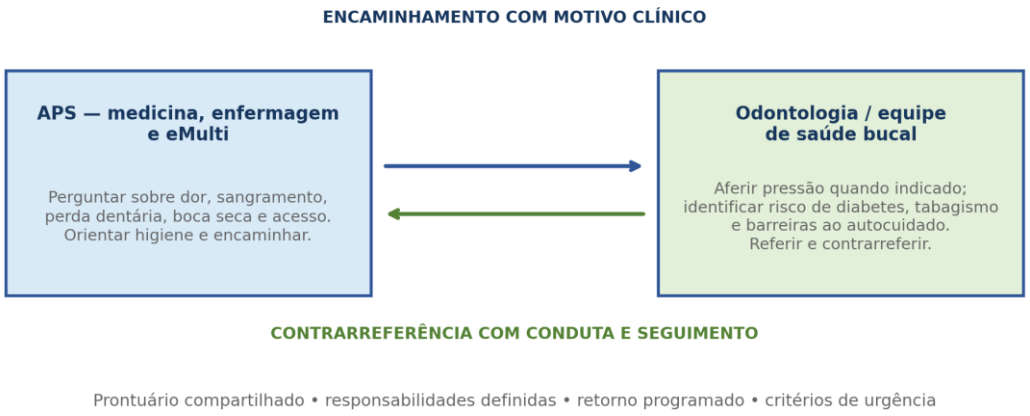
6.3 Fluxo bidirecional: do consultório geral para a odontologia e da odontologia para a APS

O modelo mais eficiente é bidirecional. Medicina, enfermagem, farmácia, nutrição e outros profissionais não precisam realizar exame periodontal

completo, mas podem perguntar sobre sangramento gengival, mobilidade dentária, dor, infecção, dificuldade mastigatória, boca seca e ausência prolongada de avaliação odontológica. Pessoas com diabetes mal controlado, tabagismo, fragilidade, perda dentária ou sinais de infecção podem ser priorizadas conforme protocolo local. A equipe de saúde bucal, por sua vez, pode aferir pressão quando apropriado, revisar tabagismo, reconhecer sinais de risco metabólico e confirmar se o paciente está vinculado à equipe responsável por suas doenças crônicas.

A Figura 7 transforma essa proposta em um circuito operacional. O ponto decisivo é a contrarreferência: o encaminhamento não termina quando a consulta é marcada. A equipe de origem precisa saber o que foi encontrado, o que foi tratado e qual seguimento é necessário.

Figura 7. Fluxo bidirecional proposto entre APS multiprofissional e odontologia.



Fonte: elaboração do capítulo a partir das evidências citadas.

A Figura 7 mostra que cada lado realiza apenas aquilo que cabe à sua competência, mas ambos participam da identificação de risco. Isso evita dois extremos: exigir que profissionais não dentais façam diagnóstico odontológico para o qual não foram treinados, ou manter a saúde bucal invisível até que o paciente procure espontaneamente o dentista. O mesmo raciocínio vale para a

odontologia: aferição de pressão ou detecção de risco de diabetes serve para encaminhar, não para substituir o manejo longitudinal da equipe de referência.

A Tabela 5 apresenta gatilhos práticos que podem ser incorporados a protocolos sem transformar a consulta em um checklist excessivo. A lógica é selecionar situações em que uma pergunta breve ou um achado rotineiro realmente muda o destino do cuidado.

Tabela 5. Gatilhos clínicos para integração da saúde bucal às linhas de cuidado de doenças crônicas.

Situação	O que a equipe geral pode observar/perguntar	O que a odontologia pode devolver à linha de cuidado
Diabetes	Sangramento gengival, dentes móveis, dor, infecção, boca seca, dificuldade de mastigar, última consulta odontológica.	Diagnóstico/tratamento oral, situação periodontal, infecção, impacto mastigatório, necessidade de manutenção e educação.
Hipertensão/doença cardiovascular	Dor/infeção oral, perda dentária, acesso odontológico; cuidado com medicamentos e procedimentos conforme risco individual.	PA elevada observada no atendimento, tabagismo, adesão referida, necessidade de avaliação médica antes de procedimentos quando indicado.
Tabagismo	Uso atual, prontidão para cessação e sintomas orais.	Reforço de cessação, lesões suspeitas, doença periodontal e necessidade de avaliação adicional.
Pessoa idosa/fragilidade	Perda dentária, prótese, engasgos, perda de peso, dieta limitada por mastigação, boca seca.	Capacidade mastigatória, condição de prótese, dor/lesões, higiene, risco de aspiração e necessidade de reabilitação.
Polifarmácia	Boca seca, alteração de paladar, dificuldade de deglutir, uso de medicamentos associados a risco oral.	Descrição de xerostomia/lesões e impacto funcional para revisão farmacoterapêutica quando pertinente.
Gestação	Dor, sangramento, acesso odontológico e medo de tratamento.	Situação oral, tratamento realizado, necessidade de seguimento e educação preventiva.

Fonte: síntese clínica baseada em Harada et al.15, Christian et al.14, Usmani et al.18 e na abordagem de fatores de risco comuns.16

A Tabela 5 deixa claro que integração não depende de equipamentos complexos. Perguntas simples podem revelar problemas com repercussão no

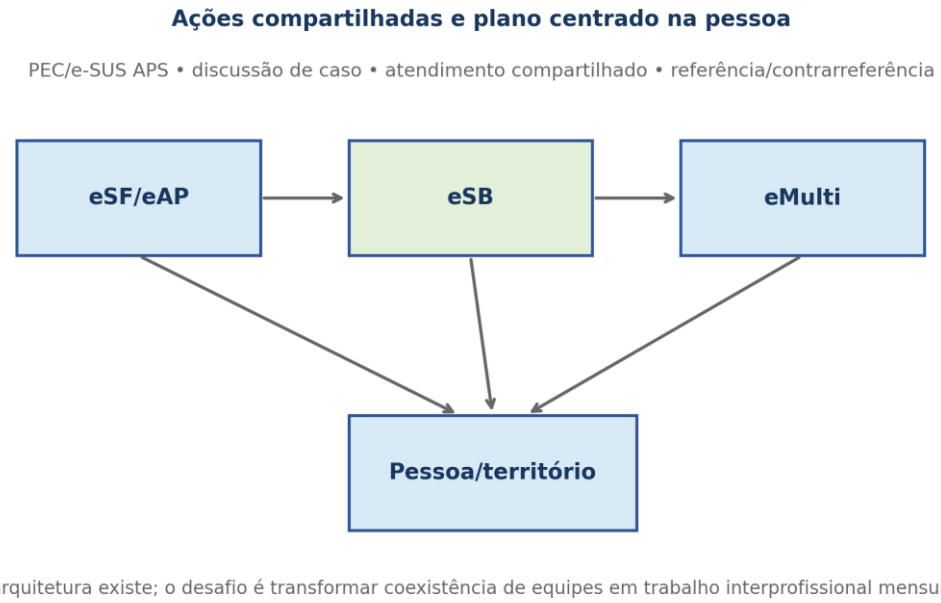
autocuidado, nutrição ou risco infeccioso. A qualidade, porém, depende da resposta disponível: rastrear sem oferecer acesso transforma necessidade identificada em mais uma responsabilidade transferida ao paciente.

7 APLICAÇÃO À ATENÇÃO PRIMÁRIA BRASILEIRA

A organização brasileira oferece uma oportunidade particular porque a saúde bucal pode estar formalmente vinculada à Estratégia Saúde da Família, e as eMulti operam de forma complementar às equipes de APS. O Ministério da Saúde descreve as eMulti como equipes de diferentes áreas que atuam de maneira complementar e integrada às demais equipes.19 A Saúde Bucal na APS é apresentada como parte da integralidade do SUS e articulada ao trabalho com famílias e comunidades.20 Em 2026, a ficha metodológica de ações interprofissionais da eMulti passou a monitorar atividades compartilhadas realizadas por pelo menos dois profissionais diferentes, tornando a interprofissionalidade um objeto de avaliação do processo de trabalho.21

A Figura 8 apresenta uma arquitetura possível. eSF/eAP, eSB e eMulti não foram organizadas em hierarquia; foram representadas como núcleos que compartilham uma pessoa e um território. O prontuário e as ações compartilhadas funcionam como infraestrutura de conexão.

Figura 8. Arquitetura proposta para integração entre eSF/eAP, eSB e eMulti na APS brasileira.



Fonte: elaboração do capítulo com base nas páginas e notas metodológicas vigentes do Ministério da Saúde em 2026.19-22

A Figura 8 evidencia que a existência administrativa das equipes não garante comunicação clínica. A transformação ocorre quando um problema identificado pela eSB modifica o plano da eSF/eAP, quando uma necessidade oral detectada pela enfermagem resulta em acesso à eSB e quando a eMulti participa de casos em que nutrição, saúde mental, reabilitação ou vulnerabilidade social precisam ser tratadas de forma conjunta.

Os indicadores atuais de Saúde Bucal na APS — como primeira consulta programada, tratamento concluído, taxa de exodontias, escovação supervisionada, procedimentos preventivos e tratamento restaurador atraumático — permitem acompanhar produção e qualidade específica da eSB.22 O próximo passo para integração é cruzar essas medidas com indicadores de cuidado de diabetes, hipertensão, gestação e pessoa idosa, evitando criar mais um conjunto de métricas paralelas. Uma gestão orientada à pessoa perguntaria, por exemplo, quantas pessoas com diabetes de maior risco permanecem sem

avaliação odontológica ou quantas pessoas com PA elevada identificada na odontologia tiveram avaliação confirmada na equipe de referência.

A Tabela 6 propõe um conjunto mínimo de ações para transformar a estrutura brasileira em integração funcional. Os itens são deliberadamente observáveis e podem ser auditados em prontuário, agenda ou reunião de equipe.

Tabela 6. Componentes mínimos para integração multiprofissional com saúde bucal na APS.

Componente	Como operacionalizar	Indicador possível
Plano compartilhado	Registrar problemas prioritários e responsabilidades por equipe/profissional.	Proporção de casos complexos com plano multiprofissional atualizado.
Triagem oral breve	Incluir perguntas de risco oral em linhas de diabetes, pessoa idosa e outras populações prioritárias.	Proporção de pessoas de risco com necessidade oral registrada e encaminhamento quando indicado.
Triagem cardiometabólica na odontologia	Aferir PA e/ou aplicar protocolo de risco quando indicado, sem substituir diagnóstico clínico.	Proporção de achados alterados com referência concluída.
Referência em circuito fechado	Registrar motivo, prioridade, confirmação de atendimento e contrarreferência.	Percentual de encaminhamentos com retorno de informação.
Atendimento compartilhado	Usar consultas, grupos ou discussão de caso com profissionais diferentes quando o problema exigir.	Ações interprofissionais registradas conforme metodologia vigente.
Revisão periódica	Reavaliar casos com diabetes descompensado, fragilidade, depressão, dificuldade de adesão ou doença oral ativa.	Tempo entre identificação e revisão do plano.
Educação coerente	Alinhar mensagens sobre dieta, tabagismo, higiene, medicação e autocuidado.	Registro de metas comuns e redução de orientações contraditórias.

Fonte: proposta de implementação elaborada a partir da evidência de integração e das normas operacionais brasileiras.10,14,18-22

A Tabela 6 traduz a ideia central do capítulo: integração precisa deixar rastros verificáveis. Uma equipe pode declarar trabalho conjunto, mas, se não existe plano compartilhado, referência retornada ou registro de ação conjunta, é

difícil saber se o paciente recebeu cuidado integrado ou apenas várias consultas. O uso de indicadores não deve burocratizar a clínica; deve revelar onde a comunicação está falhando.

8 DISCUSSÃO

A síntese mostrou que o cuidado multiprofissional é mais convincente quando está associado a um mecanismo capaz de modificar o tratamento. No diabetes, reduções de HbA1c próximas de meio ponto percentual apareceram em duas meta-análises independentes.6,7 Na hipertensão, o efeito médio sobre pressão sistólica foi ainda mais evidente, com redução superior a 6 mmHg em uma grande síntese de ensaios.8 Esses resultados não provam que "mais profissionais" sejam sempre melhores. Eles demonstram que determinadas configurações de equipe conseguem superar limitações do cuidado usual, sobretudo inércia terapêutica, baixa frequência de contato, dificuldade de adesão e ausência de seguimento estruturado.

A literatura de cuidado integrado amplo coloca um limite importante a interpretações otimistas. Zhang et al. encontraram benefício para acesso e continuidade, mas não para a maior parte dos desfechos amplos.10 O ensaio 3D melhorou experiência de cuidado sem melhorar qualidade de vida.24 Esses resultados mostram que reorganizar a estrutura é necessário, mas não suficiente. Uma equipe pode se reunir mais e continuar sem capacidade de alterar a conduta; pode documentar mais e não oferecer acesso; pode encaminhar mais e não saber se o paciente foi atendido. A efetividade depende da passagem entre informação e ação.

O cuidado colaborativo em depressão e diabetes reforça esse argumento. A melhora de depressão foi consistente, enquanto o efeito glicêmico não foi.11 A equipe deve aceitar que resultados pertencem a diferentes cadeias causais. Psicoterapia e gestão de caso podem reduzir sintomas depressivos antes de modificar comportamentos de autocuidado; mudanças de HbA1c podem

depender de titulação e acesso a medicamentos; qualidade de vida pode responder a fatores sociais que nenhuma intervenção clínica isolada controla. Avaliar multiprofissionalidade com um único desfecho produz uma visão incompleta.

A odontologia amplia essa discussão porque historicamente foi organizada em silo. A revisão de Christian et al. mostrou que estratégias de integração melhoram processos como encaminhamento e documentação.14 Harada et al. mostraram que profissionais não odontológicos podem promover saúde bucal entre pessoas com diabetes, mas a pequena base e a heterogeneidade impediram meta-análise.15 A revisão de Usmani et al., mais recente, consolidou quatro caminhos de integração e classificou como alta a confiança para melhoria de saúde bucal e acesso, mas baixa para benefícios cardiovasculares.18 Esse perfil é exatamente o que se espera de um campo de implementação ainda jovem: os efeitos aparecem primeiro perto da intervenção e tornam-se mais incertos à medida que o desfecho se distancia.

Essa incerteza não deve levar à exclusão da saúde bucal da APS. Ao contrário, ajuda a formular argumentos mais sólidos. Integrar odontologia é justificável porque doença oral é prevalente, interfere em função e qualidade de vida, compartilha determinantes com doenças crônicas e pode ser identificada em linhas de cuidado já existentes. Não é necessário utilizar uma narrativa de causalidade cardiovascular forte para defender uma referência odontológica adequada. Da mesma forma, o ambiente odontológico pode ampliar detecção de hipertensão e risco metabólico porque recebe uma população que nem sempre consulta outros profissionais com a mesma frequência.18

A abordagem dos fatores de risco comuns, proposta há décadas, continua útil: tabagismo, dieta, álcool, higiene, vulnerabilidade social e acesso afetam simultaneamente condições bucais e sistêmicas.16 Entretanto, integração contemporânea precisa ir além de educação genérica. O paciente já recebe múltiplas mensagens de autocuidado. O ganho real ocorre quando essas mensagens são coerentes, priorizadas e acompanhadas. Um nutricionista, um

dentista e um enfermeiro podem discutir alimentação por razões diferentes; o plano precisa evitar recomendações incompatíveis e transformar informação em uma meta viável para aquela pessoa.

A diversidade profissional também pode gerar efeitos indesejados. Mais consultas aumentam deslocamento, custos indiretos e carga terapêutica; orientações discordantes aumentam incerteza; registros separados podem duplicar exames; múltiplos encaminhamentos podem responsabilizar o paciente por coordenar o sistema. Por isso, a unidade de qualidade não deve ser o número de profissionais vistos, mas a proporção de necessidades resolvidas com o menor atrito possível. Essa perspectiva ajuda a interpretar por que algumas intervenções amplas de multimorbidade melhoram experiência, mas não necessariamente desfechos clínicos.

No SUS, a combinação eSF/eAP, eSB e eMulti cria uma infraestrutura promissora.19-22 A nota metodológica de 2026 para ações interprofissionais da eMulti é especialmente relevante porque reconhece que ações compartilhadas precisam ser registradas e monitoradas.21 O risco é transformar esse indicador em objetivo burocrático: duas assinaturas em uma atividade não garantem integração. O indicador ganha significado quando a ação compartilhada resolve um problema que exigia competências diferentes e quando a responsabilidade posterior permanece clara.

A saúde bucal pode ser usada como teste de maturidade da integração. Se uma pessoa com diabetes e periodontite passa pela consulta de enfermagem, medicina e odontologia sem que as equipes saibam o que as demais fizeram, o sistema é multiprofissional apenas no organograma. Se o risco oral é identificado, a consulta odontológica é priorizada, o tratamento é registrado, a equipe geral entende o impacto funcional e o seguimento é coordenado, existe integração clínica. O mesmo vale no sentido inverso quando uma pressão muito elevada ou um risco de diabetes identificado na odontologia gera avaliação e retorno confirmados.

Outro ponto é a equidade. A fragmentação prejudica mais pessoas com menor renda, baixa alfabetização em saúde, mobilidade reduzida, dificuldades de transporte e múltiplas condições. Elas têm menos capacidade de administrar referências abertas, telefonar para serviços diferentes e reconstruir informações clínicas entre consultas. A integração pode, portanto, produzir benefício distributivo mesmo quando o efeito médio de um biomarcador é pequeno. Essa dimensão raramente aparece nos forest plots, mas deve influenciar desenho e avaliação de programas.

A interpretação dos achados deve ainda considerar que a literatura de equipe frequentemente vem de contextos com recursos, escopos profissionais e regras de prescrição distintos. Resultados de enfermeiros e farmacêuticos em um sistema não são automaticamente transferíveis a outro. O mecanismo, entretanto, é transportável: ampliar frequência de contato, compartilhar informação, protocolizar resposta e reduzir inércia. No Brasil, a adaptação precisa respeitar atribuições profissionais, disponibilidade local, acesso à atenção especializada e capacidade do prontuário eletrônico.

A odontologia enfrenta desafio adicional de financiamento e cobertura em muitos países, razão pela qual revisões internacionais encontram experiências heterogêneas.14,18 A meta global da OMS de aumentar a disponibilidade de serviços essenciais de saúde bucal na atenção primária até 2030 posiciona integração como agenda de cobertura universal, e não apenas como inovação clínica.25 Essa orientação é coerente com modelos de APS que tratam saúde bucal como parte da saúde geral, mas sua implementação exige força de trabalho, financiamento, materiais e acesso real ao tratamento, não apenas triagem.

9 LIMITAÇÕES

Este capítulo apresenta limitações inerentes ao desenho de síntese de segunda ordem. A busca principal foi concentrada no PubMed/MEDLINE e complementada por rastreamento, textos completos e documentos institucionais;

portanto, não corresponde a uma revisão de novo com deduplicação de todas as bases bibliográficas. Para evitar falsa precisão, não foram apresentados números globais de registros recuperados nem um fluxograma PRISMA com contagens não auditáveis. As estimativas quantitativas foram extraídas de meta-análises publicadas e não recalculadas a partir de cada ensaio.

Outra limitação é a sobreposição potencial entre revisões. Essa característica impediu combinar Levengood, Siaw, Akasaki e outras sínteses em um único efeito. Os gráficos foram deliberadamente apresentados como painéis de concordância. Além disso, conceitos de cuidado integrado variaram entre estudos, composição e escopo das profissões mudaram entre países e parte dos efeitos pode depender de sistemas de remuneração ou atribuições profissionais específicas.

A literatura de integração da saúde bucal permaneceu heterogênea e incluiu ensaios, estudos quase experimentais e avaliações de serviço. Por isso, não foi realizada meta-análise oral própria. Resultados de programas de triagem em consultórios odontológicos e associações com utilização de emergência foram tratados como evidência de implementação, não como prova causal. A confiança em benefícios cardiovasculares de longo prazo foi mantida como baixa, em concordância com a revisão mais recente.18

10 IMPLICAÇÕES PARA PRÁTICA, GESTÃO E PESQUISA

Para a prática clínica, a primeira implicação é substituir encaminhamentos abertos por circuitos fechados. Toda referência relevante deveria responder quatro perguntas: por que o paciente foi enviado, qual a prioridade, se o atendimento ocorreu e qual informação voltou. A segunda é escolher tarefas que realmente podem ser compartilhadas. Triagem oral breve, aferição de pressão, educação para cessação do tabagismo e identificação de barreiras de adesão são exemplos de atividades que podem ser distribuídas sem diluir responsabilidade clínica.

Para gestão, indicadores devem medir conexão entre equipes. Além de produção odontológica e acompanhamento de hipertensão ou diabetes, serviços podem monitorar tempo de resposta, retorno de contrarreferência, ações compartilhadas e resolução de necessidades prioritárias. Casos complexos devem ser usados como unidade de aprendizagem: quando uma pessoa retorna repetidamente com descompensação, a equipe precisa perguntar se o problema é farmacológico, nutricional, psicológico, social, funcional ou oral, e se o fluxo atual permite responder a essa combinação.

Para pesquisa, são necessários ensaios pragmáticos e estudos de implementação que incluam odontologia como componente previamente especificado do cuidado de doenças crônicas. A agenda deveria comparar triagem isolada versus triagem com referência fechada; avaliar modelos com e sem acesso odontológico facilitado; medir desfechos orais e sistêmicos; incorporar custo, equidade e carga do paciente; e utilizar seguimento suficiente para distinguir ganhos de processo de efeitos clínicos. Estudos de componente podem esclarecer se treinamento, prontuário compartilhado, co-localização ou gestão de caso acrescentam benefício independente.

A Tabela 7 sintetiza uma agenda de pesquisa que evita repetir perguntas já suficientemente demonstradas — como se equipes podem reduzir pressão arterial — e concentra esforços em mecanismos de integração ainda incertos, especialmente na interface com saúde bucal.

Tabela 7. Prioridades para uma nova geração de estudos multiprofissionais na APS.

Pergunta	Desenho recomendado	Desfecho prioritário
A referência odontológica em circuito fechado supera o encaminhamento convencional?	Ensaio pragmático por cluster ou stepped-wedge.	Acesso concluído, resolução oral, tempo de resposta e experiência.

Pergunta	Desenho recomendado	Desfecho prioritário
A triagem cardiometabólica no ambiente odontológico aumenta diagnóstico/controle sem gerar excesso de encaminhamentos?	Ensaio pragmático ou implementação híbrida.	Confirmação diagnóstica, tratamento iniciado, PPV, carga do serviço e custo.
Integrar saúde bucal ao cuidado do diabetes melhora desfechos além da promoção isolada?	Ensaio com componentes bem definidos.	Periodonto, HbA1c, qualidade de vida oral, adesão e utilização.
Quais componentes do trabalho multiprofissional realmente importam?	Meta-análise de componentes ou desenho fatorial.	Efeito incremental de plano compartilhado, gestão de caso, titulação, prontuário e retorno.
A integração reduz iniquidades?	Coortes e ensaios com análise de heterogeneidade.	Acesso e benefício por renda, território, raça/cor, escolaridade e multimorbidade.
O ganho de processo se traduz em desfechos de longo prazo?	Seguimento ≥24 meses e vinculação a registros.	Hospitalização, eventos cardiovasculares, função, mortalidade e custos.

Fonte: elaboração do capítulo a partir das lacunas identificadas nas sínteses incluídas.10,14,15,18

A Tabela 7 mostra que a próxima etapa científica não deve ser simplesmente comparar "equipe versus não equipe". A pergunta mais útil é quais mecanismos produzem benefício, para quem e com que custo. Na saúde bucal, a prioridade é demonstrar se integração muda trajetórias clínicas e reduz barreiras, sem extrapolar de melhorias de higiene ou triagem para desfechos cardiovasculares ainda incertos.

11 CONCLUSÃO

A evidência disponível sustenta o cuidado multiprofissional como estratégia efetiva para condições crônicas quando a equipe possui mecanismos concretos de ação. No diabetes, sínteses independentes mostraram redução consistente da HbA1c; na hipertensão, grandes meta-análises mostraram redução

da pressão sistólica e maior controle. Em saúde mental associada ao diabetes, o cuidado colaborativo melhorou principalmente depressão e qualidade de vida. Em modelos amplos de integração, os benefícios foram mais claros para acesso e continuidade do que para mortalidade, hospitalização ou qualidade de vida.

Esses resultados convergem para uma conclusão organizacional: equipe multiprofissional não é sinônimo de cuidado integrado. O benefício aparece quando diferentes competências se conectam por metas compartilhadas, autoridade para responder aos achados, registro comum, seguimento e retorno entre serviços. Adicionar profissionais sem modificar essas relações pode aumentar complexidade sem melhorar a trajetória do paciente.

A odontologia deve ser incorporada a essa lógica como parte da APS, não como setor paralelo. A literatura já sustenta integração para ampliar acesso, triagem, educação, encaminhamento e resultados de saúde bucal, embora permaneça insuficiente para afirmar que esses modelos reduzem eventos cardiovasculares de longo prazo. O caminho mais seguro é bidirecional: equipes gerais reconhecem risco oral e garantem acesso; equipes odontológicas reconhecem sinais cardiometabólicos e conectam a pessoa ao cuidado longitudinal. No contexto brasileiro, eSF/eAP, eSB e eMulti oferecem uma arquitetura institucional capaz de sustentar esse modelo, desde que a interprofissionalidade seja traduzida em ações clínicas verificáveis e centradas na pessoa.

REFERÊNCIAS

1. World Health Organization; United Nations Children's Fund. Operational framework for primary health care: transforming vision into action. Geneva: WHO; 2020.

2. Starfield B. Is primary care essential? Lancet. 1994;344(8930):1129-1133. doi:10.1016/S0140-6736(94)90634-3.

3. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract. 1998;1(1):2-4.

4. Reeves S, Pelone F, Harrison R, Goldman J, Zwarenstein M. Interprofessional collaboration to improve professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2017;6(6):CD000072. doi:10.1002/14651858.CD000072.pub3.

5. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. Int J Integr Care. 2013;13:e010. doi:10.5334/ijic.886.

6. Levengood TW, Peng Y, Xiong KZ, Song Z, Elder R, Ali MK, et al. Team-Based Care to Improve Diabetes Management: A Community Guide Meta-analysis. Am J Prev Med. 2019;57(1):e17-e26. doi:10.1016/j.amepre.2019.02.005.

7. Siaw MYL, Lee JYC. Multidisciplinary collaborative care in the management of patients with uncontrolled diabetes: a systematic review and meta-analysis. Int J Clin Pract. 2019;73(2):e13288. doi:10.1111/ijcp.13288.

8. Akasaki Y, Suematsu Y, Azushima K, Shiga Y, Sakima A, Satoh M, et al. Impact of patient care teams on blood pressure control in patients with hypertension: a systematic review and meta-analysis. Hypertens Res. 2025;48(6):1827-1838. doi:10.1038/s41440-025-02152-9.

9. Bryant KB, Rao AS, Cohen LP, Dandan N, Kronish IM, Barai N, et al. Effectiveness and Cost-Effectiveness of Team-Based Care for Hypertension: A Meta-Analysis and Simulation Study. Hypertension. 2023;80(6):1199-1208. doi:10.1161/HYPERTENSIONAHA.122.20292.

10. Zhang Y, Stokes J, Anselmi L, Bower P, Xu J. Can integrated care interventions strengthen primary care and improve outcomes for patients with chronic diseases? A systematic review and meta-analysis. Health Res Policy Syst. 2025;23(1):5. doi:10.1186/s12961-024-01260-1.

11. Wang Y, Hu M, Zhu D, Ding R, He P. Effectiveness of Collaborative Care for Depression and HbA1c in Patients with Depression and Diabetes: a systematic review and meta-analysis. Int J Integr Care. 2022;22(3):12. doi:10.5334/ijic.6443.

12. Katon WJ, Lin EHB, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med. 2010;363(27):2611-2620. doi:10.1056/NEJMoa1003955.

13. Preshaw PM, Alba AL, Herrera D, Jepsen S, Konstantinidis A, Makrilakis K, Taylor R. Periodontitis and diabetes: a two-way relationship. Diabetologia. 2012;55(1):21-31. doi:10.1007/s00125-011-2342-y.

14. Christian B, George A, Veginadu P, Villarosa A, Makino Y, Kim WJ, et al. Strategies to integrate oral health into primary care: a systematic review. BMJ Open. 2023;13(7):e070622. doi:10.1136/bmjopen-2022-070622.

15. Harada Y, Prajapati D, Poudel P, Karmacharya B, Sugishita T, Rawal L. Effectiveness of oral health promotion interventions for people with type 2 diabetes delivered by non-dental health care professionals: a systematic review. Glob Health Action. 2022;15(1):2075576. doi:10.1080/16549716.2022.2075576.

16. Sheiham A, Watt RG. The common risk factor approach: a rational basis for promoting oral health. Community Dent Oral Epidemiol. 2000;28(6):399-406. doi:10.1034/j.1600-0528.2000.028006399.x.

17. Lockhart PB, Bolger AF, Papapanou PN, Osinbowale O, Trevisan M, Levison ME, et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association? Circulation. 2012;125(20):2520-2544. doi:10.1161/CIR.0b013e31825719f3.

18. Usmani W, Park C, Ogato S, Raymond M, de Courten MP, Lange PW, Hanna F. Breaking the silos: a systematic review of oral health integration strategies for improved oral health and cardiovascular outcomes. Front Public Health. 2026;14:1795955. doi:10.3389/fpubh.2026.1795955.

19. Brasil. Ministério da Saúde. Equipes multiprofissionais na Atenção Primária à Saúde (eMulti). Brasília: Ministério da Saúde; 2026 [acesso em 14 ago 2026].

20. Brasil. Ministério da Saúde. Saúde bucal na Atenção Primária à Saúde. Brasília: Ministério da Saúde; 2026 [acesso em 14 ago 2026].

21. Brasil. Ministério da Saúde. Nota metodológica: ações interprofissionais realizadas pela eMulti na Atenção Primária à Saúde. Brasília: Secretaria de Atenção Primária à Saúde; 2026.

22. Brasil. Ministério da Saúde. Indicadores da Equipe de Saúde Bucal na Atenção Primária à Saúde. Brasília: Ministério da Saúde; 2026 [acesso em 14 ago 2026].

23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

24. Salisbury C, Man MS, Bower P, Guthrie B, Chaplin K, Gaunt DM, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. Lancet. 2018;392(10141):41-50. doi:10.1016/S0140-6736(18)31308-4.

25. World Health Organization. Global strategy and action plan on oral health 2023–2030. Geneva: World Health Organization; 2024.

26. Proia KK, Thota AB, Njie GJ, Finnie RKC, Hopkins DP, Mukhtar Q, et al. Team-based care and improved blood pressure control: a Community Guide systematic review. Am J Prev Med. 2014;47(1):86-99. doi:10.1016/j.amepre.2014.03.004.

27. Prasad M, Manjunath C, Murthy AK, Sampath A, Jaiswal S, Mohapatra A. Integration of oral health into primary health care: a systematic review. J Family Med Prim Care. 2019;8(6):1838-1845. doi:10.4103/jfmpc.jfmpc_286_19.

SOBRE OS AUTORES E ORGANIZADORES

Danilo Ferreira de Sousa

Pós-Doutorado pela Universidade Federal de Alfenas. Doutor em Enfermagem pela Universidade Federal do Ceará; E-mail: prof.danilofersousa@gmail.com; ORCID: 0000-0001-9281-165X

Nélio Barreto Vieira

Doutor em Ciências da Saúde. Universidade Federal do Cariri – UFCA. Brasil; E-mail: nelio.barreto@ufca.edu.br; ORCID: 0000-0001-7100-0743

Conceição Soraya Morais Marques

Médica. Universidade Federal do Cariri; E-mail: conceicaosorayamorais@gmail.com; ORCID: 0009-0008-8596-711X

Rafaela Nonato de Menezes

Programa de Pós-Graduação em Biotecnologia pela Universidade Federal do Ceará; E-mail: rafaela.nonato@hotmail.com; ORCID: 0000-0001-7468-3391

Maria Eveline Nascimento Pereira Vieira

Nutricionista. Centro Universitário Maurício de Nassau; E-mail: eveline09-menp@hotmail.com; ORCID: 0009-0001-2230-1520

Regina Petrola Bastos Rocha

Doutorado em Ciências da Saúde pela Faculdade de Medicina do ABC. E-mail: rpetrola7@gmail.com; ORCID: 0000-0003-0626-232X

Leila Silveira Vieira Bezerra

Doutora pela Universidade de Fortaleza; E-mail: leila.silveira@ufca.edu.br; ORCID: 0000-0003-3979-4787

Paula Andrea Travecedo Ramos

Médica. Universidade Federal do Cariri; E-mail: paulatravecedo1@gmail.com ORCID: 0009-0002-8082-1011

Aline Pinto Monteiro Costa Sousa

Cirurgiã-dentista, Mestre em Saúde da Família pela Fiocruz Ceará e especialista em Saúde da Família e Ortodontia. Atua na Estratégia Saúde da Família de Fortaleza, com experiência em Saúde Bucal, Atenção Primária, gestão do cuidado, Educação Permanente e Epidemiologia. alinepmonteiro23@gmail.com

Sheyla Cristiane Xenofonte de Almeida

Doutorado em Biotecnologia em Recursos Naturais - RENORBIO (2017). CECAPE. sheylaxenofonte.prof@gmail.com

Bruno Xavier de Araújo

Cirurgião-Dentista. Programa de Pós-Graduação São Leopoldo Mandic.

Antônia Araújo Firmo

Mestrado Profissional em Gestão em Saúde. Universidade Estadual do Ceará – UECE.

Saulo de França Oliveira

Médico pela Universidade Estadual de Ciências da Saúde de Alagoas em 2014 e Residência Médica em Neurologia pela Universidade Federal Onofre Lopes da UFRN em 2019. saulo.neuro@gmail.com

Uilna Natércia Soares Feitosa

Doutora em ciências da Saúde na área de investigação clínica pela Faculdade de Medicina do ABC. Possui graduação em Médico Veterinário pela Universidade Estadual do Ceará (1998) e graduação em Enfermagem pela Universidade de Fortaleza (1991). uilna@hotmail.com

Edglê Pedro de Sousa Filho

Possui graduação em Medicina pela Universidade Federal do Ceará (1992) e doutorado em Ciências da Saúde pelo Centro Universitário Saúde ABC (2018). edglefilho@gmail.com

Davi Pedro Soares Macêdo

Acadêmico de medicina da Faculdade Paraíso Araripina-PE. davipedros123@gmail.com

Maria Isabel Pires Teixeira

Acadêmica de medicina da Faculdade de Petrolina -PE. belzinhapt@gmail.com



9786583199836ISBN 978-65-83199-83-6

thesis editora científica