

8. Decidindo até que ponto uma revisão sistemática pode ser considerada confiável

Simon Lewin^{1}, Andrew D Oxman², John N Lavis³ and Atle Fretheim⁴*

Lewin S, Oxman AD, Lavis JN, Fretheim A: SUPPORT Tools for evidence-informed health Policymaking (STP). **8. Deciding how much confidence to place in a systematic review.** Health Research Policy and Systems; 2009, 7(Suppl 1):S8 doi:10.1186/1478-4505-7-S1-S8.

<http://www.health-policy-systems.com/content/pdf/1478-4505-7-S1-s8.pdf>

1 Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Health Systems Research Unit, Medical Research Council of South Africa

2 Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

3 Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada L8N 3Z5

4 Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway

* Autor responsável por comunicações (simon.lewin@nokc.no)

Esta é a tradução de um artigo publicado no Health Research Policy and Systems, 2009; 7:Supplement 1 (www.health-policy-systems.com/supplements/7/S1).

O uso, a distribuição e a reprodução irrestritas por qualquer meio estão permitidas desde que a fonte seja citada. Podem ser encontrados links das traduções desta série para o espanhol, português, francês e chinês no website do SUPPORT (www.support-collaboration.org). Opiniões sobre como melhorar as ferramentas nesta série são bem-vindas e devem ser encaminhadas para: STP@nokc.no.

A série de artigos foi preparada como parte do projeto SUPPORT, apoiado pelo 6º Programa-Quadro INCO da Comissão Europeia, contrato 031939. A Norad (Norwegian Agency for Development Cooperation), a AHPSR (Alliance for Health Policy and Systems Research) e o Milbank Memorial Fund organizaram um encontro de revisão por pares no qual se discutiu uma versão prévia da série. John Lavis recebeu salário como Canada Research Chair in Knowledge Transfer and Exchange (Catedrático de pesquisa no Canadá para a transferência e troca de conhecimento). A Norad, o satélite norueguês do grupo EPOC (Cochrane Effective Practice and Organisation of Care), o Norwegian Knowledge Centre for the Health Services, a AHPSR, a CHSRF (Canadian Health Services Research Foundation), a EVIPNet (Evidence-Informed Policy Network) e a Organização Pan-Americana da Saúde apoiaram a tradução e difusão dos artigos. Nenhum dos financiadores atuou na elaboração, revisão ou aprovação do conteúdo.

Este artigo foi traduzido para o português por Ocean Translations e contou com o apoio da Canadian Health Services Research Foundation (CHSRF) <http://www.chsrf.ca/>; Centro Rosarino de Estudios Perinatales (CREP) www.crep.org.ar; e Organização Pan-Americana da Saúde (OPAS) (www.paho.org/researchportal).



Resumo

Este artigo faz parte de uma série escrita para as pessoas responsáveis pela tomada de decisões relacionadas a políticas e programas de saúde e para aqueles que dão apoio a estes tomadores de decisão.

A confiabilidade das revisões sistemáticas dos efeitos de intervenções de saúde é variável. Consequentemente, os formuladores de políticas e as demais pessoas precisam avaliar até que ponto tais evidências são confiáveis. O uso de processos sistemáticos e transparentes para determinar tais decisões pode ajudar a evitar a inclusão de erros e vieses em tais julgamentos. Neste artigo, sugerimos cinco perguntas que podem ser consideradas ao decidir até que ponto os resultados de uma revisão sistemática dos efeitos de uma intervenção são confiáveis. Estas perguntas são as seguintes: 1. A revisão abordou explicitamente a uma pergunta adequada sobre gestão ou política? 2. Foram usados critérios apropriados ao se considerar os estudos para a revisão? 3. A busca por estudos relevantes foi detalhada e adequadamente abrangente? 4. As avaliações da relevância dos estudos para o tópico da revisão e do seu risco de parcialidade podiam ser reproduzidas? 5. Os resultados foram similares entre os estudos?

SOBRE O STP

Este artigo faz parte de uma série escrita para as pessoas responsáveis pela tomada de decisões relacionadas a políticas e programas de saúde e para aqueles que dão apoio a estes tomadores de decisão. A série se destina a ajudar essas pessoas para assegurar que suas decisões sejam devidamente sustentadas pelas melhores evidências de pesquisa disponíveis. As ferramentas SUPPORT e como elas podem ser usadas estão descritas de maneira detalhada na Introdução desta série [1]. Um glossário para toda a série acompanha cada artigo (ver Arquivo adicional 1). Podem ser encontrados links das traduções desta série para o espanhol, português, francês e chinês no website do SUPPORT (www.support-collaboration.org). Opiniões sobre como melhorar as ferramentas nesta série são bem-vindas e devem ser encaminhadas para: STP@nokc.no.

CENÁRIOS

Cenário 1: Você é um funcionário público de alto escalão e vai apresentar uma proposta ao Ministro sobre as evidências para apoiar diversas opções de programas e políticas, as quais têm por objetivo resolver um problema prioritário de saúde. Você está preocupado em saber até que ponto as revisões sistemáticas das evidências são confiáveis para cada opção e quer ter certeza de que tais revisões foram devidamente avaliadas pela sua equipe.

Cenário 2: Você trabalha no Ministério da Saúde e está preparando um documento sobre as opções para resolver um problema prioritário de saúde. Foram identificadas diversas revisões sistemáticas dos efeitos das opções e solicitaram que você fizesse uma avaliação de até que ponto cada revisão era confiável.

Cenário 3: Você trabalha em uma unidade independente que dá suporte ao Ministério da Saúde no uso de evidências na elaboração de políticas. Você está preparando um documento para o Ministério sobre os prováveis impactos das opções para resolver um problema prioritário de saúde. Você quer orientações sobre a avaliação de até que ponto as revisões sistemáticas dos impactos de cada opção são confiáveis.

HISTÓRICO

Para os tomadores de decisão (Cenário 1), este artigo sugere uma série de perguntas que eles podem pedir que sua equipe considere ao decidir até que ponto os resultados de uma revisão sistemática dos efeitos das intervenções de saúde são confiáveis.

Para aqueles que apóiam os formuladores de políticas (Cenários 2 e 3), este artigo sugere uma série de perguntas que podem ser utilizadas para orientar uma avaliação crítica das revisões sistemáticas dos efeitos.

As revisões sistemáticas de ensaios randomizados controlados (RCT, na sigla em inglês) são amplamente aceitas como fontes das evidências mais confiáveis sobre os efeitos das intervenções de saúde [2,3]. As revisões sistemáticas são caracterizadas pela sua abordagem sistemática e explícita para identificar, selecionar e avaliar pesquisas relevantes, bem como para coletar e analisar dados de estudos incluídos [3]. Cada vez mais, as revisões sistemáticas também estão sendo usadas para identificar, avaliar e combinar evidências sobre as consequências econômicas das intervenções [4], tais como a relação custo-benefício da promoção do aleitamento materno para crianças em unidades neonatais [5] ou os custos de diferentes estratégias de divulgação e de implantação de diretrizes [6]. Eles também são usados para resumir as evidências de estudos qualitativos, como as opiniões de consumidores ou prestadores de serviço sobre as intervenções de saúde [7-10]. Neste artigo, nos concentramos em revisões sistemáticas dos efeitos das políticas ou programas de saúde. Estas medidas incluem revisões de arranjos de prestação de serviço, como os efeitos da substituição de médicos por enfermeiros no atendimento primário de saúde [11] e das estratégias para implantar mudanças, tais como os efeitos das reuniões de educação continuada para profissionais de saúde [12].

A abordagem sistemática e explícita usada em uma revisão sistemática destina-se a reduzir o risco de vieses e erros que ocorrem por acaso e a ajudar a facilitar a avaliação crítica dessas sínteses [13,14]. No entanto, o rigor com que as revisões sistemáticas são realizadas varia. Portanto, nem todas as revisões são igualmente confiáveis – ou seja, as revisões podem diferir no nível de confiança que podemos ter em seus resultados. Portanto, simplesmente se basear no fato de que uma avaliação é chamada de uma “revisão sistemática” (ou uma meta-análise) não é suficiente quando se usa os resultados para fundamentar decisões políticas.

Ao utilizar revisões sistemáticas de efeitos para fundamentar as decisões políticas, os formuladores de políticas e as demais pessoas precisam julgar até que ponto esta evidência pode ser considerada confiável. O uso de processos sistemáticos e transparentes pode ajudar a evitar a inclusão de erros e vieses em seus julgamentos. Um processo sistemático e transparente permite também que outras partes interessadas, incluindo o público, entendam e avaliem esses julgamentos. Isto é particularmente importante nos casos em que tais avaliações influenciem as recomendações ou decisões relativas a intervenções clínicas ou serviços [15] ou decisões de implantar ou suspender programas ou políticas. A Figura 1 descreve os passos envolvidos na descoberta e avaliação de revisões sistemáticas para fundamentar políticas.

A confiança nos resultados de uma revisão sistemática pode ser limitada por uma série de razões, incluindo a falta de:

- Especificação da questão e métodos da revisão antes da realização da revisão, por exemplo, em um protocolo de revisão publicado
- Especificação de critérios claros para estudar a inclusão e a exclusão
- Descrição adequada dos estudos incluídos na revisão
- Avaliação do risco de parcialidade para estudos incluídos na revisão
- Avaliação do risco de parcialidade em uma publicação, ou seja, a possibilidade de que alguns estudos, geralmente aqueles com resultados positivos (“estatisticamente significativos”) estejam mais propensos do que outros a serem publicados e, portanto, a serem incluídos em uma revisão
- Uso de métodos apropriados para combinar os resultados dos estudos incluídos (em uma meta-análise), quando relevantes
- Exame adequado das diferenças nos resultados dos estudos incluídos em uma revisão (ou seja, a “heterogeneidade” dos resultados)
- Fundamentação das conclusões da revisão nos dados incluídos

Outras possíveis limitações de revisões sistemáticas incluem os conflitos de interesse (os quais podem afetar a confiabilidade de uma revisão em qualquer uma das formas listadas acima) e a desatualização das revisões.

As variações na confiabilidade, por exemplo, foram observadas em um estudo comparando a metodologia e os componentes de divulgação das revisões Cochrane com as revisões publicadas em jornais impressos. Este estudo verificou que as revisões Cochrane incluíam componentes que as tornavam menos propensas a viés. Verificou-se que esta redução geral no risco de viés nas revisões Cochrane resultava especificamente das suas descrições claras dos critérios para inclusão e exclusão, e da avaliação formal do risco de viés dos estudos incluídos em cada revisão [16]. De maneira similar, outro estudo comparou a qualidade metodológica e as conclusões nas revisões Cochrane de ensaios de medicamentos com aqueles em revisões apoiadas pela indústria relacionadas aos mesmos medicamentos. Este estudo verificou que as revisões Cochrane tinham uma nota maior em avaliações de qualidade. Isto ocorria porque as revisões Cochrane consideravam a possibilidade de parcialidade com mais frequência do que as revisões que tinham o apoio do setor. Também se verificou que as revisões apoiadas pelo setor estavam muito mais propensas a recomendar os medicamentos em questão, sem ressalvas [17]. Diversos outros estudos sobre revisões também relataram diferenças nas suas qualidades e conclusões [18-21].

Diversas ferramentas foram concebidas para avaliar a qualidade das revisões sistemáticas, incluindo a AMSTAR (A Measurement Tool to Assess Reviews = Uma Ferramenta de Medição para Avaliar Revisões) [22], CASP (Critical Appraisal Skills Programme = Programa de Competências de Avaliação Crítica) [23] e um desenvolvido por Oxman and Guyatt [24] (ver também [25,26]), mas todas contêm critérios similares. (A ferramenta AMSTAR é descrita na Tabela 1.) Várias ferramentas também incluem escalas de classificação para indicar até que ponto uma revisão é confiável. Um

número crescente de revisões agora inclui estas avaliações. Em geral, uma alta classificação sugere que os resultados das revisões são mais confiáveis. Por outro lado, uma baixa classificação indica que os resultados das revisões são menos confiáveis e que as revisões devem ser cuidadosamente examinadas para que suas principais limitações sejam identificadas. No entanto, três pontos devem ser observados: em primeiro lugar, uma classificação ou nota global não indica necessariamente quais aspectos particulares de uma revisão foram realizados de maneira confiável – alguns podem ter sido realizados de forma mais confiável do que outros. Em segundo lugar, o processo de pontuação por si só também envolve a atribuição de pesos a itens diferentes na ferramenta de avaliação. Pode ser difícil justificar porque alguns itens devem receber um peso maior [27]. Finalmente, as ferramentas de classificação só podem avaliar a confiabilidade do que é relatado. Quando as informações essenciais sobre os métodos utilizados na revisão não são informadas, pode não ficar claro o que foi feito ou até que ponto o que foi feito constitui uma limitação importante.

A avaliação de até que ponto os resultados de uma revisão são confiáveis precisa ser diferenciada da compreensão dos resultados da revisão em si. A Tabela 2 fornece orientações sobre o que procurar nos resultados de uma revisão dos efeitos. Na Tabela 3 são apresentadas orientações sobre como avaliar até que ponto são confiáveis os resultados das revisões de estudos qualitativos e de revisões de estudos econômicos.

Uma avaliação da confiabilidade dos resultados de uma revisão também precisa ser diferenciada de qualquer avaliação que poderia ser feita da relevância das revisões para perguntas específicas da política. As considerações relevantes incluem, por exemplo, se uma revisão fornece evidências dos efeitos das diferentes opções de programa ou política em questão, e se os resultados de uma revisão são aplicáveis ao contexto no qual a política será implantada. O processo de avaliar a aplicabilidade dos resultados das revisões sistemáticas é discutido mais detalhadamente no artigo 9 desta série [28].

Neste artigo, sugerimos cinco perguntas que podem ser consideradas ao decidir até que ponto os resultados de uma revisão sistemática dos efeitos de opções são confiáveis.

QUESTÕES A SEREM CONSIDERADAS

As seguintes perguntas podem orientar os formuladores de políticas no momento de decidir até que ponto os resultados de uma revisão sistemática dos efeitos de uma opção são confiáveis:

1. A revisão abordou explicitamente a uma pergunta adequada sobre gestão ou política?
2. Foram usados critérios apropriados ao se considerar os estudos para a revisão?
3. A busca por estudos relevantes foi detalhada e adequadamente abrangente?
4. As avaliações da relevância dos estudos para o tópico da revisão e do seu risco de parcialidade podiam ser reproduzidas?

5. Os resultados foram similares entre os estudos?

1. A revisão abordou explicitamente a uma pergunta adequada sobre gestão ou política?

Um primeiro passo fundamental para avaliar a confiabilidade das conclusões de uma revisão sistemática é examinar a *pergunta* que está sendo respondida. A concepção técnica e a realização de uma revisão podem ser excelentes, mas os resultados desta revisão podem não ser úteis na tomada de decisão se não tiverem sido explicitamente direcionados para uma pergunta de gestão ou de política que seja sensata, apropriada e relevante para o problema que o formulador de política está considerando.

Uma pergunta apropriada de gestão ou de política:

- *Será explícita*: em outras palavras, ela será expressa de maneira detalhada e não de forma implícita no material apresentado. Se a pergunta da revisão não tiver sido expressa de forma explícita ou formulada com clareza, será difícil avaliar a realização da revisão de forma adequada. Isto ocorre porque a realização da revisão terá de ser considerada, pelo menos em parte, com relação à própria pergunta [29]. Por exemplo, é preciso avaliar, com relação à pergunta da revisão que os estudos se destinavam a responder, se os critérios utilizados para selecionar tais estudos para uma revisão foram adequados. Uma pergunta clara também ajuda os leitores a avaliar se uma revisão é relevante para o seu trabalho [29]
- *Será estabelecida a priori*: em outras palavras, antes de a revisão ser realizada. É importante que a pergunta da revisão seja especificada antes de uma revisão ser realizada, de preferência em um protocolo de revisão ou plano. Todas as revisões Cochrane, por exemplo, são precedidas por um protocolo de revisões publicadas e exemplos de tais revisões podem ser encontradas na Biblioteca Cochrane (<http://www3.interscience.wiley.com/cgi-bin/mrwhome/1065687/HOME>). Se a pergunta da revisão não for especificada antes da revisão ser realizada, existirá o risco da pergunta ser alterada para se adequar às evidências encontradas, minando assim a confiança nos resultados
- *Responderá a uma questão de relevância para a gestão ou formulação de políticas*. Esta pergunta terá de ser avaliada em um contexto específico, com base no conjunto de questões que são importantes em uma jurisdição específica em um determinado momento. A pergunta da revisão pode não ser relevante caso:
 - Seja muito restrita: por exemplo, uma revisão pode considerar os efeitos de um programa em uma faixa etária específica de participantes apenas, localizados em um contexto específico, ou para um leque restrito de resultados. Não seria possível, neste caso, generalizar os resultados para outras populações, contextos ou resultados
 - Seja muito ampla: uma revisão, por exemplo, pode definir um programa como sendo um tipo de programa que inclui um leque muito amplo de práticas e é possível que nem todas elas sejam relevantes para um determinado local. Ou

uma revisão pode fazer uma pergunta muito ampla que não é útil do ponto de vista da tomada de decisões. Uma pergunta, por exemplo, se as enfermeiras podem efetivamente implantar programas de promoção da saúde, não será útil para decidir se um grupo especial de enfermeiros, tais como enfermeiros registrados, pode efetivamente implantar um programa de promoção de saúde para um problema de saúde específico, como a prevenção do HIV/AIDS

- Não especifique um grupo de comparação apropriado: se, por exemplo, um programa for comparado com um cenário “sem programas” em vez de ser comparado ao melhor tratamento atual para uma determinada condição

Uma pergunta de revisão bem formulada deve especificar *todos* os seguintes itens: os tipos de população e os contextos que a revisão vai abranger (por exemplo, crianças com idades entre um mês e seis anos que vivem em uma área endêmica de malária); os tipos de programas e comparações considerados (por exemplo, medicamentos contra a malária administrados em intervalos regulares (a intervenção) em comparação com placebo ou nenhum medicamento (a comparação)); e os tipos de resultados que são de interesse (por exemplo, malária clínica e anemia grave) [30,31]. A sigla PICO (Population, Intervention, Comparison, Outcomes = População, Intervenção, Comparação, Resultados) é por vezes utilizada para resumir estes quatro componentes fundamentais para uma pergunta de revisão.

Embora possa parecer óbvia a necessidade de se ter uma pergunta de revisão bem formulada, muitas revisões narrativas não têm uma. Uma revisão de uma amostra de tais revisões publicadas nas principais revistas médicas revelou que 20% não indicavam claramente a sua finalidade [32].

2. Foram usados critérios apropriados ao se considerar os estudos para a revisão?

Critérios de inclusão e exclusão de uma revisão são as listas detalhadas dos tipos de população, as intervenções, as comparações e os resultados que uma revisão irá considerar. Estes critérios, especificados em um protocolo de revisão, irão determinar quais estudos foram incluídos em uma revisão. Eles irão, portanto, influenciar fortemente os resultados de uma revisão. É importante que estes critérios sejam adequados em relação à pergunta da revisão.

As seguintes perguntas devem ser examinadas ao se analisar se os critérios utilizados para considerar os estudos de revisão são adequados:

- A revisão especifica critérios claros de inclusão e de exclusão? Estes critérios são importantes como uma proteção contra a parcialidade relacionada à inclusão de estudos na revisão. Uma avaliação recente da qualidade metodológica de revisões sistemáticas em cirurgia geral, por exemplo, constatou que apenas 70% de tais revisões relataram os critérios utilizados para decidir quais estudos iriam ser incluídos em uma revisão [18]

- Os critérios de inclusão e exclusão são explícitos com relação ao seguinte: os tipos de população considerada, os tipos de intervenções e as comparações consideradas, e os tipos de resultados considerados?
- Os critérios de inclusão e de exclusão estão em linha com a pergunta de revisão? [33] Por exemplo, se uma revisão visa avaliar profilaxia e tratamento intermitente com medicamentos contra a malária para prevenir a malária em crianças que vivem em áreas endêmicas de malária, os critérios indicam a inclusão de estudos de crianças a partir de contextos adequados e eles especificam as formas de profilaxia e tratamento que serão consideradas? [31] Do mesmo modo, se o objetivo de uma revisão é examinar os efeitos das intervenções para aumentar a proporção de profissionais de saúde que trabalham na zona rural e em outras áreas carentes, os critérios indicam a gama de profissionais de saúde que serão incluídos e os tipos de intervenções educacionais ou financeiras que serão levadas em consideração? [34]

3. A busca por estudos relevantes foi detalhada e adequadamente abrangente?

Um aspecto fundamental de uma revisão sistemática é uma busca cuidadosa e passível de reprodução da literatura para estudos que atendam os critérios de qualificação de uma revisão. Esta abordagem é um dos elementos que diferenciam as revisões *sistemáticas* das revisões *narrativas*. A busca sistemática contribui para minimizar a parcialidade em uma revisão ao garantir que todas as evidências relevantes sejam levadas em consideração. Portanto, ela ajuda a alcançar as estimativas confiáveis dos efeitos da política ou do programa a ser analisado [35].

Viés de publicação – isto é, a publicação seletiva de estudos com base na direção e força dos seus resultados [36] – é um caminho através do qual o viés pode ser introduzido nas revisões. Uma revisão recente examinou até onde a publicação de ensaios aleatórios é influenciada pela descoberta ou não de resultados positivos e pela percepção de importância das constatações do ensaio. Foi demonstrado que havia uma probabilidade significativamente maior que ensaios com resultados positivos fossem publicados do que ensaios que apresentavam resultados negativos [37]. Esta revisão e outras pesquisas também demonstraram que os ensaios que apresentavam resultados positivos são publicados mais cedo do que outros [38]. O resultado disto é que as revisões podem superestimar os efeitos positivos dos programas, a menos que se tente identificar estudos que foram publicados e os que não foram.

As revisões sistemáticas variam em termos de inclusão de buscas abrangentes. Uma revisão dos relatórios de revisões publicadas sobre o tratamento da asma, por exemplo, constatou que apenas 52% das 33 revisões examinadas incluíam uma busca razoavelmente abrangente por evidências dos efeitos [20]. Por isso, é importante verificar como as buscas por estudos relevantes foram feitas.

As seguintes perguntas devem ser examinadas ao se analisar se a busca por estudos relevantes foi detalhada e razoavelmente abrangente [22]:

- Uma revisão descreve detalhadamente a estratégia usada para procurar por estudos relevantes? Esta descrição deve incluir: 1. A lista de fontes pesquisadas, 2. As palavras-chave usadas para fazer a busca em tais fontes (quando aplicável), e 3. O período em anos durante o qual as fontes foram pesquisadas. A Tabela 4 fornece exemplos da gama de fontes pesquisadas nas revisões publicadas na Cochrane Library
- A estratégia de busca incluiu bancos de dados eletrônicos de estudos publicados? Uma vasta gama de bases de dados eletrônicas de estudos publicados está disponível e pode ser pesquisada de graça ou por um custo bastante baixo. As principais bases de dados incluem a PubMed/MEDLINE (compilada pela Biblioteca Nacional de Medicina [National Library of Medicine], EUA), a CENTRAL (Cochrane Central Register of Controlled Trials – Cadastro Central da Cochrane para Ensaio Controlados – compilado pela Cochrane Collaboration), e bases de dados regionais, como a LILACS (Ciências de Saúde do Caribe e da América Latina). Os artigos 4 [39] e 5 [40] nesta série apresentam mais informações sobre como encontrar literaturas relevantes de pesquisa
- As buscas em bases eletrônicas de dados foram complementadas por buscas adicionais? Estas buscas adicionais podem incluir um exame das listas de referências de estudos relevantes, um contato com autores e especialistas na área, bem como a consulta a cadastros especializados de estudos relacionados à área que é tema da revisão. Esta busca adicional é útil como uma forma de ajudar a identificar mais estudos publicados e estudos não publicados (que podem incluir estudos disponíveis na literatura “cinzenta”, ou seja, nas fontes da literatura que não sejam revistas indexadas, examinadas por pares)
- As buscas estão atualizadas? A revisão especifica o período abrangido pelas buscas e as buscas estão atualizadas? Uma revisão publicada, embora seja relevante para uma pergunta de política, pode ter usado buscas que foram feitas há vários anos. Portanto, é possível que a revisão não inclua todas as últimas evidências relevantes e possa, portanto, apresentar uma estimativa não confiável dos efeitos da opção do programa ou política

4. As avaliações da relevância dos estudos para o tópico da revisão e do seu risco de parcialidade podiam ser reproduzidas?

Os autores das revisões sistemáticas precisam fazer dois julgamentos importantes sobre cada um dos estudos primários que poderão ser incluídos em uma revisão. Em primeiro lugar, o estudo atende aos critérios para inclusão em sua revisão, em outras palavras, é relevante para o tópico de revisão? Em segundo lugar, qual é o risco de haver viés nos resultados do estudo? O risco de viés refere-se ao risco de “um erro sistemático, ou desvio da verdade, nos resultados ou nas inferências” [27]. Ele também diz respeito à questão de saber se é possível presumir que os resultados de um estudo são exatos [27]. Visto que julgamentos irão afetar os resultados de uma revisão, é

importante que eles sejam apresentados de uma forma que seja transparente e que possa ser reproduzida. Outras pessoas precisam ser capazes de compreender como esses julgamentos foram feitos e ser capazes de repetir estas avaliações.

Conforme mencionado acima, as revisões precisam especificar critérios claros de inclusão e de exclusão para evitar a parcialidade no processo de seleção de estudos a serem incluídos. Estes critérios e julgamentos irão necessariamente afetar os resultados da revisão ao influenciarem os estudos selecionados para a inclusão. O viés ou os erros nesses julgamentos podem ser minimizados das seguintes maneiras: primeiro, dois revisores devem decidir, de forma independente, quais estudos devem ser incluídos em uma revisão. Debates adicionais com outros revisores também podem ser usados para resolver desacordos relacionados à inclusão de um determinado estudo. Em segundo lugar, as razões para a inclusão de um estudo (e para a exclusão de um estudo que parece relevante) devem ser registradas na revisão publicada. Isto permitirá que os leitores façam seus próprios julgamentos com relação às decisões de qualificação para inclusão ou exclusão. Isto também oferece uma “trilha de auditoria” transparente para a revisão, garantindo que o processo possa ser reproduzido.

A capacidade de uma revisão sistemática de chegar a conclusões sobre os efeitos de uma política ou programa depende também da validade dos dados obtidos a partir de cada estudo incluído. O agrupamento dos resultados dos estudos, ou a criação de um resumo de tais resultados em uma revisão, pode levar a um resultado ilusório, caso a validade dos estudos individuais incluídos na revisão seja baixa. A avaliação do risco de viés nos resultados dos estudos incluídos é, portanto, um elemento importante de uma revisão sistemática. Estas avaliações devem ser levadas em consideração na interpretação e conclusões de uma revisão [27].

Diversas abordagens diferentes para avaliar a qualidade ou o risco de viés foram desenvolvidas para os ensaios aleatórios [27, 41, 42]. Embora não discutamos estas diferentes abordagens aqui, é importante notar que as revisões devem ser explícitas com relação às abordagens utilizadas e devem aplicá-las de forma consistente.

Ao avaliar a relevância dos estudos incluídos para o tópico da revisão e o possível risco de viés, as seguintes perguntas devem ser consideradas:

- Foi usada uma abordagem explícita e transparente para avaliar a relevância dos estudos para o tópico da revisão? Uma revisão deverá descrever como a relevância foi avaliada e fornecer uma lista de estudos incluídos e excluídos
- Foi usada uma abordagem explícita e transparente para avaliar o risco de viés/parcialidade nos estudos incluídos? Uma revisão deve informar a ferramenta utilizada para avaliar o risco de viés, como a avaliação foi realizada e os resultados da avaliação
- Os resultados da avaliação de risco de viés foram levados em consideração na interpretação dos resultados de uma revisão? Quando há um alto risco de viés nos

estudos incluídos, por exemplo, poderemos ter menos confiança nos resultados de uma revisão

5. Os resultados foram similares entre os estudos?

Os resultados dos estudos incluídos em uma revisão podem ser muito similares – ou podem variar – em termos dos efeitos do programa sobre um determinado resultado. Esta variabilidade entre os estudos incluídos na revisão é normalmente chamada de “heterogeneidade” [27]. A variabilidade entre os estudos incluídos em uma revisão depende em parte do escopo da revisão. Quando o escopo é amplo, também é possível esperar que o intervalo e, portanto, a variabilidade dos estudos incluídos, também sejam grandes. Em contrapartida, se o escopo de uma revisão for restrito, os estudos incluídos tenderão a ser mais semelhantes entre si.

Se os participantes, as intervenções ou os resultados dos estudos incluídos na revisão forem muito diferentes, isso pode causar uma variação ou heterogeneidade caso o efeito de intervenção seja afetado por estes fatores. Visto que o verdadeiro efeito da intervenção será diferente em todos estes estudos, nestes casos, o efeito médio em todos os estudos não será útil.

Dependendo do nível de variabilidade, as revisões podem usar diferentes abordagens para resumir informações dos estudos incluídos, por exemplo:

- *Calculando o efeito médio (ou agrupado) de todos os estudos:* esta abordagem é útil quando a variabilidade entre os estudos é baixa. Por exemplo, uma revisão sistemática de programas de “alta hospitalar precoce combinada com atendimento hospitalar em casa” (ou seja, programas nos quais o tratamento ativo é prestado por profissionais de saúde na casa de um paciente para um problema de saúde que iria exigir internação hospitalar com cuidados intensos) constatou que os estudos incluídos eram suficientemente semelhantes para poder estimar o efeito médio do programa. A revisão encontrou evidências insuficientes de benefícios econômicos ou de saúde resultantes dos programas “alta precoce, atendimento hospitalar em casa” [43]
- *Calculando o efeito médio para subgrupos de estudos incluídos em uma revisão:* este cálculo pode ser útil quando a variabilidade geral dos estudos incluídos na revisão é alta (e é, portanto, inútil para calcular um efeito médio), exceto onde a variabilidade for baixa entre os subgrupos de estudos. Por exemplo, uma revisão de intervenções de trabalhadores de saúde leigos em atendimentos de saúde primário e comunitário agrupou estudos de acordo com os problemas de saúde tratados pelos trabalhadores de saúde leigos. Para alguns grupos, como os trabalhadores da saúde leigos, para promover a vacinação e o aleitamento materno, foi possível calcular um efeito médio em todos os estudos relevantes. A revisão encontrou evidências de que trabalhadores de saúde leigos podem melhorar a vacinação e o aleitamento materno [44]

- *Descrevendo a variedade de tamanhos dos efeitos:* quando os estudos não são suficientemente semelhantes para permitir o cálculo de um efeito médio que seja útil, pode ainda ser possível descrever a gama de efeitos encontrados nos estudos. Por exemplo, uma análise dos efeitos de auditorias e feedbacks sobre a prática dos prestadores de serviços de saúde mostraram que a conformidade com a prática desejada variou de uma redução de 16% para um aumento de 70%, com uma mediana de 5%. A revisão indicou que a auditoria e o feedback podem tornar a prática mais eficaz, mas que os efeitos são geralmente de pequenos a moderados [45]
- *Catálogo dos tipos de intervenções para solucionar um problema específico:* o amplo escopo de algumas revisões e, portanto, a variabilidade dos estudos dentro de tais revisões indicam que não é sensato tentar combinar quantitativamente os resultados dos estudos incluídos – ou mesmo descrever a variedade dos tamanhos dos efeitos. Nestes casos, pode ser feita uma revisão narrativa. Por exemplo, uma revisão sistemática da eficácia das intervenções dos serviços de saúde destinadas a reduzir as desigualdades na saúde incluiu estudos que avaliaram os programas destinados a reduzir as desigualdades na saúde e que poderiam ser implantados no sistema de saúde de maneira isolada ou em colaboração com outras agências. A variedade de estudos incluídos foi ampla, estendendo-se desde programas para melhorar o controle da pressão arterial até de intervenções de promoção da saúde. Portanto, não se tentou fazer qualquer agrupamento estatístico [46]

Quando os resultados diferem de um estudo a outro, as seguintes perguntas devem ser consideradas:

- *Existe uma explicação convincente para as diferenças que foram encontradas?* Tal explicação pode incluir diferenças nos participantes, nas intervenções, nos grupos de comparação, nos resultados, nos contextos ou nos períodos de tempo entre todos os estudos incluídos. Por exemplo, alguns estudos podem ter incluído participantes que tinham uma faixa etária mais ampla ou diferentes condições de saúde pré-existent
- *Se uma estimativa combinada foi feita, há a probabilidade de que seja significativa?*
- Se os estudos incluídos em uma revisão são variados, uma estimativa combinada pode não ser significativa
- Explorações complementares dos dados, mediante análise de subgrupos, podem ser realizadas, mas os resultados destas análises exploratórias podem não ser confiáveis

À medida que o número de revisões sistemáticas disponíveis aumenta, fica cada vez mais comum encontrar mais de uma revisão sistemática para uma pergunta específica de política. Às vezes, os resultados ou as conclusões destas revisões podem ser diferentes. A Tabela 5 dá orientações sobre como os formuladores de políticas poderão abordar tais situações.

CONCLUSÃO

Existem variações evidentes no rigor com que as revisões sistemáticas dos efeitos são conduzidas. Por isso, é importante avaliar a confiabilidade das revisões usadas para fundamentar decisões relacionadas a políticas para que seja possível avaliar até que ponto esta evidência é confiável. Deverá ser usada uma abordagem sistemática e transparente para tais avaliações e foram desenvolvidas diversas ferramentas com esta finalidade. No entanto, estas ferramentas somente podem ser usadas para avaliar o que é relatado. É por este motivo que qualquer avaliação que for feita usando essas ferramentas precisa ser bem planejada e ser realizada com cuidado.

Quando a confiabilidade de uma revisão sistemática for fraca, os formuladores de política deverão ter menos confiança nos resultados e deverão ser cautelosos, caso usem tais resultados para fundamentar as decisões políticas (conforme resumido na Figura 2). Ao tomar decisões fundamentadas pelas evidências apresentadas em uma revisão, os formuladores de políticas precisam considerar as avaliações da confiabilidade de uma revisão juntamente com outras informações, tais como a utilidade da revisão no que diz respeito à pergunta da política e às evidências dentro do contexto local.

RECURSOS

Documentos úteis e leituras adicionais

Higgins JPT, Altman DF: **Chapter 8: Assessing risk of bias in included studies.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008. Disponível em: www.cochrane-handbook.org

Counsell C: **Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews.** *Ann Intern Med* 1997, **127**: 380-387

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C *et al.*: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.** *BMC Med Res Methodol* 2007, **7**: 10. Disponível em: <http://www.biomedcentral.com/1471-2288/7/10>

Links para websites

A Rx for Change database: www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions – resume evidências de pesquisas atuais sobre os efeitos das estratégias para melhorar a prática de prescrição de medicamentos e o uso de medicamentos. Esta base de dados inclui resumos, incluindo avaliações de

confiabilidade, de revisões sistemáticas que avaliam os efeitos das estratégias voltadas para profissionais, para a organização de saúde e para os consumidores

O Grupo de Revisão EPOC (Cochrane Effective Practice and Organisation of Care, prática efetiva e organização do atendimento): www.epoc.cochrane.org/en/index.html – o Grupo de Revisão dá orientações sobre a avaliação da confiabilidade dos diversos tipos de estudos de eficácia

A Colaboração SUPPORT (SUPporting Policy relevant Reviews and Trials, suporte a ensaios e revisões relevantes para as políticas): <http://www.support-collaboration.org/index.htm> – este projeto produz resumos das revisões de alta prioridade para países de baixa e média renda per capita. Estas revisões incluem avaliações de confiabilidade

CONFLITO DE INTERESSES

Os autores declaram que não têm interesses conflitantes.

CONTRIBUIÇÕES DOS AUTORES

SL preparou o primeiro esboço deste artigo. ADO, JNL e AF contribuíram no esboço e na revisão do artigo.

AGRADECIMENTOS

Veja na Introdução desta série os agradecimentos àqueles que financiaram e aos que contribuíram com este trabalho. Além disto, gostaríamos de agradecer a Duff Montgomerie pelos comentários úteis em uma versão prévia deste artigo.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Lavis JN, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction.** Health Res Policy Syst. 2009 7(Suppl 1):I1.
2. Lavis JN, Posada FB, Haines A, Osei E: **Use of research to inform public policymaking.** *Lancet* 2004, 364:1615-21.
3. Oxman AD, Lavis JN, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 1. What is evidence-informed policymaking.** Health Res Policy Syst. 2009 7(Suppl 1):S1
4. Oxman AD, Fretheim A, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs.** Health Res Policy Syst. 2009 7(Suppl 1):S12

5. Renfrew MJ, Craig D, Dyson L, McCormick F, Rice S, King SE, Misso K, Stenhouse E, Williams AF: **Breastfeeding promotion for infants in neonatal units: a systematic review and economic analysis.** *Health Technol Assess* 2009, 13:1-iv.
6. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, *et al*: **Effectiveness and efficiency of guideline dissemination and implementation strategies.** *Health Technol Assess* 2004, 8:iii-72.
7. Carlsen B, Glenton C, Pope C: **Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines.** *Br J Gen Pract* 2007, 57:971-8.
8. Mays N, Pope C, Popay J: **Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field.** *J Health Serv Res Policy* 2005, 10 Suppl 1:6-20.
9. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J: **Patient adherence to tuberculosis treatment: a systematic review of qualitative research.** *PLoS Med* 2007, 4:e238.
10. Pound P, Britten N, Morgan M, Yardley L, Pope C, Daker-White G, Campbell R.: **Resisting medicines: a synthesis of qualitative studies of medicine taking.** *Soc Sci Med* 2005, 61:133-55.
11. Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B: **Substitution of doctors by nurses in primary care.** *Cochrane Database Syst Rev* 2005, 2:CD001271.
12. Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, Davis D, Odgaard-Jensen J, Oxman AD: **Continuing education meetings and workshops: effects on professional practice and health care outcomes.** *Cochrane Database Syst Rev* 2009, 2:CD003030.
13. Mulrow CD: **Rationale for systematic reviews.** *BMJ* 1994, 309:597-9.
14. Oxman AD, Schunemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006, 4:20.
15. Oxman AD, Glasziou P, Williams JW, Jr.: **What should clinicians do when faced with conflicting recommendations?** *BMJ* 2008, 337:a2530.
16. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D: **Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals.** *JAMA* 1998, 280:278-80.
17. Jorgensen AW, Hilden J, Gotzsche PC: **Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review.** *BMJ* 2006, 333:782.
18. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C: **Evaluating meta-analyses in the general surgical literature: a critical appraisal.** *Ann Surg* 2005, 241:450-9.
19. Jadad AR, Cook DJ, Browman GP: **A guide to interpreting discordant systematic reviews.** *CMAJ* 1997, 156:1411-6.
20. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R: **Systematic reviews and meta-analyses on treatment of asthma: critical evaluation.** *BMJ* 2000, 320:537-40.
21. Linde K, Willich SN: **How objective are systematic reviews? Differences between reviews on complementary medicine.** *J R Soc Med* 2003, 96:17-22.
22. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM: **Development of AMSTAR: a measurement**

- tool to assess the methodological quality of systematic reviews.** *BMC Med Res Methodol* 2007, 7:10.
23. Critical Appraisal Skills Programme: *10 questions to help you make sense of reviews*. United Kingdom, Public Health Resource. 2006.
 24. Oxman AD, Guyatt GH: **Validation of an index of the quality of review articles.** *J Clin Epidemiol* 1991, 44:1271-8.
 25. Canadian Coordinating Office for Health Technology Assessment: *Proposed Evaluation Tools for COMPUS*. Ottawa, Ottawa: Canadian Coordinating Office for Health Technology Assessment. 2005.
 26. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L: *Systems to rate the strength of scientific evidence [Evidence report/technology assessment no 47]*. Publication No. 02-E016. Rockville, MD, USA, Agency for Healthcare Research and Quality. 2002.
 27. Higgins JPT, Altman DF: **Chapter 8: Assessing risk of bias in included studies.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
 28. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 9. Assessing the applicability of the findings of a systematic review.** *Health Res Policy Syst.* 2009 7(Suppl 1):S9.
 29. Counsell C: **Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews.** *Ann Intern Med* 1997, 127:380-7.
 30. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]*. The Cochrane Collaboration; 2008.
 31. Meremikwu MM, Donegan S, Esu E: **Chemoprophylaxis and intermittent treatment for preventing malaria in children.** *Cochrane Database of Systematic Reviews* 2008, 2:CD003756.
 32. Mulrow CD: **The medical review article: state of the science.** *Ann Intern Med* 1987, 106:485-8.
 33. Oxman AD: **Checklists for review articles.** *BMJ* 1994, 309:648-51.
 34. Grobler LA, Marais BJ, Mabunda S, Marindi P, Reuter H, Volmink J: **Interventions for increasing the proportion of health professionals practising in underserved communities.** *Cochrane Database of Systematic Reviews* 2009, 1:CD005314.
 35. Lefebvre C, Manheimer E, Glanville J, on behalf of the Cochrane Information Retrieval Methods Group: **Searching for studies.** In *Cochrane Handbook for systematic reviews of interventions. Version 5.0.1 [updated September 2008]*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
 36. Dickersin K, Min YI: **Publication bias: the problem that won't go away.** *Ann N Y Acad Sci* 1993, 703:135-46.
 37. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: **Publication bias in clinical trials due to statistical significance or direction of trial results.** *Cochrane Database Syst Rev* 2009, 1:MR000006.
 38. Hopewell S, Clarke M, Stewart L, Tierney J: **Time to publication for results of clinical trials.** *Cochrane Database Syst Rev* 2007, 2:MR000011.
 39. Lavis JN, Wilson M, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 4. Using research evidence to clarify a problem.** *Health Res Policy Syst.* 2009 7(Suppl 1):S4.
 40. Lavis JN, Wilson MG, Oxman AD, Grimshaw J, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 5. Using**

- research evidence to frame options to address a problem.** *Health Res Policy Syst.* 2009 7(Suppl 1):S5.
41. Katrak P, Bialocerkowski A, Massy-Westropp N, Kumar VS, Grimmer K: **A systematic review of the content of critical appraisal tools.** *BMC Medical Research Methodology* 2004, 4:22.
 42. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S: **Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists.** *Control Clin Trials* 1995, 16:62-73.
 43. Shepperd S, Doll H, Broad J, Gladman J, Iliffe S, Langhorne P, Richards S, Martin F, Harris R: **Early discharge hospital at home.** *Cochrane Database Syst Rev* 2009, 1:CD000356.
 44. Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, Bosch-Capblanch X, Patrick M: **Lay health workers in primary and community health care.** *Cochrane Database Syst Rev* 2005, 1:CD004015.
 45. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD: **Audit and feedback: effects on professional practice and health care outcomes.** *Cochrane Database Syst Rev* 2006, 2:CD000259.
 46. Arblaster L, Lambert M, Entwistle V, Forster M, Fullerton D, Sheldon T, Watt I: **A systematic review of the effectiveness of health service interventions aimed at reducing inequalities in health.** *J Health Serv Res Policy* 1996, 1:93-103.
 47. Moher D, Jadad AR, Klassen TP: **Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions?** *Arch Pediatr Adolesc Med* 1998, 152:915-20.
 48. Oxman AD, Cook DJ, Guyatt GH: **Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group.** *JAMA* 1994, 272:1367-71.
 49. Oxman AD, Lavis JN, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 10. Taking equity into consideration when assessing the findings of a systematic review.** *Health Res Policy Syst.* 2009 7(Suppl 1):S10.
 50. Oxman AD, Lavis JN, Fretheim A, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence.** *Health Res Policy Syst.* 2009 7(Suppl 1):S17.
 51. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A: **Synthesising qualitative and quantitative evidence: a review of possible methods.** *J Health Serv Res Policy* 2005, 10:45-53.
 52. Noyes J, Popay J, Pearson A, Hannes K, Booth A: **Chapter 20: Qualitative research and Cochrane reviews.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
 53. Lavis JN. **Supporting the Use of Systematic Reviews in Policymaking.** *PLoS Med.* In Press.
 54. Jefferson T, Demicheli V, Vale L: **Quality of systematic reviews of economic evaluations in health care.** *JAMA* 2002, 287:2809-12.
 55. Siegfried N, Muller M, Deeks JJ, Volmink J: **Male circumcision for prevention of heterosexual acquisition of HIV in men.** *Cochrane Database Syst Rev* 2009, 2:CD003362.
 56. McGuinness B, Craig D, Bullock R, Passmore P: **Statins for the prevention of dementia.** *Cochrane Database Syst Rev* 2009, 2:CD003160.

FIGURA 1. COMO ENCONTRAR E AVALIAR AS REVISÕES SISTEMÁTICAS PARA INFORMAR DECISÕES SOBRE OPÇÕES DE PROGRAMA E POLÍTICAS

<i>Perguntas</i>	<i>Razões por confiar menos em uma revisão</i>
A revisão abordou de forma explícita uma política adequada ou questão de gerenciamento?	<ul style="list-style-type: none"> - Revisões que abordam questões pouco claras.....tendem a oferecer respostas vagas - Uma revisão que aborda uma questão distinta da política ou da questão de gerenciamento apresentada.....pode ser enganosa
Foram usados critérios adequados quando se consideraram estudos para a revisão?	<ul style="list-style-type: none"> - Critérios implícitos.....podem resultar em seleção tendenciosa dos estudos para inclusão - Critérios incongruentes com a questão abordada pela revisão.....podem causar a inclusão de estudos inapropriados para responder a questão
A pesquisa de estudos relevantes foi detalhada e razoavelmente abrangente?	<ul style="list-style-type: none"> - Pesquisas mal elaboradas (e descritas).....correm o risco de não obter evidências relevantes significativas - Pesquisas pouco abrangentes.....correm o risco de serem tendenciosas (por exemplo, poderão encontrar estudos com 'resultados positivos') e não poderem encontrar evidências relevantes significativas
As avaliações da relevância dos estudos para o tópico da revisão e seus riscos de má interpretação foram reproduzíveis?	- Se critérios explícitos não forem usados ou se apenas um indivíduo realizar todas as avaliações necessárias em uma revisão..... além de poderem ser tendenciosas, tais avaliações podem não ser reproduzíveis
Os resultados obtidos em diferentes estudos foram semelhantes entre os estudos?	Se houver diferenças significativas nos resultados obtidos em estudos distintos e não houver uma explicação convincente para tais diferenças..... é difícil saber quais são os resultados reais ou quais expectativas criar no caso de implementação de uma política semelhante àquelas que foram avaliadas

FIGURA 2. COMO AS REVISÕES PODEM SER NÃO CONFIÁVEIS E ENGANOSAS

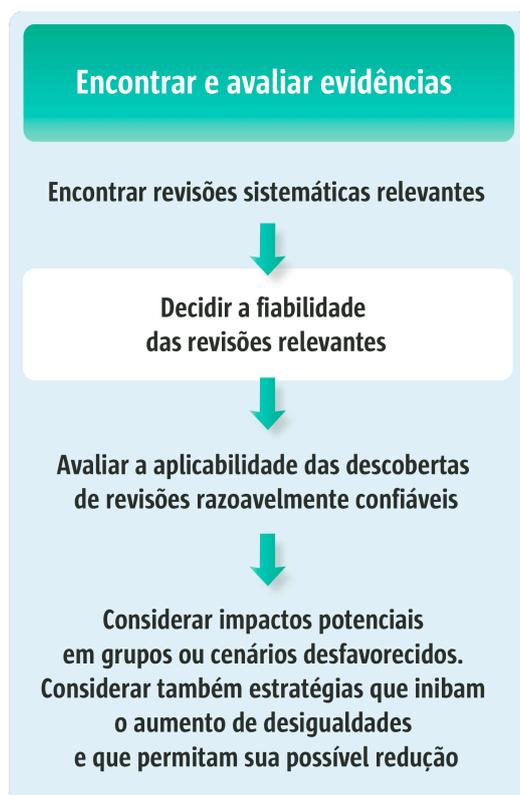


TABELA 1. AMSTAR (A MEASUREMENT TOOL TO ASSESS REVIEWS ,UMA FERRAMENTA DE MEDIÇÃO PARA AVALIAR REVISÕES) (DE [22])

Para cada questão na ferramenta abaixo, os autores de revisões precisam selecionar “Sim/Não/Não sei dizer/Não aplicável”

1. Foi fornecido um projeto “a priori”?

Os critérios de inclusão e a pergunta da pesquisa devem ser estabelecidos antes da realização da revisão

2. Houve uma duplicação na seleção de estudos e extração de dados?

Deve haver pelo menos dois extratores de dados independentes e deve estar estabelecido um procedimento de consenso para discordâncias

3. Foi realizada uma pesquisa/busca bibliográfica abrangente?

Pelo menos duas fontes eletrônicas devem ser pesquisadas. O relatório deve incluir os anos e as bases de dados usadas (por exemplo, Central, EMBASE e MEDLINE). As palavras-chave e/ou os termos MESH devem ser informados e, quando possível, a estratégia de busca deve ser fornecida. Todas as buscas devem ser complementadas por meio de consulta a conteúdos, revisões, livros-texto, cadastros especializados atualizados ou especialistas no campo de estudo específico e por meio de revisão das referências dos estudos encontrados

4. O status da publicação (ou seja, literatura cinzenta) foi utilizado como um critério de inclusão?

Os autores devem declarar que procuraram por relatórios independentemente do seu tipo de publicação. Os autores devem declarar se excluíram ou não quaisquer relatórios (da revisão sistemática), com base no seu status de publicação, idioma etc.

5. Foi fornecida uma lista de estudos (incluídos e excluídos)?

Deve ser fornecida uma lista de estudos incluídos e excluídos

6. As características dos estudos incluídos foram fornecidas?

Em uma forma agregada, como uma tabela, devem ser fornecidos dados dos estudos originais sobre os participantes, as intervenções e os resultados. Devem ser relatadas as diversas características em todos os estudos analisados, como idade, raça, sexo, dados sócio-econômicos relevantes, o estado da doença, duração, gravidade ou outras doenças

7. A qualidade científica dos estudos incluídos foi avaliada e documentada?

Devem ser fornecidos métodos “a priori” de avaliação (por exemplo, para estudos de eficácia, caso o[s] autor[es] opte[m] por incluir apenas estudos controlados de placebo, duplo-cego e aleatórios, ou ocultação de alocação como critérios de inclusão). Para outros tipos de estudos, é importante que existam itens alternativos

8. A qualidade científica dos estudos incluídos foi utilizada de forma adequada na formulação das conclusões?

O rigor metodológico e a qualidade científica dos estudos deverão ser considerados na análise e conclusões da revisão e explicitamente informados na formulação de recomendações

9. Os métodos foram usados para combinar os resultados de estudos adequados?

Para os resultados combinados deve ser feito um teste para garantir que os estudos podiam ser combinados e para avaliar a sua homogeneidade (ou seja, qui ao quadrado para homogeneidade, I^2). Se houver heterogeneidade, deverá ser usado um modelo de efeitos aleatórios e/ou a adequação clínica da combinação deverá também ser levada em consideração (isto é, combinar os resultados era apropriado?)

10. A probabilidade de viés de publicação foi avaliada?

Uma avaliação de viés de publicação deve incluir uma combinação de ferramentas gráficas (por exemplo, um gráfico de funil e outros exames disponíveis) e/ou testes estatísticos (por exemplo, teste de regressão Egger)

11. O conflito de interesses foi informado?

Possíveis fontes de apoio devem ser claramente reconhecidas, tanto na revisão sistemática quanto nos estudos incluídos

TABELA 2. COMO INTERPRETAR OS RESULTADOS DE REVISÕES SISTEMÁTICAS DOS EFEITOS

As perguntas a seguir podem ajudar a orientar os formuladores de políticas na interpretação dos resultados das revisões sistemáticas de efeitos (adaptado de [33,47,48]*):

- *Qual estimativa de efeito é apresentada?* Muitas revisões apresentam uma estimativa média do efeito em todos os estudos incluídos. Isto ocorre muitas vezes sob a forma de uma relação de risco, relação de probabilidade, ou uma diferença média padronizada
- *Uma estimativa média do efeito em todos os estudos é apropriada?* As revisões usam métodos estatísticos para sintetizar e combinar dados de resultados dos estudos incluídos na revisão. Para garantir que a combinação dos dados de resultado é apropriada, deve-se considerar se os estudos incluídos eram suficientemente semelhantes em termos de população, intervenção, comparação e resultados medidos. Sempre que uma estimativa média do efeito não for possível, as revisões geralmente apresentam uma visão geral narrativa dos dados disponíveis
- *Os limites de confiança para a estimativa do efeito são apresentados?* A revisão deverá apresentar intervalos de confiança ao redor da estimativa média do efeito. Quanto maior o intervalo de confiança menor será a certeza sobre a verdadeira magnitude do efeito
- *Se os resultados das análises de subgrupos são informados, eles são adequados?* Uma revisão pode apresentar resultados para um determinado subgrupo de participantes em todos os ensaios ou para um subgrupo de estudos [49]. Por exemplo, uma revisão das intervenções para reduzir as doenças diarreicas em crianças com menos de 5 anos de idade pode também considerar os efeitos das intervenções em crianças com menos de 1 ano de idade. Da mesma forma, uma revisão poderá incluir uma análise de subgrupo de estudos julgados como tendo um baixo risco de viés. A análise de subgrupos deve fazer sentido tanto em relação à pergunta da revisão geral quanto em relação ao conhecimento prévio dos fatores que podem ter influenciado ou moderado os efeitos da intervenção. Por exemplo, é possível prever que uma intervenção de maior intensidade pode produzir efeitos maiores. As análises de subgrupos devem ser planejadas antes que uma revisão seja realizada e estes resultados específicos sejam considerados menos confiáveis. Isto ocorre porque eles são menos confiáveis do que análises baseadas em todos os ensaios incluídos e porque análises estatísticas múltiplas podem produzir resultados positivos por acaso somente
- *Se existe uma “ausência de evidência de efeito”, toma-se cuidado para não interpretar tal fato como uma “evidência de ausência de efeito”?* “Ausência de evidência de efeito” não é igual a “evidência de ausência de efeito”. O primeiro sugere que não existe evidência suficiente disponível para tirar conclusões sobre os efeitos da intervenção em questão. O segundo sugere que há uma clara evidência nos estudos incluídos que a intervenção não tem os efeitos previstos [50]

- As conclusões e recomendações (se houver) decorrem tanto da pergunta da revisão original quanto da evidência que é apresentada na revisão? É importante analisar se as conclusões apresentadas pelos autores da revisão resultam diretamente dos dados coletados a partir da revisão e não vão além destas evidências
- As evidências são aplicáveis à pergunta de política sendo considerada? As diferenças nos sistemas de saúde podem significar que um programa ou uma intervenção que funciona em um contexto pode não funcionar da mesma maneira em outro. Os formuladores de políticas precisam avaliar se as evidências de pesquisa de uma revisão se aplicam ao seu contexto. Orientações sobre este assunto são apresentadas no Artigo 9 desta série [28]

* Existe alguma sobreposição entre as perguntas listadas aqui e as que se destinam a orientar a avaliação da confiabilidade de revisões sistemáticas. Isto ocorre porque a confiabilidade é um elemento importante na avaliação e compreensão dos resultados de uma revisão sistemática

TABELA 3. COMO AVALIAR ATÉ QUE PONTO OS RESULTADOS DE REVISÕES SISTEMÁTICAS DE ESTUDOS QUALITATIVOS E REVISÕES SISTEMÁTICAS DE ESTUDOS ECONÔMICOS SÃO CONFIÁVEIS

Existem cada vez mais revisões sistemáticas de estudos qualitativos sendo realizadas. Estas revisões usam uma variedade de abordagens, incluindo sínteses narrativas, meta-etnografia e revisão realista. Além de fornecer informações importantes, por si só, as revisões de estudos qualitativos também podem fundamentar e complementar revisões sistemáticas de efeitos [51,52]. No entanto, é importante que o leitor avalie a confiabilidade dessas revisões. Até hoje, poucas ferramentas foram concebidas para este fim específico. Entretanto, muitas das perguntas utilizadas para orientar os formuladores de políticas também são úteis para revisões de estudos qualitativos no momento da avaliação da confiabilidade de revisões sistemáticas de efeitos. Estas perguntas incluem, entre outras:

1. *A revisão abordou uma pergunta adequada sobre gestão ou política?* Deve ser possível responder a pergunta da revisão usando dados qualitativos e a pergunta da revisão deve ser relevante para a formulação de políticas. Revisões de estudos qualitativos podem proporcionar discernimentos sobre a opinião dos interessados e experiências sobre saúde e atendimento de saúde e, assim, ajudar a esclarecer um problema [39]. Revisões de estudos qualitativos também podem fornecer informações sobre como as opções funcionam ou por que elas funcionam (por exemplo, através do exame das avaliações de processo realizadas juntamente com a implantação de uma política ou programa) e sobre os pontos de vista das partes interessadas com relação às opções e suas experiências relevantes [40,53]
2. *Os critérios usados para selecionar os estudos eram apropriados?* A descrição de como os estudos foram selecionados deve ser apropriada em relação à pergunta da pesquisa

3. *Foi dada uma explicação clara e adequada para a abordagem de pesquisa utilizada?* Algumas revisões de estudos qualitativos realizam buscas abrangentes na literatura, ao passo que outras podem usar abordagens com amostragem. A abordagem escolhida deve ser claramente descrita e justificada
4. *A abordagem utilizada para avaliar a confiabilidade dos estudos incluídos foi adequada?* A revisão deve descrever como a confiabilidade dos estudos incluídos foi levada em consideração
5. *Foi utilizada uma abordagem adequada para analisar os resultados dos estudos incluídos?* A revisão deve usar uma abordagem aceita para a síntese e deve descrever a justificativa para a abordagem escolhida

As questões a serem consideradas durante a avaliação da confiabilidade das revisões de estudos econômicos incluem (de [54]):

1. É improvável que estudos relevantes importantes não tenham sido considerados?
2. Os critérios de inclusão usados para selecionar artigos eram apropriados?
3. A avaliação dos estudos podia ser reproduzida?
4. A concepção e/ou os métodos e/ou o tópico dos estudos incluídos eram comparáveis em linhas gerais?
5. Até que ponto os resultados gerais podem ser reproduzidos?
6. Os resultados ajudarão na alocação de recursos no serviço de saúde?

TABELA 4. EXEMPLOS DE FONTES PESQUISADAS NAS REVISÕES SISTEMÁTICAS

Revisão	Fontes pesquisadas
<p>Revisão de sistemas de saúde</p> <p>Exemplo: Revisão sistemática de intervenções de trabalhadores de saúde leigos em atendimento de saúde primário e comunitário [44]</p>	<ol style="list-style-type: none"> 1. Bases eletrônicas de dados de estudos publicados: <ul style="list-style-type: none"> • MEDLINE • CENTRAL (Cochrane Central Register of Controlled Trials – Cadastro Central Cochrane de Ensaios Controlados) e Cadastros Cochrane especializados (EPOC e Consumers and Communication Review Groups [Grupos de Revisão de Comunicação e Consumidores]) • Citações científicas • EMBASE • CINAHL (Cumulative Index to Nursing and Allied Health Literature – Índice Acumulativo para Literatura de Saúde Aliada e de Enfermagem) • Healthstar • AMED (Allied and Complementary Medicine Database – Base de Dados de Medicina Complementar e Aliada) • Leeds Health Education Effectiveness Database (Banco de Dados da Leeds sobre Eficácia da Educação de Saúde) 2. Bibliografias de estudos que foram avaliados para serem incluídos 3. Pediu-se a todos os autores que foram contatados que dessem detalhes de estudos adicionais
<p><i>Revisão de saúde pública</i></p> <p>Exemplo: Revisão sistemática da circuncisão masculina para prevenir que homens heterossexuais contraíam HIV [55]</p>	<ol style="list-style-type: none"> 1. Bases eletrônicas de dados de estudos publicados: <ul style="list-style-type: none"> • MEDLINE • EMBASE • Cochrane Central Register of Controlled Trials (CENTRAL) 2. Bases de dados eletrônicas de resumos de conferências: Bases de dados da Conferência da AIDSearch 3. Bases de dados eletrônicas de ensaios em andamento: <ul style="list-style-type: none"> • ClinicalTrials.gov • Ensaios Controlados Atuais 4. Pesquisadores e organizações relevantes no campo foram contatadas 5. As listas de referência de todos os estudos identificados

Revisão	Fontes pesquisadas
<p>Revisão clínica Exemplo: Revisão sistemática das estatinas para a prevenção da demência [56]</p>	<p>pelos métodos acima foram verificadas. Além disto, quaisquer revisões sistemáticas, meta-análises ou orientações de prevenção identificadas durante o processo de pesquisa foram examinadas</p> <ol style="list-style-type: none"> 1. Bases eletrônicas: <ul style="list-style-type: none"> • The Specialized Register of the Cochrane Dementia and Cognitive Improvement Group (O Cadastro Especializado do Grupo Cochrane de Demência e Melhoria Cognitiva) • Cochrane Central Register of Controlled Trials (CENTRAL) • MEDLINE • EMBASE • PsycINFO (uma base de dados da literatura de psicologia) • CINAHL • SIGLE (Literatura cinzenta na Europa) • LILACS (Latin American and Caribbean Health Science Literature = Literatura de Ciência de Saúde do Caribe e da América Latina) 2. Bases de dados eletrônicas de resumos de conferências: <ul style="list-style-type: none"> • ISTP (Index to Scientific and Technical Proceedings = Índice para Processos Técnicos e Científicos) • INSIDE (Banco de Dados da Biblioteca Britânica de Revistas e Processos de Conferência) 3. Bases eletrônicas de dados de teses: <ul style="list-style-type: none"> • Index to Theses (Índice para Teses) (ex-ASLIB) (Teses do Reino Unido e da Irlanda) • Programa Digital de Teses da Austrália • Teses e Dissertações do Canadá • DATAD (Database of African Theses and Dissertations = Banco de Dados de Teses e Dissertações Africanas) • Dissertation Abstract Online (Resumo de Dissertações Online) (EUA) 4. Bancos de dados eletrônicos de ensaios em andamento: foi realizadas buscas em uma ampla variedade destes bancos de dados

TABELA 5. O QUE OS FORMULADORES DE POLÍTICAS DEVEM FAZER QUANDO DIFERENTES REVISÕES SISTEMÁTICAS QUE ABORDAM A MESMA PERGUNTA TÊM RESULTADOS DIFERENTES?

Quando se procura por evidências para fundamentar uma decisão de política específica, não é incomum identificar mais de uma revisão sistemática relevante. Às vezes, os resultados destas revisões podem ser diferentes, e isto pode fazer com que os autores de revisões cheguem a diferentes conclusões sobre os efeitos de uma intervenção. Este cenário difere daquele em que os resultados de duas ou mais revisões concordam, mas onde os pesquisadores ou outras pessoas discordam sobre a interpretação destes resultados [19].

Existem muitas razões pelas quais os resultados de diferentes revisões sistemáticas podem ser diferentes. Isto inclui diferenças nos seguintes itens: as perguntas abordadas pelas revisões, os critérios de inclusão e exclusão utilizados, quais dados foram extraídos dos estudos, como a qualidade dos estudos foi avaliada e as decisões a respeito da (e os métodos para a) análise estatística dos dados [19].

A seguinte série de perguntas elaboradas por Jadad e colegas pode ser usada para ajudar a identificar e abordar as causas das discordâncias [19]:

- As revisões são direcionadas para a mesma pergunta? Caso não, a revisão que é escolhida deve ser aquela que responda a uma questão mais próxima àquela da pergunta da política para a qual as evidências são necessárias. Por outro lado, ela deve avaliar os resultados mais relevantes para a pergunta da política
- Se as revisões abordam a mesma pergunta, elas incluem os mesmos ensaios ou estudos primários? Se elas não incluem os mesmos ensaios, deve ser selecionada a revisão que inclui os estudos mais relevantes para a pergunta da política a ser considerada
- Se as revisões incluem os mesmos estudos, elas têm a mesma qualidade? Caso não, a revisão de melhor qualidade deve ser usada

Quando ambas as revisões forem relevantes, por exemplo, quando elas abordarem diferentes aspectos da mesma pergunta, pode ser útil coletar evidências de ambas.