

Compreendendo o GRADE: PICO e qualidade dos estudos

Understanding GRADE system: PICO and study quality

Leonardo Roeber¹, Mansueto Gomes-Neto², André Rodrigues Durães², Paulo Eduardo Ocke Reis³, Priscila Pollo-Flores³, Rose Mary Lisboa da Silva⁴, Elmiro Santos Resende¹

RESUMO

A saúde baseada em evidências se refere ao uso criterioso do conhecimento científico existente, oriundo de pesquisas clínicas, utilizando metodologias específicas que garantam solidez e clareza nas informações a serem aplicadas na tomada de decisão clínica. Dessa forma, reduzem-se as incertezas no julgamento clínico. O objetivo deste artigo foi descrever a metodologia PICO e a qualidade dos estudos com base no sistema GRADE.

Palavras-chave: Abordagem GRADE; Metodologia; Prática clínica baseada em evidências

ABSTRACT

Evidence-based health refers to the judicious use of existing scientific knowledge from clinical research, using specific methodologies that ensure solidity and clarity to the information to be applied in clinical decision-making, thus reducing uncertainties in clinical judgment. The objective of this article is to describe PICO methodology and the quality of studies in the GRADE system.

Keywords: GRADE approach; Methodology; Evidence-based practice

INTRODUÇÃO

A saúde baseada em evidências se refere ao uso criterioso do conhecimento científico, oriundo de pesquisas clínicas. Para isso, utilizam-se metodologias específicas, que garantam uma solidez e clareza às informações a serem aplicadas na tomada de decisão clínica, reduzindo as incertezas na tomada de decisão.^{1,2}

Essas metodologias de comparação são aplicáveis em situações, por exemplo, em que se deseja saber se as consequências desejáveis superam as indesejáveis.

Dessa forma, essas ferramentas de comparação favorecem a aplicação do conhecimento da forma mais imparcial possível.^{3,4}

O sistema denominado GRADE (do inglês *Grading of Recommendations Assessment, Development and Evaluation*) apresenta uma abordagem padronizada, que visa classificar o valor da evidência e a força das recomendações, sendo classificados como: críticas, essenciais e não críticos.⁵ Muitas sociedades internacionais adotam esse sistema como padrão no desenvolvimento de diretrizes para suas publicações (Tabela 1).⁵

¹ Universidade Federal de Uberlândia, Uberlândia, MG, Brasil.

² Universidade Federal da Bahia, Salvador, BA, Brasil.

³ Universidade Federal Fluminense, Rio de Janeiro, RJ, Brasil.

⁴ Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

Data de submissão: 18/2/2019. **Data de aceite:** 1/4/2019.

Autor correspondente: Leonardo Roeber. Avenida Pará, 1.720 – Umuarama – CEP 38400-902, Uberlândia, MG, Brasil – Tel.: 55 (34) 98803-9878
E-mail: leonardoroever@hotmail.com

Fonte de auxílio à pesquisa: nenhuma. **Conflitos de interesse:** nenhum.

Contribuição dos autores: Concepção e delineamento do projeto: LR, MGN, ARD, PEOR, PPF, RMLS e ESR.

Coleta, análise e interpretação de dados: LR, MGN e ARD.

Redação e revisão crítica do manuscrito: LR, MGN, ARD, PEOR, PPF, RMLS e ESR.

Aprovação da versão final do manuscrito a ser publicada: LR, MGN, ARD, PEOR, PPF, RMLS e ESR.

O SISTEMA PICO

PICO é uma acrônimo para paciente/população; intervenção (diagnóstica ou terapêutica), alternativa intervenção (comparação) e os resultados de interesse (desfechos, ou *outcomes*).

O primeiro passo para a elaboração de uma revisão sistemática é a formulação de uma pergunta de pesquisa de forma bem delimitada, além da identificação dos recursos apropriados, visto que a busca por evidências relevantes pode ser demorada.

Para isso, a pergunta deve apresentar os quatro componentes do acrônimo PICO: paciente/população; intervenção (diagnóstica ou terapêutica), alternativa intervenção (comparação) e os resultados de interesse (desfechos), visualizados na tabela 2.

Tabela 1. Qualidade da evidência^{8,9}

Fatores que podem reduzir a qualidade da evidência	
Fator	Consequência
Limitações no desenho ou execução do estudo (risco de viés)	↓ 1 ou 2 níveis
Inconsistência de resultados	↓ 1 ou 2 níveis
Indireto de evidência	↓ 1 ou 2 níveis
Imprecisão	↓ 1 ou 2 níveis
Viés de publicação	↓ 1 ou 2 níveis
Fatores que podem aumentar a qualidade das evidências	
Fator	Consequência
Grande magnitude de efeito	↑ 1 ou 2 níveis
Toda confusão plausível reduziria o efeito demonstrado ou aumentaria o efeito se nenhum efeito fosse observado	↑ 1 nível
Gradiente dose-resposta	↑ 1 nível

Fonte: adaptado de Puhan et al.⁸ e Iorio et al.⁹

Tabela 2. Elementos do PICO

PICO	Característica	Etapas para a formulação da questão de pesquisa
Desfecho	Estimar o efeito para cada desfecho	Delimitar a questão da pesquisa Escrever as informações necessárias na forma de uma pergunta Incluir na pergunta os elementos que compõem o acrônimo PICO: - P: Problema ou paciente ou população - I: Intervenção/indicador - C: Comparação - O: Resultado de interesse
Taxa de qualidade	Alta Moderada Baixa Muito baixa	Planeje uma estratégia de pesquisa identificando os elementos significativos da pergunta Traduza em termos de linguagem para descritores de assunto ou termos MeSH Para iniciar a pesquisa e manter os resultados da pesquisa inicial amplos, comece com os tópicos "P" e "I"
Recomendações	Forte: as vantagens de uma determinada conduta claramente suplantam suas desvantagens, ou então as desvantagens claramente suplantam as vantagens Frac: há certo grau de incerteza sobre a relação entre vantagens e desvantagens de uma dada conduta	Após os resultados da pesquisa inicial, faça uma pesquisa restrita com os termos da comparação, resultado, fatores de tempo ou tipo de estudo Veja resumos e textos completos de artigos para analisar a comparação e o resultado Faça o <i>download</i> da planilha booleana para planejar a pesquisa no banco de dados

MeSH: Medical Subject Headings.

Um dos potenciais problemas relacionados às perguntas norteadoras da revisão sistemática pode ser a falta da avaliação de todas as alternativas relevantes. Isso pode ocorrer, por exemplo, em casos em que as diretrizes de opções de tratamento variam em diferentes países.

Dois outros erros intimamente relacionados são a exclusão de resultados relevantes e a ênfase excessiva nos resultados substitutos com importância questionável.³ O quadro PICO facilita o processo de busca, identificando os fundamentos para uma estratégia eficaz.⁶⁻⁸

A magnitude de efeito dos resultados em termos relativos, como o risco relativo, a taxa de risco e a razão de chances, é apresentada na maioria dos estudos originais e revisões sistemáticas. No entanto, a diferença entre as consequências desejáveis e indesejáveis de uma recomendação requer o conhecimento dos efeitos em uma população específica.

O mérito do sistema PICO é a sua transparência sobre a lógica por trás da escolha, bem como equilíbrio de resultados desejáveis e indesejáveis, que determina a direção e a influência a força de uma recomendação. Esse equilíbrio depende da magnitude dos efeitos desejáveis e indesejáveis esperados e de como os resultados são valorizados pela população-alvo.¹¹

Embora valha a pena especificar os resultados críticos e essenciais, antes de começar a revisão das evidências, é preciso entender que eles podem influenciar nos julgamentos e na importância dos resultados.¹² As análises da revisão de evidências podem modificar a seleção de resultados relevantes ou sua importância relativa.

Quando um benefício potencial de um resultado, inicialmente julgado criticamente deixar de ser crítico

em uma revisão e dados outros benefícios estabelecidos, a intervenção pode ser julgada viável na ausência de um benefício demonstrado. Na triagem de uma doença, por exemplo, é provável que a avaliação de uma droga considere o impacto da intervenção em todas as causas de mortalidade como “crítica”.^{1,2} No entanto, o resumo das evidências estabelece uma redução essencial na mortalidade específica por causa de câncer de próstata, mas falha em estabelecer uma redução definitiva da mortalidade por todas as causas. Assim, a redução na mortalidade específica por causa pode ser julgada suficientemente convincente, mesmo na ausência de uma redução da mortalidade por todas as causas. A mortalidade por todas as causas então se torna menos relevante e deixa de ser um resultado crítico.⁶

Por outro lado, qualquer nova intervenção poderia estar associada com efeitos adversos que não estavam inicialmente presentes. Uma inesperada e significativa toxicidade poderia surgir em drogas aprovadas, e a não descoberta da toxicidade deve ser considerada como um efeito adverso consequente de qualquer novo medicamento. Dessa forma, a toxicidade se torna crítica somente quando evidências suficientes de sua existência emergem.⁶

AVALIAÇÃO DA QUALIDADE DOS ESTUDOS

Após a busca do PICO, o GRADE faz uma classificação preliminar dos resultados e garante que as diretrizes abordem as questões e a qualidade da evidência de maneira sistemática.^{6,11}

Desenvolvedores de diretrizes consideram recursos em ensaios controlados, como randomização, ocultação de alocação e cegamento. A análise de intenção de tratamento pode ser usada para gerar uma classificação. Em estudos observacionais, a medição apropriada da exposição e o controle adequado dos fatores de confusão devem ser avaliados. Em ambos os ensaios, controlados e observacionais, os desenvolvedores devem levar em conta a perda de acompanhamento, bem como outros aspectos da concepção, condução e análise da investigação que pode influenciar no risco de viés.⁷

Cada abordagem se destina ao uso dos dados disponíveis para verificar a probabilidade de um viés de publicação. Um registro prospectivo de todos os ensaios clínicos randomizados permite aos autores das revisões a avaliação quando os ensaios relevantes foram realizados, porque é possível pedir aos investigadores responsáveis que forneçam as informações dos dados do estudo. O registro obrigatório desses ensaios clínicos representa um dos métodos confiáveis para evitar um viés de publicação.^{7,8}

Consequentemente, a pesquisa de registros de ensaios clínicos deve ser considerada ao avaliar o risco de viés de publicação.⁸ A categorização final de evidências, baseadas no GRADE, não se relaciona necessariamente com a qualidade metodológica global de qualquer ensaio clínico controlado randomizado ou revisão. Porém, de certa forma, relaciona-se com a qualidade da evidência em um resultado específico na população de interesse.^{1,2}

Na literatura, a qualidade de uma pesquisa comumente se refere a um julgamento sobre a validade interna de um estudo individual. Embora outros significados tenham sido atribuídos à palavra “qualidade” (tipicamente risco de viés), esse significado deve corresponder mais de perto ao padrão e à compreensão não técnica da qualidade.¹¹

O GRADE não se refere a estudos individuais, mas a um corpo de evidências (por exemplo: alguns ensaios bem desenhados e executados), que pode estar associado a um baixo risco de viés, mas as confianças estimativas podem ser comprometidas por outros fatores (por exemplo: imprecisão, inconsistência indireta e viés de publicação). Também existem fatores particularmente relevantes para estudos observacionais, que permitem a avaliação da qualidade, incluindo a magnitude do efeito do tratamento e a presença de um gradiente dose-resposta. Estudos bem conduzidos podem fazer parte de um conjunto de evidências classificadas como de baixa qualidade porque fornecem apenas evidências indiretas ou imprecisas para a questão de interesse.¹¹

Os resultados podem ser classificados, na abordagem GRADE, como de interesse importante, não crítico ou não relevante. Essas recomendações auxiliam especialistas, clínicos e pacientes a interpretar os diferentes valores ou resultados dos estudos. Por exemplo, os resultados como mortalidade ou qualidade de vida podem ser considerados sintomas críticos, a julgar por um profissional de saúde, ou valioso, mas não crítico, ou não relevante, mas talvez informativo, para fazer uma recomendação.³

No sistema GRADE, é importante identificar o desenho do estudo para julgar a qualidade da evidência. Dessa forma, um corpo de evidências obtidas a partir de ensaios clínicos, inicialmente, é classificado como de alta qualidade. Por sua vez, aqueles obtidos a partir de observações de estudos são classificados como sendo de baixa qualidade.^{1,2}

Estudos bem projetados e executados, fornecem qualidades diferentes de evidências dos estudos mal conduzidos, sendo ensaios clínicos randomizados ou estudos observacionais. Assim sendo, não é possível confiar apenas nas limitações pelo desenho do estudo.

O GRADE fornece critérios adicionais de qualidade que servem para superar essa lacuna, como será descrito nos próximos tópicos.^{1,2}

LIMITAÇÕES NO DESENHO DO ESTUDO E/OU NA EXECUÇÃO

A qualidade dos estudos diminui na presença de limitação metodológica, que podem influenciar nas estimativas do efeito do tratamento. Citam-se como limitações: a falta ou ocultação de alocação ou cegamento (resultados subjetivo e avaliação altamente suscetível a vieses); a contabilização de uma proporção significativa de pacientes que iniciaram o estudo (perda considerável no seguimento ou o resultado não medido em uma proporção significativa de pacientes); a falha na adesão

ao princípio da intenção de tratar durante a análise; o pareamento para obter benefícios; o relato seletivo dos resultados que mostram um aparente efeito de tratamento e falha a em relatar resultados que não mostram nenhum efeito evidente.^{1,2}

Limitações metodológicas implicam em uma maior propensão a vieses, diminuindo a confiança na estimativa de efeito. Na presença de limitações, elas podem ser classificadas em graves ou muito graves, reduzindo, respectivamente, em 1 ou 2 níveis sua graduação de qualidade, de acordo com sua propensão qualitativa de enviesar os resultados.

Na tabela 3, são descritas as potenciais limitações observadas em ensaios clínicos.

Na tabela 4, são elucidadas as principais limitações encontradas em estudos do tipo observacionais.

Tabela 3. Potenciais limitações em ensaios clínicos

Tipo de limitação	Descrição
Sigilo de alocação ou randomização inadequada	Estudos não randomizados: os quasi-experimentos, apesar de iniciarem com pontuação máxima devido ao delineamento, são obrigatoriamente penalizados nesse critério. Métodos de randomização não aleatórios, como a data de nascimento, ordem de chegada. Alocação não sigilosa, como listas abertas com sequência de randomização, que permita ao investigador inferir sobre o grupo de intervenção destinado ao participante antes do mesmo ser incluído no estudo.
Mascaramento inadequado ou ausente	O mascaramento inadequado ou ausente pode implicar em viés, especialmente quando referente a avaliadores de desfechos, como pode ocorrer, por exemplo, na ausência de mascaramento dos avaliadores em desfechos avaliador-dependentes, como qualidade de vida e escalas psicométricas.
Descrição inadequada de seguimento de todos os pacientes inicialmente randomizados	Ocorre nos casos em que há perda de seguimento (em especial quando superior a 10%-20%) e na ausência de relato adequado do motivo e dos grupos das perdas quando presentes. Também é descrita nos casos em que a análise não é realizada por intenção de tratamento, quando indicada.
Relato seletivo de desfechos	Identificado nos casos em que os desfechos são relatados de acordo com os resultados encontrados, por exemplo, com hipóteses não definidas a priori.
Fenômeno <i>carry-over</i> em estudos cruzados (<i>cross-over</i>)	A intervenção realizada na primeira fase de um estudo cruzado interfere na segunda fase, como, por exemplo, as medidas educacionais ou uso de antibióticos que podem alterar a flora bacteriana.
Interrupção precoce do estudo	Ocorre quando há interrupção precoce de ensaio clínico devido a algum benefício, que pode ocorrer em especial na ausência de regras predefinidas para interrupção dos estudos.
Viés de seleção em ensaios clínicos com randomização de <i>clusters</i>	Ocorre nos casos em que há randomização de grupos de <i>clusters</i> não comparáveis. Por exemplo, a randomização de UTIs em vez de pacientes, com os grupos resultantes não adequadamente comparáveis entre si.

Fonte: Bertolaccin et al.,¹ Brasil.²
UTI: unidade de terapia intensiva.

Tabela 4. Tipos de vieses encontrados em estudos observacionais

Tipo de viés	Descrição
Vieses de seleção	Ocorre quando há pareamento ou seleção inadequada em estudos de caso-controle ou quando há seleção de indivíduos expostos e não expostos procedentes de diferentes populações em estudos de coorte.
Vieses de aferição	Diferenças nas medidas da exposição nos diferentes grupos, como os vieses de recordação em estudos de caso-controle. Vigilância diferente de desfechos em expostos e não expostos.
Vieses de confusão	Presença de potenciais confundidores não avaliados. Análise estatística sem adequado ajuste de confundidores.
Seguimento inadequado	Perdas de seguimento (em especial superiores a 20%). Período de seguimento pequeno para a associação em avaliação.

Fonte: Bertolaccin et al.,¹ Brasil.²

INCONSISTÊNCIA DOS RESULTADOS

A qualidade das evidências diminui se forem amplamente diferentes as estimativas de efeito do tratamento em estudos individuais (variabilidade ou heterogeneidade dos resultados) e sugerem diferenças exatas nos efeitos do tratamento. A variabilidade nos resultados individuais do estudo pode apresentar diferenças clínicas nas populações, intervenções e resultados medidos ou de limitações metodológicas (problemas com randomização ou término antecipado dos ensaios).^{1,2}

É importante, ao realizar metanálises ou em conjuntos de estudos individuais, avaliar a inconsistência dos resultados, o que significa que estudos com hipóteses semelhantes possuem estimativas diferentes do que se esperaria simplesmente pelo acaso. Quando há inconsistência importante nos resultados e as razões para essas diferenças não são adequadamente exploradas e explicadas, o efeito encontrado pode não corresponder à realidade, devendo a evidência ter seu nível de qualidade reduzido em um ou dois níveis.^{1,2}

Essa variabilidade dos resultados geralmente provém de heterogeneidade (diferença) entre os estudos referentes a:

- População: por exemplo, o ácido acetilsalicílico (AAS) possui maior benefício na prevenção de eventos cardiovasculares em indivíduos mais velhos.
- Intervenções: diferenças no tipo, na dose ou na via de administração de uma determinada intervenção. Como exemplo, citam-se as medicações de liberação prolongada, que têm melhores benefícios, pois implicam em maior aderência ao tratamento.
- Desfechos: critérios diferentes para avaliação de um mesmo desfecho, diminuição do efeito do tratamento com o aumento do tempo de observação. Por exemplo, ocorre um benefício maior no uso da sibutramina na perda de peso nas 8 semanas iniciais de administração.

Em uma metanálise, há testes que auxiliam na identificação da variabilidade entre os estudos. O teste de heterogeneidade detecta a presença de heterogeneidade estatística na amostra (presente quando $p < 0,1$) e o teste de inconsistência (I²), considerado inconsistência alta, em geral, quando superior a 50%. Esses testes podem auxiliar na análise da presença de diversidade entre os estudos.^{1,2}

A inconsistência, por vezes, pode ser explicada pela análise de subgrupo e metarregressão. Contudo, a presença de inconsistência não explicada ou parcialmente explicada pode prejudicar a confiança na medida de associação agregada encontrada, diminuindo a qualidade da evidência.^{1,2}

INDIREÇÃO DE EVIDÊNCIA

O GRADE distingue dois tipos de indireção, conforme tabela 5.

IMPRECISÃO DOS RESULTADOS

As estimativas de efeito geralmente têm amplos intervalos de confiança que incluem tanto benefícios essenciais como efeitos não essenciais (ou mesmo danos) quando os estudos incluem, relativamente, que poucos eventos ocorrem.

Essa incerteza na indeterminação dos resultados pode diminuir a qualidade da evidência e corresponde ao erro aleatório, que gera imprecisão na estimativa (intervalos de confiança amplo) devido ao pequeno tamanho amostral ou ao pequeno número de eventos. Essa imprecisão dos resultados prejudica a qualidade da evidência pela incerteza na magnitude do benefício ou do risco da intervenção.^{1,2}

A decisão de penalizar os estudos devido à imprecisão se deve a uma série de fatores, como o tamanho de amostra, a raridade do evento etc., devendo ser avaliada qualitativamente para cada situação. Entretanto, em geral, sugere-se imprecisão nos resultados quando: em uma metanálise, o tamanho cumulativo da amostra é inferior ao tamanho ótimo da informação calculado, o número de eventos totais é pequeno (inferior a 300); em desfechos contínuos (por exemplo: qualidade de vida), a

Tabela 5. Tipos de indireção baseados no sistema GRADE

Comparação indireta	Uma comparação indireta fornece evidências de menor qualidade do que uma comparação frente a frente e é comum quando se escolhe entre as drogas da mesma classe Exemplo: não há estudos comparando diretamente a intervenção A versus B, contudo há estudos comparando intervenção A versus placebo e intervenção B versus placebo. Esses estudos podem ser agregados para comparar a intervenção A e B (como é o caso se “multiple treatment comparisons” e “network metaanalysis”), contudo há perda substancial na confiança dos efeitos estimados
População, intervenção, comparadores ou desfechos indiretos	As evidências que apoiam a recomendação também são indiretas no caso de estudos em que população, intervenção, intervenção alternativa ou resultados de interesse são diferentes daqueles que a recomendação se refere Corresponde à transposição das evidências disponíveis de recomendações correlatas para a recomendação em questão proposta, ou seja, a evidência não corresponde exatamente ao mesmo cenário que se quer transpor

Fonte: Bertolaccin et al.,¹ Brasil.²

amostra total for inferior a 400 indivíduos; um intervalo de confiança de 95% apresenta ausência de efeito e considerável benefício ou risco (esse sendo geralmente considerado de aumento ou diminuição do risco relativo da ordem de 25%).^{1,2}

VIÉS DE PUBLICAÇÃO

A qualidade da evidência é reduzida caso os pesquisadores não informarem estudos em que nenhum efeito é apresentado, ou quando há resultados negativos. O risco de viés é maior se forem avaliados apenas poucos estudos ou estudos com pequena amostragem. Dessa forma, é preciso ter cautela quando apenas estudos pequenos relatam benefícios clínicos e estatisticamente significativos, pois podem estar associados a vieses.^{1,2}

Salienta-se que os estudos com resultados positivos são publicados em forma de artigo com maior frequência, em revistas de maior impacto, geralmente em língua inglesa e são publicados em um período menor (e por vezes gerando também mais do que uma publicação com a mesma base de dados). Dessa forma, tendem a compor com maior frequência as análises de revisões sistemáticas.

O relato seletivo de desfechos não é incomum, de forma que, muitas vezes, são publicados apenas os desfechos cujos resultados se mostraram favoráveis, desconsiderando os demais.^{1,2}

Em função disso, a metanálise está propensa a apresentar benefícios espúrios a favor de determinadas intervenções ou, então, são propensas a aumentar artificialmente sua magnitude, prejudicando as conclusões a respeito da questão de pesquisa.

Diversas alternativas são utilizadas para avaliar o viés de publicação, como os testes estatísticos (teste de Egger) e análises gráficas (*funnelplot*). Na presença de viés de publicação com potencial impacto nas conclusões dos resultados, a qualidade da evidência deve ser reduzida.^{1,2}

GRANDE MAGNITUDE DE EFEITO

O sistema GRADE define um risco relativo $<0,5$ ou >2 como um “efeito significativo”, e um risco relativo $<0,2$ ou >5 , como um “efeito substancial”, quando a análise se baseia em estudos de modelagem que fornecem estimativas da magnitude do efeito não explicado por um viés.^{1,2}

Quando os estudos, inclusive os observacionais, mas com forte rigor metodológico, estimam associações consistentes e com magnitude importante, a confiança dos resultados aumenta.^{1,2}

Quanto maior o tamanho do efeito, mais forte se torna a evidência. É importante enfatizar que os estudos que tiveram seus níveis de evidência reduzidos provavelmente possuem validade interna comprometida, sendo desaconselhável aumentar sua qualidade.^{1,2}

Para análise do efeito, é possível adotar os critérios descritos na tabela 6.

REDUÇÃO OU AUMENTO DO EFEITO DEMONSTRADO POR CONFUNDIMENTO PLAUSÍVEL

Todos os vieses plausíveis podem subestimar o efeito exato do tratamento, considerando as evidências de estudos observacionais de baixa qualidade. Uma situação paralela existe quando estudos observacionais não demonstram uma associação, mas todos os vieses aumentaram um efeito de intervenção e levantou efeitos prejudiciais aparentes.^{1,2}

A presença de confundidores que, se ajustados, provavelmente levariam a resultados na direção contrária à observada, aumenta a confiança a respeito da direção do efeito encontrado. Se o grupo recebendo uma determinada intervenção (ou exposição) for composto por pacientes mais doentes do que o respectivo grupo controle e mesmo assim obtiver melhores resultados, espera-se que, caso os grupos fossem iguais, o benefício da intervenção seria ainda maior. Nesse caso, o nível da qualidade da evidência, caso não haja comprometimento importante de sua validade interna, pode ser aumentado.^{1,2}

GRADIENTE DOSE-RESPOSTA

Um gradiente dose-resposta também pode aumentar a qualidade das evidências. Além disso, quanto maior a exposição, maior o risco.³ Se o grupo recebendo uma determinada intervenção (ou exposição) for composto de pacientes mais doentes do que o respectivo

Tabela 6. Tamanho do efeito

Tamanho do efeito	Medida de efeito	Qualidade
Grande	RR >2 ou $<0,5$ (baseado em evidências consistentes de pelo menos dois estudos, sem possíveis confundidores)	Aumentar um nível
Muito grande	RR >5 ou $<0,2$ (baseado em evidência direta, sem importante comprometimento da validade)	Aumentar dois níveis

Fonte: Bertolaccin et al.; Brasil.²
RR: risco relativo.

grupo controle e mesmo assim obtiver melhores resultados era esperado que, caso os grupos fossem iguais, o benefício da intervenção seria ainda maior. Nesse caso, o nível da qualidade da evidência caso não haja comprometimento importante de sua validade interna poderá ser aumentado.^{1,2}

A EVIDÊNCIAS DO GRADE

As revisões sistemáticas e metanálises devem avaliar as evidências de forma sistemática e reproduzível, reduzindo o viés de seleção. A metanálises aumenta a precisão dos efeitos estimados, explora as inconsistências entre estudos e, finalmente, fornece inferências aplicáveis a uma gama de pacientes. As metanálises também podem descrever a qualidade das evidências, incluindo medidas de precisão. É preciso atentar que os ensaios clínicos randomizados, e estudos observacionais bem conduzidos nem sempre estão disponíveis para orientar todas as questões clínicas. Portanto, uma metanálise comparativa de estudos poderia estar associada a evidências de menor qualidade.¹

Em particular, o GRADE oferece um sistema para avaliar a força das recomendações. O sistema foi projetado para revisões e diretrizes que avaliem estratégias ou intervenções de gestão alternativas, que também podem incluir nenhuma intervenção ou a melhor gestão

corrente.^{1,2} No entanto, o GRADE é mais do que um sistema de classificação, oferecendo um processo transparente e estruturado para a realização das etapas envolvidas no desenvolvimento de recomendações. O GRADE especifica uma abordagem para afiar questões, escolhendo os resultados de interesse e classificando sua importância e avaliando e incorporando evidências com considerações de valores e preferências dos pacientes para chegar a recomendações específicas.¹² O sistema GRADE (Tabela 7) define os quatro graus de evidência: alta, moderada, baixa e muito baixa qualidade, além dos níveis de evidência.^{1,2}

A qualidade da evidência é um contínuo temporal ou espacial, qualquer categorização envolve imprevisibilidade. A compreensão dessas quatro categorias supera essas limitações.³ Por outro lado, a opinião de especialistas não é um tipo de qualidade de evidência, mas uma interpretação das evidências existentes. Assim sendo, quase sempre é necessário o aconselhamento especializado para integrar e contextualizar as evidências de um estudo clínico ou metodológico. Por exemplo, em um estudo randomizado e bem desenhado, o julgamento pode fornecer evidências de qualidade contraditória a um ensaio clínico malconduzido.

A tabela 8 resume a abordagem GRADE para classificar a qualidade das evidências.

Tabela 7. Qualidade da evidência e evidência do sistema GRADE

Nível	Significado	Fonte dos Resultados	Exemplo
A/alta	Muito improvável que trabalhos adicionais irão modificar a confiança na estimativa do efeito	Ensaio clínicos randomizados bem conduzidos e com achados consistentes Em algumas situações, estudos observacionais bem conduzidos, cujos resultados mostram efeitos muito fortes de intervenções terapêuticas que não podem ser explicados por potenciais vieses	Redução de mortalidade com o uso de inibidores da enzima conversora da angiotensina em indivíduos com cardiomiopatia (diversos ensaios clínicos com alto rigor metodológico, em diferentes populações, apresentando consistência nos achados)
B/moderada	Trabalhos futuros poderão modificar nossa confiança na estimativa de efeito, podendo, inclusive, modificar a estimativa	Ensaio clínicos randomizados com importantes problemas na condução, inconsistência nos resultados, achados não dirigidos para o alvo ou desfecho de interesse, imprecisão nas estimativas, vieses de publicação. Estudos observacionais, quando relatam benefício forte em delineamento sem viés	Redução de mortalidade cardiovascular com o uso de aspirina em pacientes com insuficiência cardíaca (maioria dos estudos avaliava benefício através de desfechos compostos)
C/baixa	Trabalhos futuros muito provavelmente terão um importante impacto em nossa confiança na estimativa de efeito	Ensaio clínicos randomizados com desfechos substitutos ou outras importantes limitações. Estudos observacionais, mais especificamente estudos de coorte e caso-controle	Melhora cognitiva com inibidores da colinesterase em indivíduos com demência (ensaio clínicos com fraco rigor metodológico e inconstância nos resultados)
D/muito baixa	Qualquer estimativa de efeito deve ser vista como incerta	Estudos observacionais não controlados e observações clínicas não sistematizadas (como os relato de casos e serie de casos)	Recuperação de função motora com descompressão cirúrgica nervo mediano em indivíduos com síndrome do túnel do carpo (evidências escassas na literatura)

Fonte: Bertolaccin et al.¹ Brasil.²

Tabela 8. Resumo da abordagem GRADE para classificar a qualidade das evidências

Desenho de estudo	Qualidade inicial	Rebaixar	Elevar	Qualidade
Ensaio randomizados	Superior	Limitações no desenho e/ou execução do estudo	Grande magnitude do efeito	Alta
		Inconsistência dos resultados	Gradiente de dose-resposta moderado	Moderada
Estudos observacionais	Mais baixo	Indireta de evidência	Se nenhum efeito observado, todos os confundidores plausíveis, reduzir o efeito demonstrado ou aumentar o efeito	Baixa
		Imprecisão dos resultados Viés de publicação		Muito baixa

CONCLUSÃO

O GRADE é um critério transparente, explícito e abrangente para graduação da qualidade da evidência, apesar de ser um método complexo para não investigadores e profissionais de saúde de forma geral.

REFERÊNCIAS

- Bertolaccini L, Rocco G. The forest of methodology and the writing of evidence-based medicine papers. *Eur J Cardiothorac Surg*. 2018;1:54(4):615-21. doi: <https://doi.org/10.1093/ejcts/ezy264>
- Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy*. 2009;64(5):669-77. doi: <https://doi.org/10.1111/j.1398-9995.2009.01973.x>
- Brozek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Ueffing E, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guideline's part 3 of 3. The GRADE approach to developing recommendations. *Allergy*. 2011;66(5):588-95. <https://doi.org/10.1111/j.1398-9995.2010.02530.x>
- Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-25. doi: <https://doi.org/10.1016/j.jclinepi.2012.03.013>
- Brasil. Ministério da Saúde. Secretaria de Ciência Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes metodológicas: Sistema GRADE – Manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde. Brasília, DF: Ministério da Saúde; 2014.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-15. doi: <https://doi.org/10.1016/j.jclinepi.2010.07.017>
- Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*. 2011;64(12):1277-82. doi: <https://doi.org/10.1016/j.jclinepi.2011.01.011>
- Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al.; GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630. doi: <https://doi.org/10.1136/bmj.g5630>
- Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ*. 2015;350:h870. doi: <https://doi.org/10.1136/bmj.h870>
- Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-6. doi: <https://doi.org/10.1016/j.jclinepi.2010.07.015>
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction- GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-94. <https://doi.org/10.1016/j.jclinepi.2010.04.026>