

MINISTÉRIO DA SAÚDE

DIRETRIZES METODOLÓGICAS: Sistema GRADE – manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde

Brasília – DF
2014



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
2014



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
2014





Esta obra é disponibilizada nos termos da Licença Creative Commons – Atribuição – Não Comercial – Compartilhamento pela mesma licença 4.0 Internacional. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte.

A coleção institucional do Ministério da Saúde pode ser acessada, na íntegra, na Biblioteca Virtual em Saúde do Ministério da Saúde: <www.saude.gov.br/bvs>.

Este trabalho foi desenvolvido no âmbito do termo de cooperação nº 47 entre o Departamento de Ciência e Tecnologia e a Organização Panamericana da Saúde

Tiragem: 1ª edição – 2014 – 1.000 exemplares

Elaboração, distribuição e informações:

MINISTÉRIO DA SAÚDE

Secretaria de Ciência, Tecnologia e Insumos Estratégicos

Departamento Ciência e Tecnologia

Coordenação-Geral de Gestão do Conhecimento

SQN Quadra 2 Projeção C, térreo sala 04

CEP: 70712-902 – Brasília/DF

Tel: (61) 3410-4199

Site: www.saude.gov.br

E-mail: ats.decit@saude.gov.br

Supervisão Geral:

Carlos Augusto Gabrois Gadelha (SCTIE/MS)

Antônio Carlos Campos de Carvalho (Decit/SCTIE/MS)

Jorge Otávio Maia Barreto (Decit/SCTIE/MS)

Elaboração:

Maicon Falavigna (IATS/UFRGS)

Airton Tetelbom Stein (GHC; UFCSPA, ULBRA)

Sérgio Sirena (GHC)

Marisa Santos (INC/MS)

Revisão de Especialista:

Anna Buehler (HAOC)

Organização:

Kathiaja Miranda Souza (Decit/SCTIE/MS)

Maria Augusta Rodrigues de Oliveira (GHC)

Roberta Moreira Wichmann (Decit/SCTIE/MS)

Revisão Técnica:

Betânia Ferreira Leite (Decit /SCTIE/MS)

Carlos de Andrade (INI/Fiocruz)

Evelinda Trindade (InCOR)

Ivan Ricardo Zimmermann (DGITS/SCTIE/MS)

Júlia Souza Vidal (Anvisa)

Marcus Tolentino Silva (UFAM)

Sônia Venâncio (IS/SES-SP)

Taís Galvão (UFAM)

Editoração:

Eliana Carlan (Decit/SCTIE/MS)

Jessica Alves Rippel (Decit/SCTIE/MS)

Design Gráfico:

Gustavo Veiga e Lins (Decit/SCTIE/MS)

Normalização:

Amanda Soares Moreira (CGDI/ Editora MS)

Impresso no Brasil/Printed in Brazil

Ficha Catalográfica

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 / Ministério da Saúde, Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia. – Brasília : Ministério da Saúde, 2014.

72 p. : il.

ISBN 978-85-334-2186-8

1. Avaliação de tecnologias em saúde. 2. Pesquisa em saúde. 3. Tecnologias em saúde. I. Título.

CDU 614:001.8

Catalogação na fonte – Coordenação-Geral de Documentação e Informação – Editora MS – OS 2014/0600

Títulos para indexação

Em inglês: Methodological guideline: GRADE System – Manual graduation quality of evidence and strength of recommendation for decision making process in health.

Em espanhol: El Sistema GRADE – Manual de calidad de la graduación de la evidencia y la fuerza de la recomendación para la toma de decisiones en salud.

/// Lista de siglas e abreviaturas

AVC – Acidente Vascular Cerebral

CEBM – Oxford Centre for Evidence-based Medicine

CDC – Centers for Disease Control and Prevention

Decit – Departamento de Ciência e Tecnologia

GHC – Grupo Hospitalar Conceição

GRADE – Grading of Recommendations Assessment, Development and Evaluation

HMV – Hospital Moinhos de Ventos

IATS – Instituto de Avaliação de Tecnologias em Saúde

INC – Instituto Nacional de Cardiologia

MD – Mean difference

MID – Minimal important difference

MS – Ministério da Saúde

NATS – Núcleo de Avaliação de Tecnologia em Saúde

NICE – National Institute for Health and Clinical Excellence

NNT – Número necessário tratar

OMS – Organização Mundial da Saúde

PICO – População, intervenção, comparação e desfechos

PTC – Pareceres técnico-científicos

RNI – Razão normalizada internacional

SBE – Saúde Baseada em Evidências

SCTIE – Secretaria de Ciência, Tecnologia e Insumos Estratégicos

SMD – Standardized mean difference

SIGN – Scottish Intercollegiate Guidelines Network

TVP – Trombose venosa profunda

UCS – Universidade de Caxias do Sul

UFCSPA – Universidade Federal de Ciências da Saúde de Porto Alegre

UFRGS – Universidade Federal do Rio Grande do Sul

ULBRA – Universidade Luterana do Brasil

/// LISTA DE TABELAS

Tabela 1 – Exemplo de “Sumário de Resultados” para a apresentação das evidências **49**

Tabela 2 – Exemplo de “Perfil de Evidências” para a apresentação dos resultados e avaliação da qualidade **50**

LISTA DE QUADROS

Quadro 1 – Recomendação para o uso de anticoagulação oral em pacientes com fibrilação atrial e doença reumática valvar mitral.....	45
Quadro 2 – Níveis de evidências de acordo com o sistema GRADE	45
Quadro 3 – Implicação dos graus de recomendação de acordo com o sistema GRADE	46
Quadro 4 – Fatores que reduzem a qualidade da evidência	46
Quadro 5 – Fatores que elevam a qualidade da evidência	46
Quadro 6 – Limitações do estudo em ensaios clínicos randomizados de superioridade	47
Quadro 7 – Limitações dos estudos observacionais	47
Quadro 8 – Exemplos de redução na qualidade da evidência devido à evidência indireta	48
Quadro 9 – Exemplos de fonte de evidência indireta relativo ao uso de desfechos substitutos.....	48
Quadro 10 – Consequências da elevada magnitude do efeito na avaliação da qualidade da evidência	48
Quadro 11 – Implicação dos graus de recomendação de acordo com o sistema GRADE	52
Quadro 12 – Situações-modelo na qual uma recomendação forte pode ser indicada a despeito de baixa ou muito baixa confiança em estimativas efetivas	52
Quadro 13 – Exemplos de determinantes de recomendações fortes e fracas	53
Quadro 14 – Qualidade geral da evidência de acordo com o nível de evidência dos desfechos críticos	53
Quadro 15 – Processo de avaliação da qualidade da evidência e da graduação da recomendação segundo o sistema GRADE	54

LISTA DE FIGURAS

Figura 1 – Graduação dos níveis de evidência de acordo com o sistema GRADE	61
Figura 2 – Exemplo de questões de pesquisa no formato PICO	61
Figura 3 – Importância relativa dos desfechos	62
Figura 4 – Processo de avaliação da qualidade da evidência segundo o sistema GRADE	63
Figura 5 – Diferenças em direção de efeito com mínima heterogeneidade	64
Figura 6 – Heterogeneidade substancial com importância questionável (mesma direção de efeito)	64
Figura 7 – Heterogeneidade substancial, com inequívoca importância	65
Figura 8 – Avaliação da imprecisão em diretrizes	65
Figura 9 – Gráfico em funil (<i>funnel plot</i>). (A) Gráfico em funil na ausência de assimetria. (B) Gráfico em funil assimétrico	66
Figuras 10 e 11 – Guideline development tool	67

SUMÁRIO

Apresentação	15
1 Saúde baseada em evidências: sistemas para avaliação da qualidade da evidência e para a graduação da força da recomendação	17
2 Sistema GRADE	19
2.1 Níveis de evidência	19
2.2 Força da recomendação	20
3 Formulação de questão clínica e decisão de desfechos importantes	21
4 Avaliação da qualidade da evidência	23
4.1 Limitações metodológicas (risco de viés)	23
4.2 Inconsistência	25
4.3 Evidência Indireta	26
4.4 Imprecisão	27
4.5 Viés de publicação	28
4.6 Fatores que aumentam a qualidade da evidência	29
4.6.1 Grande Magnitude de Efeito	31
4.6.2 Presença de gradiente dose-resposta	31
4.6.3 Fatores de confusão residuais que aumentam a confiança na estimativa	31
4.7 Considerações quanto a elevar o nível de evidência	32
5 Síntese da Evidência	33
5.1 Lista dos desfechos	33
5.2 Estimativa relativa de efeito	34
5.3 Estimativa absoluta de risco basal	34
5.4 Estimativa absoluta de risco no grupo submetido à intervenção	34
5.5 Estudos incluídos	35
5.6 Nível de evidência	35
5.7 Comentários	35
6 Uso do GRADE para o desenvolvimento de recomendações	37
6.1 Importância do Problema	37
6.2 Qualidade geral da evidência	37
6.3 Balanço entre riscos e benefícios	38
6.4 Valores e preferências	38
6.5 Utilização de recursos	38
6.6 Equidade	39
6.7 Aceitabilidade	39
6.8 Viabilidade	39

7 Considerações finais	41
Quadros e tabelas	43
Figuras	59
Referências	69

Apresentação

O presente manual consiste em um documento para auxiliar pesquisadores na elaboração de recomendações com um paradigma que se denomina Saúde Baseada em Evidências. Serão apresentados os princípios para a formulação de recomendações abordando a qualidade de evidência, em especial sobre intervenções terapêuticas.

O manual apresenta o sistema GRADE – Grading of Recommendations Assessment, Development and Evaluation como forma de avaliação da qualidade da evidência e da força da recomendação pelo sistema GRADE.

O público alvo deste documento são os elaboradores de revisões sistemáticas e avaliações de tecnologias em saúde (ATS), incluindo pareceres técnicos científicos, assim como elaboradores de diretrizes clínicas. A aplicação principal deste manual está na avaliação da evidência e na formulação de recomendações para diretrizes clínicas, revisões sistemáticas e avaliações de tecnologias em saúde.

1 Saúde baseada em evidências: sistemas para avaliação da qualidade da evidência e para a graduação da força da recomendação

Devido ao grande volume de informações e variabilidade na qualidade, há necessidade de elaboração de sínteses que facilitem o acesso e possibilitem conclusões baseadas em diversas fontes de evidência, fornecendo subsídio científico para a tomada de decisão, tanto para o profissional de saúde quanto para o gestor. Nesse contexto, fontes primárias e secundárias de evidência são utilizadas. As revisões sistemáticas, fontes secundárias de evidência, têm um papel de destaque no desenvolvimento de diretrizes clínicas; as recomendações devem ser idealmente baseadas na melhor evidência disponível, sendo processos sistemáticos de revisão da literatura, que se caracteriza como métodos abrangentes e transparentes, permitindo adequado embasamento para a avaliação da evidência¹.

As práticas baseadas em evidência, apesar de não estarem isentas de críticas, podem contribuir para a fundamentação de uma decisão clínica ou de saúde pública. A Saúde Baseada em Evidências (SBE) foi a denominação mais popular do que foi inicialmente denominado “clinimetria”, e é caracterizada pela utilização da epidemiologia clínica como ferramenta para melhor apoio à tomada de decisão individual ou coletiva. Devendo contemplar além do conhecimento técnico, os valores e preferências dos pacientes e a experiência clínica², a Saúde Baseada em Evidências deve integrar a prática clínica às evidências científicas, auxiliando na tomada de decisão.

Como instrumento para a prática da SBE, desenvolveram-se sistemas para a avaliação da qualidade da evidência e para a graduação da força da recomendação, com o objetivo de informar respectivamente a confiança nas evidências apresentadas e a ênfase para que seja adotada ou rejeitada uma determinada conduta.

A Saúde Baseada em Evidências vem sendo utilizada há mais de 25 anos, e inúmeros sistemas apresentam-se como alternativas para classificar a informação. Dentre elas, merecem destaques o sistema Oxford Centre for Evidence-based Medicine (CEBM)³ e o sistema desenvolvido pelo Scottish Intercollegiate Guidelines Network (SIGN)⁴.

Como crítica aos sistemas disponíveis, esses instrumentos se caracterizam como pouco abrangentes e focados especialmente no delineamento de pesquisa, não explicitando, por exemplo, que ensaios clínicos baseados em desfechos substitutos, mesmo quando bem delineados, devem receber um nível consideravelmente mais baixo, em comparação aos estudos observacionais bem conduzidos e embasados em desfechos clínicos relevantes. Além disso, inexistente padronização entre os diferentes sistemas em relação à caracterização da evidência entre as organizações. O Quadro 1 exemplifica a falta de consistência dos códigos utilizados entre os diferentes sistemas de classificação. Frente à necessidade de uniformização desse processo destaca-se o trabalho do GRADE *working group*, grupo colaborativo para o desenvolvimento de um sistema padronizado de classificação.

O GRADE (Grading of Recommendations Assessment, Development and Evaluation) é um sistema desenvolvido por um grupo colaborativo de pesquisadores que visa à criação de um sistema universal, transparente e sensível para graduar a qualidade das evidências e a força das recomendações. Atualmente mais de 80 instituições internacionais utilizam o GRADE, entre elas a Organização Mundial da Saúde (OMS), o National Institute for Health and Clinical Excellence (NICE), a SIGN, o Centers for Disease Control and Prevention (CDC) e a colaboração Cochrane⁵.

2.1 Níveis de evidência

Nível de evidência representa a confiança na informação utilizada em apoio a uma determinada recomendação. No sistema GRADE, a avaliação da qualidade da evidência é realizada para cada desfecho analisado para uma dada tecnologia, utilizando o conjunto disponível de evidência⁶. No GRADE, a qualidade da evidência é classificada em quatro níveis: alto, moderado, baixo, muito baixo, conforme mostrado no Quadro 2. Esses níveis representam a confiança que possuímos na estimativa dos efeitos apresentados⁷.

A classificação inicial da qualidade da evidência, conforme mostrado na Figura 1, é definida a partir do delineamento dos estudos. O ensaio clínico randomizado é o delineamento de estudo mais adequado para questões relacionadas à intervenção, e quando esses são considerados, a qualidade da evidência pelo sistema GRADE inicia-se como alta. Quando apenas estudos observacionais são incluídos, a qualidade da evidência se inicia como baixa.

A partir da classificação inicial, critérios são definidos e o julgamento desses aspectos permitem reduzir ou elevar o nível de evidência. Os fatores responsáveis pela redução no nível de evidência são:

- Limitações metodológicas (risco de viés)⁸;
- Inconsistência⁹;
- Evidência indireta¹⁰;
- Imprecisão¹¹;
- Viés de publicação¹².

Adicionalmente, caso o nível não tenha sido rebaixado devido aos fatores acima apresentados, a evidência procedente de estudos observacionais pode ser elevada considerando três fatores¹³:

- Grande magnitude de efeito;
- Gradiente dose-resposta;
- Fatores de confusão residuais, os quais aumentam a confiança na estimativa.

Recomendações oriundas de opiniões de especialistas são classificadas como nível de evidência “muito baixo”. Opiniões de especialista não é caracterizada formalmente como evidência, devendo preferencialmente buscar outras fontes de informação, como por exemplo estudos observacionais não comparados (séries e relatos de casos).

Dessa forma, o GRADE caracteriza-se como um instrumento abrangente no processo de avaliação das evidências, compreendendo diversos fatores em sua análise. O foco de avaliação não é apenas no delineamento, como em outros sistemas de avaliação de evidências.

No entanto, o GRADE apresenta algumas limitações, como por exemplo, a sua complexidade na avaliação, assim como a necessidade de um julgamento qualitativo do avaliador para realizar julgamentos a respeito de cada um dos domínios avaliados. Deste modo, avaliações independentes podem por vezes graduar os níveis de evidências de forma diferente. Decorrente dessa situação, o GRADE preconiza a apresentação no formato de tabelas (perfis de evidências e sumários de resultados) que apresentam para cada questão. Apresenta o delineamento e o julgamento sobre cada aspecto do processo que pode elevar ou diminuir a classificação final, bem como a referência às publicações utilizadas. Essa medida é importante uma vez que fornece transparência do processo, não podendo ser omitida em sua avaliação¹⁴.

2.2 Força da recomendação

A força da recomendação expressa a ênfase para que seja adotada ou rejeitada uma determinada conduta, considerando potenciais vantagens e desvantagens. São consideradas vantagens os efeitos benéficos na melhoria na qualidade de vida, aumento da sobrevida e redução dos custos. São consideradas desvantagens os riscos de efeitos adversos, a carga psicológica para o paciente e seus familiares e os custos para a sociedade. O balanço na relação entre vantagens e desvantagens determina a força da recomendação.

A força da recomendação (forte ou fraca) pode ser a favor ou contra a conduta proposta. Geralmente, esse processo é realizado no desenvolvimento de diretrizes clínicas e pareceres técnicos, podendo ser específicas para o cenário avaliado (por exemplo, condicionadas à disponibilidade de recursos)¹⁵. No Quadro 3, apresentamos as definições da força de recomendação, com a implicação para gestores, pacientes e profissionais de saúde.

3 Formulação de questão clínica e decisão de desfechos importantes

O sistema GRADE necessita de uma clara definição de população, intervenção, comparação e desfechos (PICO), assim como o cenário no qual a recomendação será implementada;

- Os desfechos de interesse a serem considerados devem ser aqueles importantes para os pacientes. Os desfechos substitutos proporcionam qualidade de evidência inferior, devido ao seu caráter indireto;
- O GRADE é aplicado para cada um dos desfechos considerados na questão de pesquisa;
- Em diretrizes, a avaliação da importância dos desfechos deve preceder a da qualidade da evidência¹⁶.

A avaliação pelo sistema GRADE indica a necessidade de especificar de forma clara a população, intervenção, comparador e desfechos. A questão de pesquisa deve ser estruturada considerando os domínios do acrônimo PICO, onde cada letra representa um componente da questão: P – paciente; I – intervenção; C – comparação; e O – *outcome* (desfecho). Tal estruturação é útil no processo de desenvolvimento de recomendações específicas, adicionalmente auxilia na estratégia de busca (Figura 2). Além disso, no desenvolvimento de diretrizes, ATS e na elaboração de pareceres técnico-científicos (PTC), é importante especificar o cenário no qual a questão está sendo avaliada. Dessa forma, será avaliado o nível de evidência para cada questão – em especial para cada desfecho de interesse – mesmo que a recomendação final englobe diferentes questões de pesquisa. Por exemplo, anticoagulantes podem ter efeito na redução de mortalidade e na redução de eventos tromboembólicos; contudo seu uso pode estar associado a aumento de eventos hemorrágicos. Assim, em uma revisão sistemática, a qualidade da evidência é avaliada para cada um dos desfechos de uma forma independente.

No entanto, a recomendação em relação ao uso da intervenção leva em conta todo o conjunto, considerando o balanço entre os riscos e os benefícios da intervenção. Assim, no processo do desenvolvimento de uma diretriz clínica ou de um PTC, faz-se necessário a identificação dos desfechos a serem considerados, assim como a importância relativa entre eles no processo de tomada de decisão. A escolha por desfechos deve ser orientada por sua relevância para o paciente, e não pela informação disponível na literatura. Por exemplo, na avaliação de um novo tratamento para diabetes mellitus, a ocorrência de eventos macrovasculares e microvasculares são desfechos relevantes e devem guiar as questões clínicas; caso haja evidências apenas no impacto sobre o nível de hemoglobina glicosilada, essa será considerada como um desfecho substituto.

Além disso, deve-se avaliar sua importância relativa do desfecho no processo de tomada de decisão. Idealmente, antes da identificação da evidência, devem ser

priorizados os desfechos importantes para os pacientes, preferencialmente incluindo-os na decisão. Para definir a importância, pode-se utilizar sistema de votação ou métodos de consenso (ex.: Delphi).

O sistema GRADE classifica a importância relativa dos desfechos em uma escala de 1 a 9, dividindo-os em três categorias (Figura 3):

- Desfechos críticos para o processo de decisão;
- Desfechos importantes para o processo de decisão;
- Desfechos pouco importantes para o processo de decisão.

A definição dos tipos de desfechos (crítico, importante ou pouco importante) é essencial no processo de desenvolvimento de recomendações e possibilita a hierarquia dos desfechos avaliados, assim como uma avaliação mais pragmática do balanço entre riscos e benefícios.

4 Avaliação da qualidade da evidência

- O sistema GRADE classifica a qualidade da evidência em quatro níveis: alto, moderado, baixo e muito baixo;
- A evidência proveniente de ensaios clínicos randomizados inicia com nível de evidência alto; a evidência proveniente de estudos observacionais, com nível de evidência baixo;
- A qualidade da evidência é avaliada para o conjunto de evidências disponíveis para a questão de pesquisa (PICO), não para estudos individuais⁷.

O nível de evidência representa a qualidade da evidência científica disponível e define a confiança na informação utilizada, o que possibilita a definição de uma determinada recomendação. No sistema GRADE, a qualidade da evidência é classificada em quatro níveis: alto, moderado, baixo, muito baixo.

O processo de avaliação da evidência é apresentado na Figura 4. O ponto de partida na avaliação da qualidade da evidência é o delineamento de pesquisa. A evidência proveniente de ensaio clínico randomizado é considerada menos propensa a vieses, em especial, presença de fatores de confusão e iniciam com nível “alto”. As evidências procedentes de estudos observacionais são consideradas inicialmente com nível “baixo”.

São cinco os fatores que podem diminuir a qualidade da evidência (vide Quadro 4). Existe a possibilidade de elevação do nível da evidência, caso estejam presentes fatores que aumentem a confiança nos resultados apresentados (Quadro 5). Assim, estudos observacionais podem gerar evidência com nível alto, e ensaios clínicos randomizados podem indicar evidência de nível inferior.

Geralmente a elevação do nível de evidência não é aplicável caso tenha sido reduzido anteriormente, especialmente na presença de limitações metodológicas. Dessa forma, os critérios para elevação da qualidade da evidência são aplicáveis principalmente a estudos observacionais bem delineados.

4.1 Limitações metodológicas (risco de viés)

- No sistema GRADE, ensaios clínicos randomizados (começam com nível de evidência alto) e estudos observacionais (começam com nível de evidência baixo) podem ter a qualidade da evidência reduzida, devido à presença de limitações metodológicas referentes ao seu delineamento ou execução;
- Risco de viés pode variar de acordo com diferentes desfechos quando cada um deles é informado por um conjunto de estudos específico (ex.: mortalidade em alguns estudos, qualidade de vida em outros);

Conclusão

- As revisões sistemáticas geralmente possuem limitações na avaliação de risco de viés. Elas, usualmente, avaliam a qualidade dos diferentes estudos individualmente e não a qualidade do conjunto dos estudos⁸.

As limitações metodológicas indicam uma maior propensão a vieses, diminuindo assim a confiança na estimativa de efeito. Na presença de limitações relevantes, elas podem ser classificadas em graves ou muito graves, reduzindo respectivamente em um ou dois níveis, dependendo de como os mesmos aumentam as incertezas sobre a estimativa apresentada. A avaliação do risco de viés no sistema GRADE é realizada para o conjunto da evidência, não para estudos individuais. Contudo, ferramentas para a avaliação da qualidade de estudos individuais podem auxiliar no processo. Entre elas, destacamos o instrumento para avaliação do risco de vieses desenvolvido pela Colaboração Cochrane para estudos de intervenção¹⁷ e a escala de Newcastle-Ottawa para estudos observacionais¹⁸.

No Quadro 6 apresentamos as principais limitações em ensaios clínicos randomizados; no Quadro 7, apresentamos os principais fatores a serem avaliados em estudos observacionais. Apesar de limitações metodológicas existirem em praticamente todos os conjuntos de estudos, nem sempre sua presença é motivo para diminuir a qualidade da evidência.

Por exemplo, se em um conjunto de oito estudos, três apresentam limitações importantes em sua condução, devemos primeiramente avaliar o impacto desses estudos na estimativa de efeito. Caso sejam estudos com amostras pequenas, de modo que sua exclusão não altere substancialmente os resultados encontrados, é improvável que eles possam diminuir a confiança nos resultados. Contudo, caso esses estudos representem os estudos mais relevantes, do ponto de vista de tamanho de amostra e peso entre as estimativas, o comprometimento em sua metodologia passa a ser relevante e esse domínio deve ser rebaixado devido ao risco de viés, a depender do grau desse comprometimento.

É importante salientar que os vieses podem impactar diferentemente em cada desfecho. Ao avaliarmos o efeito de uma determinada cirurgia em pacientes com lesão espinhal aguda, foram encontrados três ensaios clínicos randomizados sem mascaramento adequado dos avaliadores de desfecho. A ausência de mascaramento, nesse caso, pode impactar em desfechos mais propensos à subjetividade, como qualidade de vida e melhora da função motora, sendo necessário reduzir o nível de evidência nesses casos. Contudo, para mortalidade geral, o mascaramento é irrelevante, uma vez que o fato de o pesquisador saber a qual tratamento o paciente foi submetido não implicará em uma classificação inadequada para esse desfecho.

4.2 Inconsistência

- O sistema GRADE sugere reduzir o nível de evidência caso seja observado inconsistência (heterogeneidade) importante nos resultados, mesmo após a análise de sensibilidade das hipóteses desenvolvidas *a priori*;
- O julgamento da inconsistência é baseado na similaridade das estimativas de efeito, na sobreposição dos intervalos de confiança e em critérios estatísticos, como o I^2 ;
- Efeitos de subgrupo devem ser interpretados com cuidado, devendo ser baseados em hipóteses determinadas *a priori*, com direção de efeito definida, plausibilidade biológica e principalmente se referendada por outros estudos⁹.

Estudos, com o objetivo de responder uma mesma questão de pesquisa, geralmente apresentam diferenças clínicas e metodológicas em sua execução. Por exemplo, as populações avaliadas nos estudos podem ser distintas, ou então, o tempo de seguimento pode ser diferente. Como resultante, podemos observar diferentes efeitos nos estudos individuais (inconsistência nos resultados), e a medida sumária pode não representar adequadamente o conjunto dos estudos.

Diversos critérios podem nos auxiliar no julgamento sobre a inconsistência entre os estudos, entre eles:

- Diferenças elevadas nas estimativas dos efeitos (ex.: risco relativo) dos estudos individuais;
- Sobreposição dos intervalos de confiança;
- Inconsistência (I^2) e teste de inconsistência (I^2) e do teste de heterogeneidade [TG1].

Na presença de inconsistência, devemos realizar análises de sensibilidade (ex. análises de subgrupo e/ou meta-regressão), com o objetivo de identificar fatores clínicos e metodológicos responsáveis por essas diferenças. Idealmente essas análises devem ser realizadas de acordo com hipóteses definidas *a priori* e com direção de efeito definida.

Quando observamos inconsistência nos resultados dos estudos individuais, não explicada por análise de sensibilidade, nossa confiança na estimativa do efeito diminui. Essa situação é acentuada quando diferentes estimativas observadas podem levar a diferentes conclusões como, por exemplo, indicar ou não a intervenção. Nesses casos, o nível de evidência deve ser reduzido por inconsistência.

É importante salientar que o resultado de testes estatísticos isoladamente, como um valor elevado para o I^2 , não indica a necessidade de reduzir o nível de evidência, devendo sempre ser avaliada qualitativamente a diferença entre as estimativas de efeito dos estudos individuais e o quanto essas estimativas alteram nossas conclusões.

Caso a análise de sensibilidade explique adequadamente a diferença nos resultados, também não se deve reduzir o nível de evidência, e apresentam-se as estimativas independentemente para cada subgrupo.

Por exemplo, na Figura 5 observamos resultados consistentes dos estudos individuais, todos apresentando efeito semelhante e com sobreposição nos intervalos de confiança. Na Figura 6 observamos estudos com presença inequívoca de heterogeneidade, contudo, com implicação possivelmente baixa, uma vez que é observada direção de efeito em todos os estudos. Na Figura 7, observamos estudos heterogêneos, com diferentes direções de efeito, o que poderia levar a diferentes conclusões; nesse caso a redução no nível de evidência faz-se necessária caso não seja adequadamente explicada por análise de sensibilidade.

4.3 Evidência indireta

- A confiança nas estimativas de efeito (qualidade da evidência) é reduzida em situações nas quais os participantes, intervenções ou desfechos avaliados nos estudos são substancialmente diferentes daqueles considerados na questão de pesquisa (PICO) da revisão sistemática ou da diretriz clínica.
- A qualidade da evidência é reduzida em situações nas quais não há comparações diretas entre as intervenções (*head-to-head*). Nessas circunstâncias, é necessário realizar comparações indiretas nas quais, por exemplo, estimamos o efeito relativo entre duas intervenções não com base na comparação direta dessas duas alternativas, mas com base na comparação com uma terceira alternativa ou controle⁹.

Evidência direta é aquela proveniente de estudos que comparam diretamente as intervenções de interesse, que avalia desfechos importantes para os pacientes. Em síntese, a evidência é considerada direta quando responde a questão de pesquisa (PICO).

A evidência pode ser considerada indireta por quatro motivos, e os exemplos são apresentados no Quadro 8.

- População: os participantes avaliados nos estudos não representam adequadamente a população de interesse (ex.: gravidade diferente).
- Intervenção: as intervenções avaliadas nos estudos diferem daquelas de interesse (ex.: diferentes doses de um fármaco, ou diferentes técnicas psicoterápicas).
- Comparação: conceitualmente é diferente dos demais, uma vez que não é relacionado à questão de pesquisa (PICO). Ocorre na ausência de comparações diretas entre as alternativas (estudos *head-to-head*), como no caso de estudos de simulação e de *multiple treatment comparisons*. Nesses casos, são apresentadas estimativas de efeito geradas por base em comparações indiretas, mais susceptíveis a vieses.
- Desfecho: os desfechos avaliados não são aqueles de interesse primário para a decisão, como por exemplo, o uso de desfechos substitutos (Quadro 10).

É comum encontramos evidência indireta para a maioria das questões de pesquisa, uma vez que dificilmente encontraremos estudos avaliando especificamente as alternativas de interesse em nossa população alvo. Isso não necessariamente implica em redução do nível de evidência, devendo ser qualitativamente avaliado o quanto isso reduz nossa confiança nas estimativas de efeito.

Por exemplo, o efeito de antidepressivos no tratamento de depressão maior foi avaliado principalmente em populações de países desenvolvidos, contudo não é esperado que o efeito em países em desenvolvimento seja substancialmente diferente. Nesse caso, não seria necessário reduzir o nível de evidência. No entanto, se a maioria dos estudos sobre a efetividade de antidepressivos foi realizado em população adulta, e queremos generalizar esses resultados para crianças, teremos menor confiança nesses resultados e devemos reduzir o nível de evidência por evidência indireta de população.

Em geral, desfechos substitutos implicam em redução do nível de evidência. Todavia, caso o desfecho avaliado seja fortemente ligado à ocorrência de desfechos importantes aos pacientes, podemos considerar não reduzir a qualidade da evidência. Por exemplo, no tratamento do HIV/AIDS, triglicerídeos a contagem de linfócitos T CD4 está fortemente associada à morbidade e letalidade da doença, podendo ser utilizado como um desfecho substituto que nos fornece uma grande confiança no impacto sobre eventos importantes para os pacientes. Além disso, não recomendamos reduzir o nível da evidência na avaliação de efeitos adversos ou de toxicidade.

4.4 Imprecisão

- O principal critério utilizado pelo sistema GRADE para julgar a precisão das estimativas é a amplitude do intervalo de confiança de 95%;
- Idealmente, o intervalo de confiança a ser avaliado é o referente ao efeito absoluto, e não ao relativo;
- Quando observamos magnitude de efeito, baseado em um pequeno número de eventos, pode-se considerar reduzir o nível de evidência, mesmo se o intervalo de confiança for aparentemente estreito¹².

Para o desenvolvimento de recomendações (por exemplo, em diretrizes clínicas), a confiança na estimativa do efeito está relacionada ao grau de certeza em que ela apoia uma determinada decisão (por exemplo, adotar ou não uma intervenção).

Uma forma adequada de definir se a estimativa é precisa em revisões sistemáticas, é através do cálculo do tamanho ótimo da informação. Geralmente, para tomada de decisão, o julgamento sobre a precisão das estimativas é baseado no intervalo de confiança de 95%. Nesse processo, observamos os limites superior e inferior do intervalo, o que possibilita avaliar se caso o real efeito fosse um desses valores, as conclusões seriam modificadas em relação à recomendação. Por exemplo, em uma revisão sistemática

hipotética, determinada intervenção reduziu em 26% o risco de Acidente Vascular Cerebral (AVC) (RR=0,74; IC95% 0,60 – 0,88). Consideramos que o risco de ocorrência do evento em uma população seja de 5%, em termos absolutos a redução seria de 1,3% (IC95% de 0,6 a 2%). Assim, o número necessário tratar (NNT) para prevenir um evento é 77 pessoas (NNT=100/1,3). O intervalo de confiança do NNT vai de 50 (100/2) a 167 (100/0,6). Isso significa que para prevenirmos um AVC precisamos tratar um número de pacientes que pode ser de apenas 50, ou pode chegar a 167.

Consideramos que a intervenção tenha poucos efeitos adversos, boa aceitabilidade e custo moderado. Podemos definir que, para indicar um medicamento com essas características, é necessário que ela previna um AVC a cada 200 pacientes tratados. Uma vez que, de acordo com o pior cenário gerado a partir do intervalo de confiança, o maior NNT estimado é de 167, portanto apresenta uma precisão em que o NNT é menor do que 200 (Figura 6).

Por outro lado, se o custo do medicamento for alto e ela possuir eventos adversos importantes, podemos indicar a implementação no caso de prevenir um evento a cada 100 pacientes tratados. Como o NNT estimado é 77, é possível que ela satisfaça esse critério, contudo como a margem superior do intervalo de confiança é 167, então não podemos com precisão afirmar que o real efeito é satisfatório (NNT=100) e devemos reduzir seu nível de evidência (Figura 8).

Como apresentado acima, idealmente a avaliação deverá ser feita com base no valor absoluto de risco, considerando que é o determinante para a tomada de decisão. Entretanto, a avaliação da precisão relacionada às medidas relativas de risco pode ser utilizada como uma alternativa.

Um fato importante a ser mencionado é que, frente a estudos iniciais, com pequeno número de eventos, geralmente com questão de pesquisa restrita e com tamanho de efeito grande, é possível que tenhamos uma associação espúria, mesmo quando for estatisticamente significativa. Nesse caso, é aconselhável reduzir o nível de evidência.

4.5 Viés de publicação

- Evidência empírica sugere que, em geral, estudos estatisticamente significativos têm maior probabilidade de serem publicados do que estudos sem significância estatística (estudos “negativos”);
- Revisões sistemáticas precoces, que são realizadas apenas quando poucos e iniciais estudos estão disponíveis, podem superestimar a estimativa de efeito uma vez que estudos “negativos”, geralmente, levam mais tempo para serem publicados (*lag-time bias*). Estudos iniciais com resultados positivos, especialmente com pequeno tamanho amostral, devem ser considerados suspeitos;
- A não publicação de estudos com resultados “negativos”, quando parece ser uma

Conclusão

prática frequente. O pesquisador deve suspeitar de viés de publicação quando os estudos são uniformemente pequenos, em especial quando em sua maioria possuem conflitos de interesse importante (por exemplo, patrocínio da indústria farmacêutica);

- A avaliação empírica de padrão de resultados (por exemplo, gráfico em funil) pode sugerir viés de publicação, no entanto deve ser interpretado com cautela¹².

No sistema GRADE, a evidência procedente de ensaios clínicos randomizados inicia com nível alto; evidência procedente de estudos observacionais inicia com nível baixo. Em ambos, a qualidade da evidência pode ser reduzida caso haja alto risco de viés de publicação. Mesmo em situações nas quais o conjunto da evidência possua poucas limitações metodológicas, o viés de publicação pode resultar numa superestimação do efeito.

Em geral, artigos com resultados favoráveis tendem mais frequentemente a serem publicados em inglês, em revistas indexadas e em um curto espaço de tempo. Os estudos “negativos” (sem efeito identificado) tendem mais frequentemente a serem publicados na língua nativa dos pesquisadores, em revistas com menor visibilidade e após maior intervalo de tempo. Assim, o resultado sumário de uma metanálise pode ser propenso a viés, caso a amostra de estudos incluídos não seja representativa da totalidade dos estudos publicados ou não. O pesquisador deve avaliar as situações nas quais o conjunto da evidência pode não representar a totalidade dos estudos publicados sobre a questão de pesquisa. Caso haja alta probabilidade de viés de publicação, potencialmente alterando as conclusões a respeito da recomendação, devemos reduzir o nível de evidência.

Entre os principais fatores que devemos considerar na avaliação de viés de publicação estão:

- Estratégia de busca pouco abrangente (ex.: poucas bases de dados, estratégia de busca inadequada, com potencial de não captar estudos relevantes);

Inclusão de artigos condicionada ao idioma de publicação (ex.: busca apenas em língua inglesa pode conferir viés quando a questão de pesquisa envolver acupuntura, onde grande parte dos estudos é publicada em mandarim, contudo pode ser pouco relevante, por exemplo, em questões relacionadas ao manejo medicamentoso de infarto agudo do miocárdio);

- Critérios estatísticos e gráficos podem indicar possibilidade de viés de publicação. Entre os mais utilizados estão a inspeção visual do gráfico em funil (*funnel plot*)¹⁹ e o teste de Egger²⁰. É esperado que estudos maiores apresentem medida de efeito mais próxima ao efeito real da intervenção, enquanto estudos menores apresentem resultados variáveis, podendo, pelo acaso, apresentar efeito superior ou inferior ao efeito real, esperando que essa distribuição de estudo seja simétrica. No entanto, em certas circunstâncias, observamos a publicação e identificação da maioria dos estudos com grande tamanho amostral, enquanto que estudos pequenos apresentando resultados “positivos” tendem a ser publicados e identificados com maior frequência do que estudos “negativos”. Como resultado, podemos obter um valor estatisticamente significativo no teste de Egger e assimetria identificada visualmente no gráfico em funil (Figura 9);
- Estudos que apresentem conflitos de interesse, como no caso de estudos patrocinados pela indústria farmacêutica, em especial quando são tamanhos de amostras pequenas, tendem a apresentar resultados uniformemente positivos. É comum nesses casos, que os estudos não sejam publicados (ou descontinuados precocemente), quando apresentam resultados desfavoráveis. Devido ao acaso, mesmo em situações quando não há efetividade de uma determinada intervenção, alguns estudos pequenos podem apresentar resultados positivos, sendo estes mais frequentemente publicados. Além disso, frente a um conjunto de pesquisas com tamanho de amostra pequena, é comum esperar que alguns, por acaso, apresentem resultados “negativos”, ou ao menos, com resultados menos expressivos do que a média dos demais; a inexistência desses estudos com resultados negativos – com os demais apresentando efeito positivo e uniforme – também indica possível viés de publicação.

4.6 Fatores que aumentam a qualidade da evidência

- O sistema GRADE inclui três critérios para elevar o nível de evidência, particularmente aplicáveis para estudos observacionais;
- Elevar o nível de evidência por um ou, até mesmo, dois níveis, é possível quando a magnitude do efeito identificado em estudos observacionais é grande, em especial nos casos em que ocorre em um curto período de tempo;
- Gradiente dose-resposta e presença de potenciais confundidores residuais, que, aumentam a confiança na estimativa, também pode elevar o nível de evidência¹¹.

A qualidade da evidência procedente de estudos observacionais inicia com nível baixo.

A ausência de alocação aleatória (randomização) em ensaios clínicos indica forte possibilidade de viés, sendo atribuído nível de evidência inicialmente inferior. Contudo, estudos observacionais podem ter seu nível de evidência elevado caso satisfaçam critérios que aumentem a confiança na estimativa de efeito.

O sistema GRADE considera três critérios para elevar o nível de evidência:

1. Grande magnitude de efeito;
2. Gradiente dose-resposta;
3. Fatores de confusão residuais que aumentam a confiança na estimativa.

4.6.1 Grande Magnitude de Efeito

Para determinadas intervenções (por exemplo, adrenalina para prevenir mortalidade em anafilaxia, ou insulina para prevenir mortalidade em cetoacidose diabética), os profissionais de saúde estão, corretamente, confiantes de sua efetividade. Assim, nesses casos onde há redução dramática na incidência de um desfecho, ocorrendo em um intervalo curto de tempo após a intervenção, o nível de evidência deve ser considerado alto, mesmo caso a evidência tenha origem a partir de estudos observacionais.

Em outras situações, o sistema GRADE considera elevada magnitude de efeito quando for observado $RR > 2$ ou $RR < 0,5$, o que pode elevar a evidência em um nível, e considera tamanho de efeito muito grande quando for observado $RR > 5$ ou $RR < 0,2$, o qual pode elevar a evidência em dois níveis (Quadro 10). Esse critério é considerado um fator para elevar a qualidade da evidência uma vez que, quando encontrado tamanho grande de efeito em estudos observacionais, é provável que a intervenção possua efeito importante, mesmo na presença de fatores de confusão.

4.6.2 Presença de gradiente dose-resposta

A presença de gradiente dose-resposta é um achado que reforça a probabilidade da ocorrência de relação causa-efeito, assim aumentando a confiança na estimativa. Dessa forma, frente a um consistente aumento do efeito associado a um aumento na exposição, a evidência do efeito torna-se mais robusta.

Por exemplo, os estudos observacionais demonstram um aumento no risco de sangramento em pacientes com níveis de anticoagulação supratrapêuticos, associados a maiores níveis de RNI²¹. Existe uma maior confiança nestes resultados, sendo possível elevar a qualidade da evidência em um nível.

4.6.3 Fatores de confusão residuais que aumentam a confiança na estimativa

Estudos observacionais com forte rigor metodológico irão contabilizar, de forma acurada, fatores associados com os desfechos de interesse, e conduzir análise ajustando para diferenças desses fatores entre os grupos de intervenção e controle. A razão principal para considerar estudos observacionais com nível de evidência inicialmente baixo é que os fatores não mensurados ou desconhecidos provavelmente não estejam adequadamente balanceados entre os diferentes grupos, ao contrário do que acontece com o uso de randomização, o que se caracteriza como confundimento residual. Em

geral, o confundimento residual superestima a estimativa de efeito. Contudo, quando o mesmo potencialmente subestima o efeito encontrado, aumentamos a confiança no efeito, ou na ausência do mesmo. Por exemplo, a revisão sistemática de forte rigor metodológico demonstrou maiores taxas de mortalidade em hospitais privados com fins lucrativos comparados a hospitais privados sem fins lucrativos²². Uma fonte potencial de viés é que hospitais sem fins lucrativos tendem a atender pacientes mais graves, o que colaboraria para uma taxa de mortalidade maior. Entretanto, observou-se o contrário, e o confundimento residual possivelmente esteja reduzindo o tamanho de efeito encontrado, o que aumenta nossa confiança na menor mortalidade, considerando que o estudo foi realizado em hospitais privados e sem fins lucrativos.

Uma situação semelhante pode ocorrer quando não há efeito observado. Por exemplo, a fenformina é um medicamento hipoglicemiante com potencial para causar acidose láctica. Por esse motivo, suspeitava-se que a metformina, hipoglicemiante oral da mesma classe farmacológica, pudesse causar efeito semelhante. Devido ao alerta desse potencial efeito adverso, médicos provavelmente eram mais propensos a monitorarem e reportarem acidose láctica associada à metformina do que a outros medicamentos, causando potencial viés de aferição. Contudo, estudos observacionais com forte rigor metodológico não demonstraram associação entre esse efeito adverso com doses terapêuticas do fármaco. O fato do potencial viés de aferição não mostrar associação positiva reforça as conclusões de que não há risco aumentado de acidose láctica com o uso de metformina em doses terapêuticas, podendo ser elevado à qualidade da evidência.

4.7 Considerações quanto a elevar o nível de evidência

O uso desses critérios em situações nas quais o nível de evidência for reduzido previamente precisa ser avaliado com cautela; em geral não é recomendado elevar o nível de evidência nessas situações, sendo esses critérios aplicáveis predominantemente para o conjunto de evidência procedente de estudos observacionais com resultados substanciais.

Em especial, não se deve elevar o nível de evidência na presença de limitações metodológicas, uma vez que estaríamos aumentando a força da associação de um efeito incerto devido à alta probabilidade de viés. Além disso, é desaconselhado elevar o nível de evidência devido à magnitude de efeito grande quando temos medidas de efeito imprecisas.

Assim, o uso desses critérios em ensaios clínicos é limitado. Como exceção da regra acima, os critérios apresentados podem ser aplicados a estudos randomizados na presença de limitações metodológicas quando essas se referem a problemas relacionados à randomização e/ou a alocação dos participantes. Nesse caso, o nível de evidência iniciaria como baixo, podendo o conjunto da evidência ser avaliado como estudos observacionais.

5 Síntese da Evidência

Ressaltamos que o julgamento sobre os domínios apresentados no sistema GRADE, apesar de serem baseados em critérios definidos, são sujeitos a uma avaliação qualitativa do pesquisador e conduz em uma diferença entre os avaliadores. Assim, é necessário que ao longo do processo, o avaliador realize a graduação da evidência de forma transparente, e apresente os motivos que o fizeram atribuir determinado nível de evidência.

Além disso, para auxiliar o processo de tomada de decisão e de formulação de recomendações, sugere-se sintetizar os resultados obtidos com a revisão da literatura, apresentando medidas de riscos relativas e absolutas para a população à qual se quer aplicar a intervenção. Para esse propósito, foram desenvolvidos dois formatos padronizados para a apresentação das evidências: os sumários de resultados (Tabela 1) e os perfis de evidências (Tabela 2).

O sumário de resultados apresenta um formato mais conciso, o que favorece o processo editorial, no qual é apresentado o menor número possível de colunas e são descritos os julgamentos relativos ao processo de avaliação da evidência, assim como outros comentários, na roda de rodapé.

Os perfis de evidências são tabelas que incluem o processo detalhado sobre o julgamento da avaliação da evidência, apresentando uma coluna para cada um dos domínios avaliados no processo, além de comentários adicionais. Permite mais fácil visualização do processo de avaliação, sendo utilizado principalmente em diretrizes clínicas. Apesar de apresentarem diferentes formatos, a informação contida no mesmo é semelhante, incluindo os sete elementos seguintes¹⁴:

1. Lista dos desfechos;
2. Estimativa relativa de efeito;
3. Estimativa absoluta de risco basal;
4. Estimativa absoluta de risco no grupo submetido à intervenção (ou diferença de risco em relação ao grupo controle);
5. Estudos incluídos;
6. Nível de evidência;
7. Comentários.

5.1 Lista dos desfechos

Devem ser apresentados os desfechos importantes para os pacientes, não devendo ser incluídos desfechos substitutos nas tabelas. Em especial para o processo de desenvolvimento de recomendações, como no caso de diretrizes clínicas, não é aconselhado apresentar mais de sete desfechos, definido pelo GRADE Working Group. Estes desfechos devem ser escolhidos de acordo com os mais importantes para tomada de decisão.

Deve ser apresentada sempre a informação referente ao maior nível de evidência. Quando o nível for o mesmo para a evidência procedente de ensaios clínicos e de estudos observacionais, o mesmo desfecho pode ser apresentado em duas linhas diferentes, resumizando os achados a partir do conjunto de evidências distintas para cada classe de estudos.

5.2 Estimativa relativa de efeito

Deve ser apresentada a estimativa sumária de efeito, idealmente obtida por revisão sistemática. Quando não for disponível medida sumária, a informação pode ser omitida, incluindo-se nos comentários considerações sobre a estimativa de risco a partir de estudos individuais. Para desfechos binários, é sugerida a apresentação do risco relativo; para desfechos contínuos, se disponível, diferença mínima importante (*minimal importante difference* – MID) ou então, diferença de médias (*mean difference* – MD), ou diferença padronizada de médias (*standardized mean difference* – SMD).

5.3 Estimativa absoluta de risco basal

Para a tomada de decisão, o impacto absoluto na redução de eventos é mais importante do que o efeito relativo que a intervenção possui. Por exemplo, ensaio clínico avaliando o efeito da rosuvastatina mostrou redução relativa de 43% na prevenção de trombose venosa profunda (TVP)²³. Contudo a taxa anual de eventos no grupo controle foi de apenas 0,32%, o que leva a um impacto absoluto pequeno com a intervenção, mesmo frente a um efeito relativo grande.

A estimativa de risco deve ser idealmente estimada para a população em que se está propondo a intervenção. Essa estimativa de risco pode ser obtida a partir de estudos prognósticos e de incidência, ou então, pode ser utilizada como estimativa a taxa de eventos no grupo controle de ensaios clínicos randomizados. Quando a intervenção está sendo proposta para diferentes cenários (por exemplo, internação hospitalar regular e internação em unidade de tratamento intensivo), podem ser apresentadas diferentes estimativas de risco basal e calculado o impacto para cada um dos cenários.

5.4 Estimativa absoluta de risco no grupo submetido à intervenção

Para comparar com o risco basal – sem o uso da intervenção – deve ser apresentada a taxa de eventos com o uso da intervenção, ou então a redução absoluta de risco com o uso da mesma. A taxa de eventos é obtida multiplicando a estimativa de risco basal pelo risco relativo, e geralmente os valores absolutos são apresentados para cada 1.000 pacientes como denominador, para facilitar o processo de formulação de recomendações.

Utilizando o exemplo acima apresentado, a rosuvastatina reduz em 43% (RR=0,57) a ocorrência de TVP, sendo 0,32% a incidência anual estimada na população avaliada. Com o uso de rosuvastatina, nessa população a taxa anual de eventos seria cerca de 0,18% ($0,32\% \times 0,57$), representando uma redução absoluta de apenas 0,14% de TVP (cerca de um evento a cada 700 pacientes tratados). Se considerarmos uma população com risco basal de 5% (por exemplo, população de alto risco, como pacientes com neoplasias), a taxa de eventos com o uso do medicamento seria de 2,85%, ou então uma redução de 2,15% em TVP (cerca de um evento a cada 47 pacientes tratados, ou de 21 eventos a cada 1000).

5.5 Estudos incluídos

Para cada desfecho, deve ser apresentado o delineamento de pesquisa, o número de estudos incluídos, o número total de participantes nesses estudos e as referências correspondentes.

5.6 Nível de evidência

Deve ser apresentado o nível de evidência para cada desfecho, incluindo o julgamento sobre cada domínio na avaliação da evidência (sendo esses apresentados em colunas específicas nos perfis de evidência).

5.7 Comentários

Geralmente são apresentados como notas de rodapé, podendo ser destinada uma coluna específica para essa finalidade nos perfis de evidência. Devem ser sempre relatadas as razões para o julgamento referentes ao processo de avaliação da evidência (motivos para reduzir ou não o nível de evidência, assim como razões para elevar o nível de evidência). Além disso, podem ser relatadas considerações referentes às estimativas de risco basal e aos achados de estudos individuais.

Para o desenvolvimento dos sumários de evidência está disponível em um aplicativo *online* gratuito: Guideline Development Tool (<www.guidelinedevelopment.org> e uma lista de checagem para a elaboração do *guideline* (<<http://cebgrade.mcmaster.ca/guidecheck.html>>). Além disso, o aplicativo permite gerenciamento do processo de desenvolvimento de diretrizes clínicas, incluindo gerenciamento de grupos de trabalho, declaração de conflitos de interesse, formulação de recomendações e redação de texto. (Figuras 10 e 11).

6 Uso do GRADE para o desenvolvimento de recomendações

Além da avaliação da qualidade da evidência, o sistema GRADE se propõe a auxiliar no processo de formulação de recomendações, fazendo uma ligação estruturada entre esses dois processos, considerando fatores adicionais relevantes para o processo de tomada de decisão (Quadro 11).

Como visto anteriormente, o sistema GRADE classifica as recomendações em fortes e fracas, também denominadas condicionais, a favor ou contra a intervenção proposta, representando a confiança que temos que as vantagens sobrepõem às desvantagens do curso de ação proposto. Como visto na seção 3, as recomendações fracas e fortes tem diferentes interpretações e implicações para gestores, pacientes e profissionais de saúde (Tabela 13). Apesar de não ser usualmente recomendado, situações nas quais o balanço entre vantagens e desvantagens é semelhante ou então incerto, não fornecer uma recomendação é justificável. Contudo essa prática é desestimulada uma vez que essa ausência de recomendação não auxilia no processo de tomada de decisão, causando confusão ao usuário da diretriz.

Como apêndice, apresentamos a tabela GRADE para desenvolvimento de recomendações, que inclui os seguintes fatores determinantes da força da recomendação²⁴:

- Importância do problema;
- Qualidade geral da evidência;
- Balanço entre benefícios e riscos;
- Valores e preferências;
- Utilização de recursos;
- Equidade;
- Aceitabilidade;
- Viabilidade.

6.1 Importância do Problema

- O problema é prioritário?

No processo inicial de desenvolvimento de uma diretriz deve-se analisar a magnitude e a transcendência da condição que está sendo avaliada.

6.2 Qualidade geral da evidência

- Qual é a qualidade da evidência (nível de evidência para o conjunto da evidência)?

A qualidade da evidência é definida para o conjunto dos desfechos? Qual será o nível de evidência para a recomendação? Nesse caso, são considerados apenas os desfechos críticos para tomada de decisão; a qualidade da evidência geralmente é

dada pelo menor nível de evidência entre os mesmos. Caso todos os desfechos críticos demonstrem efeito seguindo a mesma direção (benefício ou risco), pode ser considerado como qualidade geral da evidência o maior nível de evidência (Quadro 1).

Os níveis de evidência elevados representam maior confiança nas estimativas de efeito, sendo mais propensas a gerarem recomendações fortes. Por sua vez, níveis de evidência baixo e muito baixo, geralmente geram recomendações fracas. Em certas circunstâncias, em especial frente a riscos ou benefícios grandes, níveis de evidência baixo ou muito baixo podem gerar recomendações fortes, como apresentado no quadro 12.

6.3 Balanço entre riscos e benefícios

- Qual é o balanço entre benefícios e riscos?

Será julgado pelo grupo o balanço entre o risco e benefício clínico, considerando todos os desfechos importantes para a tomada de decisão. Por consenso, será definido se, e quanto o benefício sobrepõe os riscos, para o conjunto de desfechos da intervenção avaliada. As situações nas quais os benefícios claramente sobrepõem os riscos, e nas quais os riscos claramente sobrepõem os benefícios, geralmente levam a recomendações fortes. As recomendações fracas são procedentes de situações nas quais há certo equilíbrio entre os riscos e os benefícios.

6.4 Valores e preferências

- Há similaridade em relação à importância dada para os principais desfechos?

Para a definição do balanço entre riscos e benefícios, os desfechos não possuem peso equivalente. Em algumas situações, pode haver divergências entre a importância atribuída a cada desfecho. Por exemplo, não há consenso sobre a importância de determinados eventos em pacientes com diabetes mellitus: alguns pacientes darão mais importância a eventos macrovasculares (ex.: infarto), enquanto outros a desfechos microvasculares (ex.: retinopatia). Quando os valores e preferências variam de forma significativa, a certeza sobre o balanço de risco e benefício diminui, enfraquecendo a recomendação.

6.5 Utilização de recursos

- Os custos (ou uso de recursos) estão associados a intervenções pequenas?
- Os custos incrementais (ou uso de recursos) são pequenos relativos aos benefícios esperados?

O custo e utilização de recursos devem ser avaliados para implementação das recomendações no cenário proposto (por exemplo, no Brasil, ou em determinada região), assim como os benefícios em saúde esperados com a implementação das recomendações. Quando possível, deve-se incorporar a análise de custo efetividade, assim como a análise do impacto orçamentário no processo²⁵.

A participação de gestores é importante nesse processo, com o intuito de avaliar os recursos disponíveis num determinado contexto, assim como as alternativas

para incorporação (incluindo investimentos em diferentes problemas de saúde). As intervenções mais baratas, com maiores benefícios em saúde, e com estrutura disponível para sua implementação, são mais propensas a gerarem recomendações fortes.

6.6 Equidade

- Quais são os impactos referentes a iniquidades em saúde?

Nesse caso deve ser avaliado o potencial impacto das recomendações no combate a desigualdades sociais, sendo estas mais propensas a gerarem recomendações fortes. Por exemplo, a construção de centros de radiologia em grandes centros pode aumentar as iniquidades entre regiões urbanas e rurais, enquanto estações de mamógrafos móveis podem diminuir essa desigualdade. Outro exemplo é a implantação de unidades básicas de saúde em regiões vulneráveis, promovendo equidade e proporcionando um maior acesso à saúde das populações desassistidas.

6.7 Aceitabilidade

- A opção é aceitável para os principais atores interessados?

As preferências dos atores (gestores, profissionais de saúde e pacientes interessados) em relação à intervenção devem ser avaliadas, assim como as suas alternativas. Por exemplo, a criação de novos leitos para internação psiquiátrica é uma alternativa que desagrada alguns gestores e defensores do movimento antimanicomial; contudo há um clamor social devido à epidemia do uso de *crack*, assim como a dificuldade do manejo ambulatorial.

6.8 Viabilidade

- A opção é viável para a implementação?

Com o intuito de implementar uma recomendação na prática clínica é necessário identificar os recursos necessários, assim como a estrutura disponível. Por exemplo, na assistência farmacêutica é necessário um número de profissionais de saúde disponíveis (incluindo médicos e farmacêuticos) e com habilidades para realizarem determinadas tarefas, assim como definir as potenciais barreiras para a sua implementação.

Outro exemplo é a realização de colonoscopia, que é uma intervenção com benefício bem estabelecido na prevenção e diagnóstico precoce de câncer colorretal em indivíduos acima de 50 anos. Contudo, a carência de estrutura e de profissionais para a realização do exame preventivo, leva à dificuldade de implementação em certos locais, sendo o mesmo restrito para casos com maior suspeição nessas localidades.

No Quadro 13 apresentamos alguns exemplos de como esses determinantes podem gerar recomendações fortes ou fracas. Para padronizar a redação das recomendações, deve-se utilizar para as recomendações fortes o termo “recomendamos”, enquanto recomendações fracas o termo “sugerimos”.

No processo de desenvolvimento de uma diretriz clínica, geralmente é fornecido previamente aos membros do painel os perfis de evidência para benefícios e riscos, assim como a evidência referente aos demais fatores quando disponíveis, como valores e preferência e utilização de recursos. Os julgamentos relacionados a cada um dos determinantes da força de recomendação apresentados devem ser realizados pelo grupo, durante os painéis para consenso, assim como a redação das recomendações.

7 Considerações finais

O GRADE consiste em um sistema abrangente e transparente para a avaliação da evidência e para a formulação de recomendações, consistindo em uma ferramenta útil na elaboração de diretrizes em saúde, auxiliando o processo de tomada de decisão. Como limitante, o sistema é complexo, necessitando treinamento prévio dos avaliadores, assim como um conhecimento robusto em epidemiologia clínica. Devido a isso, algumas instituições, como a Organização Mundial da Saúde, solicitam que no grupo de trabalho de suas diretrizes sejam incluídos metodologistas com treinamento nesse sistema, aptos a auxiliarem no processo de avaliação de evidências e formulação e recomendações.

O GRADE apresenta uma forma lógica e estruturada de avaliação da evidência e de desenvolvimento de recomendações, contudo é dependente do julgamento qualitativo do avaliador, sendo esta uma crítica comum ao sistema. Esse sistema não pretende eliminar a subjetividade, mas sim estruturar o julgamento, mantendo a transparência no processo. Apesar desse grau de subjetividade, o instrumento é reprodutível, apresentando boas taxas de concordância ($\kappa=0,72$ IC95% 0,61 a 0,79 considerando avaliadores experientes e $\kappa=0,66$, IC95% 0,56 a 0,75 considerando somente estudantes recém-treinados)²⁶.

Por fim, o instrumento está em constante evolução e desenvolvimento. Apresentamos aqui a aplicação para intervenções, contudo há trabalhos sendo realizados utilizando o GRADE em outras áreas, como acurácia diagnóstica, avaliação econômica e prognóstico²⁷.



Quadros e Tabelas

Quadro 1 – Recomendação para o uso de anticoagulação oral em pacientes com fibrilação atrial e doença reumática valvar mitral

Instituição	Nível de evidência	Grau de Recomendação
American Heart Association	B	Class I
American College of Clinical Pharmacy	A	I
Scottish Intercollegiate Guidelines Network	IV	C

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>

Nota: Apesar das recomendações serem embasadas nas mesmas evidências, o julgamento sobre a qualidade da evidência e o grau de recomendação diferem em forma e força.

Quadro 2 – Níveis de evidências de acordo com o sistema GRADE

Nível	Definição	Implicações	Fonte de informação
Alto	Há forte confiança de que o verdadeiro efeito esteja próximo daquele estimado.	É improvável que trabalhos adicionais irão modificar a confiança na estimativa do efeito.	- Ensaios clínicos bem delineados, com amostra representativa. - Em alguns casos, estudos observacionais bem delineados, com achados consistentes*.
Moderado	Há confiança moderada no efeito estimado.	Trabalhos futuros poderão modificar a confiança na estimativa de efeito, podendo, inclusive, modificar a estimativa.	- Ensaios clínicos com limitações leves**. - Estudos observacionais bem delineados, com achados consistentes*.
Baixo	A confiança no efeito é limitada.	Trabalhos futuros provavelmente terão um impacto importante em nossa confiança na estimativa de efeito.	- Ensaios clínicos com limitações moderadas**. - Estudos observacionais comparativos: coorte e caso-controle.
Muito Baixo	A confiança na estimativa de efeito é muito limitada. Há importante grau de incerteza nos achados.	Qualquer estimativa de efeito é incerta.	- Ensaios clínicos com limitações graves**. - Estudos observacionais comparativos presença de limitações**. - Estudos observacionais não comparados***. - Opinião de especialistas.

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>

*Estudos de coorte sem limitações metodológicas, com achados consistentes apresentando tamanho de efeito grande e/ou gradiente dose resposta.

**Limitações: vieses no delineamento do estudo, inconsistência nos resultados, desfechos substituídos ou validade externa comprometida.

***Séries e relatos de casos.

Quadro 3 – Implicação dos graus de recomendação de acordo com o sistema GRADE

Público alvo	Forte	Fraca (condicional)
Gestores	A recomendação deve ser adotada como política de saúde na maioria das situações	É necessário debate substancial e envolvimento das partes interessadas
Pacientes	A maioria dos indivíduos desejaria que a intervenção fosse indicada e apenas um pequeno número não aceitaria essa recomendação	Grande parte dos indivíduos desejaria que a intervenção fosse indicada; contudo alguns indivíduos não aceitariam essa recomendação
Profissionais de saúde	A maioria dos pacientes deve receber a intervenção recomendada	O profissional deve reconhecer que diferentes escolhas serão apropriadas para cada paciente para definir uma decisão consistente com os seus valores e preferências

Fonte: Elaboração própria.

Quadro 4 – Fatores que reduzem a qualidade da evidência

Fator	Consequência
Limitações metodológicas (risco de viés)	↓ 1 ou 2 níveis
Inconsistência	↓ 1 ou 2 níveis
Evidência indireta	↓ 1 ou 2 níveis
Imprecisão	↓ 1 ou 2 níveis
Viés de publicação	↓ 1 ou 2 níveis

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Quadro 5 – Fatores que elevam a qualidade da evidência

Fator	Consequência
Elevada magnitude de efeito	↑ 1 ou 2 níveis
Fatores de confusão residuais que aumentam a confiança na estimativa	↑ 1 nível
Gradiente dose-resposta	↑ 1 nível

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Quadro 6 – Limitações do estudo em ensaios clínicos randomizados de superioridade

1 – Ausência de sigilo da alocação
Pesquisadores envolvidos no recrutamento dos pacientes podem prever para qual grupo o próximo paciente incluído na pesquisa será alocado devido à ausência de sigilo da lista de alocação ou a de randomização inadequada (por exemplo, alocação pelo dia da semana, data de nascimento ou alternância)
2 – Ausência de mascaramento (cegamento)
Pacientes, cuidadores, coletadores de dados, avaliadores de desfecho e analistas de dados cientes ao grupo ao qual os pacientes são alocados
3 – Seguimento incompleto
Perda no seguimento de pacientes randomizados e ausência de análise por intenção de tratar (<i>intention-to-treat</i>)
4 – Relato seletivo de desfechos
Relato incompleto ou ausente de alguns desfechos com base nos resultados.
5 – Outras limitações
Interrupção precoce do estudo por benefício Utilização de medidas de desfecho sem validação (ex.: desfecho relatado pelo paciente)

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Quadro 7 – Limitações dos estudos observacionais

1 – Seleção e inclusão inadequada de participantes
Pareamento inadequado em estudos de casos e controles (sub ou sobrepareamento) Seleção de grupos expostos e não expostos em estudos de coorte a partir de diferentes populações
2 – Ausência de mascaramento (cegamento)
Diferenças na avaliação da exposição em estudos de caso e controles (ex: viés de recordação) Diferenças na avaliação dos desfechos em estudos de coorte (ex: diferente vigilância nos desfechos em expostos e não-expostos)
3 – Falhas para controlar adequadamente os fatores de confusão
Ausência de medidas acuradas para os fatores prognósticos conhecidos Pareamento inadequado e/ou ajustamento inadequado na análise estatística para os fatores prognósticos
4 – Seguimento incompleto

Fonte: Elaboração própria.

Quadro 8 – Exemplos de redução na qualidade da evidência devido à evidência indireta

Questão de pesquisa	Evidência indireta
Uso de oseltamivir na profilaxia de gripe aviária causada por influenza A	População: Ensaios clínicos disponíveis avaliaram a efetividade da profilaxia de gripe sazonal, não da gripe aviária
Rastreamento com cápsula endoscópica/ colonoscopia virtual para prevenção de mortalidade associada a câncer colorretal	Intervenção: ensaios clínicos que avaliam sangue oculto nas fezes e colonoscopia convencional proporcionam evidência indireta para o potencial efeito benéfico de novas tecnologias diagnósticas
Rosiglitazona na prevenção de eventos macrovasculares em pacientes com diabetes mellitus tipo 2	Desfecho: estudos mostravam efeito benéfico na redução dos níveis de HbA1C, sendo indicado seu uso. Posteriormente, evidência direta demonstrou aumento no risco de infarto do miocárdio e de mortalidade cardiovascular com o uso da intervenção
Escolha entre diferentes estatinas na prevenção de eventos cardiovasculares	Comparação: poucos estudos compararam diretamente estatinas; o efeito relativo entre elas pode ser acessado indiretamente, uma vez que em sua maioria foram testadas frente ao uso de placebo

Fonte: Elaboração própria.

Quadro 9 – Exemplos de fonte de evidência indireta relativo ao uso de desfechos substitutos

Condição clínica	Desfechos importantes para o paciente	Desfechos substitutos
Diabetes mellitus	Complicações macrovasculares e microvasculares, qualidade de vida, internação hospitalar	Hemoglobina glicosilada, perfil glicêmico
Osteoporose	Fraturas	Densidade óssea
Glaucoma	Perda da visão	Pressão intra-ocular
Doença pulmonar obstrutiva crônica	Mortalidade, exarcebações, qualidade de vida	Função pulmonar (espirometria)

Fonte: Elaboração própria.

Quadro 10 – Consequências da elevada magnitude do efeito na avaliação da qualidade da evidência

Tamanho de efeito	Interpretação	Consequência
RR > 2 ou RR < 0,5	Magnitude de efeito grande	↑ 1 nível
RR > 5 ou RR < 0,2	Magnitude de efeito muito grande	↑ 2 níveis

Fonte: Elaboração própria.

Tabela 1 – Exemplo de “Sumário de Resultados” para a apresentação das evidências**Sumário de resultados****Autor (es): - / Data:****Questão: A heparina deve ser usada para anticoagulação em pacientes com câncer sem nenhuma outra indicação terapêutica ou profilática ?****Cenário: Ambulatorial****Referências:** Akl EA, Gunukula S, Barba M, Yosucio VE, van Doornmaal FF, Kuipers S, Middeldorp S, Dickinson HO, Bryant A, Schünemann H. Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation. *Cochrane Database Syst Rev.* 2011 Apr 13;(4):CD006652

Desfechos	Participantes (estudos)	Qualidade da evidência (GRADE)	Efeito Relativo (IC 95%)	Efeito absoluto antecipado para 1 ano (número de eventos por 1000)	
				Risco sem anticoagulação	Diferença de Risco com uso de Heparina (IC 95%)
Morte Seguimento: 12 meses	2 531 (8)	MODERADO ^{1,2,3}	RR 0.93 (0.85 a 1.12)	649 por 1000	35 a menos por 1000 (de 97 a menos a 13 a mais)
Tromboembolismo venoso sintomático Seguimento: 12 meses	2 264 (7)	ALTO ¹	RR 0.55 (0.37 a 0.82)	29 por 1000	13 a menos por 1000 (de 5 a 18 a menos)
Sangramento Grave Seguimento: 12 meses	2 843 (9)	MODERADO ^{1,4}	RR 1.30 (0.59 to 2.88)	7 por 1000	2 a mais por 1000 (de 3 a menos a 13 a mais)
Sangramento Leve Seguimento: 12 meses	2 345 (7)	MODERADO ⁴ Devido a imprecisão	RR 1.05 (0.75 a 1.46)	27 por 1000	1 a menos por 1000 (de 8 a menos a 12 a mais)
Qualidade de Vida Seguimento: 12 meses	138 (1)	BAIXO ⁵	Não	Ver Comentários ⁶	

¹ Grande maioria dos estudos com adequado sigilo de alocação, e cegamento de avaliadores de desfecho. Não houve redução da qualidade, embora tenha havido alguma preocupação sobre a falta de cegamento em alguns estudos, o risco global de viés foi considerado baixo.

² Existe uma heterogeneidade moderada entre os estudos incluídos na análise da morte em 12 meses (I² = 35%). A análise de subgrupo para a mortalidade em 12 meses foi estatisticamente significativa e sugere benefício de sobrevida em pacientes com câncer de pequenas células, mas não em pacientes com câncer avançado. Em geral, foi decidido fazer o rebaixamento de um nível ao considerar estas questões, juntamente com imprecisão.

³ IC inclui efeitos sugerindo benefícios, assim como danos.

⁴ IC inclui a possibilidade de danos e benefícios.

⁵ Alto risco de viés em apenas 138 pacientes recrutados.

⁶ Os escores para as duas escalas de qualidade de vida utilizados no estudo são semelhantes para os dois grupos estudados, no início do estudo (*baseline*) e no seguimento.

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Tabela 2 – Exemplo de “Perfil de Evidências” para a apresentação dos resultados e avaliação da qualidade

Questão: A levocabastina intranasal comparada ao placebo deve ser utilizada em adultos com rinite alérgica sazonal?

Referências: 1. Dahl R., Pedersen B., Larsen B. Intranasal levocabastine for the treatment of seasonal allergic rhinitis: a multicentre, double-blind, placebo-controlled trial. *Rhinology*, 1995;33:121-125. 2. Di Lorenzo G., Gervasi F., Drago A., Esposito Pellitteri M., Di Salvo A., Cosentino D., Potestio M., Colombo A., Candore G., Mansueto S., Caruso C. Comparison of the effects of fluticasone propionate, aqueous nasal spray and levocabastine on inflammatory cells in nasal lavage and clinical activity during the pollen season in seasonal rhinitis. *Clin Exp Allergy*, 1999;29:1367-1377. 3. Hampel F., Jr., Martin B.G., Dolen J., Travers S., Karcher K., Holton D. Efficacy and safety of levocabastine nasal spray for seasonal allergic rhinitis. *Am J Rhinol*, 1999;13:55-62. 4. Ortolani C., Foresi A., Di Lorenzo G., Bagnato G., Bonifazi F., Crimi N., Emmi L., Prandini M., Senna G.E., Tursi A., Mirone C., Leone C., Fina P., Testi R. A double-blind, placebo-controlled comparison of treatment with fluticasone propionate and levocabastine in patients with seasonal allergic rhinitis. FLNCO2 Italian Study Group. *Allergy*, 1999;54:1173-1180. 5. Schata M., Jorde W., Richarz-Barthauer U. Levocabastine nasal spray better than sodium cromoglycate and placebo in the topical treatment of seasonal allergic rhinitis. *The Journal of allergy and clinical immunology*, 1991;87:873-878. 6. Svensson C., Andersson M., Greiff L., Blychert L.O., Persson C.G. Effects of topical budesonide and levocabastine on nasal symptoms and plasma exudation responses in seasonal allergic rhinitis. *Allergy*, 1998;53:367-374.

Avaliação da Qualidade										Resumo dos Achados				Importância	
Estudos	Delimitamento	Limitações metodológicas	Inconsistência	Evidência Indireta	Imprecisão	Outras Considerações	Levocabastina intranasal		Efeito		Qualidade (GRADE)	Importância			
							Sintomas nasais (seguimento 4 a 6 semanas)	Placebo	Relativo (IC 95%)	Absoluto					
6	Ensaio Randomizados	Sem limitações graves ¹	Sem inconsistência grave	Sem evidências indiretas importantes	Sem imprecisão grave ²	Nenhuma	242	233	-	DMP* 0.48 a menos (0.15 a 0.81 a menos) ³	AAAA ALTO	CRÍTICA			
Sintomas nasais (avaliação global da eficácia) (seguimento 2 a 4 semanas; classificação: excelente ou boa)															
2	Ensaio Randomizados	Sem limitações graves ⁴	Sem inconsistência grave	Sem evidências indiretas importantes	grave ⁵	Nenhuma	89/150 (59.3%)	53/149 (35.6%)	RR 1.83 (1.16 a 2.89)	30 a mais por 100 (de 6 a mais a 67 a mais)	AAAO MODERADO	CRÍTICA			
Sintomas Oculares (seguimento 2 a 6 semanas)															
4	Ensaio Randomizados	Sem limitações graves ⁶	Sem inconsistência grave	Sem evidências indiretas importantes	grave ⁷	Nenhuma	262	253	-	Resultados não sumarizados ⁸	AAAO MODERADO	IMPORTANTE			
Qualidade de vida – não avaliada															
0	-	-	-	-	-	-	0	0	-	-	-	CRÍTICA			
Eventos Adversos (seguimento 2 a 4 semanas; sonolência/radigag)															
2	Ensaio Randomizados	grave ^{9,10}	Sem inconsistência grave	Sem evidências indiretas importantes	grave ¹¹	Nenhuma	4/151 (2.6%)	1/157 (0.6%)	RR 3.17 (0.51 a 19.71)	1.4 a mais por 1000 (de 3 a menos a 119 a mais)	AAAO BAIXO	CRÍTICA			

*DMP: diferença de médias padronizada
 1 - Apenas um estudo com seguimento do paciente e dois estudos sem relato do tipo de análise (em geral 10% dos pacientes não completaram os estudos).
 2 - Os resultados não excluem efeito pequeno ou grande.
 3 - Outros três estudos não relataram variabilidade nos resultados, por isso não foi possível combinar seus resultados. No entanto, todos apresentaram diferença estatisticamente significativa, favorecendo levocabastina (um mostrou uma diferença de 0.2% e o outro de -28%, o terceiro ensaio não relatou a escala ou variabilidade nos resultados). Esse está correto?
 4 - Em um estudo, 25 % dos pacientes não completaram o tratamento.
 5 - Os resultados não excluem um efeito grande ou um insignificante. Apenas 1.42 eventos.
 6 - Apenas um estudo com seguimento dos pacientes, dois não relataram o tipo de análise e um estudo relatou a variabilidade nos resultados.
 7 - Não foi possível verificar a precisão dos resultados, uma vez que apenas um estudo forneceu variabilidade nos resultados.
 8 - Um estudo relatou -18% diferença versus placebo, um segundo estudo relatou 7,1% das línguas de sintomas oculares, em comparação com 56,1% no braço placebo. Um terceiro estudo relatou uma alteração nos sintomas -1,92 vs -0,19 numa escala de 4 pontos e o último estudo relatou um tamanho do efeito de -0,28 (IC 95%: -1,10 a 0,54), favorecimento (não seria favorecimento?) a levocabastina.
 9 - Outros eventos adversos foram mal relatados. Em um estudo de 4 vs 3 pacientes abandonaram o estudo devido a eventos adversos não especificados, um estudo relatou que não houve eventos adversos, um forneceu uma tabela detalhada, mas não houve diferenças e 2 estudos de não relataram eventos adversos em geral.
 10 - 15% dos pacientes não completaram os ensaios
 11 - Resultados não excluem um grande dano ou um benefício. Muito poucos eventos.

Quadro 11 – Implicação dos graus de recomendação de acordo com o sistema GRADE

Público alvo	Forte	Fraca (condicional)
Gestores	A recomendação deve ser adotada como política de saúde na maioria das situações	É necessário debate substancial e envolvimento das partes interessadas.
Pacientes	A maioria dos indivíduos desejaria que a intervenção fosse indicada e apenas um pequeno número não aceitaria essa recomendação	Grande parte dos indivíduos desejaria que a intervenção fosse indicada; contudo considerável número não aceitaria essa recomendação
Profissionais de saúde	A maioria dos pacientes deve receber a intervenção recomendada.	O profissional deve reconhecer que diferentes escolhas serão apropriadas para cada paciente para definir uma decisão consistente com os seus valores e preferências.

Fonte: Elaboração própria.

Quadro 12 – Situações-modelo na qual uma recomendação forte pode ser indicada a despeito de baixa ou muito baixa confiança em estimativas efetivas

Condição	Exemplo
1 - Quando baixa qualidade de evidência sugere benefício numa situação de risco de vida.	Plasma congelado fresco ou vitamina K em pacientes recebendo varfarina com RNI elevado e um sangramento intracraiano. Apesar da confiança na estimativa de efeito do benefício ser baixa, a recomendação é forte quanto ao uso da intervenção.
2 - Quando baixa qualidade de evidência sugere benefício e alta qualidade sugere dano ou um custo muito elevado.	Ressonância magnética de corpo inteiro para o rastreamento de câncer. Baixa qualidade de evidência de benefício de detecção precoce, mas alta qualidade de evidência de possível dano e/ou alto custo (recomendação forte contra essa estratégia)
3 - Quando evidência de baixa qualidade sugere equivalência de eficácia de duas alternativas, mas evidência de alta qualidade sugere menos danos com uma das alternativas	Erradicação de <i>Helicobacter pylori</i> (HP) em pacientes com linfoma MALT gástrico em estágio precoce. Baixa qualidade de evidência sugere que a erradicação de HP resulta em taxas semelhantes de resposta completa em comparação com alternativas de tratamento de radiação ou gastrectomia; alta qualidade de evidência sugere menos dano ou morbidade (recomendação forte em relação à erradicação do HP)
4 - Quando alta qualidade de evidência sugere equivalência de duas alternativas e baixa qualidade de evidência sugere dano em uma alternativa.	Hipertensão em mulher em fase de planejamento de anticoncepção ou gravidez. Recomendação forte para o uso de labetalol e nifedipina e forte recomendação contra o uso de inibidor de enzima de conversão de angiotensina (IECA) e de bloqueador do receptor de angiotensina (BRA). Todos os agentes têm alta qualidade de evidência de desfechos benéficos, com baixa qualidade de evidência para os potenciais efeitos adversos com IECA e BRA.
5 - Quando a evidência sugere benefício modesto e há baixa/ muito baixa qualidade de evidência sugerindo a possibilidade de dano catastrófico.	Testosterona em homens com ou em risco de câncer de próstata. Evidências de alta qualidade para benefício moderado de tratamento com testosterona em homens com deficiência de andrógeno sintomático para melhora de densidade óssea mineral e força muscular. Baixa evidência de qualidade para dano em pacientes com/ ou risco de câncer de próstata.

Fonte: Elaboração própria.

Quadro 13 – Exemplos de determinantes de recomendações fortes e fracas

FATOR	EXEMPLO DE RECOMENDAÇÃO FORTE	EXEMPLO DE RECOMENDAÇÃO FRACA
Balanco entre consequências desejáveis e indesejáveis de estratégias de manejos alternativos. Quanto mais equilibrado o balanço, menos provável de termos uma recomendação forte.	Aspirina indicada em infarto do miocárdio reduz à mortalidade com um mínimo de toxicidade, em desconforto e custo.	Anticoagulação versus aspirina em pacientes com fibrilação atrial com um escore 1 no CHADS ₂ (risco de AVC moderado); benefício na redução do AVC, com um equilíbrio no balanço em relação ao aumento de risco de sangramento.
Confiança na estimativa do efeito (qualidade de evidência). Quanto mais baixa a confiança, menos provável de termos uma recomendação forte.	Muitos ensaios clínicos randomizados de qualidade tem mostrado o benefício de esteróides inalados em asma.	Somente série de casos tem examinado a utilidade de pleurodese em pneumotórax.
Incerteza ou variabilidade em valores e preferências. Quanto menor a confiança na estimativa de valores e preferências típicas, e uma maior variabilidade, menos provável em termos uma recomendação forte.	Confiança relativa: evidência a partir de estudos empíricos mostram que pacientes referem um valor substancialmente mais elevados em evitar um AVC incapacitante, do que evitar sangramento gastrointestinal grave.	Incerteza: não existe evidência empírica em relação ao valor relativo na opinião dos pacientes em evitar um sangramento pós-operatório que requeira reoperação em comparação com gravidade pós operatório, mas não fatal de embolia pulmonar.
	Variabilidade pequena: pacientes jovens com linfoma irão colocar invariavelmente um valor mais elevado nos efeitos da quimioterapia em prolongar a vida, em vez de evitar a toxicidade do tratamento.	Grande variabilidade: alguns pacientes mais idosos com linfoma terão a opinião com um alto valor nos efeitos de prolongamento da vida com uso de quimioterapia em comparação a evitar toxicidade no tratamento, no entanto outros têm uma opinião contrária.
Uso de recursos. Quanto mais elevado o uso de recursos, menos provável de termos uma recomendação forte.	O baixo custo da aspirina versus nenhuma profilaxia antitrombótica contra o AVC em pacientes com ataque isquêmico transitório.	O alto custo de clopidogrel e a combinação de dipiridamol e aspirina versus aspirina com profilaxia contra o AVC em pacientes com ataque isquêmico transitório.

Fonte: Elaboração própria.

Quadro 14 – Qualidade geral da evidência de acordo com o nível de evidência dos desfechos críticos

Exemplo 1:

Mortalidade: RR 0,90 (IC 95% 0,68 – 1,13) – qualidade baixa

Trombose venosa: RR 0,56 (0,40 – 0,69) – qualidade alta

Sangramento grave: RR 1,30 (0,85 – 2,10) – qualidade baixa

Qualidade geral da evidência: baixa – apesar de termos alta certeza na estimativa de benefício para prevenção de trombose venosa, a qualidade da evidência é baixa para potenciais efeitos adversos (sangramento grave).

Exemplo 2:

- Mortalidade: RR 0,82 (0,70 – 0,90) – qualidade baixa

- Acidente vascular cerebral (AVC): RR 0,87 (0,72 – 0,99) – qualidade baixa

- Infarto agudo do miocárdio: RR 0,71 (0,58 – 0,81) – qualidade alta

Qualidade geral da evidência: alta – apesar de termos baixa confiança na estimativa de efeito para mortalidade e para AVC, é provável que haja benefício e a intervenção não aumente o risco dos mesmos. Uma vez que há benefício documentado, com qualidade de evidência alta, para prevenção de IAM, o nível geral de evidência será alto, uma vez que IAM (o que é IAM?) foi considerado desfecho crítico para tomada de decisão.

Fonte: Elaboração própria.

Quadro 15 – Processo de avaliação da qualidade da evidência e da graduação da recomendação segundo o sistema GRADE

54

Questão (PICO)				
População: Intervenção: Comparação: Cenário:		Introdução sobre o problema:		
DOMÍNIO		JULGAMENTO	EVIDÊNCIA DA PESQUISA	CONSIDERAÇÕES ADICIONAIS
P R O B L E M A	O problema é prioritário?	Não		
		Provavelmente não		
		Incerto		
		Provavelmente sim		
		Sim		
		Há variabilidade		
B E N E F Í C I O S E R I S C O S	Qual a qualidade da evidência (nível de evidência para o conjunto das evidências)?	Sem estudos concluídos		
		Muito baixo		
		Baixo		
		Moderado		
		Alto		
		Desfechos críticos	1...	2...
Alto				
Moderado				
Baixo				
Muito Baixo				

Continua

Continuação

Questão (PICO)																																	
População: Intervenção: Comparação: Cenário:		Introdução sobre o problema:																															
DOMÍNIO	JULGAMENTO	EVIDÊNCIA DA PESQUISA	CONSIDERAÇÕES ADICIONAIS																														
	<p>Benefícios sobrepõem os riscos</p> <p>Benefícios sobrepõem levemente os riscos</p> <p>Há equilíbrio entre riscos e benefícios</p> <p>Riscos sobrepõem levemente os benefícios</p> <p>Riscos sobrepõem os benefícios</p>	<table border="1"> <tr> <td>1...</td> <td>1...</td> <td>2...</td> <td>3...</td> <td></td> <td></td> </tr> <tr> <td>2...</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>3...</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>4...</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>5...</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>	1...	1...	2...	3...			2...						3...						4...						5...						
1...	1...	2...	3...																														
2...																																	
3...																																	
4...																																	
5...																																	
V A L O R E S E P R E F E R Ê N C I A S	<p>Há similaridade em relação à importância dada para os principais desfechos?</p> <p>Não</p> <p>Provavelmente não</p> <p>Incerto</p> <p>Provavelmente sim</p> <p>Sim</p>																																

Continua

Continuação

56

Questão (PICO)																
População: Intervenção: Comparação: Cenário:		Introdução sobre o problema:														
DOMÍNIO	JULGAMENTO	EVIDÊNCIA DA PESQUISA	CONSIDERAÇÕES ADICIONAIS													
R E C U R S O S	Os custos ou uso de recursos associados à intervenção são pequenos? (valores unitários – perspectiva focada no indivíduo)	<table border="1"> <tr><td>Não</td><td></td></tr> <tr><td>Provavelmente não</td><td></td></tr> <tr><td>Incerto</td><td></td></tr> <tr><td>Provavelmente sim</td><td></td></tr> <tr><td>Sim</td><td></td></tr> </table>	Não		Provavelmente não		Incerto		Provavelmente sim		Sim					
	Não															
Provavelmente não																
Incerto																
Provavelmente sim																
Sim																
Os custos incrementais (ou uso de recursos) são pequenos em comparação aos benefícios esperados?	<table border="1"> <tr><td>Não</td><td></td></tr> <tr><td>Provavelmente não</td><td></td></tr> <tr><td>Incerto</td><td></td></tr> <tr><td>Provavelmente sim</td><td></td></tr> <tr><td>Sim</td><td></td></tr> <tr><td>Há variabilidade</td><td></td></tr> </table>	Não		Provavelmente não		Incerto		Provavelmente sim		Sim		Há variabilidade				
Não																
Provavelmente não																
Incerto																
Provavelmente sim																
Sim																
Há variabilidade																
E Q U I D A D E	Quais os impactos referentes às inequidades em saúde?	<table border="1"> <tr><td>Aumento</td><td></td></tr> <tr><td>Provável aumento</td><td></td></tr> <tr><td>Incerto</td><td></td></tr> <tr><td>Provável redução</td><td></td></tr> <tr><td>Redução</td><td></td></tr> <tr><td>Há variabilidade</td><td></td></tr> </table>	Aumento		Provável aumento		Incerto		Provável redução		Redução		Há variabilidade			
Aumento																
Provável aumento																
Incerto																
Provável redução																
Redução																
Há variabilidade																

Continua

Continuação

Questão (PICO)															
População: Intervenção: Comparação: Cenário:		Introdução sobre o problema:													
DOMÍNIO	JULGAMENTO	EVIDÊNCIA DA PESQUISA	CONSIDERAÇÕES ADICIONAIS												
A C E I T A B I L I D A D E	A opção é aceitável para as principais partes interessadas?	<table border="1"> <tr><td>Não</td><td></td></tr> <tr><td>Provavelmente não</td><td></td></tr> <tr><td>Incerto</td><td></td></tr> <tr><td>Provavelmente sim</td><td></td></tr> <tr><td>Sim</td><td></td></tr> <tr><td>Há variabilidade</td><td></td></tr> </table>	Não		Provavelmente não		Incerto		Provavelmente sim		Sim		Há variabilidade		
Não															
Provavelmente não															
Incerto															
Provavelmente sim															
Sim															
Há variabilidade															
V I A B I L I D A D E	A opção é viável para implementar?	<table border="1"> <tr><td>Não</td><td></td></tr> <tr><td>Provavelmente não</td><td></td></tr> <tr><td>Incerto</td><td></td></tr> <tr><td>Provavelmente sim</td><td></td></tr> <tr><td>Sim</td><td></td></tr> <tr><td>Há variabilidade</td><td></td></tr> </table>	Não		Provavelmente não		Incerto		Provavelmente sim		Sim		Há variabilidade		
Não															
Provavelmente não															
Incerto															
Provavelmente sim															
Sim															
Há variabilidade															

Continua

Conclusão

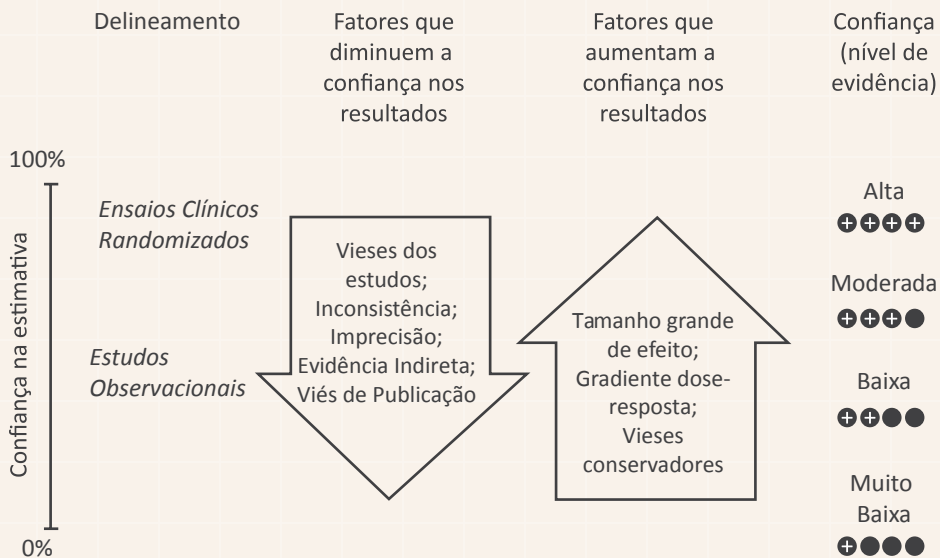
RECOMENDAÇÃO						
QUESTÃO DE PESQUISA						
Balanco entre riscos e benefícios.	As desvantagens claramente superam as vantagens.	As desvantagens provavelmente superam as vantagens.	O balanço entre vantagens e desvantagem é incerto*.	O balanço entre vantagens e desvantagem indica que eles são semelhantes*.	Os benefícios provavelmente superam as consequências indesejáveis.	As vantagens claramente superam as desvantagens.
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Recomendação forte contra adoção da intervenção.	Recomendação fraca contra a não adoção da intervenção.			Recomendação fraca A FAVOR da adoção da intervenção.	Recomendação forte A FAVOR da adoção da intervenção.
	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Decisão						
Recomendação (texto)						
Observações e justificativa						
Considerações para implementação						
Prioridades em pesquisa						
*Nessas situações, não fornecer recomendação é uma alternativa justificável.						

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.



Figuras

Figura 1 – Graduação dos níveis de evidência de acordo com o sistema GRADE



Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Figura 2 – Exemplo de questões de pesquisa no formato PICO

1 - A N-acetilcisteína previne a nefropatia induzida por contraste em pacientes de alto risco submetidos a procedimento com contraste?

População: *pacientes de alto risco para nefropatia induzida por contraste submetidos a procedimento com contraste*

Intervenção: *N-acetilcisteína*

Comparação: *Placebo ou ausência de medicação*

Outcome (desfecho): *nefropatia induzida por contraste*

2 - Qual é a eficácia da angioplastia guiada por ultrassonografia intra-coronária versus angioplastia guiada por angiografia quantitativa na incidência de eventos cardiovasculares maiores?

População: *pacientes submetidos à angioplastia*

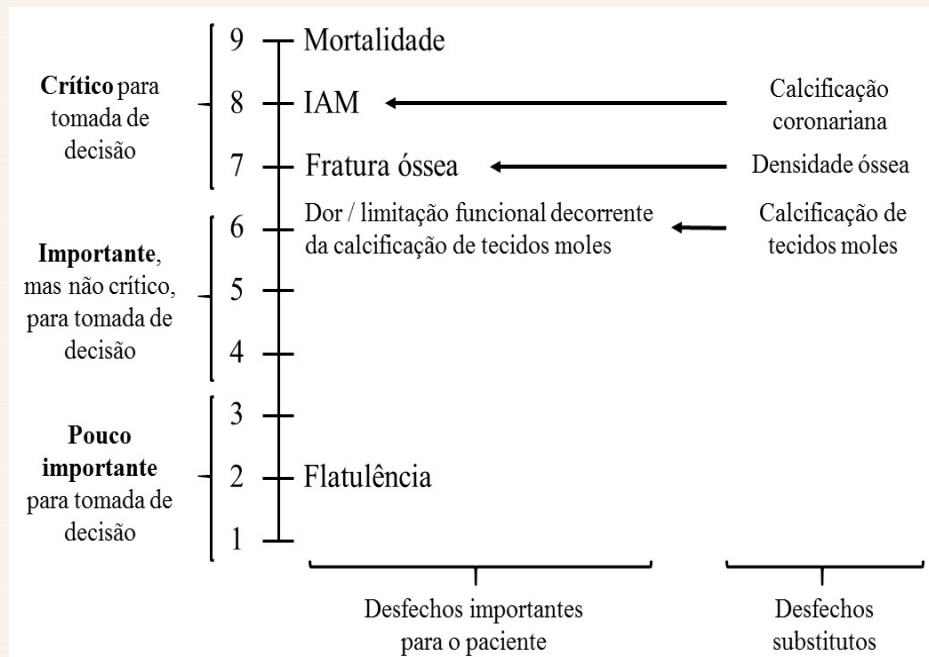
Intervenção: *angioplastia guiada por ultrassom*

Comparação: *angioplastia guiada por angiografia quantitativa*

Outcome (desfecho): *eventos cardiovasculares maiores*

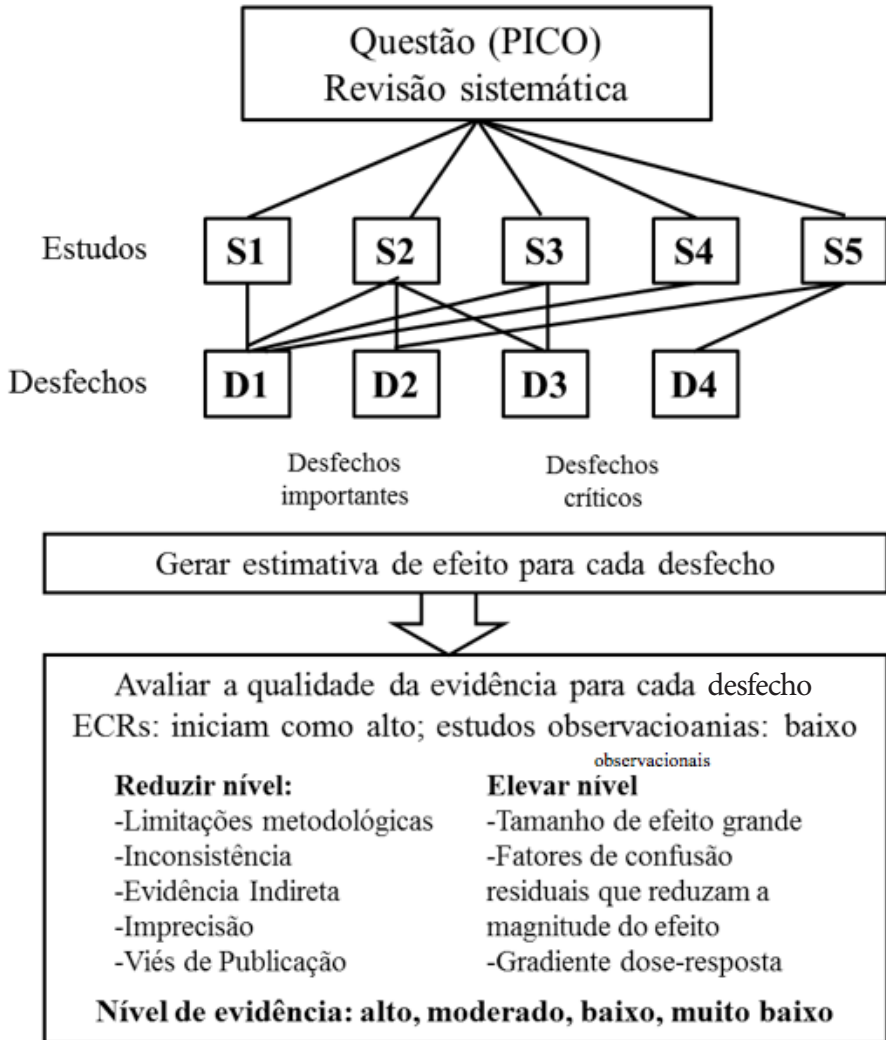
Fonte: Elaboração própria.

Figura 3 – Importância relativa dos desfechos



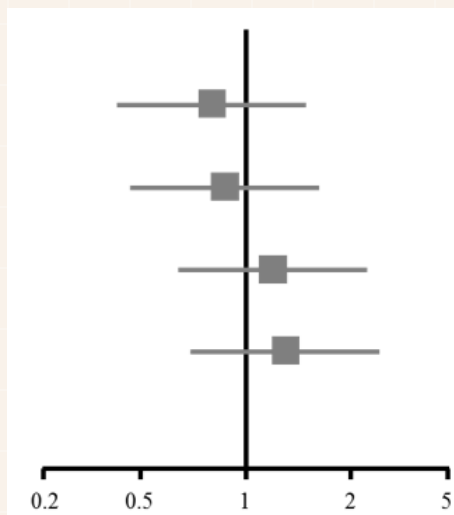
Fonte: (GUYATT et al., 2011, adaptado)¹⁶.

Figura 4 – Processo de avaliação da qualidade da evidência segundo o sistema GRADE



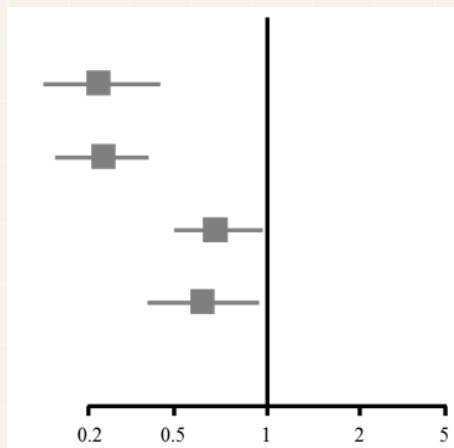
Fonte: Elaboração própria.

Figura 5 – Diferenças em direção de efeito com mínima heterogeneidade



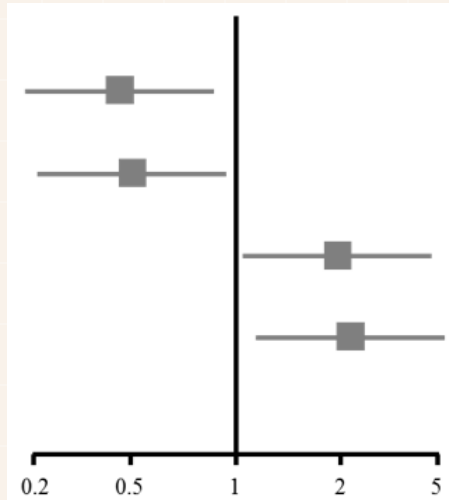
Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Figura 6 – Heterogeneidade substancial com importância questionável (mesma direção de efeito)



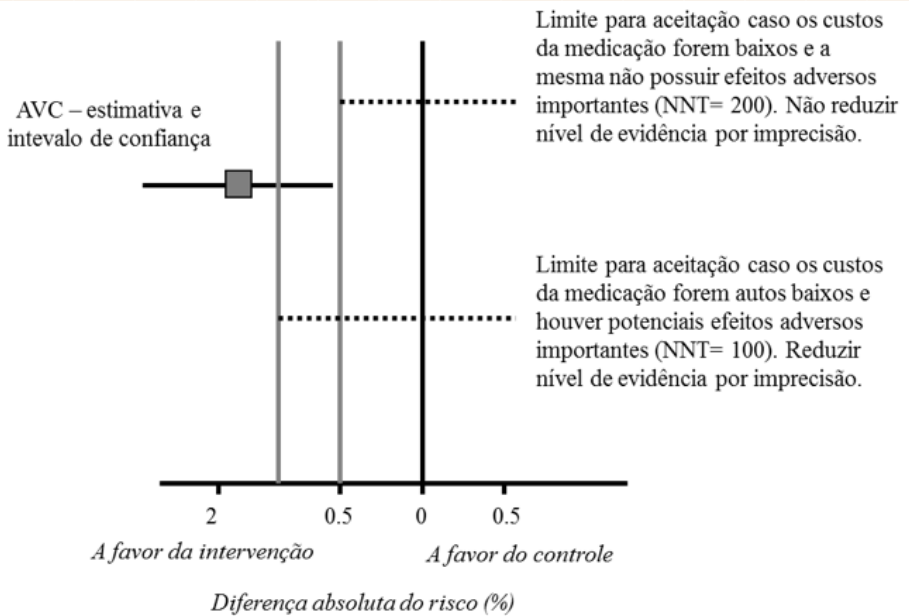
Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Figura 7 – Heterogeneidade substancial, com inequívoca importância



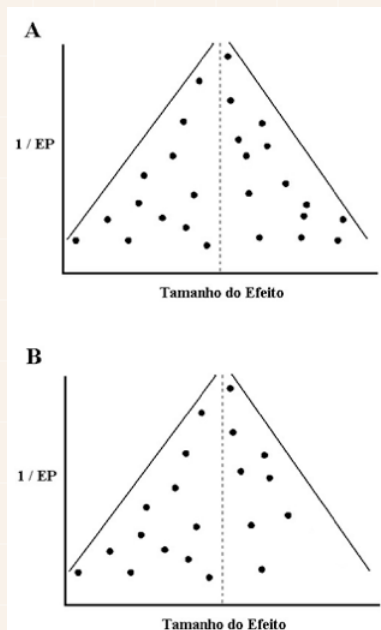
Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Figura 8 – Avaliação da imprecisão em diretrizes



Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

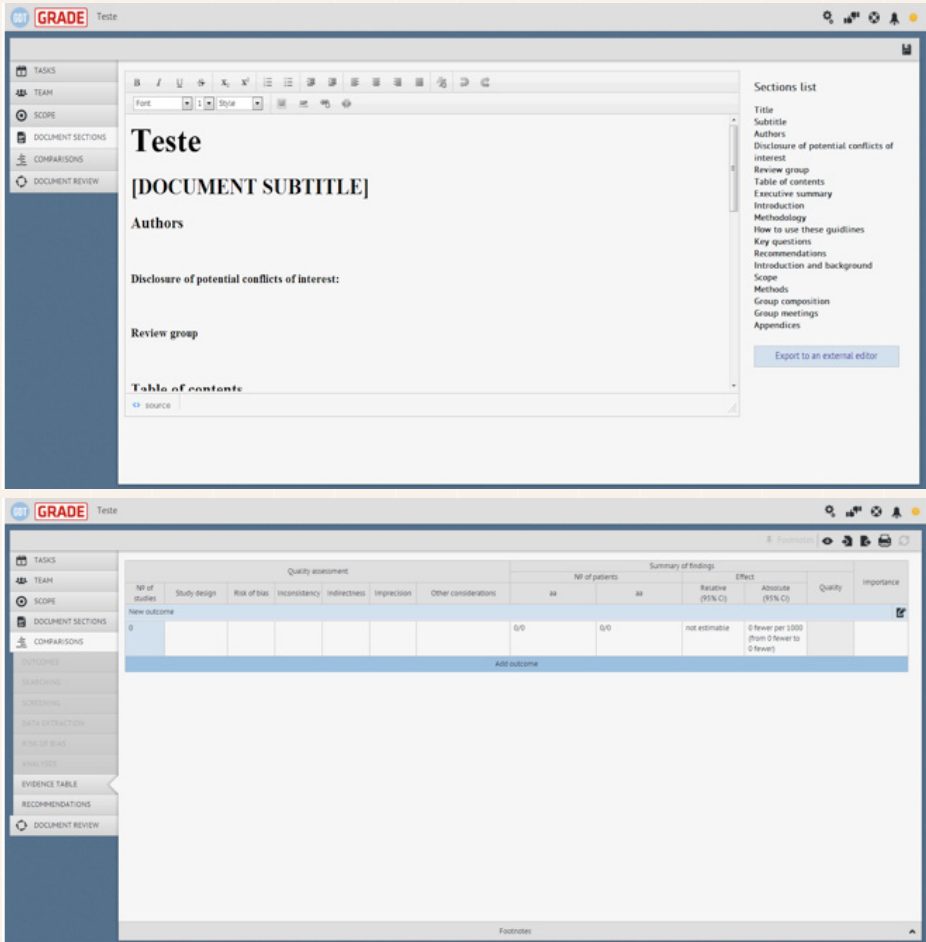
Figura 9 – Gráfico em funil (*funnel plot*). (A) Gráfico em funil na ausência de assimetria. (B) Gráfico em funil assimétrico



Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Na figura são apresentados dois gráficos em funil, o primeiro deles seguindo uma distribuição ideal (A) e o segundo apresentando forte assimetria (B). Cada ponto no gráfico representa um estudo e a sua totalidade constitui o conjunto de estudos de uma metanálise. O eixo X do gráfico representa o tamanho do efeito, enquanto o eixo Y corresponde ao inverso do erro padrão (EP), que por sua vez está relacionado ao tamanho amostral. A linha pontilhada corresponde ao verdadeiro tamanho do efeito para determinada intervenção.

No primeiro gráfico (A) observamos que os estudos se distribuem simetricamente, tendo como centro a linha pontilhada, com os estudos maiores mais próximos à real medida de associação e com os menores se distribuindo em ambos os lados da linha, com maior dispersão quanto menor for o EP mesmos. No segundo gráfico (B) isso não acontece, havendo a impressão de que foram captados na revisão os estudos maiores – assim como no primeiro gráfico – e estudos pequenos com resultados favoráveis, porém não aqueles com amostra pequena e com ausência de resultados positivos.

Figuras 10 e 11 – *Guideline development tool*


The figure displays two screenshots of the GRADE Guideline Development Tool (GDT) interface. The top screenshot shows the document editing view for a document titled "Teste". The document content includes sections for "Title", "Subtitle", "Authors", "Disclosure of potential conflicts of interest:", "Review group", and "Table of contents". A "Sections list" panel on the right provides a navigation menu for these sections. The bottom screenshot shows the "Summary of findings" table, which is used for grading evidence and determining the strength of recommendations. The table includes columns for quality assessment, number of patients, and effect estimates.

New outcome	Quality assessment						Summary of findings						
	Nº of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nº of patients	Relative (95% CI)	Effect	Absolute (95% CI)	Quality	Importance
0								0/0	0/0	not estimable	0 fewer per 1000 (from 0 fewer to 0 fewer)		
Add outcome													

Fonte: Elaboração GRADE working group - <<http://www.gradeworkinggroup.org>>.

Referências

1. BRASIL. Ministério da Saúde. **Diretrizes metodológicas:** elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados. Brasília: Ministério da Saúde, 2012.
2. ALMEIDA FILHO, Naomar de. A clínica, a epidemiologia e a epidemiologia clínica. **Physis**, Rio de Janeiro, v. 3, n. 1, 1993. Disponível em: <http://www.cebm.net/mod_product/design/files/CEBM-Levels-of-Evidence-2.1.pdf> Acessado em: out. 2013.
3. CENTRE FOR EVIDENCE BASED MEDICINE. **Oxford centre for evidence-based medicine 2011 levels of evidence.** Disponível em: <http://www.cebm.net/mod_product/design/files/CEBM-Levels-of-Evidence-2.1.pdf>. Acessado em: out. 2013.
4. SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK (SIGN). **Grading System.** Disponível em: <<http://www.sign.ac.uk/guidelines/fulltext/50/annexoldb.html>>. Acessado em: out. 2013.
5. GRADE WORKING GROUP. Disponível em: <<http://gradeworkinggroup.org/society/index.htm>>. Acessado em: out. 2013.
6. GUYATT, G. H. et al. Rating quality of evidence and strength of recommendations: What is “quality of evidence” and why is it important to clinicians? **British Medical Journal**, London, v. 336, n. 7651, p. 995-998, 2008.
7. BALSHEM, H. et al. Grade guidelines: 3 rating the quality of evidence. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 4, p. 401-406, 2011.
8. GUYATT, G. H. et al. GRADE guidelines: 4 rating the quality of evidence-study limitations (risk of bias). **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 4, p. 407-415, 2011.
9. GUYATT, G. H. et al. GRADE guidelines: 7 rating the quality of evidence-inconsistency. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 12, p. 1294-1302, 2011.
10. GUYATT, G. H. et al. GRADE guidelines: 8 rating the quality of evidence-indirectness. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 12, p. 1303-1310, 2011.
11. GUYATT, G. H. et al. GRADE guidelines: 6 rating the quality of evidence-imprecision. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 12, p. 1283-1293, 2011.
12. GUYATT, G. H. et al. GRADE guidelines: 5 rating the quality of evidence-publication bias. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 12, p. 1277-1282, 2011.

13. GUYATT, G. H. et al. GRADE guidelines: 9 rating up the quality of evidence. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 12, p. 1311-1316, 2011.
14. GUYATT, G. H. et al. GRADE guidelines: 12 preparing summary of findings tables-binary outcomes. **Journal of Clinical Epidemiology**, Maryland Heights, v. 66, n. 2, p. 158-172, 2013.
15. GUYATT, G. H. et al. Rating quality of evidence and strength of recommendations: going from evidence to recommendations. **British Medical Journal**, London, v. 336, n. 7652, p. 1049-1051, 2008.
16. GUYATT, G. H. et al. GRADE guidelines: 2 framing the question and deciding on important outcomes. **Journal of Clinical Epidemiology**, Maryland Heights, v. 64, n. 4, p. 395-400, 2011.
17. HIGGINS, J. P. T.; GREEN, S. (Ed.). **Cochrane handbook for systematic reviews of interventions version 5.1.0**. The Cochrane Collaboration, 2011. Disponível em: <www.cochrane-handbook.org>. Acessado em: out. 2013.
18. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non randomised studies in meta-analyses. Disponível em: <http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp>. Acessado em: out. 2013.
19. STERNE, J. A. C. et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. **British Medical Journal**, London, v. 342, d4002, 2011.
20. EGGER, M. et al. Bias in meta-analysis detected by a simple, graphical test. **British Medical Journal**, London, v. 315, p. 629-634, 1997.
21. CANNEGIETER, S. C.; ROSENDAAL, F. R.; BRIËT, E. Thromboembolic and bleeding complications in patients with mechanical heart valve prostheses. **Circulation**, Dallas, v. 89, n. 2, p. 635-641, 1994.
22. DEVEREAUX, P. J. et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. **Canadian Medical Association Journal**, Ottawa, v. 166, n. 11, p. 1399-1406, 2002.
23. GLYNN, R. J. et al. A randomized trial of rosuvastation in the prevention of venous thromboembolism. **The New England Journal of Medicine**, Waltham, v. 360, n. 18, p. 1851-1861, 2009.

24. ANDREWS, J. C. et al. GRADE guidelines: 15 going from evidence to recommendation-determinants of a recommendation's direction and strength. **Journal of Clinical Epidemiology**, Maryland Heights, v. 66, n. 7, p. 726-735, 2013.
25. BRASIL. Ministério da Saúde. **Diretrizes Metodológicas: estudos de avaliação econômica de tecnologias em saúde**. Brasília: Ministério da Saúde, 2009.
26. MUSTAFA, R. A. et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. **Journal of Clinical Epidemiology**, Maryland Heights, v. 66, n. 7, p. 736-742, 2013.
27. SCHÜNEMANN, H. J. et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. **British Medical Journal**, London, v. 336, n. 7653, p. 1106-1110, 2008.

Esta obra foi impressa em papel couché fosco 240 g/m² (capa) e papel off set 90 g/m² (miolo) pela Nome da Gráfica, em novembro de 2014. A Editora do Ministério da Saúde foi responsável pela normalização (OS 0600).

ISBN 978-85-334-2186-8



9 788533 421868

DISQUE SAÚDE



Operatório Geral do SUS
www.saude.gov.br

Biblioteca Virtual em Saúde do Ministério da Saúde
www.saude.gov.br/bvs



Ministério da
Saúde