

Ministério da Saúde
Secretaria de Ciência, Tecnologia e Insumos Estratégicos
Departamento de Ciência e Tecnologia

ROBIS (Risk of Bias in Systematic Reviews):
Ferramenta para avaliar o risco de viés em revisões sistemáticas
Orientações de uso

Tradução para o português (não validada)

Brasília-DF

2017

SUMÁRIO

A ferramenta ROBIS.....	4
Definição de risco de viés	4
Público-alvo	4
Fase 1: Avaliação da relevância (opcional).....	5
Exemplo de avaliação da relevância ³	6
Fase 2: Identificação dos potenciais riscos de viés durante o processo da revisão	6
Domínio 1: Critérios de elegibilidade dos estudos.....	7
Exemplos de classificação.....	9
Domínio 2: Identificação e seleção dos estudos.....	13
Exemplos de classificação.....	15
Domínio 3: Coleta de dados e avaliação do estudo.....	18
Exemplos de classificação.....	21
Domínio 4: Síntese e resultados.....	23
Exemplos de classificação.....	27
Fase 3: Avaliação do risco de viés geral	29
Exemplos de classificação.....	31
Referências.....	35
Apêndice: Glossário.....	37

Tabela

Tabela 1. Exemplos de questões-alvo e PICO/equivalente para diferentes tipos de revisões sistemáticas	5
---	---

Quadros

Quadro 1. Questões norteadoras referentes ao Domínio 1 e orientações para respostas.....	8
Quadro 2. Potencial risco de viés relacionado à especificação dos critérios de elegibilidade	9
Quadro 3. Exemplo de classificação para critérios de elegibilidade julgados como de alto potencial de viés	10
Quadro 4. Exemplo de classificação para critérios de elegibilidade julgados como de baixo potencial de viés.....	12
Quadro 5. Questões norteadoras referentes ao Domínio 2 e orientações para respostas.....	14
Quadro 6. Potencial de viés relacionado aos métodos utilizados para identificar e/ou selecionar estudos.....	15
Quadro 7. Exemplo de classificação para estratégia de busca julgada como de alto potencial de viés	16
Quadro 8. Exemplo de classificação para estratégia de busca julgada como de baixo potencial de viés	17
Quadro 9. Questões norteadoras referentes ao Domínio 3 e orientações para respostas.....	19
Quadro 10. Potencial de viés relacionado aos métodos utilizados para coletar dados e avaliar os estudos.....	20
Quadro 11. Exemplo de classificação para coleta de dados e avaliação do estudo julgado como de alto potencial de viés	21
Quadro 12. Exemplo de classificação para coleta de dados e avaliação do estudo julgado como de baixo potencial de viés	22
Quadro 13. Questões norteadoras referentes ao Domínio 4 e orientações para respostas.....	24
Quadro 14. Potencial de viés relacionado aos métodos utilizados para sintetizar resultados	27
Quadro 15. Exemplo de classificação para síntese julgada como de alto potencial de viés	27
Quadro 16. Exemplo de classificação para síntese de resultados julgada como de baixo potencial de viés.....	28
Quadro 17. Questões norteadoras referentes à fase 3 e orientações para respostas	30
Quadro 18. Riscos de viés relacionados aos métodos utilizados na identificação e/ou seleção dos estudos.....	31
Quadro 19. Quadro resumo dos potenciais vieses identificados durante a Fase 2	31
Quadro 20. Exemplo de classificação para síntese de resultados julgada como de alto potencial de viés.....	32
Quadro 21. Quadro resumo dos potenciais de vieses identificados durante a Fase 2	32
Quadro 22. Exemplo de classificação para síntese de resultados julgada como de baixo potencial de viés.....	34

A ferramenta ROBIS

A ROBIS (Risk of Bias in Systematic Reviews) é uma ferramenta para avaliar risco de viés de revisões sistemáticas estruturada em três fases: (1) avaliação da relevância (opcional), (2) identificação dos potenciais riscos de viés durante o processo da revisão e (3) avaliação do risco de viés geral. A ferramenta contém questões norteadoras que auxiliam na avaliação crítica. A partir da classificação obtida por meio destas questões, os avaliadores são capazes de julgar o risco de viés geral da revisão.

Definição de risco de viés

A ROBIS avalia tanto o risco de viés da revisão quanto sua relevância para a pergunta de pesquisa a ser respondida. A ferramenta aborda especificamente: 1) se os métodos utilizados na revisão sistemática minimizaram o risco de viés das estimativas sumarizadas e das conclusões; e 2) se a pergunta de pesquisa de cada revisão avaliada se assemelha à pergunta de pesquisa a ser respondida, por exemplo, por uma *overview* de revisões sistemáticas ou por uma diretriz. O viés ocorre quando erros sistemáticos ou limitações no delineamento, condução ou análise da revisão influenciam os resultados. A evidência de uma revisão pode ter relevância limitada se a questão da revisão não corresponder à pergunta da *overview*/diretriz.

Público-alvo

A ROBIS foi desenvolvida pensando em três grupos específicos de usuários:

- Autores de *overview* de revisões sistemáticas
- Elaboradores de diretrizes
- Revisores que desejam avaliar o risco de viés na sua revisão sistemática ou minimizar o risco de viés ao planejar os métodos de revisão na fase de protocolo.

Outros possíveis usuários da ROBIS incluem organizações que apoiam a tomada de decisões (NICE, IQWiG), clínicos com interesse em medicina baseada em evidência, editores de revistas e revisores de manuscritos.

Fase 1: Avaliação da relevância (opcional)

Os avaliadores primeiro reportam a pergunta que estão tentando responder (na *overview* ou diretriz), o que chamamos de “questão alvo”. Para revisões de eficácia ou efetividade é solicitada a definição da pergunta conforme o acrônimo PICO (participantes, intervenções, comparadores e resultados). Para revisões de diferentes tipos de perguntas (teste de diagnóstico, fatores prognósticos, etiologia ou modelos de predição) são fornecidas categorias alternativas (ver Tabela 1). Os avaliadores completam o PICO ou equivalente usando a ROBIS e, então, a pergunta a ser respondida (questão alvo) e a pergunta da revisão sistemática são comparadas a fim de verificar se são correspondentes. Se uma ou mais categorias das perguntas não forem correspondentes, esta fase deve ser classificada como "Não". Se há uma correspondência parcial entre as categorias esta fase deverá ser classificada como "Incerta". Por exemplo, se a questão-alvo se refere a adultos, mas a revisão sistemática é restrita a participantes com mais de 60 anos. Se uma revisão está sendo avaliada isoladamente e não há uma questão-alvo, esta fase da ROBIS pode ser omitida.

Tabela 1. Exemplos de questões-alvo e PICO/equivalente para diferentes tipos de revisões sistemáticas

Tipo de revisão	Pico/equivalente	Exemplo
Intervenção ¹	Pacientes/População: Intervenção: Comparador: Desfecho:	Adultos com hepatite C crônica Terapia tripla antiviral com interferon peguilhado Terapia dupla antiviral Resposta virológica
Etiologia ²	Pacientes/População: Exposição e comparador: Desfecho:	Adultos Índice de massa corporal Câncer colorretal
Diagnóstico ³	Pacientes/População: Teste <i>index</i> : Padrão de referência: Condição Alvo:	Adultos com sintomas sugestivos de câncer retal Ecografia endoscópica Histologia cirúrgica Câncer retal inicial
Prognóstico ⁴	Pacientes/População: Predição de desfecho: Modelo pretendido: Momento pretendido no tempo:	Mulheres grávidas, com ou sem restrição de crescimento fetal, sem evidência de ruptura prematura das membranas e ausência de anomalias congênitas ou estruturais. Desfechos adversos na gravidez (baixo ou alto peso ao nascer, morte neonatal e mortalidade perinatal) Predição Gravidez tardia (>37 semanas)

Exemplo de avaliação da relevância³

Para revisão de estudo sobre diagnóstico:

Categoria	Questão-alvo	Revisão que está sendo avaliada³
Pacientes:	Adultos com sintomas sugestivos de câncer retal	Incerta
Teste <i>index</i> (S):	Ecografia endoscópica	Ecografia endoscópica
Referência Padrão:	Histologia cirúrgica	Cirurgia
Condição Alvo	Câncer retal precoce	Câncer retal precoce

Avaliação da relevância	Justificativa
A questão abordada pela revisão corresponde à pergunta que você está tentando responder (por exemplo, na <i>overview</i> ou diretriz)?	Incerta A pergunta coincide com as categorias teste <i>index</i> , padrão de referência e condição alvo, mas a população não foi definida na revisão e, portanto, não está claro se a questão abordada pela revisão corresponde à questão-alvo.

Fase 2: Identificação dos potenciais riscos de viés durante o processo da revisão

A Fase 2 tem como objetivo identificar áreas onde o viés pode ser introduzido na revisão sistemática. Envolve a avaliação de quatro domínios para cobrir os principais processos da revisão: (1) critérios de elegibilidade dos estudos; (2) identificação e seleção; (3) coleta de dados e avaliação dos estudos; e (4) síntese e resultados. Nesta fase a ROBIS identifica as principais limitações para auxiliar a avaliação do risco de viés na fase final. Cada um dos quatro domínios da ferramenta compreende três seções: 1) informação usada para subsidiar a avaliação, 2) questões norteadoras e 3) avaliação do potencial risco de viés. Os domínios devem ser considerados sequencialmente e não avaliados isoladamente. Isso significa que, ao avaliar o domínio 2 (identificação e seleção dos estudos), o avaliador deve considerar a estratégia de busca em relação à questão de pesquisa apresentada no domínio 1.

As questões norteadoras são respondidas como “Sim”, “Provavelmente Sim”, “Provavelmente Não”, “Não” e “Sem informações”, com “Sim” indicando baixo potencial de risco de viés. O risco potencial de viés associado com cada domínio é avaliado em seguida como “Baixo”, “Alto” ou “Incerto”. Se as respostas de todas as questões norteadoras para um domínio são “Sim” ou “Provavelmente Sim”, então o potencial risco de viés pode ser avaliado como baixo. Se alguma questão norteadora for respondida como “Não” ou “Provavelmente Não” o risco de viés poderá ser classificado

como alto. A categoria “Sem Informação” deve ser usada somente quando os dados reportados forem insuficientes para permitir um julgamento. Ao registrar as informações utilizadas para subsidiar a avaliação, pretendemos tornar a classificação transparente e, quando necessário, facilitar a discussão entre os autores da revisão. É provável que os usuários da ROBIS necessitem de conteúdo e de conhecimentos metodológicos para concluir uma avaliação.

Domínio 1: Critérios de elegibilidade dos estudos

Este primeiro domínio objetiva avaliar se os critérios de elegibilidade dos estudos foram especificados previamente, esclarecidos e apropriados para a pergunta da revisão. Uma revisão sistemática deve, primeiramente, apresentar uma pergunta ou um objetivo claro⁽⁵⁾. A pergunta e/ou o objetivo do estudo devem estar refletidos nos critérios de inclusão da revisão. Assim, haverá uma maior probabilidade de as decisões sobre a inclusão de estudos serem realizadas de maneira mais consistente, evitando-se que sejam feitas com base em conhecimentos prévios sobre os resultados ou características dos estudos a serem incluídos. Normalmente, só é possível avaliar se os critérios de inclusão foram previamente especificados se houver registro de protocolo da revisão com data anterior à realização e elaboração da revisão sistemática.

Os critérios de elegibilidade combinam aspectos da pergunta de pesquisa com os delineamentos dos estudos que a revisão irá considerar. Idealmente, as características dos estudos elegíveis devem ser específicas e englobar a população/participantes envolvidos, a exposição ou intervenção avaliada e as comparações realizadas. Autores de revisões sistemáticas frequentemente impõem restrições nos critérios de elegibilidade, tais restrições devem ser julgadas como apropriadas ou não pelos revisores. Por exemplo, as revisões podem ter restringido a inclusão dos estudos com base em algumas características dos próprios estudos, tais quais, ter selecionado apenas aqueles com os desfechos de interesse, que atenderam a critérios específicos de qualidade ou que foram conduzidos em datas específicas. Da mesma forma, restrições às fontes de informações podem também ser impostas quando uma revisão registra que apenas estudos disponíveis na íntegra ou no idioma inglês serão selecionados. Quaisquer restrições relacionadas à inclusão dos estudos necessitam ser reportadas e justificadas no texto. Em certas circunstâncias, estas restrições podem introduzir vieses na seleção de estudos elegíveis para a revisão.

O Quadro 1 resume as questões norteadoras do Domínio1 e descreve orientações para resposta de cada questão. O Quadro 2 indica uma visão geral dos níveis de classificação para o Domínio1. Exemplos da classificação para este domínio (Potencial alto e baixo de viés) estão descritos nos Quadros 3 e 4.

Quadro 1. Questões norteadoras referentes ao Domínio 1 e orientações para respostas

Questões norteadoras	Orientações para respostas
1.1 Os objetivos e critérios de elegibilidade foram definidos previamente?	Uma revisão sistemática precisa, primeiramente, conter uma pergunta ou um objetivo claro que estejam refletidos nos critérios utilizados para decidir se os estudos são elegíveis para inclusão. As informações que devem estar detalhadas <i>a priori</i> no protocolo de uma revisão geralmente incluem delineamentos dos estudos, participantes e tipos de exposições ou intervenções elegíveis; embora estas informações variem de acordo com tipo de revisão. Quaisquer exclusões devem ser especificadas previamente. Se um protocolo com estas informações for disponibilizado, a resposta para esta questão será “Sim”. Se não for disponibilizado um protocolo, mas houver informações sobre objetivos pré-definidos, detalhamento dos critérios de elegibilidade e razões suficientes para acreditar que estes critérios foram especificados antes do início da revisão sistemática, a resposta a esta questão pode ser “Provavelmente Sim”. Qualquer mudança <i>post hoc</i> aos critérios de elegibilidade e desfechos devem manter-se alinhados aos objetivos da revisão e serem justificados e documentados. No caso de não haver um protocolo previamente publicado com informações sobre objetivos e critérios de elegibilidade pré-definidos, estando disponíveis apenas <i>post hoc</i> na revisão, a resposta a esta questão deve ser “Provavelmente não”; a menos que haja motivos para acreditar que estas informações tenham sido especificadas e obedecidas previamente. Nos casos em que todas ou algumas dessas informações estiverem faltantes, a resposta a esta questão deve ser “Não”.
1.2 Os critérios de elegibilidade eram apropriados à pergunta da revisão?	Os critérios de elegibilidade devem ser provenientes da pergunta e devem fornecer informações que permitam o julgamento sobre a adequação dos estudos incluídos à questão da revisão. É provável que a informação requerida varie de acordo com o tópico. Por exemplo, para julgar a adequação, o avaliador pode precisar de uma descrição clara da população em termos de mínimo e máximo de idades, diagnósticos dos participantes, local de realização, dosagem da droga, ou frequência a uma determinada exposição. Provavelmente o avaliador terá que ter algum conhecimento sobre o tema para responder a esta questão.
1.3 Os critérios de elegibilidade foram bem especificados (sem ambiguidade)?	Informações detalhadas sobre as características dos estudos elegíveis devem ser fornecidas para evitar quaisquer ambiguidades sobre tipos de estudo, população, intervenções, comparadores e desfechos. Os critérios devem ser detalhados a ponto de a revisão poder ser replicada com base nos critérios especificados. Algumas informações importantes geralmente são omitidas dos critérios de inclusão em revisões sistemáticas. Por exemplo, informações sobre o diagnóstico dos participantes. Os diagnósticos podem ser obtidos a partir de uma variedade de métodos, alguns mais válidos e acurados que outros. Os autores da revisão devem decidir previamente quais são os métodos diagnósticos apropriados para a pergunta de pesquisa, a fim de evitar a introdução de viés durante o processo da revisão. Da mesma forma, informações específicas sobre intervenções/exposições e comparadores devem ser fornecidas, incluindo-se dosagem da medicação, frequência de administração, tratamentos concorrentes etc. Provavelmente o avaliador terá que ter algum conhecimento sobre o tema para responder a esta questão. No entanto, se dúvidas específicas permanecerem sobre os critérios de

	elegibilidade, esta questão será geralmente respondida como “Não” ou “Provavelmente Não”.
1.4 Com relação aos critérios de elegibilidade, as restrições baseadas nas características dos estudos foram apropriadas?	Quaisquer restrições aplicadas com base em características dos estudos devem ser claramente descritas. Estas informações permitirão aos avaliadores julgarem se as restrições foram apropriadas. Exemplos de tais restrições podem ser: delineamento, data de publicação, tamanho e qualidade do estudo, além de desfechos disponíveis. Esta questão é diferente da questão 1.3 que se refere a se os critérios de elegibilidade são apropriados à pergunta da revisão. Quando há informações suficientes e o avaliador está razoavelmente convencido que as restrições são adequadas, esta questão deve ser respondida com “Sim” ou “Provavelmente Sim”. Quando as restrições relacionadas às características dos estudos não são justificadas e há informações insuficientes para julgar se essas restrições são apropriadas, esta questão deve ser respondida com “Provavelmente Não” ou “Não”. Quando os critérios de elegibilidade são suficientemente detalhados e não há restrições quanto às características do estudo, esta questão deve ser respondida como “Sim”.
1.5 Com relação aos critérios de elegibilidade, as restrições baseadas nas fontes de informações foram apropriadas?	Qualquer restrição aplicada com base nas fontes de informações deve ser claramente descrita e fornecida a razão para tanto. Estas informações permitirão aos avaliadores julgarem se as restrições foram apropriadas. Exemplos podem ser: <i>status</i> ou formato de publicação, idioma e disponibilização de dados. Esta questão é diferente da questão do Domínio2 que se refere a restrições de pesquisas. Quando os critérios de elegibilidade são suficientemente detalhados e não há restrições quanto às fontes de informação, esta questão deve ser respondida como “Sim”.

Quadro 2. Potencial risco de viés relacionado à especificação dos critérios de elegibilidade

Baixo	Nota-se que esforços foram realizados para deixar a pergunta e os objetivos da revisão bem definidos e para justificar que critérios de elegibilidade adequados foram preconizados na revisão.
Alto	Estudos que seriam importantes para responder à pergunta de pesquisa foram provavelmente excluídos da revisão pelos possíveis motivos: falta de objetivos e critérios de elegibilidade previamente especificados, imposição de restrições inapropriadas ou porque estudos inadequados para responder a pergunta da pesquisa foram incluídos.
Incerto	As informações disponíveis são insuficientes para fazer julgamento sobre o risco de viés.

Exemplos de classificação

- a. Revisão classificada como de “alto potencial de viés” em relação aos critérios de elegibilidade dos estudos.³

Descrição do texto

Resumo – Objetivo:

“Avaliar a acurácia do ultrassom endoscópico no estágio T de câncer retal inicial”

Critérios de seleção dos estudos:

“Apenas estudos com ultrassom endoscópico confirmado por histologia cirúrgica foram selecionados. O critério de ultrassom endoscópico utilizados para T0 (estágio inicial) foi tumor restrito à mucosa. Desta seleção, apenas estudos que permitissem a construção de uma tabela 2x2 para valores verdadeiro-positivo, falso-positivo, falso-negativo e verdadeiro-positivo foram incluídos”.

Quadro 3. Exemplo de classificação para critérios de elegibilidade julgados como de alto potencial de viés

Questões norteadoras	Classificação	Justificativa
1.1 Os objetivos e critérios de elegibilidade foram definidos previamente?	Sem informação	Não houve evidência de que os objetivos e critérios de elegibilidade foram previamente especificados.
1.2 Os critérios de elegibilidade eram apropriados à pergunta da revisão?	Provavelmente não	Os critérios de elegibilidade foram apropriados, mas algumas informações estavam incompletas particularmente em relação à população. Portanto, não foi possível julgar se os critérios de elegibilidade eram apropriados.
1.3 Os critérios de elegibilidade foram bem especificados (sem ambiguidade)?	Provavelmente não	Não houve detalhamento suficiente dos critérios de elegibilidade. Em particular, não houve detalhamento sobre que populações ou delineamentos de estudos foram elegíveis.
1.4 Com relação aos critérios de elegibilidade, as restrições baseadas nas características dos estudos foram apropriadas?	Sem informação	Não foram fornecidos detalhes sobre restrições relacionadas às características dos estudos. De forma geral, não há detalhamento suficiente sobre os critérios de elegibilidade.
1.5 Com relação aos critérios de elegibilidade, as restrições baseadas nas fontes de informações foram apropriadas?	Provavelmente sim	A revisão foi restrita a estudos em que a tabela 2x2 pôde ser construída. Embora nenhuma justificativa para isso tenha sido dada, isto é comum em revisões de Teste de Acurácia Diagnóstica (TAD), sendo assim uma restrição razoável.
Gradação do potencial risco de viés nas especificações dos critérios de elegibilidade do estudo	Alto	Não existiram informações suficientes sobre os critérios de elegibilidade para julgar se os estudos apropriados foram incluídos na revisão; em particular, o detalhamento sobre participantes elegíveis estava incompleto. Também não ficou claro se os critérios foram especificados previamente ou <i>post-hoc</i> . Portanto, existe alta preocupação quanto à especificação dos critérios de elegibilidade da revisão.

- b. Revisão avaliada como de “baixo potencial de viés” relacionada às especificações dos critérios de elegibilidade dos estudos.⁴

Descrições do texto:

Objetivo

“Avaliar a associação e o valor preditivo de medidas ultrassonográficas do volume de fluido amniótico para efeitos adversos da gravidez”.

Métodos – seleção dos estudos:

“1. População. Mulheres grávidas, com ou sem restrição de crescimento fetal, sem evidência de ruptura da membrana e sem evidência de anormalidades congênita ou estrutural.

2. Teste *index*. Qualquer medida de fluido amniótico reportada pelos autores incluindo AFI, volume de fluido amniótico e *maximum deepest pocket*. Foi aceito qualquer ponto de corte utilizado para definir alto ou baixo fluido amniótico reportado pelos autores dos estudos incluídos.

3. Desfecho. Qualquer padrão de referência para o comprometimento do bem-estar fetal ou neonatal; incluindo: pH anormal do cordão ao nascimento, escores de *Apgar*, óbito perinatal e desfechos relacionados. Qualquer padrão de referência para restrição do crescimento fetal ou para desenvolvimento fetal inferior ao esperado para a idade gestacional: peso ao nascer <10^o, <5^o, <3^o percentil, pontos de corte do peso absoluto ao nascimento.

4. Delineamento dos estudos. Estudos observacionais nos quais os resultados do teste de interesse são comparados com outros achados e confirmados por um padrão ouro, permitindo a geração da tabela 2x2 para computar índices de associação e testes de acurácia para cada ponto de corte disponível. Séries de casos com até 10 participantes foram excluídos.”

Métodos - extração de dados:

“Todos os estudos deveriam registrar que excluíram ruptura de membrana e anormalidades congênitas e estruturais, devido à associação de anormalidades do trato urinário e do cariótipo com anormalidades do volume do fluido”.

Quadro 4. Exemplo de classificação para critérios de elegibilidade julgados como de baixo potencial de viés.

Questões norteadoras	Classificação	Justificativa
1.1 Os objetivos e critérios de elegibilidade foram definidos previamente?	Sim	Os autores especificaram claramente no <i>abstract</i> e <i>background</i> que os objetivos do estudo foram avaliar a associação e o valor preditivo das mensurações ultrassonográficas do volume de fluido amniótico para efeitos adversos da gravidez. Um protocolo separado forneceu a pergunta de pesquisa “Qual é a acurácia do volume de fluido amniótico para prever comprometimento fetal/neonatal e restrição de crescimento fetal?” Os autores forneceram detalhes sobre os critérios de elegibilidade na seção de “Seleção dos estudos” organizados por população, teste <i>index</i> , desfecho e delineamento do estudo.
1.2 Os critérios de elegibilidade eram apropriados à pergunta da revisão?	Sim	A pergunta da revisão indicou a necessidade de identificar estudos sobre a acurácia do volume de fluido amniótico para prever comprometimento fetal/neonatal e restrição de crescimento fetal. O detalhamento no artigo sobre os estudos elegíveis para inclusão pareceu apropriado para a pergunta da revisão.
1.3 Os critérios de elegibilidade foram bem especificados (sem ambiguidade)?	Provavelmente Sim	Os tipos de delineamento dos estudos foram claramente relatados. Foram incluídos estudos observacionais. Os autores excluíram séries de casos com 10 ou menos participantes. Detalhes sobre a população de interesse (gestantes, com ou sem restrição de crescimento fetal, sem evidência de ruptura prematura da membrana, sem evidência de anormalidades congênitas ou estruturais), teste <i>index</i> (qualquer medida de volume do fluido amniótico independentemente do ponto de corte utilizado, incluindo-se AFI, volume do fluido amniótico e <i>maximum deepest pocket</i>), desfecho (qualquer padrão de referência sobre comprometimento do bem-estar fetal/neonatal e qualquer padrão de referência para restrição de crescimento fetal ou para desenvolvimento fetal inferior ao esperado para a idade gestacional).
1.4 Com relação aos critérios de elegibilidade, as restrições baseadas nas características dos estudos foram apropriadas?	Provavelmente Sim	As restrições baseadas nos delineamentos dos estudos e problemas de tamanho de amostras foram descritas de forma clara. Descrições adicionais foram fornecidas na seção de extração de dados. As restrições pareceram apropriadas, embora não tenham sido justificadas.
1.5 Com relação aos critérios de elegibilidade, as restrições baseadas nas fontes de informações foram apropriadas?	Provavelmente Sim	Não foram aplicadas restrições de idiomas. Algumas restrições podem ter sido baseadas na disponibilidade de dados, embora nenhuma justificativa tenha sido dada. Nenhuma outra restrição tocante às fontes de informações foi descrita.

		Portanto, essa questão foi respondida como “Provavelmente Sim”.
Gradação do potencial risco de viés nas especificações dos critérios de elegibilidade do estudo	Baixo	Todas as questões foram respondidas como “Sim” ou “Provavelmente Sim”, então não foi identificado potencial risco de viés sobre a especificação dos critérios de elegibilidade. Esforço considerável foi feito para especificar a pergunta e os objetivos da revisão, bem como para especificar previamente e justificar os critérios de elegibilidade.

Domínio 2: Identificação e seleção dos estudos

O objetivo deste domínio é avaliar se os estudos primários que atendiam aos critérios de inclusão não foram selecionados para a revisão. Uma busca sensível pode recuperar o maior número de artigos possíveis, sendo um dos principais componentes da revisão. É necessário que a pesquisa seja realizada ou orientada por um profissional especialista em buscas na literatura. As estratégias de busca devem ser realizadas em bases de dados e fontes eletrônicas apropriadas, as quais indexam periódicos, resumos de congresso e registros de ensaios clínicos a fim de se identificar o maior número possível de artigos publicados e não publicados. Além de uma estratégia de busca apropriada e sensível, é necessário também realizar buscas adicionais, por exemplo, verificando referências em revisões existentes, pesquisando por citações e busca manual. A estratégia de busca deve incluir palavras-chave (no título e resumo) e os termos de indexação (MeSH e Emtree). Também pode ser útil procurar por campos adicionais, por exemplo, medicamento, número de registro ou nome comercial. A estratégia de busca geralmente envolve uma combinação de termos referentes à população, intervenção, exposição e teste *index*. Para algumas revisões, tais como revisões de ensaios clínicos randomizados, podem ser utilizados filtros para metodologia ou delineamento dos estudos nas bases de dados pesquisadas. No entanto, para outros tipos de revisões, como de estudos de acurácia diagnóstica, o uso de filtros demonstrou perda de estudos relevantes e, portanto, deve ser evitado.^{6, 7} É difícil avaliar a sensibilidade da busca sem ter conhecimento técnico sobre estratégias de busca e o tema da revisão. Os avaliadores devem considerar se foram utilizados termos apropriados e abrangentes que compreendam todos os conceitos possíveis da questão de pesquisa. Em uma revisão de câncer de mama realizar a busca só pelo termo “*breast cancer*” não é suficiente. Uma busca que poderia ser utilizada para esta população é a seguinte:

Base de dados MEDLINE busca realizada via OvidSP

1 exp breast neoplasms/ or Carcinoma, Ductal, Breast/ (214761)

2 exp breast/ (31443)

3 exp neoplasms/ (2533252)

4 2 and 3 (17379)

5 (breast\$ adj5 (neoplasm\$ or cancer\$ or tumor?\$ or metastas\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intraductal\$ or lobular or medullary)).ti,ab,ot,hw. (256928)

6 (mammar\$ adj5 (neoplas\$ or cancer\$ or tumor?\$ or metastas\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intra-ductal\$ or lobular or medullary)).ti,ab,ot,hw. (34861)

7 1 or 4 or 5 or 6 (277790)

O Quadro 5 resume as questões norteadoras do Domínio 2 e descreve orientações para a resposta de cada questão. O Quadro 6 indica uma visão geral dos níveis de classificação para o Domínio 2. Exemplos da classificação para este domínio (Potencial alto e baixo de viés) estão descritos nos Quadros 7 e 8.

Quadro 5. Questões norteadoras referentes ao Domínio 2 e orientações para respostas

Questões norteadoras	Orientações para respostas
2.1 A busca incluiu uma variedade suficiente de bases de dados/fontes eletrônicas para pesquisar artigos publicados e não publicados?	O avaliador precisa julgar o que constitui uma variedade apropriada de bases de dados para a questão de pesquisa. Deve-se, no mínimo, conduzir uma pesquisa no <i>MEDLINE</i> e <i>EMBASE</i> . A busca por material de congresso também deve ser considerada no resultado da pesquisa. Orientações sobre variedade adequada das bases de dados podem ser encontradas no <i>Handbook</i> da Cochrane ⁵ ou no site do <i>Centre for Reviews and Dissemination</i> (CRD) (http://www.york.ac.uk/inst/crd/finding_studies_systematic_reviews.htm).
2.2 Além da busca nas bases de dados foram utilizados métodos adicionais para a identificação de artigos relevantes?	Métodos adicionais, tais como busca por citações, contato com especialistas, busca nas referências, pesquisa manual, etc.
2.3 Os termos e a estrutura da estratégia de busca foram adequados para obter o maior número possível de artigos?	A estratégia de busca deve ser completa mostrando todos os termos utilizados com detalhes suficientes para que possa ser replicada. Se somente alguns dados forem fornecidos, por exemplo, lista com os termos sem nenhuma indicação de como foram combinados, os avaliadores podem responder esta questão como "Provavelmente Sim" ou "Provavelmente Não". Os avaliadores devem considerar se a estratégia de busca incluiu termos adequados para o tópico, se uma combinação de termos controlados (MeSH para <i>Medline</i>) e palavras-chave no título e resumo foram usadas e se os filtros aplicados foram adequados. Por exemplo, para revisões de testes de acurácia diagnóstica, o uso de filtros mostrou perda de estudos relevantes e, portanto, essa pergunta deve ser respondida como Não para uma estratégia que incluía esses filtros. ⁶ As orientações sobre a avaliação crítica das estratégias de busca podem ser encontradas no <i>PRESS Evidence-Based Checklist</i> (http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402).

2.4 As restrições baseadas na data, formato de publicação ou idioma foram adequadas?	Se nenhuma restrição foi aplicada na estratégia de busca, então essa questão deve ser respondida como “sim”. Esta questão é diferente do domínio 1 (1.5) que avalia a restrição do critério de seleção. É necessária informação sobre os três componentes desta questão (data, formato de publicação e idioma) para poder avaliar este item. A restrição de artigos baseados em idiomas (somente artigos em inglês) ou formato de publicação (por exemplo estudos publicados na íntegra) é raramente apropriada. Dessa forma, se tais restrições forem aplicadas a resposta geralmente é "Não". Para que esta questão seja respondida como “sim”, restrições de data podem ser aplicadas, no entanto devem ser claramente explicadas. Por exemplo, se uma medicação ou teste não estava disponível antes de uma determinada data, então pode se iniciar a busca a partir da data em que a medicação ou teste se tornou disponível.
2.5 Foram realizadas tentativas para minimizar erros na seleção dos estudos?	Tanto o processo de seleção por títulos e resumos quanto a avaliação dos estudos na íntegra são contemplados nesta questão. Para uma resposta "Sim", os títulos e resumos devem ser selecionados de forma independente por pelo menos dois revisores e a avaliação do texto na íntegra também deve envolver ao menos dois revisores (independentes ou com uma pessoa para realizar a avaliação e a segunda verificar os resultados).

Quadro 6. Potencial de viés relacionado aos métodos utilizados para identificar e/ou selecionar estudos

Baixo	Dada a questão da revisão e os critérios de elegibilidade avaliados no domínio 1, um esforço substancial foi realizado para identificar o maior número de estudos possíveis. Foi utilizada uma variedade de métodos de pesquisa por meio de uma estratégia de busca sensível e apropriada, além de serem tomadas medidas para minimizar o viés e erros na seleção dos estudos.
Alto	Alguns estudos elegíveis para a revisão foram provavelmente perdidos.
Incerto	Não há informações suficientes para fazer uma avaliação sobre o risco de viés.

Exemplos de classificação

a. Revisão classificada como de alto potencial de viés em relação aos métodos utilizados para identificar e/ou selecionar estudos ¹

Descrição do texto:

“Foi realizada busca nas bases de dados: *Ovid MEDLINE* de 1947 a agosto de 2012, *Cochrane Library* (até o primeiro trimestre de 2012), *Embase* (1976 a agosto de 2012), *Scopus* (1960 a agosto de 2012) e *PsychINFO* (1806 a agosto de 2012), além de busca por registros de ensaios clínicos. Ao menos dois revisores independentes avaliaram os estudos para inclusão. Foram selecionados artigos no idioma inglês. Os resumos de congresso foram incluídos somente na análise de sensibilidade. ”

Quadro 7. Exemplo de classificação para estratégia de busca julgada como de alto potencial de viés

Questões norteadoras	Classificação	Justificativa
2.1 A busca incluiu uma variedade suficiente de bases de dados/fontes eletrônicas para pesquisar artigos publicados e não publicados?	Sim	<i>MEDLINE, EMBASE, Scopus, PsycINFO e The Cochrane Library</i> foram pesquisadas (apesar de não estar claro quais recursos foram utilizados dentro da <i>Cochrane Library</i>). Este item foi julgado como adequado, pois uma variedade de bases de dados foi analisada. Registos de ensaios clínicos também foram pesquisados.
2.2 Além da busca nas bases de dados foram utilizados métodos adicionais para a identificação de artigos relevantes?	Não	Não
2.3 Os termos e a estrutura da estratégia de busca foram adequados para obter o maior número possível de artigos?	Nenhuma informação	A estratégia de busca completa não foi reportada e não há detalhes dos termos utilizados. Não há informações suficientes para o julgamento desta questão.
2.4 As restrições baseadas na data, formato de publicação ou idioma foram adequadas?	Não	A revisão foi restrita ao idioma inglês. Portanto, existe um potencial viés de publicação.
2.5 Foram realizadas tentativas para minimizar erros na seleção dos estudos?	Provavelmente sim	A avaliação dos estudos incluídos foi conduzida de forma independente por pelo menos dois revisores. No entanto, não ficou explícito se este método foi aplicado tanto na seleção dos estudos quanto na avaliação dos artigos na íntegra. Sendo assim, esta questão foi classificada como “Provavelmente sim”.
Gradação do potencial risco de viés em relação aos métodos utilizados para identificar e/ou selecionar os estudos.	Alto	A restrição de artigos no idioma inglês significa que há um alto risco de que estudos relevantes não tenham sido incluídos nesta revisão. Não foi possível julgar se a estratégia de busca estava adequada.

b. Revisão avaliada como de baixo potencial de viés em relação aos métodos utilizados para identificar e/ou selecionar estudos ⁴

Descrição do texto:

"As seguintes bases de dados foram pesquisadas até outubro de 2011: MEDLINE; EMBASE; *Cumulative Index To Nursing and Allied Health Literature (CINAHL)*; *The Cochrane Central Register of Systematic Reviews*; *The Cochrane Central Register of*

Controlled Trials; *DARE*; *MEDION*; *SIGLE*; *Index of Scientific and Technical Proceedings*, *Web of Science* e *ClinicalTrials.gov*. A pesquisa utilizou palavras-chave e termos MeSH relacionados aos testes avaliados combinados com os termos MeSH: ‘*Prenatal Diagnosis*’, ‘*Ultrasonography*’, ‘*Amniotic Fluid*’ e ‘*Pregnancy Outcome*’. A estratégia de busca completa está disponível no Apêndice S2. As referências de todos os artigos primários e de revisões incluídas foram analisadas para identificar artigos citados que não foram encontrados por pesquisas eletrônicas. Nenhuma restrição de idioma foi aplicada. As bases de dados foram examinadas por dois revisores (RKM, CHM). Foram realizadas traduções para artigos em idiomas diferentes do inglês. Três revisores tomaram as decisões finais de inclusão/exclusão de acordo com os seguintes critérios. "

Quadro 8. Exemplo de classificação para estratégia de busca julgada como de baixo potencial de viés

Questão	Classificação	Justificativa
2.1 A busca incluiu uma variedade suficiente de bases de dados/fontes eletrônicas para pesquisar artigos publicados e não publicados?	Sim	MEDLINE, EMBASE, CINAHL, SIGLE, Index of Scientific and Technical Proceedings, ClinicalTrials.gov, Cochrane Library, DARE e MEDION foram pesquisadas. Web of Science também foi pesquisada, apesar de não estar claro qual base de dados dentro da plataforma foi utilizada. Dessa forma, foi considerada apropriada a variedade de bases de dados para encontrar artigos publicados e não publicados.
2.2 Além da busca nas bases de dados foram utilizados métodos adicionais para a identificação de artigos relevantes?	Sim	As referências de todos os artigos primários e de revisões incluídos foram examinadas para identificar artigos elegíveis que não foram encontrados na pesquisa eletrônica.
2.3 Os termos e a estrutura da estratégia de busca foram adequados para obter o maior número possível de artigos?	Sim	Os detalhes da estratégia de busca encontram-se disponíveis no apêndice. Foi utilizada para a pesquisa a combinação de termos da população (mulheres grávidas) e da intervenção (ultrassonografia), o que pareceu sensível e sem restrições inapropriadas (por exemplo, delineamento de estudo).
2.4 As restrições baseadas na data, formato de publicação ou idioma foram adequadas?	Sim	Não houve restrições para data e idioma e as pesquisas incluíram etapas para identificar a literatura cinzenta. Por não ter incluído quaisquer restrições na pesquisa esse item foi julgado como “sim”.
2.5 Foram realizadas	Provavelmente sim	O processo de seleção por títulos e resumos, bem como a avaliação do texto na íntegra foi reportado e incluiu

tentativas para minimizar erros na seleção dos estudos?		múltiplos revisores. No entanto, não está claro se os revisores atuaram de forma independente e por isso a resposta foi “provavelmente sim”.
Gradação do potencial risco de viés em relação aos métodos utilizados para identificar e/ou selecionar os estudos.	Baixo	Todas as perguntas foram respondidas como "Sim" ou "Provavelmente Sim" e, portanto, não foram identificados potenciais vieses. Esta revisão parece ter incluído uma proporção alta de estudos relevantes.

Domínio 3: Coleta de dados e avaliação do estudo

O terceiro domínio tem como objetivo avaliar se o viés foi introduzido por meio da coleta de dados ou durante o processo de avaliação do potencial risco de viés. Por isso, a coleta rigorosa de dados deve seguir o delineamento do estudo, conforme definido no protocolo de pesquisa, e utilizar um formulário de coleta de dados que tenha sido estruturado. Todos os dados que contribuirão para a síntese e interpretação dos resultados devem ser coletados, incluindo dados numéricos e estatísticos, e características gerais do estudo primário, tais como características do delineamento de estudo, financiamento, cenário, características dos participantes, critério de seleção, detalhes da intervenção/exposição (diagnóstico) e perda de participantes. O tipo de dado numérico depende do delineamento dos estudos incluídos na revisão. Por exemplo, para uma revisão de acurácia de teste diagnóstico dados da tabela 2x2 para avaliar o teste são geralmente extraídos. Para a revisão de estudos observacionais pode ser extraída uma estimativa de efeito ajustada com um intervalo de confiança de 95%, bem como detalhes das variáveis ajustadas e método de análise. Se os dados não estiverem disponíveis no formato adequado para contribuir com a síntese, os autores da revisão devem relatar como esses dados foram obtidos. Por exemplo, os autores de estudos primários podem ser contatados para obter dados adicionais. De forma similar, alterações estatísticas adequadas podem ser usadas para obter os dados necessários. Além disso, o processo de extração de dados pode gerar erros que podem surgir de equívocos ao transcrever dados ou pela falta de coleta de informações relevantes que estão disponíveis no relato de estudo. Desse modo, o viés também pode surgir do processo de extração de dados que, pela sua natureza, é subjetivo e aberto à interpretação. A extração de dados pareada (ou extração de dados única com uma

verificação rigorosa) é, portanto, essencial para evitar a existência de erros randômicos e potencial viés.⁸

A validade dos estudos incluídos deve ser avaliada usando critérios apropriados de acordo com o delineamento dos estudos primários incluídos na revisão.^{5, 8} A avaliação pode ser realizada utilizando uma ferramenta estruturada, desenvolvida especificamente para o delineamento do estudo que está sendo avaliado, ou pode ser simplesmente uma lista de critérios relevantes que pode ser uma fonte potencial de viés. Se uma ferramenta publicada ou um critério *ad hoc* são usados, o avaliador deve considerar se os critérios são suficientes para identificar todas as fontes importantes e potenciais vieses nos estudos incluídos. A decisão quanto à validade dos critérios deve ser baseada em evidências empíricas relacionadas ao risco de viés de cada delineamento dos estudos⁹⁻¹⁴. Como acontece na extração de dados, viés ou erro pode ocorrer no processo de avaliação do risco de viés. A avaliação do risco de viés deve, portanto, envolver dois revisores independentes, ou no mínimo o segundo revisor deve verificar as decisões do primeiro. O consenso e as discrepâncias devem ser relatados na avaliação do risco de viés e extração de dados.

O Quadro 9 resume as questões norteadoras incluídas no Domínio 3 com a orientação de classificação para cada questão. O Quadro 10 fornece uma síntese do nível das classificações para o Domínio 3. Exemplos da classificação para este domínio (Potencial alto e baixo de viés) estão fornecidos nos Quadros 11 e 12.

Quadro 9. Questões norteadoras referentes ao Domínio 3 e orientações para respostas

Questões norteadoras	Orientações para respostas
3.1 Foram realizadas tentativas para minimizar o erro na coleta de dados?	Para minimizar viés e erros no processo de coleta de dados é necessário envolver pelo menos dois revisores, e utilizar formulários de extração de dados estruturados. É recomendado que este processo seja realizado por dois revisores de forma independente ou a extração pode ser realizada por um revisor e confirmada por um segundo. O processo de confirmação deve envolver o segundo revisor na leitura detalhada do documento, não só para verificar os dados extraídos, mas também para garantir que nenhuma informação relevante seja perdida.
3.2 As características disponíveis do estudo foram suficientes para que os autores da revisão e leitores sejam capazes de interpretar os resultados?	As informações sobre as características do estudo são essenciais para interpretar os resultados da revisão, pois permite uma investigação adequada da heterogeneidade e da relevância da aplicabilidade dos resultados. Essas informações podem ser relatadas na seção de métodos, onde as informações sobre a coleta de dados são descritas, podem estar disponíveis nos Quadros com as características dos estudos incluídos, ou podem estar sumarizadas nos resultados. Esta questão pode ser difícil de julgar, uma vez que todas as informações coletadas como parte da revisão nem sempre são apresentadas na publicação, muitas vezes por falta

	de espaço. Os avaliadores podem, portanto, precisar acessar recursos adicionais como apêndices da internet.
3.3 Todos os resultados relevantes foram coletados para uso na síntese?	<p>Resultados suficientes do estudo devem ser extraídos para permitir que seja conduzida uma síntese adequada. Idealmente, os autores da revisão devem relatar quais dados foram necessários para a síntese e em qual formato. Por exemplo, para uma revisão de teste de acurácia diagnóstica dados da tabela 2x2 são geralmente extraídos. Para uma revisão de estudos observacionais, pode ser extraída uma estimativa de efeito ajustada com um intervalo de confiança de 95%, detalhes das variáveis ajustadas e método de análise. Se esses dados não forem explicitamente relatados na seção de métodos, os dados dos Quadros, resumos gráficos ou dados relatados no texto podem ser consultados. Também pode ser necessário acessar recursos adicionais, como apêndices da internet ou o protocolo da revisão.</p> <p>É muito raro que todos os estudos primários incluídos na revisão apresentem os dados no formato adequado para a síntese. Por exemplo, uma revisão pode exigir dados contínuos no formato de diferença de média avaliada a partir da mudança do <i>baseline</i> com intervalo de confiança de 95%. No entanto, há uma variedade de outras medidas que podem ser relatadas ao invés da diferença de média, com estudos frequentemente não relatando medidas de variação. Para uma revisão responder "Sim" a esta pergunta, informações detalhadas devem ter sido incluídas na seção de métodos para descrever como os dados dos resultados que não foram reportados no formato adequado para a síntese foram obtidos, ex. por estimativas/transformações dos dados relatados ou por meio de contato com os autores para obter informações adicionais.</p>
3.4. O risco de viés (ou qualidade metodológica) foi formalmente avaliado usando critérios apropriados?	<p>A avaliação do risco de viés é um componente essencial de qualquer revisão. Se o risco de viés não foi formalmente avaliado, então esta questão deve ser respondida como "Não". Se uma avaliação formal foi realizada, os avaliadores terão que usar o seu senso crítico para verificar se os critérios utilizados foram adequados. Se uma ferramenta publicada e reconhecida foi utilizada para o delineamento adequado, como a ferramenta <i>Cochrane Risk of Bias</i>¹⁵ para ECR ou <i>QUADAS-2</i> para estudos de teste de acurácia diagnóstica¹⁶, então a avaliação estará adequada e esta pergunta poderá ser respondida como "Sim".</p> <p>No entanto, se a revisão simplesmente lista as questões avaliadas, utiliza uma ferramenta não publicada ou que não é mais recomendada, então a avaliação se torna mais difícil. O avaliador precisa então julgar se os critérios avaliados foram capazes de identificar potenciais fontes de viés nos estudos primários, dado o delineamento do estudo. Por exemplo, a escala de Jadad tem sido usada, mas não inclui o sigilo de alocação, o que tem sido considerado um potencial risco de viés em ECR.¹⁷ Para responder "Sim" a esta pergunta, as revisões que usaram a escala de Jadad¹⁷ também devem ter o sigilo de alocação avaliado.</p>
3.5. Foram realizadas tentativas para minimizar o erro na avaliação do risco de viés?	Assim como na coleta de dados, a avaliação do risco de viés deve envolver pelo menos dois revisores de forma independente ou pode ser realizada por um revisor e confirmada pelo segundo.

Quadro 1. Potencial de viés relacionado aos métodos utilizados para coletar dados e avaliar os estudos

Baixo	Dado os estudos incluídos na revisão conforme avaliados no domínio 2, o risco de viés foi avaliado usando critérios adequados, a extração de dados e a avaliação do risco de viés envolveu dois revisores. As características dos estudos e os resultados relevantes foram extraídos.
-------	---

Alto	Algum viés pode ter sido introduzido por meio da coleta de dados ou processo de avaliação do risco de viés.
Incerto	As informações reportadas foram insuficientes para avaliar o risco de viés.

Exemplos de classificação

a) Revisão classificada como de alto potencial de viés em relação aos métodos utilizados para coletar dados e avaliar estudos³.

Descrição do texto:

"Tabelas do tipo 2x2 foram construídas com os dados extraídos de cada estudo. Dois autores buscaram e extraíram os dados independentemente. Algumas diferenças foram resolvidas por consenso... Não há consenso sobre como avaliar estudos sem um grupo controle. Por isso, este critério não se aplica a estudos com a ausência deste grupo. Portanto, para esta meta-análise e revisão sistemática, os estudos foram selecionados com base nos dados completos e nos critérios de inclusão."

Comentário: As características do Quadro dos estudos incluídos contêm apenas detalhes sobre o delineamento do estudo (retrospectivo ou prospectivo), tipo de câncer (todos os tipos de câncer retal) e procedimento confirmatório.

Quadro 11. Exemplo de classificação para coleta de dados e avaliação do estudo julgado como de alto potencial de viés

Questões norteadoras	Classificação	Justificativa
3.1 Foram realizadas tentativas para minimizar o erro na coleta de dados?	Sim	A extração de dados foi realizada de forma independente por dois revisores. As diferenças foram resolvidas por consenso.
3.2. As características do estudo foram suficientes para que os autores e leitores da revisão fossem capazes de interpretar os resultados?	Não	O Quadro de síntese incluiu apenas detalhes ou delineamento (prospectivo ou retrospectivo), tipo de câncer (todos os tipos de câncer retal) e tipo de procedimento confirmatório (todas cirúrgicas). Nenhum outro detalhe sobre o delineamento do estudo ou a população foi relatado no texto.
3.3 Todos os resultados relevantes foram coletados para uso na síntese?	Sim	Dados 2x2 foram extraídos de cada estudo, que são suficientes para calcular todas as medidas de acurácia diagnóstica, intervalos de confiança e realizar uma meta-análise.
3.4 O risco de viés (ou qualidade metodológica) foi formalmente avaliado usando critérios apropriados?	Não	A qualidade do estudo não foi formalmente avaliada. Os autores afirmam que isso ocorre porque não há consenso sobre como avaliar estudos sem um grupo controle, mas há ferramentas aceitas para avaliar os estudos de

		teste de acurácia diagnóstica que poderiam ter sido usados.
3.5 Foram realizadas tentativas para minimizar o erro na avaliação do risco de viés?	Não	A qualidade do estudo não foi formalmente avaliada.
Gradação do potencial risco de viés em relação aos métodos utilizados para coletar dados e avaliar estudos.	Alto	A ausência de avaliação formal da qualidade da evidência significa que o risco de viés nos estudos incluídos não é claro. Os detalhes do estudo foram insuficientes para permitir que o leitor interprete os resultados. Há, portanto, um alto risco de viés na coleta de dados e no processo de avaliação dos estudos incluídos.

b. Revisão classificada como de baixo potencial de viés em relação aos métodos utilizados para coletar dados e avaliar os estudos ¹

Descrição do texto:

"Um pesquisador extraiu dados sobre delineamento do estudo, população, cenário, intervenções, análise, seguimento e resultados. Um segundo pesquisador verificou os dados extraídos. Dois pesquisadores aplicaram independentemente critérios pré-definidos para avaliar a qualidade do estudo como ótimo, razoável ou ruim. As discrepâncias foram resolvidas por consenso."

Quadro 12. Exemplo de classificação para coleta de dados e avaliação do estudo julgada como de baixo potencial de viés

Questões norteadoras	Classificação	Justificativa
3.1. Foram realizados esforços para minimizar o erro na coleta de dados?	Provavelmente sim	Um revisor realizou a coleta de dados e esta foi verificada por um segundo revisor. Não está claro se o segundo revisor apenas verificou se os dados extraídos eram precisos ou se também leu o documento em detalhes para garantir que nenhuma informação relevante fosse perdida. Esta questão foi, portanto, respondida como "Provavelmente sim".
3.2. As características do estudo foram suficientes para que os autores e leitores da revisão fossem capazes de interpretar os resultados?	Sim	As características detalhadas do estudo e os Quadros de resultados foram fornecidas em apêndices que relataram informações suficientes para autores e leitores interpretarem os resultados.
3.3. Todos os resultados relevantes foram coletados para uso na síntese?	Sim	Os dados dicotômicos foram extraídos, como número de eventos e número total de pacientes em cada grupo de tratamento; estes foram usados para calcular risco relativo. Esta estratégia foi apropriada para ser utilizada na síntese.
3.4. O risco de viés (ou qualidade metodológica) foi formalmente avaliado usando critérios apropriados?	Provavelmente sim	Detalhes sobre os critérios utilizados para avaliar o risco de viés não foram fornecidos, mas os autores citam o método <i>US Preventive Services Task Force</i> (2001), que inclui alguns critérios adequados.

3.5 Foram realizadas tentativas para minimizar o erro na avaliação do risco de viés?	Sim	Dois investigadores avaliaram de forma independente a qualidade dos estudos.
Gradação do potencial risco de viés em relação aos métodos utilizados para coletar dados e avaliar os estudos	Baixo	Todas as questões foram classificadas como “Sim” ou “Provavelmente Sim” e, portanto, não foram identificados potenciais vieses. Os processos de revisão da coleta dos dados e avaliação dos estudos são improváveis de terem introduzido viés nesta revisão.

Domínio 4: Síntese e resultados

Este domínio tem como objetivo avaliar se, tendo sido tomada a decisão de combinar dados de estudos primários incluídos (na síntese qualitativa ou quantitativa), os revisores usaram métodos adequados. As abordagens para a síntese dependem da natureza da questão que está sendo discutida e da natureza dos estudos primários que estão sendo sintetizados. Para os ECR, uma abordagem comum é usar uma média ponderada das estimativas de efeito do tratamento (na escala logarítmica), ponderando pela precisão das estimativas.¹⁸ Pode ser adotado para isso, modelos de efeitos fixos ou de efeitos randômicos. No entanto, existem muitas variações com opções de modelagem de dados; por exemplo, usando uma abordagem de regressão logística para dados binários,¹⁹ modelando dois ou mais desfechos simultaneamente (meta-análise bivariada ou multivariada),²⁰ modelagem de múltiplos efeitos de tratamento simultaneamente (meta-análise de rede²¹), ou modelagem de variação nos efeitos do tratamento (meta-regressão²²). Estas variações podem ser combinadas, o que torna a síntese muito complexa. Opções semelhantes estão disponíveis para outros tipos de perguntas de revisão. Para a acurácia do teste diagnóstico, a abordagem bivariada tornou-se padrão, na qual a sensibilidade e especificidade são simultaneamente modeladas para levar em consideração sua correlação.²³ Para algumas revisões, uma síntese estatística pode não ser apropriada e, em vez disso, uma *overview* qualitativa ou narrativa de resultados deve ser relatada.

Alguns dos aspectos mais importantes a considerar em qualquer síntese (quantitativa ou qualitativa) são (i) se a abordagem analítica é adequada para a questão de pesquisa apresentada; (ii) se a variação entre os estudos (heterogeneidade) é levada em consideração; (iii) se os vieses nos estudos primários são levados em consideração; (iv) se as informações dos estudos primários que estão sendo sintetizadas estão completas (especialmente se houver o risco de que os dados ausentes sejam sistematicamente

diferentes dos dados disponíveis, por exemplo, devido a viés de publicação ou divulgação); e (v) se os revisores introduziram viés no modo como relatam os resultados. Os aspectos técnicos da meta-análise, como a escolha do método de estimativa, geralmente não são observações importantes. No entanto, a presença de erros pode ser relevante, por exemplo, ao interpretar erro padrão como desvio padrão, ao não ajustar o modelo por detalhes do delineamento tais como *matched data* ou *clustered data*, ou ao aplicar a média ponderada padrão às razões de risco ao invés de aplicar a logaritmos.

O Quadro 13 resume as questões norteadoras incluídas no Domínio 4 junto com a orientação de classificação para cada questão. O Quadro 14 fornece uma visão geral do nível de classificação para o Domínio 4. Exemplos de classificações para este domínio (Potencial alto e baixo de viés) estão fornecidos nos Quadros 15 e 16.

Quadro 13. Questões norteadoras referentes ao Domínio 4 e orientações para respostas

Questões norteadoras	Orientações para respostas
4.1 A síntese de resultados incluiu todos os estudos que deveriam ser incluídos?	<p>A síntese geralmente deve incluir todos os estudos que coletaram dados relevantes para a pergunta de pesquisa. Resultados de estudos individuais podem estar ausentes na síntese porque o estudo é desconhecido para os revisores (possivelmente devido ao viés de publicação): estes estudos ausentes são abordados dentro do Domínio 2 e pela questão 4.5. Esta questão aborda três outras situações relacionadas com os estudos que foram identificados para inclusão na revisão: (i) os resultados específicos do estudo incluído não estão disponíveis para os revisores, (ii) os revisores não conseguiram coletar ou processar os dados disponíveis, ou (iii) os revisores excluíram propositalmente os resultados. No caso (i), pode estar claro que os estudos incluídos coletaram dados, mas não relataram os resultados desejados. Por exemplo, um ensaio clínico pode coletar dados para um desfecho de interesse dos revisores, mas não reportar os resultados. A importante distinção aqui é novamente se o resultado teria sido suprimido por causa dos achados da pesquisa (por exemplo, significância estatística) ou simplesmente por não ter sido reportado (por exemplo, por razões práticas, tais como limite de páginas do estudo). Se os resultados não estiverem disponíveis em um ou mais estudos porque os achados não foram estatisticamente significantes, a omissão destes dados introduzirá viés na síntese de resultados do estudo.</p> <p>No caso (ii) [ver também o Domínio 3, questão 3.3], estudos podem ter sido omitidos por erro, ou porque os revisores não estavam familiarizados com cálculos estatísticos que permitiriam a sua inclusão. Isto seria problemático se os estudos omitidos tivessem resultados sistematicamente diferentes dos que foram incluídos. Quanto ao caso (iii), os revisores podem tomar decisões inadequadas para excluir alguns estudos da síntese. Um exemplo seria a exclusão de estudos que podem levar a uma grande heterogeneidade dos resultados baseada apenas em considerações estatísticas (embora isso possa ser razoável quando realizado como análise de sensibilidade).</p>

	<p>Para abordar estas considerações, recomendamos que o avaliador examine o número de estudos incluídos (a partir de um fluxograma) e o número de estudos sintetizados (por meio de <i>forest plot</i> ou Quadros). Em revisões com grande número de estudos (por exemplo, mais de 10 ou 20 em uma síntese), <i>funnel plots</i> e análises estatísticas podem fornecer informações sobre a possibilidade de presença de vieses de seleção pelos motivos reportados acima: ver questão 4.5.¹³</p>
<p>4.2 Todas as análises pré-definidas foram seguidas ou as perdas de participantes foram explicadas?</p>	<p>O objetivo desta questão é identificar os vieses introduzidos pelos autores por meio da seleção e dos métodos de análise, de forma que os resultados que não sejam de interesse para eles sejam suprimidos ou substituídos. Para responder "Sim", a revisão deve ter seguido um protocolo publicado ou acessível. Exemplos incluem o registro do protocolo no PROSPERO, um registro prospectivo internacional de revisões sistemáticas (http://www.crd.york.ac.uk/PROSPERO/), ou a publicação do protocolo de uma revisão Cochrane (http://www.thecochranelibrary.com/). Se há indicação que análises pré-definidas foram seguidas, por exemplo, a seção de métodos parece ser rigorosa e todas as análises mencionadas são consideradas nos resultados, então o avaliador pode responder "Provavelmente sim". Na ausência de um protocolo pré-especificado, o avaliador deve responder "Não" a esta questão. No entanto, se a revisão sistemática não faz referência à existência ou ausência de um protocolo, recomendamos que o avaliador responda a esta questão como "sem informação".</p>
<p>4.3 A síntese dos resultados foi apropriada dada a natureza e à similaridade das questões de pesquisa, dos delineamentos de estudos e desfechos dos estudos incluídos?</p>	<p>Esta questão aborda principalmente a metodologia da síntese. A metodologia deve ser orientada pela natureza dos estudos e da questão apresentada. Se uma síntese quantitativa é realizada (meta-análise), esta pergunta aborda se foi adequado realizá-la e os métodos estatísticos utilizados. Se uma abordagem narrativa é usada para a síntese, esta questão trata se uma síntese quantitativa teria sido mais adequada e se a abordagem narrativa foi apropriada.</p> <p>A decisão quanto à adequação de uma síntese quantitativa é movida principalmente pela similaridade dos estudos nas questões de pesquisa, delineamentos e desfechos (quando medidos). Isto é frequentemente atribuído como "heterogeneidade clínica" ou "diversidade clínica". É importante que o avaliador possa estar convencido de que o resultado da síntese tem significado - isto é, pode potencialmente informar políticas, práticas ou pesquisas futuras. A questão da heterogeneidade estatística é abordada na questão 4.4.</p> <p>O avaliador deve responder "Sim" a esta questão se é usado um método bem aceito de combinação estatística de estudos, incluindo a ponderação adequada de cada estudo. Por exemplo, na presença de variação entre estudos, mas onde se pode assumir que os efeitos do tratamento vieram de uma mesma distribuição, o modelo de efeito randômico e não o de efeito fixo poderá ser utilizado. No entanto note que as meta-análises de efeito randômico simples não são adequadas quando existe uma forte relação entre tamanho do estudo e tamanho do efeito ('efeito de estudo de pequenas amostras' ou 'assimetria <i>funnel plot</i>').</p> <p>Abordagens de regressão ou subgrupos podem ser consideradas quando há estudos suficientes para explorar os motivos de variação entre estudos. Geralmente não recomendamos métodos de combinação de estudos que ignorem a influência da precisão do estudo na estimativa global agrupada – ex. aqueles que ponderam todos os estudos igualmente. Se tais métodos forem usados, sugerimos que o avaliador responda "Não" a esta questão.</p> <p>Para estudos de teste de acurácia diagnóstica, um método geralmente aceito é uma abordagem bivariada para explicar a correlação entre sensibilidade e especificidade. O aconselhamento estatístico pode ser</p>

	necessário para meta-análises usando abordagens multivariada, multinível, bayesiana ou meta-análise de rede.
4.4 A variação entre os estudos (heterogeneidade) foi baixa ou abordada na síntese?	<p>Esta questão aborda a variação nos resultados dos estudos ao invés da variação nas suas características. A variação entre os estudos pode ser avaliada visualmente (por <i>forest plot</i>); usando teste estatístico (por exemplo, teste χ^2 ou Q de Cochran); usando uma medida de variância entre estudos (frequentemente referida como τ^2); ou usando uma medida de variação de intervalo de confiança (por exemplo, a estatística de inconsistência I^2).</p> <p>Se uma heterogeneidade considerável for ignorada na meta-análise, ela pode levar a conclusões erradas e/ou a uma falsa precisão. Se uma análise de efeito fixo for usada na presença de heterogeneidade é importante que os revisores reconheçam que a análise ignora a heterogeneidade. O resultado deveria ser complementado com resultados relativos à variação entre os estudos, e os resultados da análise de efeito fixo deveriam ser interpretados adequadamente.</p> <p>Na maioria dos casos, o uso de uma meta-análise de efeito fixo sem análise adicional para explorar a variação receberia uma resposta "Não" para esta questão.</p> <p>Se um modelo de efeito randômico tiver sido utilizado adequadamente para permitir a heterogeneidade e/ou outras análises de subgrupo/meta-regressão, o avaliador poderá responder "Sim". No entanto, a meta-análise de efeito randômico de estudos com características ou resultados extremamente diversos pode produzir resultado sem sentido, particularmente se os resultados dos estudos apontam em direções opostas de efeito.</p> <p>Se uma síntese narrativa foi conduzida com base em uma combinação estatística inadequada devido à heterogeneidade clínica, esta pergunta deve ser respondida "Sim".</p>
4.5 Os resultados foram robustos? Como demonstrado, por exemplo, pelo <i>funnel plot</i> ou análise de sensibilidade.	<p>Esta questão aborda a imprecisão dos resultados da síntese, e se eles mudariam caso uma abordagem diferente fosse aplicada. Os gráficos <i>funnel plot</i> são usados para verificar a relação entre o tamanho do efeito e o tamanho do estudo (frequentemente mensurado estatisticamente usando precisão). A assimetria no <i>funnel plot</i> pode ocorrer por várias razões, incluindo viés de publicação, viés relatado ou não relatado dentro de estudos, diferentes magnitudes de viés em estudos maiores versus estudos menores, ou diferenças substanciais nos efeitos dos estudos de tamanhos diferentes. Quando um gráfico <i>funnel plot</i> é assimétrico os métodos de efeito fixo e efeito aleatório produzem resultados diferentes. Geralmente podem-se empregar ambos os métodos como uma forma de análise de sensibilidade. Alternativamente, o método <i>trim-and-fill</i>, abordagens de regressão ou uso de modelos de seleção podem ser usados para examinar o impacto potencial dos diferentes tipos de viés de publicação sobre os resultados.²⁴</p> <p>Outros exemplos de análises de sensibilidade são o uso de diferentes métodos para imputação de dados ou exclusão de estudos. Alternativamente, uma análise de sensibilidade pode ser realizada para verificar se os resultados são robustos à remoção de alguns estudos, ou de estudos que foram classificados como de alto risco de viés.</p> <p>O uso da análise de sensibilidade por si só não é considerado adequado para que esta pergunta seja respondida como "Sim"; os avaliadores devem decidir se o seu uso demonstrou robustez nos resultados. Se houver poucos estudos ou estudos muito heterogêneos, pode ser claro que os resultados não são robustos, mesmo se os revisores não realizarem análises de sensibilidade.</p> <p>Se uma síntese narrativa foi conduzida, o avaliador deve considerar se diferentes abordagens para resumir os estudos poderiam ter levado a conclusões diferentes, e se os estudos isolados estão levando as</p>

	conclusões.
4.6 Os vieses nos estudos primários foram mínimos ou abordados na síntese?	Os avaliadores devem responder a esta pergunta como "Não" se julgarem haver um viés considerável nos estudos selecionados que tenham sido ignorados pelos autores. Por exemplo, se o risco de viés não foi avaliado na revisão sistemática os usuários da ROBIS devem responder "Não". Da mesma forma, se o risco de viés foi avaliado, mas os revisores não o descreveram nos resultados/conclusões, deve ser atribuída uma resposta "Não" a esta pergunta. Exemplos de quando esta pergunta poderia ser respondida como "Sim" são os casos em que todos os estudos receberam uma classificação de "baixo risco de viés" ou se análises de sensibilidade foram empregadas nos estudos que estavam com alto risco de viés. Se os vieses forem abordados apenas como parte da discussão dos resultados da síntese, o avaliador deve responder "Não"; estas discussões serão abordadas mais adiante no julgamento geral sobre o risco de viés na revisão (Fase 3).

Quadro 14. Potencial de viés relacionado aos métodos utilizados para sintetizar resultados

Baixo	É improvável que a síntese produza resultados enviesados, pois qualquer limitação nos dados já foi superada, ou os resultados foram tão convincentes que as limitações teriam pouco impacto.
Alto	A síntese é susceptível a produzir resultados enviesados, porque (i) os potenciais vieses foram ignorados, (ii) uma importante variabilidade entre os estudos não foi considerada; (iii) houve inadequações na metodologia; ou (iv) os resultados são reportados de forma incompleta a ponto de levantar uma possibilidade de viés.
Incerto	As informações reportadas são insuficientes para avaliação do risco de viés.

Exemplos de classificação

a. Revisão classificada como de alto potencial de viés em relação à síntese²⁵

Quadro 15. Exemplo de classificação para síntese julgada como de alto potencial de viés

Questões norteadoras	Classificação	Justificativa
4.1 A síntese de resultados incluiu todos os estudos que deveriam ser incluídos?	Sem informação	O autor especificou que 35 estudos foram relevantes para a revisão, dos quais 29 foram utilizados em uma meta-análise. Nenhuma outra informação foi fornecida para verificar a adequação dos estudos incluídos.
4.2 Todas as análises pré-definidas foram seguidas ou as perdas de participantes foram explicadas?	Sem informação	Nenhuma análise foi explicitamente predefinida em um protocolo. Não foram fornecidas informações adicionais no texto.
4.3 A síntese dos resultados foi apropriada quando consideradas a natureza e a similaridade das questões de pesquisa, os delineamentos e os desfechos dos estudos incluídos?	Não	O autor relatou apenas que uma "meta-análise foi realizada" (p.11). O autor agrupou os eventos e as populações e usou os resultados para gerar a medida <i>odds ratio</i> , portanto não explorou as comparações entre os estudos.
4.4 A variabilidade entre os	Provavelmente	Nenhum detalhe foi fornecido sobre a heterogeneidade

estudos (heterogeneidade) foi baixa ou abordada na síntese?	não	estatística. No entanto, o autor discutiu análise de subgrupo para estudos com alto risco de heterogeneidade. Além disso, os estudos foram agrupados de acordo com a prevalência de circuncisão na comunidade. Nenhum outro detalhe foi dado sobre os subgrupos.
4.5 Os resultados foram robustos? Como demonstrado, por exemplo, pelo <i>funnel plot</i> ou análise de sensibilidade.	Provavelmente não	O autor discutiu alguns estudos que podem ser problemáticos. "Embora estas hipóteses possam estar erradas, a inclusão ou exclusão desses estudos não afetou o resultado global" (P.11). O <i>odds ratios</i> e o intervalo de confiança para essas análises não foram reportados.
4.6 Os vieses nos estudos primários foram mínimos ou abordados na síntese?	Não	Nem a qualidade nem o risco de viés dos estudos foram explicitamente avaliados. Ademais, o viés não foi considerado na síntese.
Gradação do potencial risco de viés em relação à síntese e resultados	Alto	Os estudos individuais não foram avaliados quanto ao risco de viés, nem o viés potencial foi considerado na síntese. O autor agrupou as intervenções entre os estudos ao invés de analisar as comparações dentro dos estudos, o que é uma abordagem arriscada para a meta-análise. Não houve discussão ou avaliação da heterogeneidade na análise.

b. Revisão classificada como de baixo potencial de viés relacionada à síntese de resultados.⁴

Quadro 16. Exemplo de classificação para síntese de resultados julgada como de baixo potencial de viés

Questões norteadoras	Classificação	Justificativa
4.1 A síntese de resultados incluiu todos os estudos que deveriam ser incluídos?	Sem informação	O fluxograma de seleção reportou que 43 estudos foram incluídos na revisão. A figura com a classificação do QUADAS-2 e um apêndice com as características dos estudos também mostram um total de 43 estudos. No entanto, os <i>forest plot</i> não apresentaram achados por estudo, uma vez que estudos podem reportar mais de um desfecho. Assim, não foi possível avaliar se todos os estudos foram incluídos nesta síntese.
4.2 Todas as análises pré-definidas foram seguidas ou as perdas de participantes foram explicadas?	Provavelmente sim	O protocolo da revisão foi disponibilizado como informação suplementar. Nenhum desvio foi reportado ou observado. Embora o protocolo não tenha especificado detalhes sobre a avaliação do risco de viés nem sobre a estratégia para análise, a seção de métodos considerou rigorosamente ambos os itens. Todas as análises da seção de métodos são consideradas nos resultados.
4.3 A síntese dos resultados foi apropriada quando consideradas a natureza e a similaridade das questões de pesquisa, os delineamentos e os desfechos dos estudos incluídos?	Sim	Os autores afirmam que “os resultados foram combinados por meio do modelo de meta-análise de efeitos randômicos, onde as medidas do volume de fluido amniótico, do ponto de corte e do desfecho utilizado foram as mesmas” (página 688). Para testar a habilidade preditiva dos testes, os autores calcularam sensibilidade, especificidade e razão de probabilidade a partir de tabelas 2x2 e sintetizaram os resultados

		usando o modelo bivariado de meta-análise de efeitos randômicos.
4.4 A variabilidade entre os estudos (heterogeneidade) foi baixa ou abordada na síntese?	Sim	A heterogeneidade estatística foi antecipada nesta revisão e abordada usando modelo de efeitos randômicos. Meta-regressão ou análise de subgrupo foram realizadas para examinar os potenciais modificadores de efeito. Os intervalos preditivos também foram utilizados quando $I^2 > 0\%$.
4.5 Os resultados foram robustos? Como demonstrado, por exemplo, pelo <i>funnel plot</i> ou análise de sensibilidade.	Provavelmente não	Os autores não declararam se as análises de sensibilidade foram utilizadas para avaliar a robustez dos resultados.
4.6 Os vieses nos estudos primários foram mínimos ou abordados na síntese?	Provavelmente sim	Os vieses foram avaliados utilizando a ferramenta QUADAS-2. Esta avaliação indicou que a descrição da referência padrão e o cegamento estavam sob alto risco de viés. Outros aspectos do QUADAS-2 não indicaram grandes problemas quanto ao risco de viés. A "Qualidade" foi relatada como uma análise de subgrupo, no entanto foi reportada apenas para um dos dois desfechos primários. Para o desfecho oligodrâmnios e bem-estar fetal, os autores declararam que "Não houve diferença na análise de subgrupo" (P.691).
Gradação do potencial risco de viés em relação à síntese e resultados	Baixa	Os autores abordaram satisfatoriamente a heterogeneidade em suas análises e a exploraram usando análises de subgrupo. O risco de viés dos estudos individuais foi abordado e incluído na análise de subgrupo.

Fase 3: Avaliação do risco de viés geral

A fase final da ferramenta avalia se a revisão sistemática como um todo está sob o risco de viés. Essa avaliação utiliza a mesma estrutura do Domínio 2 uma vez que inclui questões norteadoras e justificativas; no entanto, o julgamento sobre potenciais riscos de viés foi substituído por uma avaliação geral do risco de viés da revisão. A primeira questão desta fase pergunta se a interpretação dos achados aborda todos os potenciais riscos identificados entre os domínios 1 e 4. Se não foi identificado nenhum risco potencial, a questão pode ser respondida como "Sim". Se um ou mais riscos potenciais forem identificados em quaisquer dos domínios anteriores, mas forem considerados na interpretação dos resultados e nas conclusões do artigo, esta questão pode também ser respondida como "Sim". A depender da classificação das outras questões desta fase, a revisão poderá ser classificada como "Baixo risco de viés".

Esta fase ainda inclui três questões relativas à interpretação dos achados da revisão. Essas questões abordam se as conclusões foram baseadas nas evidências apresentadas, se a relevância dos estudos incluídos foi considerada e se os autores evitaram enfatizar os resultados com base na significância estatística. Tais aspectos são importantes para a

interpretação dos achados de uma revisão, pois são onde vieses podem ter sido introduzidos no estudo.

O Quadro 17 sumariza as questões norteadoras do Domínio 4 e descreve orientações para resposta de cada questão. O Quadro 18 indica uma visão geral dos níveis de classificação para o Domínio 4.

Quadro 17. Questões norteadoras referentes à fase 3 e orientações para respostas

Questões norteadoras	Orientações para resposta
A. A interpretação dos achados considerou todos os riscos potenciais identificados na Fase 2?	Esta questão se refere à avaliação de risco potencial de viés realizada na Fase 2 e resumida no início da Fase 3. Se todos os domínios avaliados na Fase 2 foram classificados como de “baixo risco potencial de viés”, esta questão pode ser respondida como “Sim”. Se um ou mais domínios foram classificados como “alto risco potencial de viés” ou “risco potencial de viés incerto”, os revisores devem ponderar se durante a interpretação dos resultados os autores consideraram os riscos potenciais de viés identificados na Fase 2. Por exemplo, vamos supor que uma revisão tenha sido julgada como de “alto risco de viés” por ter aplicado um filtro de delineamento de estudos durante as buscas na literatura. Nesse caso, o filtro pode ter ocasionado perda de estudos relevantes. Se a revisão tiver apontado a utilização do filtro como uma potencial limitação do estudo e tiver discutido como isso pode ter impactado os achados da revisão, essa questão pode ser respondida como “Sim”.
B. A relevância dos estudos para a pergunta de pesquisa da revisão foi adequadamente considerada?	Um aspecto importante na interpretação dos achados de uma revisão é considerar a relevância (aplicabilidade/validade externa) dos estudos incluídos para a pergunta de pesquisa da revisão. Se os estudos incluídos não são diretamente aplicáveis à pergunta da revisão e se esse ponto não for abordado na interpretação dos resultados, algum viés pode ter sido introduzido na revisão sistemática. Algumas revisões podem considerar a relevância dos estudos incluídos como parte formal da avaliação de qualidade. Isto é particularmente comum em revisões de acurácia diagnóstica que usaram a ferramenta QUADAS-2 ²⁶ a qual inclui tanto uma avaliação formal da aplicabilidade dos estudos primários à pergunta da pesquisa quanto uma avaliação do risco de viés. Outras revisões podem usar uma abordagem menos formal, discutindo a relevância dos estudos na seção de discussão da revisão. Ambas abordagens podem ser apropriadas desde que as conclusões da revisão reflitam a evidência, inclusive a relevância dos estudos incluídos.
C. Os autores se preocuparam em não enfatizar os resultados com base apenas na sua significância estatística?	Se múltiplas análises forem utilizadas em apenas uma revisão, é importante que sejam reportadas de forma equilibrada pelos autores. Às vezes, os autores da revisão destacam os resultados com base na significância estatística. Por exemplo, se uma revisão avaliou dez resultados e não encontrou efeito em nove deles, mas encontrou efeito benéfico significativo em um, pode ser tentador querer supervalorizar esse resultado. Porém, isso geraria uma visão enganosa dos resultados verdadeiros da revisão; portanto, esta prática deve ser evitada.

Quadro 18. Riscos de viés relacionados aos métodos utilizados na identificação e/ou seleção dos estudos

Baixo	É provável que os resultados da revisão sejam confiáveis. Nenhum risco potencial de viés foi identificado na Fase 2 durante o processo de revisão ou, se foi identificado, este foi considerado nas conclusões da revisão. As conclusões foram baseadas na evidência e consideraram a relevância dos estudos incluídos.
Alto	Os potenciais riscos de viés identificados na Fase 2 não foram considerados nas conclusões da revisão, as conclusões ou não foram confirmadas pela evidência ou não consideraram a relevância dos estudos incluídos para a pergunta da pesquisa.
Incerto	As informações reportadas foram insuficientes para avaliar o risco de viés.

Exemplos de classificação

- a. Revisão classificada como de alto potencial de viés.³

Quadro 19. Quadro resumo dos potenciais vieses identificados durante a Fase 2

Domínio	Gradação do potencial risco de viés	Justificativa
1. Limitações relacionadas à especificação dos critérios de elegibilidade	Alto	Os critérios de elegibilidade eram ambíguos e pouco claros, não foram reportados detalhes sobre a população elegível.
2. Limitações relacionadas aos métodos usados para identificar e selecionar estudos	Alto	É provável que estudos relevantes tenham sido perdidos durante as buscas. Embora poucas informações tenham sido fornecidas, a variedade de bases de dados pesquisada parece não ter sido suficiente, nem a estratégia de busca incluiu termos para delineamentos de estudos de diagnóstico. Ademais, métodos de busca adicionais à pesquisa em bases de dados não foram reportados e a restrição das buscas a partir de 1980 não foi justificada.
3. Limitações relacionadas aos métodos usados para coletar dados e avaliar estudos	Alto	A falta de avaliação da qualidade dos artigos significa que o risco de viés dos estudos incluídos não é claro. Não houve detalhamento suficiente sobre os estudos para permitir a interpretação dos resultados. Portanto, existe um alto risco de viés tanto na coleta de dados quanto no processo de avaliação dos estudos.
4. Limitações relacionadas à síntese de resultados	Alto	Embora os métodos de análise utilizados nesta revisão não tenham sido os mais robustos, eles parecem ser confiáveis. A maior preocupação nesta revisão se refere à falta de avaliação da heterogeneidade. Como foram fornecidos poucos detalhes sobre os estudos e a avaliação formal do risco de viés não foi realizada, não foi possível avaliar se os estudos diferiram uns dos outros. Por exemplo, se um estudo for mais confiável do que outros incluídos na revisão, os resultados, como um todo, estarão com viés.

Descrição do texto:

“Esta meta-análise e revisão sistemática indicaram que a sensibilidade e a especificidade agrupadas do ultrassom endoscópico para invasão do tumor T0 é muito alta (aproximadamente 97%)... A heterogeneidade entre os estudos foi avaliada não só por um teste de heterogeneidade, mas também por sorteios das curvas SROC para encontrar a área sob a curva (AUC) uma vez que diferentes estudos podem usar diferentes critérios. Uma AUC de 1 para qualquer teste diagnóstico indica que o teste é excelente. Curvas SROC para o ultrassom endoscópico mostraram que o valor de AUC foi muito próximo a 1, indicando que o ultrassom endoscópico é um teste diagnóstico excelente para o estágio T0 do câncer retal... O ultrassom endoscópico tem sensibilidade e especificidade excelentes, o que ajuda a diagnosticar o estágio T0 do câncer retal de forma precisa...”

Quadro 20. Exemplo de classificação para síntese de resultados julgada como de alto potencial de viés

Questões norteadoras	Classificação	Justificativa
A. A interpretação dos achados considerou todos os riscos potenciais identificados na Fase 2?	Não	Nenhuma das limitações identificadas na Fase 2 foram consideradas na interpretação dos resultados pelos autores da revisão.
B. A relevância dos estudos foi apropriadamente considerada para a pergunta de pesquisa da revisão?	Não	A revisão não avaliou se os estudos incluídos eram relevantes para a pergunta de pesquisa. Ademais, havia informações insuficientes para permitir que os autores realizassem este julgamento.
C. Os autores se preocuparam em não enfatizar os resultados com base apenas na sua significância estatística?	Sim	As ênfases dadas na discussão e na conclusão foram baseadas em todas as estimativas das análises primárias dos estudos.
Risco de viés	Alto	A Fase 2 identificou algumas áreas potenciais de risco de viés no processo da revisão que não foram consideradas pelos autores. Tais áreas incluem critérios de elegibilidade pouco claros, possibilidade de perda de estudos, falta de avaliação formal da qualidade, detalhamento insuficiente dos estudos incluídos e falhas ao considerar heterogeneidade entre os estudos na síntese de resultados. A revisão também não considerou a relevância dos estudos para a pergunta de pesquisa a ser respondida. Portanto, esta revisão apresenta um alto risco de viés.

b. Revisão classificada como de baixo potencial de viés.

Quadro 21. Quadro resumo dos potenciais de vieses identificados durante a Fase 2

Domínio	Classificação	Justificativa
1. Limitações relacionadas à especificação dos critérios de	Baixo	Todas as questões foram respondidas como “Sim” ou “Provavelmente sim”, então não foram

elegibilidade		identificados riscos de potenciais vieses na especificação dos critérios de elegibilidade. A pergunta e os objetivos da revisão foram bem especificados, bem como o detalhamento e a justificativa dos critérios de elegibilidade.
2. Limitações relacionadas aos métodos usados para identificar e selecionar estudos	Baixo	Tanto o processo de leitura de títulos e resumos quanto o de textos na íntegra foram reportados e incluíram vários revisores. No entanto, não ficou claro se os revisores atuaram independentemente, assim este item foi classificado como “Provavelmente sim”.
3. Limitações relacionadas aos métodos usados para coletar dados e avaliar estudos	Baixo	Tanto a avaliação de todos os artigos quanto a coleta de dados relevantes foram realizadas de forma independente por no mínimo 2 revisores. A qualidade dos estudos foi avaliada utilizando-se ferramenta apropriada.
4. Limitações relacionadas à síntese de resultados	Baixo	Os autores avaliaram a presença de heterogeneidade e realizaram análises de subgrupo.

Descrição do texto:

Conclusões

“Evidência atualizada indica que o oligohidrânio está fortemente associado com a mortalidade e com o desenvolvimento fetal inferior ao esperado para a idade gestacional; enquanto que o polihidrânio está associado ao peso ao nascer superior ao 90º percentil. Apesar de suas fortes associações com desfechos não favoráveis, eles não predizem desfechos de risco”.

Discussão

“Nossa avaliação de qualidade revelou potenciais riscos de viés relacionados à seleção de pacientes, à realização do teste *index* e ao método padrão de referência. Não foi possível realizar análise de subgrupo para gravidez pré-termo *versus* gravidez a termo e alguns estudos reportaram dados insuficientes para determinar se limites para mensuração do líquido amniótico foram ajustados para a gestação. Sempre que possível, utilizamos os resultados mais próximos ao momento do parto e realizamos análise de subgrupo para os casos em que o teste foi feito em até sete dias após o parto. Houve pouquíssimas informações sobre os métodos exatos de padrão de referência e se algum tratamento foi realizado entre o teste *index* e o padrão. Portanto, uma grande preocupação se refere ao número de gestações com indução do parto por causa da presença de oligohidrânios, pois isto influencia os resultados dos desfechos da gestação, por exemplo: viés de intervenção. Esse viés só pode ser totalmente removido

por meio da realização de um ensaio clínico randomizado, o que seria impossível neste caso, pois mensurações do volume do líquido amniótico se tornaram padrão no monitoramento fetal e no manejo de gestações de alto risco, portanto o recrutamento para tal ensaio clínico seria muito difícil. Finalmente, as medidas de desfecho utilizadas nesta revisão foram as reportadas pelos autores dos estudos incluídos, sendo muitas delas subjetivas (admissão à unidade neonatal de cuidados intensivos e necessidade de ressuscitação). A única medida objetiva de desfecho não favorável são as amostras pareadas do pH do cordão umbilical; desfechos de longo prazo, como paralisia cerebral não foram reportados. ”

Quadro 22. Exemplo de classificação para síntese de resultados julgada como de baixo potencial de viés

Questões norteadoras	Classificação	Justificativa
A. A interpretação dos achados considerou todos os riscos potenciais identificados na Fase 2?	Sim	Não foram identificados potenciais riscos de viés durante a Fase 2.
B. A relevância dos estudos para a pergunta de pesquisa da revisão foi adequadamente considerada?	Provavelmente sim	As implicações dos achados da revisão foram discutidas em detalhe. As fontes potenciais de viés em termos de populações e desfechos nos estudos incluídos também foram detalhadamente discutidas, fato este que também influencia na avaliação de relevância.
C. Os autores se preocuparam em não enfatizar os resultados com base apenas na sua significância estatística?	Sim	As conclusões da revisão refletem tanto os achados significantes quanto os não significantes.
Risco de viés	Baixo	A avaliação da Fase 2 não identificou potenciais riscos de viés no processo de revisão. As potenciais limitações dos estudos incluídos foram discutidas em detalhe na seção de discussão. As conclusões refletem apropriadamente os resultados da revisão.

Considerações práticas ao utilizar a ROBIS

Nem sempre é fácil avaliar o risco de viés em revisões sistemáticas. A ferramenta ROBIS fornece uma abordagem estruturada para este fim. Embora a ROBIS tenha sido criada para ser usada por revisores de diferentes formações, é necessário ter alguma expertise metodológica e relacionada ao conteúdo abordado na revisão para utilização adequada da ferramenta. Recomenda-se que a avaliação com a ROBIS seja desempenhada por dois revisores idealmente independentes ou, no mínimo, por um revisor que avalie e o outro que confira a avaliação realizada. Nós recomendamos que

todas as questões norteadoras da ROBIS sejam consideradas na avaliação da revisão. A ferramenta foi desenvolvida para ser genérica e poder ser aplicada a diversos tópicos; dessa forma, as questões que fazem parte da ROBIS deveriam ser relevantes para qualquer tipo de revisão. Como a ROBIS envolve a realização de julgamentos, a estratégia adotada para deixar esse processo mais transparente foi pedir aos revisores que justificassem cada uma das respostas dadas às questões norteadoras.

A avaliação completa requer que todas as questões norteadoras de todos os domínios sejam respondidas. No entanto, em alguns casos pode ser apropriado “interromper” a avaliação quando certo nível de classificação for atingido. Isto vai depender do objetivo da avaliação. Se o objetivo de utilização da ferramenta for fornecer uma avaliação do risco de viés na revisão e revelar as razões para os vieses, então uma avaliação completa será necessária. No entanto, se o objetivo é simplesmente avaliar se a revisão está sob potencial risco de viés em apenas determinados domínios, os avaliadores podem interromper a avaliação assim que o alto risco de viés tenha sido identificado. Por exemplo, se uma revisão não utilizou uma variedade adequada de bases de dados (questão 2.1), a revisão pode ser considerada com alto risco de viés neste domínio. Avaliadores interessados apenas em identificar se a revisão está sob o risco de viés, podem optar por não avaliar o restante das questões ou domínios, e simplesmente classificar a revisão como de alto risco de viés.

Referências

- [1] Chou R, Hartung D, Rahman B, Wasson N, Cottrell EB, Fu R. Comparative effectiveness of antiviral treatment for hepatitis C virus infection in adults: a systematic review. *Ann Intern Med* 2013;158(2):114-23.
- [2] Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371(9612):569-78.
- [3] Puli SR, Bechtold ML, Reddy JB, Choudhary A, Antillon MR. Can endoscopic ultrasound predict early rectal cancers that can be resected endoscopically? A meta-analysis and systematic review. *Dig Dis Sci* 2010;55(5):1221-9.
- [4] Morris RK, Meller CH, Tamblyn J, Malin GM, Riley RD, Kilby MD, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology* 2014;121(6):686-99.
- [5] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 23.3.11]. Available from: www.cochrane-handbook.org

- [6] Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *The Cochrane database of systematic reviews* 2013;9:MR000022.
- [7] Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011;64(6):602-7.
- [8] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 23.3.11] Available from: <http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm>
- [9] Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282(11):1061-1066.
- [10] Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997;315(7109):617-9.
- [11] Rutjes AW, Reitsma JB, Di NM, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174(4):469-476.
- [12] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273(5):408-12.
- [13] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21(11):1513-24.
- [14] Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336(7644):601-5.
- [15] Higgins JPT, Altman DG, Gøtzsche PC, Moher D, Oxman AD, Savovic J, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMJ* 2011;343:d5928.
- [16] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25.
- [17] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12.
- [18] Borenstein M, Hedges L, Higgins J, Rothstein H. *Introduction to meta-analysis.* . Chichester, UK: Wiley, 2009.
- [19] Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res* 2014.
- [20] Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Stat Methods Med Res* 2013;22(2):133-58.
- [21] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23(20):3105-24.
- [22] Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-73.

[23] Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working G. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149(12):889-97.

[24] Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56(2):455-63.

[25] Van Howe RS. Circumcision and HIV infection: review of the literature and meta-analysis. *Int J STD AIDS* 1999;10(1):8-16.

[26] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-536.

Apêndice: Glossário

Esta sessão apresenta uma visão geral da definição de alguns dos termos utilizados na ROBIS no contexto em que os mesmos são empregados na ferramenta.

Termo	Definição
Viés	Erros sistemáticos ou limitações no delineamento, condução ou análise da revisão influenciam os resultados.
Coleta de dados	Processo de coleta de dados de estudos primários e outras fontes, também conhecida como extração de dados. Pode envolver a coleta de informações de fontes adicionais (ex.: contatando os autores) bem como a extração dos dados dos estudos primários.
Base de dados	Base de dados bibliográfica que indexa estudos.
Fontes eletrônicas	Fontes disponíveis na internet (que não sejam bases de dados) utilizadas para armazenar estudos para inclusão em revisões sistemáticas.
Critérios de elegibilidade	Critério utilizado para determinar quando um estudo deve ser incluído ou excluído de uma revisão sistemática. Também conhecido como critério de inclusão/exclusão.
Diretriz	Conhecimento organizado sistematicamente para auxiliar profissionais e pacientes a tomarem decisões relacionadas ao cuidado em saúde apropriado para situações clínicas específicas. Devem ser baseadas nas melhores evidências disponíveis. Isto significa que os elaboradores de diretrizes devem incluir e avaliar as revisões sistemáticas existentes.
Heterogeneidade	Variabilidade entre os estudos. Heterogeneidade pode ser usada de forma geral para descrever as diferenças entre os estudos em termos da população, intervenção, comparador, qualidade metodológica e delineamento. Heterogeneidade estatística ocorre quando o efeito estimado entre os estudos difere muito do valor esperado devido à variação da amostra (chance).
Meta- análise	Método estatístico que consiste em combinar os resultados de diferentes estudos. Isto é feito calculando-se a média ponderada das estimativas de efeito. Uma meta-análise é normalmente conduzida como parte de uma revisão sistemática, mas nem sempre é apropriado ou possível

	fazê-la.
<i>Overview</i>	É uma revisão de revisões sistemáticas.
Autores da <i>overview</i>	Autores da revisão de revisões sistemáticas.
Leitores	Pessoas que estão lendo a revisão sistemática.
Relevância	O quanto que a questão da revisão combina com a questão de interesse.
Relatórios	Síntese dos métodos, achados e conclusões de um estudo, que podem estar publicados ou não. Um único estudo pode constar de diferentes relatórios e um único relatório pode conter diferentes estudos.
Revisão	Abreviação para revisão sistemática. Também pode ser utilizado para significar uma revisão não sistemática, porém este não é o significado no contexto da ROBIS.
Autores da revisão	Autores da revisão sistemática, algumas vezes chamados também de revisores.
Estratégia de busca	Termo exato e suas combinações utilizadas na busca nas bases de dados bibliográficas.
Significância estatística	A probabilidade de que o resultado observado ou um que seja mais extremo, considerando-se que a hipótese nula é verdadeira, é causado pelo acaso. Testes de hipóteses estatísticas são usados para determinar se os resultados de um estudo podem levar à rejeição da hipótese nula. Normalmente, define-se um limiar de 5% para a rejeição da hipótese nula, porém qualquer outro valor pode ser escolhido. Assim, se o valor de p é menor do que 0,05 (ou outro nível especificado), então o resultado deve ser considerado estatisticamente significativo e a hipótese nula rejeitada.
Estudos	Estudos primários incluídos na revisão sistemática. Um único estudo pode estar presente em diferentes relatórios.
Estimativa sumária	Uma única estimativa de efeito derivada de múltiplos estudos usando meta-análise.
Síntese	Análise quantitativa e qualitativa utilizada para sumarizar os resultados de estudos primários incluídos na revisão sistemática.
Revisão sistemática	Abordagem sistemática de revisar e sumarizar as evidências de estudos. Segue uma estrutura definida para identificar, avaliar e sumarizar toda evidência disponível sobre uma questão de pesquisa específica. Pode incluir meta-análise, sem ser, no entanto, um requisito essencial.
Questão alvo	Questão que os avaliadores estão tentando responder na <i>overview</i> ou diretriz.