



4.16 Anexo da seção 4.5 – Exemplos de ferramentas de avaliação da qualidade



Forma de evidências	Exemplos de ferramentas de avaliação de qualidade
Tipos de evidências para as quais existem ferramentas de avaliação de qualidade	
Análise de dados	<p>ROBINS-I (riskofbias.info) para estudos observacionais, como os que analisam associações entre fatores selecionados (incluindo intervenções) e desfechos selecionados, em que há um risco de viés de:</p> <ul style="list-style-type: none"> • confusão (onde a relação observada entre um fator e um desfecho difere da relação verdadeira por causa de um ou mais fatores adicionais que não são levados em conta) • seleção dos participantes para o estudo • classificação de intervenção(ões) • desvios da intervenção pretendida • dados ausentes • medição dos desfechos • seleção do resultado relatado
Avaliação	<p>Risk of Bias (RoB) 2 (riskofbias.info) para ensaios clínicos randomizados, em que o risco de confusão é menor, mas há risco de viés de algumas (embora menos) das mesmas fontes acima:</p> <ul style="list-style-type: none"> • processo de randomização • desvios das intervenções pretendidas • dados (de desfechos) ausentes • medição dos desfechos • seleção do resultado relatado
Pesquisa do comportamento/ de implementação	Veja outras linhas para os tipos relevantes de estudos ou sínteses
Pesquisa qualitativa	<p>Checklist de avaliação crítica do JBI (<i>Joanna Briggs Institute</i>) para pesquisa qualitativa (bit.ly/31Lsib1), em que considerações muito diferentes são levadas em conta, como:</p> <ul style="list-style-type: none"> • congruência entre a metodologia de pesquisa e a pergunta de pesquisa, métodos de coleta de dados, representação e análise de dados, e interpretação dos resultados, bem como entre a perspectiva filosófica indicada e a metodologia • reflexividade por parte do pesquisador, como declarações localizando o pesquisador cultural e teoricamente, e discutindo a influência do pesquisador na pesquisa e vice-versa • representação dos participantes do estudo e suas vozes • fluxo de conclusões da análise e interpretação dos dados
Síntese de evidências	<p>Veja acima para os tipos relevantes de estudos considerados na síntese de evidências</p> <p>AMSTAR (amstar.ca) (do inglês <i>A Measurement Tool to Assess systematic Reviews</i> - uma ferramenta de medição para avaliar revisões sistemáticas) para a qualidade da síntese de evidências, em que o risco de viés pode surgir da:</p> <ul style="list-style-type: none"> • identificação de todos os estudos potencialmente relevantes por meio de uma busca abrangente da literatura publicada e cinzenta sem restrições de idioma • seleção de todos os estudos que abordam a pergunta de pesquisa usando critérios explícitos sobre os desenhos de estudo e sobre os participantes, intervenções/fatores, comparações e desfechos, e com pelo menos dois revisores aplicando os critérios • avaliação da qualidade e extração de dados de todos os estudos incluídos • síntese dos achados de todos os estudos incluídos <p>Observe que há duas versões da AMSTAR: 1) a versão original que pode ser aplicada em todos os tipos de sínteses, embora com alguns critérios removidos do numerador e do denominador; 2) uma segunda versão da AMSTAR que é mais especificamente relevante para sínteses de ensaios clínicos randomizados</p> <p>Gradação das Recomendações de Avaliação e Desenvolvimento (GRADE, na sigla em inglês) (bit.ly/3C9pMrx) para a certeza das evidências para os desfechos de uma intervenção, com:</p> <ul style="list-style-type: none"> • certeza rebaixada devido ao risco de viés (com evidências de ensaios clínicos randomizados começando como de “confiança alta” e evidências de estudos observacionais começando como de qualidade baixa e depois sendo ajustada com base no RoB2 ou ROBINS-I), imprecisão (p. ex., um ou dois estudos pequenos), inconsistência (p. ex., dois estudos mostrando achados muito diferentes), evidências indiretas (p. ex., medidas substitutas utilizadas ou cenários atípicos estudados) e viés de publicação (p. ex., mais comum com estudos observacionais devido à falta de registros de estudos ou com estudos financiados pela indústria devido ao incentivo comercial para publicar estudos positivos) • certeza elevada para grande magnitude de efeito, gradiente de dose-resposta e quando toda a confusão residual diminuiria a magnitude do efeito <p>Abordagem para Avaliar a Confiança em Achados de Sínteses de Evidências Qualitativas (GRADE-CERQual; cerqual.org) para a certeza das evidências para a representação qualitativa de um fenômeno de interesse, com:</p> <ul style="list-style-type: none"> • certeza rebaixada devido a preocupações sobre limitações metodológicas (porque os problemas na forma como os estudos foram desenhados ou relatados foram identificados usando uma ferramenta de avaliação crítica, como a do JBI acima), relevância (porque o contexto em que os estudos primários foram conduzidos é substancialmente diferente do contexto da pergunta de síntese), coerência (porque alguns dos dados contradizem os achados ou são ambíguos) e adequação (porque os dados não são suficientemente informativos ou provêm apenas de um pequeno número de estudos ou participantes)



Avaliação de tecnologias/ análise de custo-efetividade

Checklist da *International Network of Agencies for Health Technology Assessment* (INAHTA) (bit.ly/2YJVMVK) para a qualidade das avaliações de tecnologias, com duas das 14 questões discutindo a abordagem para sintetizar as evidências (com *prompts* semelhantes à AMSTAR) e outra questões discutindo se a avaliação foi contextualizada por meio de uma análise de custo-efetividade (com dados de custo locais – nacionais ou subnacionais), e consideração das implicações legais, éticas e sociais locais

Checklist de Drummond de análises de custo-efetividade (bit.ly/3FbnB8R), e para avaliações econômicas de forma mais geral, com questões sobre o desenho do estudo, coleta de dados, e análise e interpretação dos resultados

Checklist da Philips para análises de custo-efetividade que incluem um componente de modelagem analítica de decisão (bit.ly/3FcWBGc) com questões sobre a estrutura do modelo (p. ex., fundamentação explícita, suposições justificadas e horizonte temporal adequado), os dados usados (p. ex., probabilidades de base dos estudos observacionais, efeitos de tratamento de ensaios clínicos randomizados, e avaliações de quatro tipos de incertezas: a estrutura do modelo, as etapas metodológicas seguidas, a heterogeneidade na população estudada, e os parâmetros usados) e a consistência (interna e externa) – há ainda a ferramenta complementar TRUST para avaliar incertezas nos modelos analíticos de decisão (bit.ly/3quFSKp)



Diretrizes

Ferramenta AGREE II (bit.ly/30qyFAB) para avaliar o desenvolvimento, relato e avaliação (ou avaliação da qualidade) das diretrizes, que usa 23 itens agrupados em seis domínios, cada um sendo pontuado de forma independente:

- escopo e propósito descritos
- envolvimento das partes interessadas (cidadão/paciente e profissional)
- rigor de desenvolvimento (com sínteses de evidências usadas como entrada de dados, um processo robusto de desenvolvimento das recomendações, e recomendações vinculadas às evidências de apoio)
- clareza na apresentação
- aplicabilidade
- independência editorial (em relação ao financiador e aos conflitos de interesse dos membros do painel)

Graduação das Recomendações de Avaliação e Desenvolvimento (GRADE, na sigla em inglês) (bit.ly/3C9pMrx) para avaliar a força das recomendações, que usa quatro considerações principais:

- equilíbrio entre os desfechos desejáveis e indesejáveis (escolhas), levando em consideração as melhores estimativas da magnitude dos efeitos nos desfechos desejáveis e indesejáveis, e a importância desses desfechos (valores típicos estimados e preferências)
- confiança na magnitude das estimativas dos efeitos das intervenções sobre desfechos importantes (veja GRADE na linha anterior)
- confiança nos valores e preferências e uso de recursos de variabilidade

Tipos de evidências para as quais ainda não existem ferramentas de avaliação de qualidade



Modelagem

Ainda não existe uma ferramenta amplamente aceita para a maioria dos tipos de modelos; no entanto, há algumas perguntas gerais que podem ser feitas sobre modelos (muito semelhantes às listadas como parte do *checklist* da Philips acima), como:

- estrutura do modelo (p. ex., fundamentação explícita, suposições justificadas e horizonte temporal adequado)
- dados usados (p. ex., probabilidades de base dos estudos observacionais, efeitos de intervenção de uma variedade de fontes*, e avaliações de quatro tipos de incerteza: a estrutura do modelo, as etapas metodológicas seguidas, a heterogeneidade na população estudada, e os parâmetros usados)
- consistência (interna e externa)
- disponibilidade do *software* ou ferramenta para que possa ser avaliado por outros

*Um dos desafios com a COVID-19 era que os desenhos de estudo geralmente usados para obter os efeitos da intervenção, como os ensaios clínicos randomizados, eram ética ou logisticamente difíceis e/ou demoravam muito para serem concluídos; então, outros desenhos de estudo precisavam ser usados e era necessário buscar a opinião de *experts* (e existem abordagens que permitem fazer isso de forma sistemática e transparente, como a SHELF – veja bit.ly/30nteC4)

Abordagens usadas com alguns tipos de evidências para as quais ainda não existem ferramentas de avaliação de qualidade



Inteligência artificial

Ainda não existe uma ferramenta amplamente aceita