



Certeza das evidências, por quê?

João Pedro Lima¹, Xiajing Chu¹, Gordon H Guyatt¹,
Wimonchat Tangamornsuksan^{1,2}

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton (ON) Canada.
2. Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok, Thailand.

Recebido: 11 maio 2023.

Aprovado: 12 maio 2023.

Trabalho realizado no Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton (ON) Canada.

RESUMO

Para tomar a melhor decisão clínica, é preciso compreender as evidências a respeito dos benefícios, malefícios e ônus das alternativas de manejo. Revisões sistemáticas e meta-análises que sejam realizadas com rigor oferecem resumos precisos das evidências. No entanto, é possível que esses resumos avaliem apenas as evidências cujo grau de certeza é baixo e, ao fazê-lo, ressaltam que provavelmente não existe uma decisão única que será a melhor para todos os pacientes. O *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE) é um método sistemático e transparente para avaliar o grau de certeza das evidências em revisões sistemáticas. Neste artigo, abordaremos a importância de avaliar o grau de certeza das evidências; explicaremos como o sistema GRADE classifica o grau de certeza das evidências provenientes de revisões sistemáticas e apresentaremos o *evidence to decision framework* (quadro para a avaliação de evidências) do GRADE para decidir se as evidências se traduzem em recomendações fortes ou fracas nas diretrizes de prática clínica.

Descritores: Revisões Sistemáticas como Assunto; Metanálise como Assunto; Medicina Baseada em Evidências; Tomada de decisões.

INTRODUÇÃO

Para responder às perguntas dos pacientes sobre as opções de tratamento, os médicos precisam considerar as evidências relevantes a respeito dos benefícios, malefícios e ônus. Revisões sistemáticas e meta-análises abordam questões clínicas estruturadas e, quando bem-feitas, fornecem resumos precisos das evidências. Entretanto, quando o grau de certeza das evidências é baixo (isto é, quando se diz que as evidências são de baixa qualidade), mesmo os resumos de evidências mais rigorosos deixarão grande incerteza a respeito dos benefícios e malefícios. Neste artigo, abordaremos a importância de avaliar o grau de certeza das evidências provenientes de estudos de intervenção e diagnóstico e explicaremos a abordagem *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE) para avaliar o grau de certeza das evidências provenientes de revisões sistemáticas e a força das recomendações nas diretrizes de prática clínica.

Pacientes, médicos e formuladores de políticas serão frequentemente induzidos em erro se não considerarem o grau de certeza das evidências. Tomemos como exemplo o uso de glicocorticoides sistêmicos, que até recentemente eram amplamente usados no manejo da fibrose pulmonar idiopática. O grau de certeza das evidências a favor do benefício do uso de glicocorticoides nesses pacientes nunca foi maior que baixo, ao passo que há evidências de alto grau de certeza a respeito dos múltiplos malefícios dessa intervenção.⁽¹⁾ A melhor prática para os médicos que venham a oferecer a terapia com glicocorticoides aos pacientes seria esclarecer a natureza especulativa de quaisquer benefícios e o alto risco de malefícios substanciais.

Muitos pacientes, cientes dos benefícios incertos e do alto grau de certeza das evidências de malefícios, recusariam a intervenção. A falha em reconhecer que as evidências de benefício apresentam baixo grau de certeza resultaria no uso excessivo da intervenção.

Uma avaliação formal do grau de certeza das evidências é uma estratégia eficaz para evitar o uso excessivo de intervenções cujos benefícios são questionáveis. O GRADE é um método sistemático e transparente para avaliar o grau de certeza das evidências em revisões sistemáticas (Quadro 1) e para elaborar e determinar a força das recomendações em diretrizes de prática clínica.⁽²⁾ Mais de 110 organizações, incluindo a Organização Mundial da Saúde, o *National Institute for Health and Care Excellence* (Reino Unido), a *Cochrane Collaboration* e as principais organizações profissionais estadunidenses, incluindo a *American Thoracic Society* e o *American College of Chest Physicians*, já adotaram o método GRADE. Além disso, o principal livro didático eletrônico do mundo, o *UpToDate*, inclui mais de 10.000 recomendações GRADE. O GRADE é atualmente o padrão ouro para realizar revisões sistemáticas e elaborar diretrizes.⁽³⁾

A aplicação do sistema GRADE de avaliação do grau de certeza das evidências requer a disponibilidade de conduzir com rigor revisões sistemáticas para abordar questões clínicas. O GRADE também oferece aos grupos envolvidos na elaboração de diretrizes os *evidence to decision* (EtD) *frameworks* (quadros para a avaliação de evidências) para ajudá-los a decidir como as evidências se traduzem em recomendações.⁽⁴⁾ Depois de considerar todas as questões destacadas no quadro EtD, os painéis

Endereço para correspondência:

Wimonchat Tangamornsuksan. Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, 10210, Bangkok, Thailand. Tel.: 66 838704088. E-mail: tangamow@mcmaster.ca ou wimonchat.tan@gmail.com; ou João Pedro Lima. Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West 2C Area, Hamilton, Ontario, L8S 4K1, Canada. Tel.: 1 365 888-0631. E-mail: limaj1@mcmaster.ca
Apoio financeiro: Nenhum.

Quadro 1. Grau de certeza das evidências: critérios de avaliação.

Desenho do estudo	Confiança nas estimativas	Menor se	Maior se
Ensaio randomizados	Alta	Risco de viés -1 Grave -2 Muito grave	Efeito grande +1 Grande +2 Muito grande
	Moderada	Inconsistência -1 Grave -2 Muito grave	Resposta à dose +1 Evidências de gradiente
Estudos observacionais	Baixa	<i>Indirectness</i> -1 Grave -2 Muito grave	Todos os confundidores plausíveis +1 Reduziriam um efeito demonstrado ou
	Muito baixa	Imprecisão -1 Grave -2 Muito grave Viés de publicação -1 Provável -2 Muito provável	+1 Sugeririam um efeito espúrio quando os resultados não mostram nenhum efeito

de especialistas responsáveis pela elaboração de diretrizes emitirão uma recomendação forte ou fraca a favor ou contra um tratamento ou teste diagnóstico.

É possível que médicos ingênuos se sintam prematuramente inclinados a mudar sua prática com base nos resultados de um único ensaio randomizado, sem dar atenção a considerações de risco de viés, imprecisão em virtude do tamanho limitado da amostra e aplicabilidade caso os pacientes incluídos não sejam suficientemente semelhantes àqueles sob seus cuidados. Além disso, é possível que médicos ingênuos estejam prontos para mudar sua prática de maneira inapropriada com base em uma revisão sistemática e meta-análise cujas evidências sejam exclusivamente de baixo grau de certeza. As evidências podem apresentar baixo grau de certeza se forem exclusivamente provenientes de estudos observacionais não randomizados. As evidências também podem apresentar baixo grau de certeza mesmo que sejam baseadas em ensaios randomizados caso esses ensaios apresentem limitações quanto ao desenho do estudo e tamanho da amostra; resultados inconsistentes; ou limitações quanto à aplicabilidade aos pacientes em questão. Nas seções seguintes desta revisão, faremos mais considerações a respeito das limitações de ensaios clínicos controlados randomizados (ECR) e revisões sistemáticas e meta-análises no processo de tomada de decisão clínica, ressaltando a importância do GRADE na avaliação do grau de certeza das evidências e recomendações de tratamento e testes diagnósticos em diretrizes de prática clínica.

A ABORDAGEM GRADE EM REVISÕES SISTEMÁTICAS E RECOMENDAÇÕES CLÍNICAS

Abordagem GRADE para avaliar o grau de certeza de evidências a respeito de intervenções

A abordagem GRADE do grau de certeza das evidências começa com o reconhecimento de que decisões clínicas sólidas requerem resumos sistemáticos rigorosos das evidências da mais alta qualidade a respeito das intervenções que estejam sendo levadas em

consideração. Uma vez que uma revisão sistemática assim esteja disponível, a avaliação GRADE do grau de certeza das evidências começa com o desenho do estudo: ensaios randomizados começam como evidências de alto grau de certeza e estudos observacionais começam como evidências de baixo grau de certeza no sistema GRADE de quatro graus de certeza das evidências (alto, moderado, baixo e muito baixo; Quadro 1).⁽²⁾ Depois do desenho do estudo, o GRADE identificou cinco domínios que merecem consideração ao avaliar o grau de certeza das evidências: risco de viés, inconsistência, *indirectness*, imprecisão e viés de publicação (Quadro 1).⁽²⁾

Os revisores diminuem o grau de certeza das evidências em um nível quando identificam questões sérias e em dois níveis quando identificam questões muito sérias em qualquer um desses cinco domínios. Os revisores podem aumentar o grau de certeza das evidências provenientes de estudos observacionais, principalmente em virtude da grande ou muito grande magnitude do efeito.⁽⁵⁾ Os revisores avaliam o grau de certeza das evidências não de estudos individuais, mas da totalidade de evidências resumidas em revisões sistemáticas, e separadamente para cada desfecho. Todos os desfechos importantes para o paciente são avaliados quanto ao grau de certeza.

A seguir, faremos breves considerações a respeito dos cinco motivos pelos quais se deve diminuir o grau de certeza das evidências. No tocante ao risco de viés,⁽⁶⁾ os ensaios randomizados podem ser limitados pelos seguintes fatores: randomização não sigilosa; ausência de cegamento de pacientes, médicos, coletores de dados e julgadores; e perdas de seguimento. Os ensaios randomizados também superestimarão os efeitos do tratamento se forem interrompidos precocemente em virtude de grandes efeitos do tratamento, particularmente se o tamanho da amostra for pequeno.⁽⁷⁾

Em segundo lugar, a certeza diminui quando os resultados de diferentes estudos apresentam inconsistências que não são explicadas. Os revisores julgam a consistência pela semelhança das estimativas

pontuais e pela extensão da sobreposição dos IC. Critérios estatísticos podem dar ainda mais apoio a julgamentos de inconsistência, incluindo testes de heterogeneidade (Pode-se atribuir ao acaso as diferenças entre os resultados dos estudos?) e I^2 , que quantifica a inconsistência em uma escala de 0 a 100.^(8,9)

Em terceiro lugar, os estudos incluídos em uma revisão sistemática devem refletir a pergunta da revisão. Para avaliar *indirectness* (o termo GRADE referente à aplicabilidade das evidências à pergunta em questão), os revisores consideram se os pacientes, intervenções, comparações e desfechos diferem dos de interesse.⁽¹⁰⁾ *Indirectness* é ainda mais importante para diretrizes do que para revisões sistemáticas.

Em quarto lugar, o GRADE considera a amplitude dos IC em torno das estimativas dos efeitos absolutos do tratamento.⁽¹¹⁾ Para diminuir o grau de certeza das evidências, é preciso considerar se o IC ultrapassa um limiar de interesse. Por exemplo, se toda a confiança estiver na faixa de um efeito importante, não haverá redução por imprecisão. Se ultrapassar o limiar de importância, deixando dúvida se um efeito é trivial ou importante, os revisores irão diminuir o grau de certeza. Tomemos como exemplo a Figura 1: no caso da intervenção A, os revisores diminuiriam o grau de certeza por imprecisão, ao passo que no caso da intervenção B, não o fariam.

Finalmente, ensaios que não conseguem mostrar os efeitos positivos do tratamento podem não ser publicados e, assim, resultar em superestimação do efeito do tratamento, um fenômeno conhecido como viés de publicação.⁽¹²⁾ Os autores de uma revisão suspeitarão de viés de publicação nos casos em que uma empresa farmacêutica patrocinou todos os estudos disponíveis, particularmente se o tamanho da amostra dos estudos for pequeno.

Caso um conjunto de evidências provenientes de ensaios randomizados apresente várias dessas limitações, os revisores podem reduzir o grau de certeza das evidências para moderado, baixo ou mesmo muito

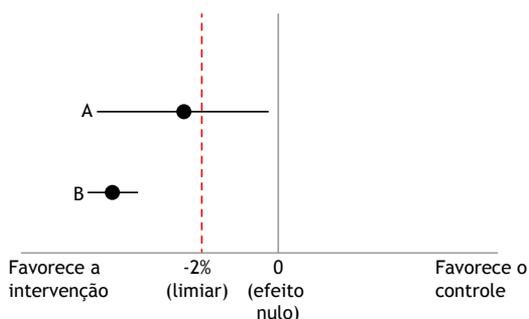


Figura 1. Avaliando a imprecisão: deve-se considerar se o intervalo de confiança ultrapassa um limiar de interesse.

Quadro 2. Domínios que afetam a força de uma recomendação.

- Desfechos desejáveis e indesejáveis (efeitos estimados)
- Grau de certeza das evidências
- Incerteza ou variabilidade de valores e preferências
- Uso de recursos (custos), viabilidade, aceitabilidade e equidade

baixo. Além disso, essas limitações também se aplicam a estudos observacionais e podem levar à redução do grau de certeza de baixo para muito baixo. Em raras ocasiões, os revisores podem aumentar o grau de certeza em virtude de efeitos grandes ou muito grandes (insulina para cetoacidose diabética e diálise para doença renal terminal, por exemplo).

Como o fazem com intervenções terapêuticas, as revisões sistemáticas devem embasar questões clínicas diagnósticas.⁽¹³⁾ A maioria dos estudos de testes diagnósticos concentra-se exclusivamente na precisão diagnóstica, e os cinco motivos pelos quais se deve diminuir o grau de certeza das evidências aplicam-se a revisões sistemáticas de estudos assim.⁽¹⁴⁾ Idealmente, porém, os estudos concentrar-se-ão no impacto de estratégias diagnósticas alternativas em desfechos importantes para o paciente (mortalidade e qualidade de vida, por exemplo) por meio de desenhos randomizados.^(15,16) Para esses estudos, o grau de certeza das evidências é avaliado da mesma maneira que a abordagem GRADE a intervenções clínicas.

Como o GRADE ajuda a traduzir as evidências em recomendações?

O GRADE usa o quadro EtD para ajudar as pessoas a usar as evidências para embasar as decisões clínicas. Esse quadro inclui considerações sobre a magnitude dos benefícios, malefícios e ônus; o grau de certeza das evidências a respeito desses benefícios, malefícios e ônus; os valores e preferências do paciente; e, às vezes, custos, viabilidade, aceitabilidade e questões de equidade (Quadro 2).⁽⁴⁾ As recomendações clínicas, após a consideração de todas essas questões, devem fornecer declarações explícitas sobre a melhor conduta.

Painéis de especialistas envolvidos na elaboração de diretrizes fazem recomendações fortes quando concluem que todos ou quase todos os pacientes devidamente esclarecidos escolheriam a intervenção proposta. Por outro lado, fazem recomendações fracas (também conhecidas como condicionais) quando consideram que os pacientes, diante das opções de tratamento, fariam escolhas diversas baseadas em diferentes valores e preferências.⁽¹⁷⁾

Desfechos desejáveis e indesejáveis (efeitos estimados)

Quando os benefícios (desfechos desejáveis) são grandes e os malefícios e ônus (desfechos indesejáveis) são de pequena magnitude, é mais provável que os painéis de especialistas responsáveis pela elaboração de diretrizes emitam uma recomendação forte. Por outro lado, quando as consequências desejáveis e indesejáveis estão bem equilibradas, é provável que uma recomendação fraca seja mais apropriada.

Grau de certeza das evidências

Quando o grau de certeza das evidências é alto ou moderado, recomendações fortes podem ser apropriadas. Quando o grau de certeza das evidências é baixo ou muito baixo, é muito pouco provável que haja alta confiança de que os benefícios superem os malefícios e ônus (ou de que estes superem aqueles), e recomendações fracas quase sempre serão apropriadas.

Incerteza ou variabilidade de valores e preferências

Avaliar uma recomendação envolve determinar o valor que se atribui aos benefícios versus malefícios e ônus. Embora os pacientes tenham opiniões diferentes a respeito desses valores, os painéis de especialistas envolvidos na elaboração de diretrizes devem, ao fazer recomendações, se concentrar nos valores e preferências típicos ou habituais dos pacientes. Sendo esse o caso, uma grande variabilidade de valores e preferências na população relevante de pacientes tornará mais provável uma recomendação fraca, assim como o fará a incerteza a respeito dos valores e preferências dos pacientes. Embora frequentemente haja evidências limitadas a respeito das preferências e valores dos pacientes, a experiência clínica pode deixar o painel confiante de que os valores e preferências dos pacientes diferem amplamente.⁽¹⁴⁾

Uso de recursos (custos), viabilidade, aceitabilidade e equidade

Intervenções de tratamento ou testes diagnósticos podem aumentar ou diminuir o uso de recursos em comparação com uma alternativa. O impacto do custo pode variar, dependendo do contexto e da situação socioeconômica do paciente. As considerações adicionais, muitas vezes secundárias, são o uso de recursos, a viabilidade, a aceitabilidade e a equidade. Embora nem sempre sejam pertinentes, essas considerações às vezes são importantes, particularmente quando as diretrizes adotam uma perspectiva de saúde pública ou de sistemas, em vez da perspectiva individual do paciente.

Como os médicos interpretam e aplicam as recomendações GRADE à assistência ao paciente?

Os médicos devem ser capazes de diferenciar uma recomendação não confiável de recomendações

confiáveis; compreender o significado da força da recomendação e compreender como aplicar a recomendação à assistência ao paciente.⁽¹⁸⁾ Um guia para auxiliar os profissionais de saúde a interpretar e usar recomendações em diretrizes elaboradas com a abordagem GRADE sugere critérios específicos para interpretar, avaliar criticamente e aplicar as recomendações GRADE (Quadro 3).⁽¹⁷⁾

Compreendendo o significado da força da recomendação

Para interpretar as recomendações GRADE, os médicos devem considerar a força da recomendação e o grau de certeza das evidências. Painéis de especialistas envolvidos na elaboração de diretrizes por meio da abordagem GRADE emitirão recomendações fortes ou fracas/condicionais. Se estiver confiante de que os efeitos desejáveis superam as consequências indesejáveis, o painel emitirá uma recomendação forte, que geralmente será expressa pelo verbo “recomendamos”.⁽¹⁷⁾ Por outro lado, se estiver menos confiante a respeito do equilíbrio entre as consequências desejáveis e indesejáveis na conduta proposta, o painel emitirá uma recomendação fraca, que geralmente será expressa pelo verbo “sugerimos”.

Os painéis emitem recomendações fracas quando acreditam que é improvável que a recomendação se aplique a todos os pacientes. Nesse caso, os médicos devem empregar tempo para garantir que cada paciente receba a opção terapêutica que reflita seus valores e preferências.⁽¹⁹⁾

Distinguindo recomendações confiáveis de recomendações não confiáveis

Os médicos não devem simplesmente compreender os conceitos de força da recomendação e grau de certeza das evidências; eles também devem ser capazes de escolher diretrizes confiáveis para nelas embasar sua prática. Considerar cinco domínios pode auxiliar nessa escolha (Quadro 3).⁽¹⁷⁾

Todos os desfechos relevantes para o paciente foram explicitamente considerados?

O equilíbrio entre desejável e indesejável na conduta proposta dependerá de quais desfechos são considerados. Os médicos devem avaliar se o painel

Quadro 3. Guia do usuário do GRADE para profissionais de saúde, incluindo a interpretação, avaliação crítica e uso das recomendações GRADE na assistência ao paciente.

Compreendendo o significado da força da recomendação

O que significa força?

O que significa a certeza das evidências?

Distinguindo recomendações confiáveis de recomendações não confiáveis

Todos os desfechos relevantes para o paciente foram explicitamente considerados?

A recomendação foi baseada nas melhores evidências disponíveis atualmente?

A força da recomendação é apropriada?

A recomendação é clara e pode ser posta em prática?

A recomendação fornece as informações adicionais necessárias?

Aplicando as recomendações à assistência ao paciente

Recomendações fortes

Recomendações fracas

GRADE: *Grading of Recommendations Assessment, Development, and Evaluation.*

responsável pela elaboração das diretrizes considerou e incluiu todos os desfechos relevantes para o paciente.

[A recomendação foi baseada nas melhores evidências disponíveis atualmente?](#)

A recomendação deve ser baseada nas melhores evidências disponíveis atualmente. Para avaliar a credibilidade do processo de elaboração das diretrizes, os médicos devem levar em conta se as recomendações foram embasadas em uma revisão sistemática. Idealmente, os painéis responsáveis pela elaboração de revisões sistemáticas devem estar atualizados.

[A força da recomendação é apropriada?](#)

Os painéis responsáveis pela elaboração de diretrizes devem considerar todas as questões no quadro EtD ao fazer suas recomendações e raramente fazer recomendações fortes quando o grau de certeza das evidências é baixo (Quadro 2).

[A recomendação é clara e pode ser posta em prática?](#)

A recomendação deve fornecer os detalhes da ação recomendada, a situação à qual as recomendações se aplicam, a quem se aplicam e a ação clínica à qual a intervenção foi comparada.

[Aplicando as recomendações à assistência ao paciente](#)

Os médicos podem aplicar recomendações fortes a todos ou quase todos os pacientes sem a necessidade

de uma discussão detalhada com o paciente. No caso de recomendações fracas, os médicos devem compreender e ser capazes de comunicar as evidências aos pacientes num processo de tomada de decisão compartilhada.

CONSIDERAÇÕES FINAIS

Nem ECR individuais nem revisões sistemáticas das melhores evidências disponíveis garantem evidências de alto grau de certeza; na verdade, ECR e revisões sistemáticas rigorosas podem produzir evidências de baixo grau de certeza. O GRADE é um sistema para avaliar o grau de certeza das evidências em revisões sistemáticas e a força das recomendações nas diretrizes clínicas. Ao aplicar as diretrizes aos cuidados clínicos, os médicos devem compreender as implicações de recomendações fortes e particularmente fracas que os obrigam a considerar os valores e preferências do paciente em seu processo de tomada de decisão.

CONTRIBUIÇÕES DOS AUTORES

JPL, XC e WT contribuíram igualmente para este estudo. JPL, XC, GHG e WT: concepção e desenho do estudo. JPL, XC e WT: redação do manuscrito. JPL, XC, GHG e WT: revisão e aprovação da versão final do manuscrito.

CONFLITOS DE INTERESSE

Nenhum declarado.

REFERÊNCIAS

- Pitre T, Mah J, Helmecci W, Khalid MF, Cui S, Zhang M, et al. Medical treatments for idiopathic pulmonary fibrosis: a systematic review and network meta-analysis. *Thorax*. 2022;77(12):1243-1250. <https://doi.org/10.1136/thoraxjnl-2021-217976>
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394. <https://doi.org/10.1016/j.jclinepi.2010.04.026>
- GRADE Working Group [homepage on the Internet]. GRADE. Available from: <https://www.gradeworkinggroup.org/>
- Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016. <https://doi.org/10.1136/bmj.i2016>
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-1316. <https://doi.org/10.1016/j.jclinepi.2011.06.004>
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-415. <https://doi.org/10.1016/j.jclinepi.2010.07.017>
- Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203-2209. <https://doi.org/10.1001/jama.294.17.2203>
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol*. 2011;64(12):1294-1302. <https://doi.org/10.1016/j.jclinepi.2011.03.017>
- Guyatt G, Zhao Y, Mayer M, Briel M, Mustafa R, Izcovich A, et al. GRADE guidance 36: updates to GRADE's approach to addressing. *J Clin Epidemiol*. 2023;158:70-83. <https://doi.org/10.1016/j.jclinepi.2023.03.003>
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol*. 2011;64(12):1303-1310. <https://doi.org/10.1016/j.jclinepi.2011.04.014>
- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision [published correction appears in *J Clin Epidemiol*. 2021 Sep;137:265]. *J Clin Epidemiol*. 2011;64(12):1283-1293. <https://doi.org/10.1016/j.jclinepi.2011.01.012>
- Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol*. 2011;64(12):1277-1282. <https://doi.org/10.1016/j.jclinepi.2011.01.011>
- Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy*. 2009;64(8):1109-1116. <https://doi.org/10.1111/j.1398-9995.2009.02083.x>
- Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies [published correction appears in *BMJ*. 2008 May 24;336(7654)]. doi: 10.1136/bmj.a139. Schünemann H, Holger J [corrected to Schünemann, Holger JJ]. *BMJ*. 2008;336(7653):1106-1110. <https://doi.org/10.1136/bmj.39500.677199.AE>
- Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ*. 1986;134(6):587-594.
- El Dib R, Tikkinen KAO, Akl EA, Gomaa HA, Mustafa RA, Agarwal A, et al. Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes. *J Clin Epidemiol*. 2017;84:61-69. <https://doi.org/10.1016/j.jclinepi.2016.12.009>
- Neumann I, Santesso N, Akl EA, Rind DM, Vandvik PO, Alonso-Coello P, et al. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. *J Clin Epidemiol*. 2016;72:45-55. <https://doi.org/10.1016/j.jclinepi.2015.11.017>
- Brignardello-Petersen R, Carrasco-Labra A, Guyatt GH. How to Interpret and Use a Clinical Practice Guideline or Recommendation: Users' Guides to the Medical Literature [published correction appears in *JAMA*. 2022 Feb 22;327(8):784]. *JAMA*. 2021;326(15):1516-1523. <https://doi.org/10.1001/jama.2021.15319>
- Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-725. <https://doi.org/10.1016/j.jclinepi.2012.03.013>