



UNESP - Universidade Estadual Paulista
“Júlio de Mesquita Filho”
Faculdade de Odontologia de Araraquara



Ellen Brito Pereira

**Análise de qualidade e risco de viés nas revisões sistemáticas da literatura:
levantamento dos instrumentos e escala existentes**

Araraquara

2022



UNESP - Universidade Estadual Paulista
Faculdade de Odontologia de Araraquara



Ellen Brito Pereira

**Análise de qualidade e risco de viés nas revisões sistemáticas da literatura:
levantamento dos instrumentos e escala existentes**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Odontologia da Faculdade de Odontologia de Araraquara, da Universidade Estadual Paulista, para a obtenção do grau de Cirurgião-dentista.

Orientadora: Prof^a. Dr^a. Lívია Nordi Dovigo

Araraquara

2022

P436a Pereira, Ellen Brito
Análise de qualidade e risco de viés nas revisões sistemáticas da literatura : levantamento dos instrumentos e escala existentes / Ellen Brito Pereira. -- Araraquara, 2022
43 p.

Trabalho de conclusão de curso (Bacharelado - Odontologia)
- Universidade Estadual Paulista (Unesp), Faculdade de Odontologia, Araraquara
Orientadora: Lívia Nordi Dovigo

1. Revisão. 2. Medicina baseada em evidências. 3. Reprodutibilidade dos testes. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Odontologia, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**UNESP - Universidade Estadual Paulista
Faculdade de Odontologia de Araraquara**

Ellen Brito Pereira

**Análise de qualidade e risco de viés nas revisões sistemáticas da literatura:
levantamento dos instrumentos e escala existentes**

Orientadora: Prof^a. Dr^a. Lívia Nordi Dovigo

Assinatura Orientadora: 

Assinatura Aluno: 

Araraquara, 30 de março de 2022.

AGRADECIMENTOS ESPECIAIS

Primeiramente a Deus, a razão de tudo, meu fôlego, alegria, meu Pai e amigo. Obrigada por sempre acreditar em mim, por ter me dado força, saúde e proteção pra chegar até aqui. Sem Ti eu nada seria nem conseguiria. Nunca conseguirei agradecer tanto amor e cuidado, obrigada! Amo muito Você!

Aos meus pais, Irandi e Nilo, que sempre me apoiaram, sonharam meus sonhos, sorriram e choraram comigo. Vocês são meus maiores exemplos de caráter, determinação e amor. Obrigada por nunca desistirem das minhas lutas, mas fazerem delas as suas. Obrigada por todo apoio financeiro, que eu sei que nunca foi fácil para vocês, mas nunca mediram esforços para que eu conquistasse meus objetivos. Principalmente, obrigada por todo apoio afetivo durante essa caminhada, vocês foram e sempre serão essenciais em toda minha vida. Nunca conseguirei agradecer-vos o suficiente, obrigada! Amo muito vocês!

À minha irmã Naila, por ser os meus “pés no chão”, sempre me auxiliando nas tomadas de decisões e me apoiando independentemente delas. Obrigada por estar sempre ao meu lado, e por compartilhar suas experiências comigo, me dando forças para prosseguir. Saber que sempre posso contar contigo foi essencial para chegar até aqui, obrigada! Amo muito você!

Ao meu namorado e melhor amigo Gustavo, por sempre me apoiar e não me deixar desistir! Obrigada também por toda paciência e companheirismo, mesmo à distância eu sempre te senti pertinho de mim e isso me deu forças para continuar! Te amo muito meu amor!

À minha ex babá e atual amiga Elenice, por sempre acreditar em mim, vibrar com todas minhas conquistas e ser um ombro amigo sempre que precisei. Seu apoio, carinho e amor comigo e com a minha família foram essenciais para mais essa conquista, obrigada! Amo muito você!

À minha tia Maraisa, por ter me ensinado tanto com seu jeito simples e alegre de viver e sempre comemorar minhas raras voltas para casa, obrigada tia Mara! Amo muito você!

A todos meus familiares e amigos que sempre torceram por mim, obrigada! Amo muito vocês!

AGRADECIMENTOS

À Universidade Estadual Paulista – Júlio de Mesquita Filho, na figura do senhor Reitor, Professor Doutor Sandro Roberto Valentini, obrigada por sempre prezar por um ensino de excelente qualidade.

À professora Doutora Elaine Maria Sgavioli Massucato que, quando ingressei, era a diretora dessa instituição. Obrigada por sua excelente gestão, carinho e cuidado com os estudantes.

Ao Professor Doutor Edson Alves de Campos, atual diretor dessa faculdade. Obrigada por sua liderança e ensino, sempre visando o melhor para a instituição.

À Professora Doutora Patrícia Petromilli Nordi Sasso Garcia, atual vice-diretora dessa faculdade. Obrigada por sempre me transmitir coisas boas e me ajudar a passar pela graduação com tanta leveza.

À minha orientadora, Professora Doutora Livia Nordi Dovigo, que sempre me acolheu com tanto carinho e compressão e sempre acreditou no meu potencial. Obrigada por sempre estar disposta a me ajudar no que for preciso, sempre com toda gentileza. Obrigada por todo ensino como Professora e Orientadora, você foi fundamental para que eu chegasse até aqui!

A todos os funcionários da FOAr, que diariamente permitem com que a gente coloque em prática tudo que aprendemos.

Ao Departamento de Odontologia Social, obrigada pelo uso de suas dependências e pela oportunidade de ingressar no programa de Iniciação Científica. Em especial, muito obrigada à secretária Neli, sempre disposta a nos ajudar com muito carinho.

A todos os pacientes que confiaram a mim seu atendimento. Vocês foram essenciais para minha formação e desenvolvimento pessoal. Cada um de vocês é único e cada história me tocou de forma particular e intensa, contribuindo para que eu me tornasse um ser humano melhor. Obrigada pela confiança! Ver a alegria no rosto de vocês com certeza anulava qualquer cansaço ou tristeza, obrigada!

Aos meus colegas da turma 92, obrigada por cada momento que passamos juntos, aprendi e evolui muito com vocês!

Pereira EB. Análise de qualidade e risco de viés nas revisões sistemáticas da literatura: levantamento dos instrumentos e escala existentes [Trabalho de Conclusão de Curso – Graduação em Odontologia]. Araraquara: Faculdade de Odontologia da UNESP; 2022.

RESUMO

O aumento rápido e expressivo de publicações na área da saúde vem preocupando pesquisadores com a qualidade das informações bem como a dificuldade de se acompanhar tudo o que é divulgado a respeito de determinado problema. A revisão sistemática da literatura (RS; do inglês, *systematic review*) é uma investigação científica retrospectiva que tem o intuito de fornecer uma síntese abrangente e imparcial dos estudos existentes sobre uma questão específica, por meio do seguimento de protocolos formais para garantir a reprodutibilidade e reduzir os erros nas conclusões. Por ser um tipo de estudo que reúne todas as evidências a respeito de um determinado problema e realiza uma avaliação crítica das informações publicadas, as RS têm sido colocadas no topo da hierarquia de evidências, e são os principais componentes das diretrizes clínicas e políticas publicadas por organizações como a Organização Mundial de Saúde. No entanto, para que possam cumprir esse papel, as RS devem ser conduzidas com rigor metodológico, fato que tem sido negligenciado. Esse estudo realizou um levantamento de literatura acerca dos instrumentos e escalas existentes para avaliação de qualidade e risco de viés de RS. A criação da declaração PRISMA melhorou a redação e, conseqüentemente, qualidade das RS. Porém, ainda assim foi necessário desenvolver ferramentas que avaliassem, especificamente, a qualidade metodológica e o risco de viés das RS. Entre as ferramentas que avaliam a qualidade das RS, destacam-se AMSTAR, R-AMSTAR e AMSTAR 2. A ferramenta ROBIS foi desenvolvida especificamente para avaliar o risco de viés em RS. Apesar de AMSTAR 2 e ROBIS serem consideradas uma melhoria em relação ao AMSTAR, ainda são necessárias pesquisas e evidências a respeito dessas duas ferramentas, visto que são bem recentes. Atualmente, não há um consenso a respeito de qual instrumento usar para medição de qualidade metodológica de RS, destacando-se a necessidade de revisão e atualização dos atuais instrumentos.

Palavras – chave: Revisão. Medicina Baseada em Evidências. Reprodutibilidade dos testes.

Pereira EB. Analysis of quality and risk of bias in systematic literature reviews: survey of existing instruments and scale [Trabalho de Conclusão de Curso – Graduação em Odontologia]. Araraquara: Faculdade de Odontologia da UNESP; 2022.

ABSTRACT

The rapid and expressive increase in publications in the health area has been worrying researchers about the quality of information as well as the difficulty of following up everything that is published about a given problem. A systematic review of the literature (RS) is a retrospective scientific investigation that aims to provide a comprehensive and unbiased synthesis of existing studies on a specific issue, by following formal protocols to ensure reproducibility and reduce errors in conclusions. As a type of study that gathers all the evidence regarding a given problem and performs a critical assessment of published information, SRs have been placed at the top of the evidence hierarchy, and are the main components of clinical guidelines and policies published by organizations such as the World Health Organization. However, in order to fulfill this role, SRs must be conducted with methodological rigor, a fact that has been neglected. This study carried out a literature survey about the existing instruments and scales to assess the quality and risk of bias of RS. The creation of the PRISMA statement improved the wording and, consequently, the quality of the SRs. However, it was still necessary to develop tools that specifically assess the methodological quality and risk of bias of the SRs. Among the tools that assess the quality of SR, AMSTAR, R-AMSTAR and AMSTAR 2 stand out. The ROBIS tool was developed specifically to assess the risk of bias in SR. Although AMSTAR 2 and ROBIS are considered an improvement over AMSTAR, research and evidence is still needed regarding these two tools, as they are quite recent. Currently, there is no consensus on which instrument to use to measure the methodological quality of RS, highlighting the need to review and update the current instruments.

Keywords: Review. Evidence-Based Medicine. Reproducibility of Results.

SUMÁRIO

1 INTRODUÇÃO	8
2 PROPOSIÇÃO	11
3 REVISÃO DA LITERATURA	12
3.1 Histórico das Revisões Sistemáticas e Meta Análise	12
3.2 Instrumentos de Avaliação de Qualidade e Risco de Viés	19
4 DISCUSSÃO	36
5 CONCLUSÃO	40
REFERÊNCIAS	41

1 INTRODUÇÃO

A melhor maneira de tomar decisões em saúde, sejam elas quanto ao diagnóstico ou tratamento dos pacientes, é baseando-se em evidências científicas¹. Por isso, os profissionais da área da saúde precisam manter-se atualizados com os resultados publicados em diferentes estudos¹. Porém, o grande volume de informações e a rápida velocidade com que novos dados são disponibilizados tornou-se um obstáculo para que um profissional, seja clínico ou pesquisador, consiga reunir todas as informações relevantes dentro de seus campos de atuação². Revisões sistemáticas (RS) são estudos secundários que reúnem todas as informações disponíveis na literatura sobre determinado assunto, definido a partir de uma pergunta específica e bem formulada, que o pesquisador pretende responder³. Essa pergunta deve ser formulada a partir da estratégia PICO, acrônimo de: identificação da população (P), da intervenção (I), do que se pretende comparar (C), e do desfecho (O)³. A partir dessa questão específica, é feita uma busca extensa, por toda a literatura, de evidências científicas que possam responder a essa questão⁴. Essa busca é feita sob análise crítica e, os resultados encontrados, são reunidos e resumidos, seguindo uma metodologia rigorosa⁵.

Técnicas estatísticas de análise dos dados oriundos dos estudos primários podem ser utilizadas nas RS³. Esse tipo de análise vem sendo atribuída ao matemático britânico Karl Pearson que desenvolveu, em 1904, mecanismos que tinham como objetivo, reunir resultados de estudos diferentes³. Na década de 60 e início da década de 80, ocorreram dois acontecimentos relevantes: a publicação da primeira RS com objetivo clínico e o surgimento do termo meta-análise (MA)³. O grande destaque das RSs com MA na área da saúde ocorreu a partir da década de 90, contemporaneamente com o estabelecimento da Colaboração Cochrane, uma rede colaborativa para condução de RS nas ciências médicas. Milhares de protocolos e RS foram cadastrados na Colaboração Cochrane desde sua fundação³.

Por ser um tipo de estudo que reúne todas as evidências a respeito de um determinado problema e realiza uma avaliação crítica das informações publicadas, as RS têm sido colocadas no topo da hierarquia de evidências, e são os principais componentes das diretrizes clínicas e políticas publicadas por organizações como a Organização Mundial de Saúde (OMS)⁶. Na Odontologia, à medida que os princípios e a prática baseada em evidências se tornaram mais aceitos e difundidos, houve um

grande crescimento no número de RS e no reconhecimento mais amplo de seu valor⁷. Um estudo analisou as RS publicadas em 2017 e verificou que ocorreu a publicação de 495 RS relacionadas à Odontologia, ou seja, 6,05% de todas as RS publicadas em 2017, sendo o Brasil o maior contribuinte (117 RS ou 23,6%)⁸. Dos 495 estudos, foram identificadas apenas 13 revisões Cochrane (2,6% das RS relacionadas à Odontologia)⁸. Somente quatro estavam atualizadas (0,8%), o restante se tratava de RS originais⁸. Apesar das várias revisões publicadas em Odontologia em 2017, algumas especialidades tiveram maior destaque, como a Cirurgia Oral (75 artigos publicados)⁸. Sacks et al.¹⁰ demonstraram que 1188 RS foram publicados entre 1991 e 2012 e as características variaram entre as especialidades odontológicas⁸. Porém, a grande quantidade de revisões publicadas não significa que todas possuam excelente qualidade e podem ser utilizadas na prática baseada em evidências⁸.

Muitas RS possuem problemas como, não relatar o método de extração de dados utilizado (26,4% em 2017) ou o método utilizado para avaliar o risco de viés (40,5% em 2017)⁸. Além disso, muitas não relatam a fonte de financiamento (38,8% em 2017)⁸. Essa grande variação das características de relato e condução das RS que ocorreu entre as diferentes especialidades odontológicas, muitas vezes, produz resultados não concretos e tendenciosos⁸. Provavelmente, esse cenário tem ocorrido porque muitas revisões estão sendo conduzidas e publicadas sem que seus autores se preocupem em conhecer o processo metodológico ou avaliar corretamente os estudos que estão incluindo⁸. Portanto, há necessidade de adotar medidas para melhorar a qualidade das revisões e condução das mesmas em Odontologia⁸. Uma das formas de obter essa melhoria é ampliar o conhecimento metodológico dos autores e leitores e auxiliá-los quanto à maneira correta de aplicar as ferramentas utilizadas na avaliação da qualidade metodológica e risco de viés das RS⁸. Foram desenvolvidas algumas ferramentas que avaliam desde a qualidade metodológica de estudos primários incluídos nas RS, até a qualidade das próprias RS⁵. A avaliação realizada a partir dessas ferramentas produz resultados que, posteriormente, poderão ser utilizados para basear tomadas de decisões em saúde, como diagnósticos e/ou tratamentos⁹.

Diante da importância que as RS possuem para a tomada de decisões na área de saúde, bem como para nortear futuras pesquisas nas diferentes áreas do

conhecimento, é fundamental que esses estudos sejam conduzidos com todo o rigor metodológico. Além disso, os leitores das RS precisam estar preparados para avaliar criticamente as informações publicadas e, nesse sentido, a utilização de instrumentos que auxiliam a análise de qualidade e risco de viés nas RS pode trazer grande benefício.

2 PROPOSIÇÃO

O objetivo deste trabalho foi realizar um levantamento de literatura sobre as ferramentas disponíveis para avaliação crítica de RS, como a qualidade metodológica e risco de viés nos resultados, bem como realizar uma discussão dos dados de aplicabilidade e confiabilidade dos instrumentos encontrados.

3 REVISÃO DA LITERATURA

Foi realizada uma busca na literatura para localização dos artigos de interesse para esse trabalho, de acordo com a proposição. A busca eletrônica ocorreu em duas diferentes bases de dados (MEDLINE via PubMed e LILACS via BVS) e também na página da internet *Google Scholar*. Na construção das buscas, foram utilizados termos relacionados à avaliação crítica (*critical appraisal*), revisões sistemáticas (*systematic reviews*), bem como o nome das ferramentas já conhecidas pelos autores quando do delineamento do estudo (AMSTAR e ROBIS). A lista de referências dos trabalhos localizados também foi consultada em busca de trabalhos de interesse que não foram localizados por meio das buscas nas bases de dados. Todos os trabalhos foram lidos e elaborados resumos, os quais estão separados nas subseções descritas a seguir, conforme seu assunto principal abordado.

3.1 Histórico das Revisões Sistemáticas e Meta Análise

Sacks et al.¹⁰, em 1987, relataram que Light e Smith deram início à ideia de agrupar dados de vários estudos diferentes, o que hoje é conhecido por MA (termo conferido por Glass). Segundo esse estudo, a MA é um instrumento que revisa criticamente e combina dados estatísticos de vários estudos diferentes. Os autores relataram que a MA é mais aplicável a ensaios clínicos randomizados (ECRs). O objetivo desse estudo foi levantar o estado atual das MA; propor alguns critérios àqueles que realizam e utilizam MA e propor estudos mais aprofundados em algumas áreas. Para isso utilizaram estudos publicados de janeiro de 1966 a outubro de 1986, que combinavam os resultados de ECRs. Alguns estudos foram encontrados no Harrem Contents e outros nas bases de dados da National Library of Medicine e Bibliographic Retrieval Services, principalmente. Como critério de inclusão, os dados de mais de um estudo deviam estar combinados e, pelo menos um, deveria ser ECR. Os estudos foram avaliados por dois pesquisadores, de forma independente. Dos 177 estudo encontrados, 86 atenderam aos critérios de inclusão. Concluíram que “as qualidades importantes de qualquer MA podem ser divididas em seis áreas principais: desenho do estudo, combinabilidade, controle de viés, análise estatística, análise de sensibilidade e problemas de aplicabilidade”.

Em 1997, Egger e Smith¹¹ abordaram a história da MA e a importância de MA de ECRs. Karl Pearson, em 1904, foi o primeiro pesquisador a combinar dados de

diversas amostras. Em 1955, foi publicada a primeira MA avaliando o efeito de uma intervenção terapêutica. Devido ao grande avanço das MA, a necessidade de sistematizar RS e comprovar intervenções de forma sistemática, tomou grande proporção no campo científico. Ao reunir resultados de vários ensaios clínicos independentes, as MA permitem melhor avaliação das intervenções disponíveis. Além disso, a análise das MA acaba evidenciando as questões de pesquisa mais urgentes ou que precisam ser melhores estudadas e também evidencia, estatisticamente, intervenções benéficas, impedindo novas pesquisas desnecessárias sobre o assunto.

Oxman et al.¹², em 1994, relataram a complexidade que envolve a confecção de uma RS, desde a escolha da questão específica a ser respondida, até a avaliação da validade dos estudos durante a etapa de inclusão. Segundo os autores, uma lista de verificação pode auxiliar tanto os desenvolvedores de RS, por exemplo, identificando erros antes que o processo avance, quanto os leitores, ajudando-os a identificar possíveis erros e, assim, descartar ou não aquele estudo. Quanto mais clara estiver a forma com que os desenvolvedores tomaram as decisões, mais facilmente os leitores conseguirão avaliar aquele estudo isolado ou RS. Evitar erros, por exemplo, utilizando listas de verificação, evita tomadas de decisões erradas em saúde, que poderiam prejudicar toda uma população. Segundo os autores, revisões realizadas de forma não sistemática, apresentam alto índice de erros em comparação com revisões realizadas sistematicamente, podendo levar a decisões equivocadas. Ainda relataram que os erros sistemáticos (vieses) são mais perigosos em uma RS do que os erros aleatórios. Por isso, a validade é o principal fator a ser avaliado em uma RS. Outros fatores que devem ser avaliados: “escolha da questão específica, grau de inovação na abordagem, impacto potencial em desenvolvimentos científicos futuros, qualidade literária e tratamento de questões éticas pertinentes”. Foram desenvolvidas várias listas de verificação para avaliar a validade de uma RS, com algumas diferenças nos itens incluídos e como esses itens são abordados. Se o desenvolvedor optar por fazer uma lista de verificação, a primeira pergunta a ser respondida é se o assunto escolhido possui relevância e a segunda diz respeito à validade dos resultados. Segundo os autores, se a questão específica não for bem formulada e se os métodos para responderem a essa questão não forem claros, não convém avançar no processo. Também destacaram como possíveis fontes de vieses: a

formulação da questão específica; a identificação, seleção e avaliação dos estudos; a coleta de dados e interpretação dos resultados. Abordaram a respeito da inclusão e avaliação da validade de estudos, que deve ser realizada de forma bem clara, para evitar seleção e avaliação tendenciosa. Ressaltaram a importância da qualidade das MA, para evitar falsos positivos. Também descreveram como deve ser realizada a análise de dados e a conclusão.

Atallah et al.¹³, em 1996, relataram que as RS são usadas, frequentemente, para avaliar, cientificamente, se determinada terapia médica é mais eficiente que outra, por exemplo. Segundo esse estudo, desde que as RS sejam realizadas sob metodologia rigorosa e com evidências suficientes a respeito da melhor intervenção terapêutica, o pesquisador ou profissional da saúde pode tomar decisões com base nelas. Se houverem ECRs, porém, com resultados inconsistentes, deve-se fazer MA. Ao analisar a RS, o resultado final pode ser a favor, contra ou inconclusivo a respeito da terapêutica proposta. Além disso, a análise de RS sobre determinado assunto, oferece um parâmetro no sentido de já existirem estudos suficientes sobre ou não. Esse estudo também relata a importância da sistematização das RS, visto que as não sistematizadas são incompletas, imprecisas, não reproduzíveis e, conseqüentemente, inconclusivas. Outro ponto relatado nesse estudo foi a criação da Colaboração Cochrane, em Oxford, pelo Dr. Iain Chalmers, com o objetivo de evitar a duplicação desnecessária de estudos a respeito do mesmo assunto.

As RS como alternativas para a tomada de decisões em saúde, mediante o crescente número de estudos independentes publicados, foi um dos assuntos destacados por Jadad et al.¹⁴, em 2000. Idealmente, a condução de uma RS deveria ser descrita de forma clara e adotar estratégias para minimizar vieses e aumentar a precisão, podendo ou não incluir MA. Segundo os autores, muitas RS possuem falhas na metodologia, o que prejudica a validade das mesmas. Destacaram a importância da criação da Colaboração Cochrane como forma de melhorar a qualidade metodológica das RS e MA. O objetivo desse estudo foi avaliar a qualidade das RS e MA publicadas sobre o tratamento da asma e comparar as características das RS Cochrane com as de periódicos revisados por pares. Realizaram uma busca na literatura através das bases de dados MEDLINE, CINAHL, HealthSTAR e EMBASE, além de buscas na Cochrane Library, buscas em coleções pessoais e em listas de referências. Os estudos encontrados foram revisados por

dois revisores independentes. Os dados foram extraídos por três revisores treinados, de forma independente. A qualidade metodológica foi avaliada com o índice de Oxman e Guyatt, uma ferramenta validada que pontua as revisões em uma escala de sete pontos (pontuações de 1 a 3, significam falhas graves ou extensas e pontuações de 4 a 7, significam falhas mínimas ou menores). Revisões publicadas na Cochrane Library foram separadas daquelas publicadas em periódicos revisados por pares. Como resultado, foram incluídas 50 RS e MA, sendo que a maioria foi publicada nos últimos dois anos referentes a esse estudo. Doze revisões foram publicadas na Cochrane Library e 38 em 22 periódicos revisados por pares. Quarenta revisões avaliadas continham falhas graves ou extensas. O número de RS e MA avaliando o tratamento da asma está aumentando. A maioria dessas revisões tem relevância clínica, mas tem limitações metodológicas que poderiam ter sido evitadas facilmente. Os autores discutiram sobre a necessidade de futuras RS incluírem itens como: estratégias de pesquisa abrangentes e bem relatadas, tentativa de induzir viés durante a seleção dos estudos, avaliação da validade dos estudos e descrição clara dos métodos utilizados para sintetizar os dados disponíveis. Segundo eles, a partir desse estudo, foi possível confirmar o maior rigor das revisões Cochrane em comparação com os periódicos revisados por pares, visto que as revisões Cochrane seguem instruções padronizadas, são revisadas por pares em vários níveis e recebem contribuições de diferentes grupos em diferentes estágios. Relataram a possibilidade de melhorar a qualidade metodológica das RS publicadas pelos periódicos revisados por pares, desde que deixem claro os critérios adotados no relatório. Concluíram que a maioria das revisões publicadas em periódicos revisados por pares ou financiadas pela indústria apresentam sérias falhas metodológicas que prejudicam sua validade.

Em seu estudo em 2001, Castro et al.¹⁵ apresentaram o passo a passo para elaborar uma RS, de acordo com a Colaboração Cochrane (que recomenda a confecção da RS em sete passos) e pelo NHS Centre for Reviews and Dissemination, University of York (que recomenda a confecção da RS em nove passos, agrupados em três estágios). As duas formas são semelhantes, porém, uma enfatiza mais alguns pontos e a outra enfatiza pontos diferentes. O estudo é finalizado considerando as possíveis conclusões que podem ser obtidas ao final da confecção e/ou análise de uma RS.

Chalmers et al.¹⁶, em 2002, relataram resumidamente sobre a história da síntese em pesquisa. De acordo com esse estudo, o histórico de publicações passadas deve embasar as novas sínteses. Esse assunto assumiu grande importância a partir de 1884. Em 1904, Karl Pearson conferiu ainda mais visibilidade a esse tema. Três anos depois, Pearson, Goldberger, entenderam como era importante reduzir vieses. Com o fim da Segunda Guerra Mundial, houve um aumento considerável de pesquisas científicas, fazendo com que os pesquisadores pensassem uma forma de organizar, analisar e agrupar as diversas pesquisas publicadas. O termo MA foi introduzido por Glass, em 1976 (não obteve grande sucesso). Edward Kass, em 1981, foi o primeiro a se preocupar com a qualidade dos estudos primários. David Pillemer, em 1984, declarou as sínteses em pesquisa como subjetivas e baseadas em julgamentos incertos. Em 1991, foi criado o NHS Centre for Reviews and Dissmination e o Centro Cochrane britânico. Na década de 90, a síntese em pesquisa, ainda era vista como “de segunda classe” e “indigna de ser mencionada”. Em 1995, foi criada a The Cochrane Database of Systematic Reviews, primeira base online para agrupar RS. Devido ao grande aumento no número de publicações, clínicos e pesquisadores passaram a consultar RS. No final da década de 90, Cochrane consolidou o termo “RS”. Com isso, clínicos e pesquisadores passaram a usar os resultados das RS para tomada de decisões.

O objetivo do trabalho de Thomson et al.¹⁷, em 2010, foi discutir sobre os desafios logísticos e metodológicos na confecção de *overviews* (sínteses de RS). *Overviews* são mais amplas e RS aprofundam mais em uma intervenção ou grupo de intervenções relacionadas no contexto de uma condição específica. Alguns objetivos das *overviews*: fornecer uma lista abrangente das RS relevantes atualmente disponíveis, resumir as evidências disponíveis e suas implicações clínicas e / ou políticas e destacar áreas onde é necessária mais pesquisa. Esse estudo relatou que os desafios metodológicos surgiram devido à dificuldade de agrupar informações já sintetizadas ou não comuns entre as RS. Entre os desafios encontrados, destaca-se a dificuldade de classificar a qualidade das evidências incluídas, visto que *overviews* podem ser afetadas por vieses presentes nas RS incluídas. Outro desafio encontrado relaciona-se ao fato de que duas ou mais intervenções podem ser aplicadas simultaneamente para várias condições, dificultando a classificação de intervenções únicas. Segundo esse estudo, *overviews*

não são imunes a falhas, portanto, é importante usar ferramentas que avaliem sua validade. Concluíram que *overviews* são tão úteis quanto as RS e estudos primários em que se baseiam, porém, justamente por se basearem nesses estudos, vieses podem ser introduzidos. Concluíram ainda que há uma perda de informações no processo de síntese de RS.

O grande número de artigos e RS publicadas diariamente foi demonstrado por Bastian et al.¹⁸, em 2010. Segundo os autores, havia um crescimento exponencial de ECRs, o que mostrava a necessidade de agrupá-los. Apesar de demonstrarem isso em seus estudos, ainda havia muita publicação de revisões narrativas, não sistemáticas, que não constituem bases ideais para intervenções terapêuticas. Desde as décadas de 1970 e 1980, RS e MA ganharam força no campo científico, porém, só em 1988 estabeleceu-se a primeira base eletrônica atualizada regularmente para publicação de RS e MA. Em 1993, a colaboração internacional Cochrane foi criada inspirada nessa primeira base eletrônica. Os autores desse estudo concluíram que, para atender às necessidades de pacientes, clínicos e formuladores de políticas, era preciso focar na produção de RS, afim de reduzir ensaios desnecessários.

O estudo publicado por Bown e Sutton¹⁹, em 2010, teve como objetivo orientar pesquisadores responsáveis pela produção das RS e leitores das mesmas. Conferem uma visão geral dos métodos mais utilizados para confecção e interpretação das MA. Segundo os autores, grandes ECRs podem ter maior relevância estatística para assegurar a eficácia de certa intervenção do que as MA isoladamente. Também relataram que, ao planejar uma RS, a questão específica a ser respondida deve estar relacionada à necessidade científica da revisão. Para definir essa questão específica, seria interessante responder ao método PICO (população, intervenção, comparador e outcome). Relataram ainda que novas RS devem ser produzidas caso surjam muitas evidências sobre determinado assunto, para que elas sejam agrupadas. Isso porque seria mais interessante atualizar as RS já existentes, o que não ocorre com tanta frequência. Ainda sobre o planejamento das RS, recomenda-se uma busca da literatura em bases de dados diversas, portais adicionais e, inclusive, uma busca na literatura cinzenta. Deve-se estar bem claro os critérios de inclusão e exclusão durante a seleção dos estudos. Esses critérios não devem ser tão abrangentes e nem muito restritos, porque ambas as formas afetarão

a confecção posterior da MA. Também recomendaram a extração de dados em duplicata e, posteriormente, comparação dos resultados e resolução das discrepâncias em consenso. Fazer uma seleção criteriosa do modelo de MA.

Silva et al.¹, em 2012, relataram a importância e dificuldade dos profissionais da saúde em manter-se atualizados em relação às intervenções clínicas baseadas em evidências científicas de qualidade, devido à alta carga horária de trabalho e afazeres diários. Com isso, o objetivo desse trabalho foi destacar o papel das *overviews* para profissionais da saúde, desenvolvidas para sintetizar informações de RS sobre certa condição, considerando todas as intervenções disponíveis. Como metodologia, usaram três filtros para desenhos de estudos em duas bases de dados. Analisaram o crescimento do número de ensaios clínicos randomizados (ECRs), RS e *overviews* publicados, porque são desenhos de estudo com os níveis mais altos na hierarquia de evidências para a tomada de decisão clínica. Uma equação de predição do número esperado de publicações foi desenvolvida e aplicada. O resultado foi um crescimento exponencial dos ECRs, RS e *overviews*. Conclusão: com o crescimento exponencial das publicações, demonstrado neste estudo, um novo tipo de estudo foi proposto, a OoRs, o qual poderá reduzir incertezas para a tomada de decisão e gerar uma nova hierarquia na pirâmide de evidências. Concluíram que a qualidade dos trabalhos publicados pode ser comprometida por falhas metodológicas, conflitos de interesse, manipulação de dados e outros. Com isso, além de manter-se atualizados, os profissionais da saúde devem saber julgar a qualidade desses trabalhos, baseados em ferramentas como AMSTAR (que avalia a qualidade metodológica das RS) e ROBIS (que avalia o risco de viés). Além disso, concluíram que as *overviews* foram desenvolvidas para facilitar a tomada de decisões clínicas.

Segundo Gurevith et al.²⁰, em 2018, agrupar os resultados de vários estudos é fundamental para compreender um problema e identificar variações nos resultados. Embora a MA tenha facilitado a tomada de decisões baseada em evidências científicas, também gera críticas e controvérsias. A MA tem como objetivos: avaliar a eficácia de certa intervenção ou a associação de causas para certa condição e fornecer uma visão mais geral do que a obtida em um estudo primário individual. Também relataram as limitações da MA como ferramenta para síntese de pesquisas e para embasar decisões. Por outro lado, a introdução das MA

foi o primeiro passo para síntese em pesquisas e, conseqüentemente, facilitar a tomada de decisões por pesquisadores e clínicos.

3.2 Instrumentos de Avaliação de Qualidade e Risco de Viés

Oxman e Guyatt²¹, em 1991, relataram em um estudo que, assim como era importante avaliar a qualidade de estudos primários, também seria importante avaliar a qualidade de Overviews, para auxiliar editores e leitores, por exemplo. Com esse objetivo, desenvolveram critérios para avaliar a qualidade das Overviews e, a partir desses critérios, desenvolveram o questionário Overview Quality Assessment Questionnaire (OQAQ). Em estudos anteriores, os autores descreveram como desenvolveram esses critérios e comprovaram a confiabilidade deles. Nesse artigo, avaliaram a validade do OQAQ para avaliação da qualidade das Overviews. Para isso, avaliaram 36 artigos de revisão, utilizando o OQAQ. Ao todo, esses artigos foram avaliados por 9 examinadores. A sensibilidade do OQAQ foi avaliada por um questionário de 13 itens. Inicialmente, utilizaram 7 hipóteses para avaliar a validade de construto. Para a maioria das 13 questões utilizadas para avaliar a sensibilidade, a pontuação média foi de 5 ou mais, indicando satisfação geral com o instrumento. A principal limitação observada foi a necessidade de julgamento na aplicação do questionário. Seis das 7 hipóteses usadas para testar a validade de construto se confirmaram. Concluíram que o OQAQ tem validade para avaliar a qualidade de Overviews.

Em seu estudo em 2001, Shea et al.²², relataram a importância das MA para tomada de decisões clínicas e como base política para diretrizes práticas baseadas em evidências. Segundo os autores, há um crescente número de MA, o que é bom por um lado e pode ser ruim por outro. Isso porque, muitas vezes, são desenvolvidas sem o rigor necessário, aumentando o risco de incorporação de viés. De acordo com esse estudo, os métodos usados para analisar os artigos a serem incluídos na MA e os métodos utilizados no próprio artigo, devem ser descritos de forma clara. Para isso, já foram sugeridas várias diretrizes. Esse estudo relata o desenvolvimento do QUOROM (Quality of Reporting of Meta-analyses) para avaliar a qualidade de MA de ECRs. Como métodos, foi planejada uma conferência QUOROM em Chicago, de 2 a 3 de outubro de 1996, com 30 indivíduos entre eles: epidemiologistas clínicos, clínicos, estatísticos, pesquisadores e editores. Eles foram divididos em pequenos

grupos e cada um deveria identificar itens que pensavam ser importantes para incluir em uma lista de verificação útil para pesquisadores, editores e revisores. Depois de determinar esses itens mais importantes, cada subgrupo utilizou uma técnica Delphi modificada. As bases de dados pesquisadas incluíram o MEDLINE e a Cochrane Library. Artigos de relevância potencial foram recuperados e avaliados criticamente pelo comitê diretor do QUOROM. Entre os conteúdos discutidos estavam: a busca de evidências; tomada de decisão sobre quais evidências incluir; descrição das características dos estudos primários; síntese de dados quantitativos; confiabilidade e questões relacionadas à validade interna (ou qualidade); e implicações clínicas relacionadas à validade externa (ou generalização). Essa conferência teve como resultado a declaração QUOROM, que consiste em uma lista de verificação para avaliar a qualidade de MA de ECRs. Ela diz respeito a como seria a forma ideal, segundo os autores, de apresentar o resumo, introdução, métodos, resultados e discussão em uma MA. Essa lista é organizada em 21 títulos e subtítulos. Os autores devem fornecer um diagrama de fluxo com informações sobre o número de ECRs identificados, incluídos e excluídos e as razões para excluí-los. Os itens incluídos em um relatório de MA foram escolhidos baseados em evidências, sempre que possível. Os autores relataram que ainda não sabe-se todos os fatores que incluem viés em uma MA, necessitando-se mais pesquisas para melhorar a qualidade de desenvolvimento das MA.

Shea et al.², em 2007, abordaram sobre o desenvolvimento, há mais de uma década, das primeiras ferramentas de avaliação da qualidade metodológica de RS, por Oxman e Guiatt e Sacks. Citaram a existência de mais de 24 instrumentos que realizam essa avaliação, porém, poucos são usados (a maioria constitui-se de instrumentos longos e/ou complicados de aplicar). Na busca por um instrumento adequado de avaliação da qualidade de RS, esse estudo objetivou o desenvolvimento de uma nova ferramenta, com base em ferramentas e evidências prévias, além de considerar opinião de especialistas. Inicialmente, formularam uma ferramenta contendo 37 itens, de dois instrumentos pré-existentes: o Questionário de Avaliação Geral da Qualidade aprimorado (OQAQ), contendo 10 itens e uma lista de verificação criada por Sacks, contendo 24 itens. Aplicaram essa ferramenta a 99 RS em papel e 52 eletrônicas do Banco de Dados de RS Cochrane. Dois revisores, de forma independente, avaliaram a qualidade metodológica de cada revisão incluída,

com base nos 37 itens da ferramenta. A partir de uma análise fatorial, a ferramenta passou para 29 itens. O estudo resultou no desenvolvimento de uma ferramenta de medição para avaliação de múltiplas RS, denominado de AMSTAR, composto por 11 itens.

A importância das RS foi discutida por Shea et al.²³, em 2009. Além disso, destacaram a importância da avaliação da qualidade metodológica das RS, visto que a mesma influencia na interpretação e aplicação dos resultados. Em um estudo anterior, testaram ferramentas atuais de avaliação da qualidade metodológica de RS e concluíram que elas precisavam ser revisadas e atualizadas. De acordo com esse estudo, atualmente, NÃO há um consenso a respeito de qual instrumento usar para medição de qualidade metodológica de RS, destacando a necessidade de revisão e atualização dos atuais instrumentos. Considerando a necessidade de um instrumento de avaliação da qualidade metodológica das RS com boa validade, confiabilidade e viabilidade, criaram o AMSTAR, se baseando nos primeiros instrumentos de medição de qualidade metodológica, criados por Oxman e Guyatt²¹, 1991 e Sacks et al.¹, 1987. O objetivo desse estudo foi avaliar a validade interna do AMSTAR. Usaram uma amostra aleatória de 30 RS das 151 usadas no desenvolvimento do instrumento. Dois revisores (um sem treinamento formal) aplicaram o AMSTAR e as duas ferramentas de avaliação de qualidade, o Overview of Quality Assessment Questionnaire aprimorado (OQAQ, desenvolvido por Oxman e Guyatt²¹) e a lista de verificação desenvolvida por Sacks et al.¹, às 30 RS. AMSTAR já possuía alta validade de face e conteúdo, devido ao método em que foi construído. Com relação à validade do construto, os resultados do AMSTAR foram semelhantes aos resultados dos demais instrumentos. Além disso, mostrou-se um instrumento com viabilidade melhor que o OQAQ e a lista de verificação de Sacks. Como conclusão, sugeriram que o AMSTAR possui boa concordância, confiabilidade, validade interna e viabilidade. Além disso, concluíram que é um instrumento mais abrangente e viável em relação aos existentes. Lembrando que, nesse estudo, AMSTAR foi aplicado apenas a RS de ensaios clínicos randomizados que avaliaram intervenções de tratamento.

O objetivo do estudo de Higgins et al.²⁴, em 2013, foi desenvolver uma ferramenta para melhor avaliação das MA publicadas. Isso porque, segundo os autores, há um crescimento exponencial de publicações de MA e não existem

ferramentas adequadas para avaliar a qualidade metodológica das mesmas. Com esse objetivo, os autores desenvolveram uma ferramenta envolvendo, principalmente, questões estatísticas e interpretativas, utilizando materiais pré-existentes e opiniões de especialistas. Utilizaram o coeficiente Kappa ponderado para medir a concordância entre os itens da ferramenta. A ferramenta era composta por 43 itens, divididos em 4 áreas-chave. Desenvolveram orientações para utilização da ferramenta. A confiabilidade inter-examinador (CIE) foi de razoável a moderada. Os autores consideraram importante que os avaliadores da qualidade de MA, saibam sobre metodologia e sobre o assunto abordado. Concluíram que algumas orientações da ferramenta podem ser melhoradas, que algumas perguntas podem ser desnecessárias e que alguns problemas estão faltando, mas é uma iniciativa para que novas ferramentas de avaliação da qualidade de MA sejam desenvolvidas.

A importância das RS na tomada de decisões clínicas e, conseqüentemente, a importância de se identificar e eliminar possíveis vieses (que podem aparecer em qualquer etapa de confecção de uma RS), foi destacada por Whiting et al.²⁵, em 2016. O desenvolvimento e adoção da declaração PRISMA, contribuiu para melhorias na elaboração de RS. Falhas ou limitações sistemáticas no design ou na condução de uma revisão têm o potencial de influenciar os resultados. Existem várias ferramentas para avaliar a qualidade das RS (nenhuma universalmente aceita). AMSTAR: provavelmente ferramenta de avaliação de qualidade metodológica das RS mais usada. O objetivo desse estudo foi desenvolver uma nova ferramenta (ROBIS), específica para avaliar o risco de viés em RS. Método: Para isso, utilizaram uma abordagem em 4 estágios: Definição do escopo; Revisão da base de evidências; Realização de uma reunião presencial; Refinamento da ferramenta por meio de um piloto. Resultados: ROBIS está atualmente voltado para quatro grandes categorias de revisões, principalmente em ambientes de cuidados de saúde: intervenções, diagnóstico, prognóstico e etiologia. O público-alvo do ROBIS são principalmente desenvolvedores de diretrizes, autores de overviews e autores de revisões que podem querer avaliar ou evitar o risco de viés. A ferramenta é concluída em três fases: (1) avaliar a relevância (opcional), (2) identificar preocupações com o processo de revisão e (3) avaliar o risco de parcialidade. A Fase 2 cobre quatro domínios por meio dos quais o viés pode ser introduzido em

uma RS: critérios de elegibilidade do estudo; identificação e seleção de estudos; coleta de dados e avaliação de estudos; e síntese e resultados. A fase 3 avalia o risco geral de viés na interpretação dos resultados da revisão e se isso considera as limitações identificadas em qualquer um dos domínios da fase 2. Concluíram que o uso de uma abordagem baseada em domínio suportada por perguntas de sinalização segue os métodos mais recentes para o desenvolvimento de ferramentas de risco de viés. ROBIS é a primeira ferramenta desenvolvida com rigor, projetada especificamente para avaliar o risco de viés em RS.

A primeira experiência com a ferramenta ROBIS e a comparação entre ROBIS e AMSTAR quanto a validade, reprodutibilidade e aplicabilidade, foi relatada por Bühn et al.²⁶, em 2017. Esse estudo foi baseado em 16 RS, analisadas por 4 revisores. Quanto aos resultados, é possível observar uma CIE maior para o domínio 2 (identificação e seleção de estudos) e menor para o domínio 4 (síntese e descobertas), sendo essa confiabilidade bem maior no caso do par de revisores experientes. A porcentagem de respostas positivas para cada revisão com o ROBIS foi fortemente correlacionada com as avaliações com a AMSTAR. Concluíram que a ferramenta ROBIS tem boa confiabilidade e validade de constructo quanto à avaliação do risco de viés de RS. Porém, são necessários mais estudos de validade pra analisar a confiabilidade e aplicabilidade da ferramenta. Apontou ainda que a ferramenta ROBIS seria mais indicada para revisores mais experientes, enquanto a AMSTAR pode ser aplicada por revisores não experientes.

Pieper et al.⁹, em 2017, relataram a importância das medições dos estudos serem confiáveis e válidas, visto que serão usadas posteriormente para tomada de decisões. Uma das formas de medições, é a avaliação da qualidade metodológica (ferramenta AMSTAR) e risco de viés (ferramenta ROBIS). A ferramenta AMSTAR foi desenvolvida com base no Overview Quality Assessment Questionnaire (OQAQ) e na lista de verificação criada por Sacks e consiste em 11 itens. Posteriormente, outro grupo de pesquisa criou o R- AMSTAR (uma revisão do AMSTAR), cujo objetivo é quantificar a qualidade metodológica de cada RS, atribuindo um escore de qualidade a cada uma. O objetivo desse estudo foi analisar as diferenças na confiabilidade entre avaliadores do AMSTAR e R-AMSTAR, dependendo do par de revisores. Método: Cinco revisores aplicaram independentemente AMSTAR e R-AMSTAR a 16 RS sobre saúde ocupacional. As respostas foram dicotomizadas e as

medidas de confiabilidade foram calculadas aplicando o método de Holsti (r) e o kappa de Cohen (κ) para todos os pares de revisores em potencial. Dado que cinco revisores participaram do estudo, havia dez pares de revisores possíveis. Resultados: semelhantes – pares de revisores com alta confiabilidade interexaminador no AMSTAR normalmente também tinham alta confiabilidade interexaminador no R-AMSTAR e vice-versa, e a mesma relação foi observada para todas as medidas de confiabilidade. O mesmo par de revisores gerou a maior confiabilidade para ambos os instrumentos. As medidas de confiabilidade kappa pairwise de Cohen mostraram uma correlação moderada entre AMSTAR e R-AMSTAR. A confiabilidade interexaminador média para AMSTAR foi maior para o item 1, enquanto os valores mais baixos foram encontrados para os itens 3, 8, 9 e 11, que apresentaram apenas concordância razoável. Conclusão: A confiabilidade entre avaliadores varia amplamente dependendo do par de revisores. Pode haver algumas deficiências associadas à realização de estudos de confiabilidade com apenas dois revisores.

Segundo relatado por Shea et al.²⁷, em 2017, as RS são as principais fontes de evidência científica para tomada de decisões em saúde. Também relataram o alto crescimento na publicação de RS. Como citado nesse estudo, as RS estão muito sujeitas a vieses, logo, aceitar os resultados delas sem uma avaliação crítica, pode acarretar em tomada de decisões incorretas. Destacaram a importância do AMSTAR (publicado em 2007) como ferramenta de avaliação metodológica de RS de ensaios clínicos randomizados e do AMSTAR 2, que abrangeu estudos randomizados e não randomizados de intervenções em saúde (estudos não randomizados estão mais sujeitos a vieses, exigindo uma avaliação mais rigorosa do risco de viés). Metodologia de desenvolvimento do AMSTAR 2: convocaram um grupo de especialistas, autores do AMSTAR. Usaram uma técnica de grupo nominal para propor e priorizar mudanças específicas no instrumento e chegar a um acordo sobre a redação dos itens. As ideias foram então enunciadas em formato round-robin. Entraram em acordo a respeito da alteração de sete itens no instrumento original, entre eles, simplificar as categorias de resposta. O teste piloto inicial foi realizado por membros do grupo. As versões preliminares foram apresentadas em workshops, onde o feedback direcionou outras modificações e reformulação do instrumento. Resultados: AMSTAR 2 manteve 10 domínios originais do AMSTAR (com mudança

na redação dos itens) e tem 16 itens (enquanto AMSTAR tem 11). Foram adicionadas ao AMSTAR 2 considerações mais detalhadas e separadas do risco de viés para estudos randomizados e não randomizados. As principais modificações incluem categorias de resposta simplificadas; consideração mais detalhada do risco de viés com os estudos incluídos; melhor alinhamento com a estrutura PICO para questões de pesquisa; uma justificativa mais detalhada da seleção de desenhos de estudo para inclusão em uma revisão; e mais informações sobre estudos que foram excluídos das revisões. Os tempos de conclusão para as 20 revisões usadas pelos revisores 1 e 2 variaram de 15 a 32 minutos (o que não inclui o tempo de leitura das revisões). Isso é quase o dobro do tempo necessário para completar o AMSTAR original (10-15 minutos) – aplicado apenas a RSs de ensaios clínicos randomizados. Esse estudo também destacou o surgimento da declaração PRISMA (Itens de relatório preferidos para RSs e meta-análises), criada para melhorar os padrões de relatórios das RSs. Segundo esse estudo, ainda existem relativamente poucos instrumentos que avaliem todas as etapas importantes de uma RS.

O objetivo do estudo de Banzi et al.⁴, em 2018, foi avaliar a confiabilidade entre avaliadores, validade e tempo necessário para aplicação dos instrumentos AMSTAR e ROBIS. Metodologia: estudo transversal. Cinco avaliadores aplicaram AMSTAR e ROBIS a 31 RS. Combinaram domínios semelhantes nas duas ferramentas e pontuações finais, para avaliar a validade (usando a correlação de Kendall tau). Como resultados, a confiabilidade entre avaliadores variou de razoável a perfeita para AMSTAR e de moderada a substancial para ROBIS. Kappa para qualidade geral / risco de viés foi de 0,73 (intervalo de confiança de 95% [IC] 0,65-0,81) para AMSTAR e 0,64 (IC de 95% 0,54-0,74) para ROBIS. A maioria das revisões, quando avaliadas com AMSTAR, apresentaram qualidade intermediária (53%). Já quando foram avaliadas com ROBIS, ficaram divididas em alto (53%) e baixo (47%) risco de viés. O tempo médio de conclusão da ferramenta ROBIS foi cerca do dobro do AMSTAR. Concluíram que AMSTAR e ROBIS apresentam semelhante confiabilidade entre avaliadores, mas são diferentes quanto à construção e aplicabilidade (o tempo médio de conclusão do ROBIS mostra que ele tem menos aplicabilidade do que o AMSTAR – porém, também é preciso considerar que todos os avaliadores estavam mais familiarizados com o AMSTAR). Alguns autores sugeriram que a qualidade explica apenas parte dos vieses dos resultados,

assim, AMSTAR e ROBIS podem ser usados como ferramentas complementares para avaliar RSs.

Dosenovic et al.⁵, em 2018, relataram sobre uma das etapas da metodologia rigorosa usada na confecção de RSs, que consiste em avaliar a qualidade dos estudos primários incluídos, através de ferramentas. Também relatou sobre a existência de ferramentas usadas para avaliar as próprias RSs, como AMSTAR (2007) e R-AMSTAR (2010). AMSTAR é considerada uma ferramenta confiável e válida para avaliar a qualidade metodológica de RSs, mais curta e simples do que R-AMSTAR. O objetivo principal desse estudo era analisar a qualidade metodológica das RSs no campo da NeuP e comparar essas duas ferramentas diferentes para avaliação da qualidade (AMSTAR e R-AMSTAR). A confiabilidade inter-examinadores também foi avaliada nesse estudo. Como metodologia, AMSTAR e R-AMSTAR foram aplicados a todas as RSs incluídas. Aplicou-se a ferramenta AMSTAR e, após 4 a 6 semanas, avaliou-se a qualidade metodológica dessas revisões com a ferramenta R-AMSTAR. As pontuações AMSTAR e R-AMSTAR de cada RS foram posteriormente convertidas em percentuais para cada lista de verificação para permitir a comparação das pontuações de qualidade entre as diferentes ferramentas. Resultados: no geral, ambas as ferramentas produziam classificações de qualidade comparáveis das RSs incluídas. A pontuação média no AMSTAR foi de 5,47, indicando qualidade moderada. Concluíram que, para melhorar a qualidade metodológica das RSs, é necessário um melhor planejamento em sua confecção, com autores, revisores e editores de RSs trabalhando em conjunto. Segundo esse estudo, já existem evidências que mostram que a qualidade metodológica das RSs aumentou com a aplicação do AMSTAR e considerando que ambas ferramentas produziram classificações de qualidade semelhantes, é mais recomendado usar o AMSTAR, por ser mais curto e apresentar melhores resultados que o R-AMSTAR.

O grande crescimento no uso da ferramenta AMSTAR foi relatado por Pieper et al.²⁸, em 2018. Segundo esse estudo, a confiabilidade, validade e aplicabilidade entre avaliadores do AMSTAR foram consideradas satisfatórias. Os objetivos desse estudo foram: analisar como o AMSTAR é aplicado pelos autores e analisar se os autores prestam atenção ao objetivo original do AMSTAR. Como metodologia, realizaram uma pesquisa sistemática na literatura para identificar publicações

relevantes que aplicaram o AMSTAR em pelo menos um estudo incluído. O primeiro grupo consistiu em overviews, nas quais o AMSTAR é frequentemente usado para avaliar a qualidade metodológica das revisões incluídas e o segundo grupo consistiu em estudos sobre a qualidade metodológica das RSs. Os autores usaram vários métodos diferentes para categorizar as revisões com base em sua pontuação no AMSTAR. Resultados: na maioria dos casos (66%), o AMSTAR foi aplicado a avaliações incluídas, enquanto 29% dos autores não relataram o procedimento. A confiabilidade interavaliadores da avaliação AMSTAR foi relatada em 58 dos 247 (23%) estudos. Na maioria dos estudos (86%), não ficou claro qual versão do AMSTAR foi aplicada. A maioria dos autores (31%), usou a seguinte categorização em relação à qualidade das RSs: baixa (pontuação 0 a 4), média (pontuação 5 a 8) e alta (pontuação 9 a 11). Na discussão, relataram que existe uma grande diferença quando o AMSTAR é aplicado pelos autores para avaliar a qualidade metodológica das revisões. Também relataram que deve-se evitar usar uma pontuação geral. Segundo esse estudo, os autores de revisões metodológicas parecem estar muito mais conscientes das possíveis limitações ao aplicar o AMSTAR. As diferenças na aplicação do AMSTAR podem estar relacionadas aos níveis de conhecimento e experiência dos usuários. Por fim, relataram o desenvolvimento recente de duas novas ferramentas para a avaliação de RSs. ROBIS (2016), para avaliar o risco de viés nas RSs. AMSTAR 2 (AMSTAR modificada): além de analisar a qualidade metodológica de RSs de ensaios clínicos randomizados, também avalia de ensaios clínicos não randomizados. Concluíram que autores, revisores e editores de RSs devem prestar mais atenção ao uso correto das ferramentas de avaliação das RSs e que os autores das overviews devem ter um especialista em metodologia em sua equipe de revisão.

O objetivo do estudo de Gates et al.²⁹, em 2018, foi investigar a confiabilidade entre avaliadores, usabilidade e aplicabilidade das seguintes ferramentas de avaliação de RSs: AMSTAR, AMSTAR 2 e ROBIS. Métodos: usaram uma amostra aleatória de 30 RSs de ensaios clínicos randomizados. Para comparabilidade com as avaliações anteriores, também calcularemos as estatísticas kappa de Cohen ponderadas e kappa de Fleiss. Para estimar a usabilidade, calcularemos o tempo médio para concluir a avaliação e chegar a um consenso para cada ferramenta. Para informar as aplicações das ferramentas, testaremos as associações estatísticas

entre os escores de qualidade e os julgamentos de risco de viés, e os resultados e conclusões das RSs. Destacou a importância de documentos de orientação para condução de RSs, como o Manual Cochrane para RSs de Intervenções e o PRISMA. Segundo esse estudo, a baixa qualidade de muitas RSs dificulta o processo de conduzir overviews. Também destacou AMSTAR como ferramenta mais frequentemente utilizada para classificar RSs. Embora se possa presumir que RSs de baixa qualidade correm alto risco de viés e vice-versa, nem sempre é o caso. Além disso, mesmo RSs bem conduzidas podem apresentar alto risco de viés. Para avaliar o risco de viés em RSs, foi desenvolvida a ferramenta ROBIS. Resultados: AMSTAR, AMSTAR 2 e ROBIS exibem boa validade de face e conteúdo. Com relação à confiabilidade entre avaliadores, Pieper et al. revisou estudos que examinaram as propriedades de medição do AMSTAR e encontraram confiabilidade substancial para a maioria dos itens. Os desenvolvedores do AMSTAR 2 relataram confiabilidade moderada ou melhor para a maioria dos itens em uma amostra de 20 RSs de intervenções de saúde. Com relação à usabilidade, os relatórios indicam que o AMSTAR leva de 10 a 20 minutos para concluir e o AMSTAR 2 leva de 15 a 32 minutos. ROBIS leva ainda mais tempo. Destacou a necessidade de mais pesquisas e evidências em relação ao AMSTAR 2 e ROBIS, que são ferramentas bem recentes. Ainda segundo esse estudo, análises descritivas de overviews sugerem que os autores não assumem o risco de viés e / ou avaliações de qualidade de RS antes de tirar conclusões. Esse bloqueio em relação ao uso de ferramentas para avaliar RSs, pode ser devido ao tempo real ou percebido para aplica-las e os recursos necessários para concluí-las, além da confiança dos revisores em suas próprias avaliações.

Pieper et al.³⁰, em 2019, destacaram a importância de uma avaliação crítica das RSs, empregando ferramentas como AMSTAR, AMSTAR 2 e ROBIS. AMSTAR possui boas propriedades de medição em termos de validade, confiabilidade e aplicabilidade. AMSTAR 2 foi desenvolvido como forma de avaliar a qualidade metodológica de RSs que incluem ensaios clínicos randomizados e não randomizados. O objetivo desse estudo foi relatar uma primeira experiência com o AMSTAR 2 e compará-lo com o ROBIS, em relação à validade, confiabilidade e aplicabilidade. Métodos: analisaram 30 RS que incluíram ensaios clínicos de estudos randomizados e não randomizados. Quatro revisores avaliaram todas as 30 RSs de

forma independente, usando AMSTAR 2 e, depois, ROBIS. Nenhum revisor foi formalmente treinado em uma das ferramentas. Todas as RSs foram avaliadas na mesma ordem pelos quatro revisores. A ordem foi aleatória. Para avaliar a confiabilidade entre avaliadores, calcularam o Kappa de Fleiss. Resultados: confiabilidade entre avaliadores razoável, tanto no AMSTAR 2 ($\kappa = 0,30$, 95% - intervalo de confiança) quanto no ROBIS ($\kappa = 0,28$, 95%). O tempo médio de conclusão foi um pouco maior para AMSTAR 2 (18 min) do que para ROBIS (16 min), com muita diferença entre os revisores. Concluíram que ambas ferramentas podem ser aplicadas a RSs que incluem ensaios clínicos randomizados e não randomizados. Ambos, AMSTAR 2 e ROBIS, são considerados uma melhoria em comparação com o AMSTAR no que diz respeito à avaliação da qualidade metodológica / risco de viés nas RSs.

Lorenz et al.³¹, em 2019, determinaram a confiabilidade inter-examinadores na avaliação de múltiplas RS com a AMSTAR 2. Compararam AMSTAR e ROBIS para avaliar a validade convergente entre essas ferramentas. Para isso, utilizaram duas amostras de 30 RSs, avaliadas por dois grupos de quatro avaliadores. Cada grupo aplicou AMSTAR ou ROBIS à sua amostra, além do AMSTAR 2. Nos resultados, observa-se um acordo moderado para AMSTAR 2, substancial para AMSTAR e justo para ROBIS. Quanto à validade, os resultados mostraram uma associação positiva entre AMSTAR e AMSTAR 2 e entre ROBIS e AMSTAR 2. Para a avaliação geral, AMSTAR 2 mostrou alta concordância com ROBIS e baixa com AMSTAR. Em conclusão, a confiabilidade inter-examinadores da AMSTAR 2 foi um pouco menor do que o AMSTAR, e maior do que ao ROBIS. As medições de validade indicam que o AMSTAR 2 está intimamente relacionado ao ROBIS e ao AMSTAR.

Lecerq et al.³², em 2019, relataram o recente estudo de Lorenz et al, comprovando a validade e CIE da ferramenta AMSTAR 2. O objetivo desse estudo foi avaliar a confiabilidade entre AMSTAR e AMSTAR 2 e medir a CIE do AMSTAR 2. Seguiram a metodologia adotada no estudo de Lorenz et al, calculando o coeficiente Kappa de Fleiss, para itens dicotômicos e Kappa ponderado linear para itens categóricos, para determinar a CIE do AMSTAR 2. Para medir a validade convergente entre AMSTAR e AMSTAR 2, primeiro, avaliaram a concordância da classificação geral para AMSTAR e AMSTAR 2. Depois, a quantidade de itens

preenchidos por cada MA com AMSTAR e AMSTAR 2, com uma correlação de Spearman. Como resultados, de acordo com a classificação recomendada por AMSTAR 2, a maioria das amostras de MA, foi classificada com qualidade criticamente baixa. Já com a ferramenta AMSTAR, seguindo a classificação recomendada por Lorenz, a maioria das amostras de MA foi classificada com qualidade média. A correlação entre AMSTAR e AMSTAR 2 indicou forte associação positiva. Concluíram que o AMSTAR 2 possui uma CIE moderada, assim como identificado no estudo de Lorenz. Segundo os autores, a capacidade discriminativa do AMSTAR 2 não é tão boa, necessitando de mais investigações a respeito.

Puga et al.³³, em 2020, relataram que uma das formas de melhorar a qualidade de publicações científicas, é utilizando os instrumentos disponíveis para avaliação e confecção de estudos científicos, para ajudar a buscar informações e, assim, desenvolver uma melhor estratégia de busca. Segundo os autores, esses instrumentos colaboram na síntese de informações sobre determinado estudo, conferindo a ele mais qualidade metodológica, rigor, confiabilidade e reprodutibilidade. Além disso, orientam editores, revisores, pesquisadores e leitores. Nesse estudo, foram listados quatro instrumentos com esse fim: MECIR (Methodological Expectations for Cochrane Intervention Reviews); PRISMA; AMSTAR e PRESS (Peer Review of Electronic Search Strategies). O instrumento MECIR tem como objetivo checar e guiar, de maneira rigorosa, as buscas na literatura, sendo obrigatório na confecção de RS Cochrane. Segundo os autores desse estudo, “os instrumentos PRESS e MECIR fornecem mais detalhes para a realização de buscas do que PRISMA e AMSTAR, incluindo orientações detalhadas para esta etapa e maior rigor”. Uma boa busca na literatura deve envolver busca em múltiplas bases de dados, literatura cinza, experts no assunto, refazer a busca antes de publicar a RS, usar corretamente os termos de buscas para cada base de dados, checar a lista de referência dos artigos incluídos na RS, documentar a estratégia de busca juntamente com o protocolo da RS junto ao PROSPERO, justificar restrições de busca como línguas e ano de publicação, estruturar a pergunta da RS e sua estratégia de busca utilizando a estrutura PICO, usar corretamente os filtros nas bases de pesquisa, investigar se os termos utilizados são muito amplos ou restritos, entre outros fatores. Segundo os autores, bibliotecários podem ajudar nesse processo.

Em seu estudo publicado em 2020, Martinez-Zapata et al.³⁴ analisaram a CIE em relação ao AMSTAR 2, entre metodologistas recentes e experientes na avaliação de RS. As bases de dados utilizadas para busca de estudos relacionados ao assunto foram MEDLINE e Cochrane Library. Essa busca foi realizada por dois revisores independentes. Calculou-se o coeficiente Kappa para os 16 itens do AMSTAR 2, variando de excelente (Q1) a ruim (Q3, Q5 e Q7). Concluíram que o AMSTAR 2 tem CIE razoável entre metodologistas recentes e experientes na avaliação de RS. Porém, algumas questões específicas exigem mais treinamento, para melhorar a CIE.

Em seu estudo divulgado em 2020, Chapman³⁵ alertou a respeito do crescente número de RS e MA publicadas, principalmente devido ao surgimento da “medicina baseada em evidências”. Segundo os autores, quando bem realizadas, as RS oferecem uma visão abrangente, visto que as mesmas reúnem vários estudos a respeito da questão específica a ser respondida; fornecem várias citações relevantes, principalmente quando trata-se de um tema novo, e apresentam uma visão sobre o risco de viés dos estudos isolados. Porém, como houve uma “proliferação repentina” das RS e MA, foi criada a declaração PRISMA (Preferred Reporting Items for Systemic Reviews and Meta-Analyses), com o objetivo de melhorar a qualidade das mesmas. Essa declaração PRISMA foi criada em 1996, a partir do QUORUM (Quality Of Reporting Of Meta-analyses) e a partir de definições usadas pela Colaboração Cochrane. Diferente do QUORUM, a declaração PRISMA passou a incluir RS com o objetivo de fornecer aos autores uma ferramenta de checklist para a redação de uma RS ou MA de qualidade. Segundo esse estudo, PRISMA, AMSTAR e AMSTAR 2 são complementares na avaliação da qualidade das RS. Além disso, relataram que as ferramentas de avaliação tratam-se de um campo ainda em evolução.

Luchini et al.³⁶, em 2020, destacaram a importância das RS para a tomada de decisões em saúde. Segundo esse estudo, a qualidade, validade e confiabilidade da RS depende da qualidade dos estudos incluídos e da metodologia aplicada na confecção das RS. Nesse estudo, os autores abordam as vantagens, desvantagens e principais características das principais ferramentas de avaliação de qualidade de estudos individuais, RS e MA, sendo essas ferramentas: escala Newcastle-Ottawa (NOS) e a Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE), para estudos observacionais; Consolidated Standards of Reporting Trials (CONSORT), a escala Jadad, e a ferramenta da Cochrane sobre o risco de viés (RoB2) para ensaios clínicos randomizados (ECRs); PRISMA; AMSTAR 2 e AMSTAR- PLUS (Assessment of Multiple Systematic Reviews PLUS, para RS e MA. Relatam a importância das RS como o mais alto nível de evidência científica, porém, também alertam sobre o grande risco de viés. Segundo os autores, a maioria dos vieses é incorporada nas RS a partir dos estudos incluídos, destacando a importância de se avaliar a qualidade desses estudos a partir de ferramentas validadas e padronizadas. De acordo com esse estudo, a ferramenta NOS tem como principais vantagens versatilidade e ampla aplicabilidade. Entre suas limitações, destacam-se total conhecimento a respeito do tema abordado para sua interpretação e baixa concordância entre os coautores. A respeito da ferramenta STROBE, os autores desse estudo relatam não ser possível aplicá-la a 100% dos estudos observacionais, porém, quando é aplicável, é confiável e reproduzível para avaliação da qualidade de estudos observacionais incluídos em RS. Para os autores, é uma ferramenta útil para avaliar se o estudo atende a todos os itens exigidos. O instrumento CONSORT não foi criado especificamente para avaliar qualidade, mas é fundamental em RS que incluem ensaios clínicos randomizados. É uma ferramenta em constante evolução, logo, exige pesquisa constante por parte dos avaliadores que a utilizarão. A escala Jadad é uma ferramenta para avaliar qualidade de ECRs. Sua avaliação é ampla, sendo necessário outros instrumentos para uma avaliação mais detalhada de ECRs. Não é um instrumento muito preciso. O instrumento RoB2 é o padrão-ouro para avaliar a qualidade e a presença de viés em ECRs. Logo, é o mais recomendado para RS de alta qualidade. A declaração PRISMA foi criada para servir como um checklist na redação de RS, porém, também pode servir como uma medida de qualidade da RS. Não é recomendada para avaliar qualidade de MA. A ferramenta AMSTAR 2 é mais indicada e válida para avaliar a qualidade de MA, porém, deve-se utilizar outros instrumentos para complementar essa avaliação. AMSTAR- PLUS avalia a qualidade metodológica da MA em si e dos estudos incluídos na MA. Concluíram que, as ferramentas mais indicadas para avaliação da qualidade ao se tratar de um estudo individual ou RS, são STROBE, CONSORT e PRISMA. Para avaliação de estudos observacionais em RS, a ferramenta NOS é a mais indicada, visto que é precisa e exige menos tempo para

ser completada. Já quando for avaliar a qualidade de ECRs, a ferramenta RoB2 é a mais recomendada. Para avaliar a qualidade metodológica das MA, a ferramenta AMSTAR2 é a mais indicada, mas, se além disso, quiser avaliar a qualidade do conteúdo da MA, AMSTAR-PLUS é mais indicado. Segundo os autores, o ideal seria uma ferramenta para avaliar estudos observacionais e uma para avaliar qualidade e risco de viés dos estudos de intervenção, porém, como isso não ocorre, o pesquisador deve analisar qual a ferramenta mais adequada para seu estudo.

Em um estudo em 2021, Pieper et al.³⁷ também relataram a importância de RS de alta qualidade para a tomada de decisões na área da saúde. Além disso, também relataram que o aumento exponencial de RS publicadas, tem prejudicado a qualidade das mesmas, reforçando a necessidade de seguir uma metodologia rigorosa para a confecção das revisões. Os autores citaram a respeito do desenvolvimento de diversos instrumentos para avaliar a qualidade metodológica das RS, porém, poucos com validade e confiabilidade testada. O objetivo desse estudo foi analisar se os pesquisadores descrevem como aplicam o AMSTAR2, se aplicam diferente do proposto pelos desenvolvedores e se a forma de aplicar pode interferir nos resultados. Em relação aos métodos, trata-se de um estudo transversal, dividido em duas partes. A primeira parte comparou três formas de aplicação diferentes, para avaliar a confiança geral nos resultados da RS. Para generalizar mais os resultados desse estudo, usaram duas amostras. Uma de 60 RS sobre “intervenções farmacológicas ou psicológicas no tratamento de depressão maior em adultos”, e outra de 58 RS sobre “terapia cognitivo-comportamental em transtornos psiquiátricos em adultos, publicadas de 2011 a 2015”. A avaliação da confiança geral das amostras foi realizada por dois revisores, de forma independente. A segunda parte analisou a aplicação do AMSTAR2 pelos autores. Foram incluídos todos os artigos de língua inglesa que aplicaram o AMSTAR2 em pelo menos uma RS. Essa seleção foi realizada por dois revisores, de forma independente. Considerando a primeira parte desse estudo, como resultados, a classificação de confiança geral nos resultados das revisões foi criticamente baixa na maior parte das 60 RS da primeira amostra, para as três formas de aplicação realizadas. Isso também ocorreu na segunda amostra, de 58 RS. Considerando a segunda parte desse estudo, 70% usou a classificação original com quatro classificações diferentes de confiança geral. Dessas, menos da metade relatou como

chegaram à classificação geral. A maioria relatou usar os sete domínios críticos do AMSTAR 2. Cinco estudos relataram que desenvolveram suas próprias regras para obter uma classificação geral. Um artigo adotou a classificação original juntamente com uma classificação própria. Os autores discutiram sobre a importância de relatar como obteve-se a classificação geral, visto que, segundo eles, isso pode afetar a classificação geral do AMSTAR 2. Como a maioria das RS foi classificada com qualidade metodológica criticamente baixa, reforçaram a necessidade dos autores de RS relatarem melhor seus métodos de aplicação. A classificação de confiança geral com AMSTAR 2, também depende da escolha de itens críticos. Concluíram que o método de obter confiança geral deve ser melhor relatado pelos autores, visto que isso auxilia a tomada de decisões e influencia diretamente na classificação geral do AMSTAR 2 e esse relatório “deve permitir a reprodução das classificações gerais para editores, revisores e leitores”.

Buhn et al.³⁸, em 2021, também relataram a importância de RS de alta qualidade para a tomada de decisões baseada em evidências. O objetivo desse estudo foi avaliar a confiabilidade do teste reteste (TRR) do AMSTAR, para avaliar a importância de sua medição visto que, segundo os autores, essa ferramenta parece ter validade e confiabilidade inter observadores (IRR) já testada. Quanto aos métodos, a avaliação da qualidade metodológica das RS foi realizada com o AMSTAR, por sete avaliadores, em três instituições, de forma independente, sendo essas RS sobre saúde ocupacional. As respostas foram dicotomizadas e calculou-se o TRR de todos os avaliadores e itens do AMSTAR usando o coeficiente AC1 de Gwet. Nos resultados, o menor TRR foi encontrado para o item 4 do AMSTAR, que diz respeito ao status de publicação como critério de inclusão. O item 1 apresentou concordância perfeita. Concluíram que é importante considerar o TRR ao avaliar as propriedades de medição do AMSTAR, mas reforçaram a necessidade de mais evidências e também a necessidade de avaliar o TRR em outros instrumentos de classificação, como AMSTAR 2 e ROBIS.

Em uma carta ao editor, no Journal of Clinical Epidemiology, em 2021, Franco et al.³⁹ expuseram suas opiniões à respeito da importância de uma maior transparência ao aplicar o AMSTAR 2. Segundo os autores, as ferramentas RoB2, ROBINS-I e ROBIS não possuem boa confiabilidade inter-examinadores, assim como o AMSTAR 2. Porém, diferente do AMSTAR 2, essas outras ferramentas

incluem um “suporte para julgamento”, permitindo que editores, revisores e leitores, julguem se concordam ou não com os critérios dos avaliadores, o que torna as avaliações mais transparentes. Portanto, os autores desse estudo destacam a relevância de incluir o “suporte para julgamento” no AMSTAR 2, para tornar as avaliações mais claras.

Lorenz et al.⁴⁰, também em uma carta ao editor, no *Journal of Clinical Epidemiology*, em 2021, reforçaram o mesmo relatado por Franco e Meza's no mesmo ano, acrescentando que “o julgamento também pode ocorrer na classificação geral de confiança” e, portanto, os autores devem informar de forma clara como obtiveram a classificação geral a partir do AMSTAR 2. A aplicação do AMSTAR 2 pode envolver a inclusão de vários vieses, por isso os autores alertaram que as avaliações de qualidade devem ser descritas da forma mais clara possível.

4 DISCUSSÃO

Com o surgimento da medicina/ odontologia baseada em evidências, as RS com ou sem MA vem ganhando destaque na literatura científica nos últimos anos¹¹. Isso porque elas agrupam dados de vários estudos a respeito de uma questão específica, facilitando a tomada de decisões em saúde, por exemplo, sobre qual intervenção terapêutica seria mais indicada em determinada situação ou diante de um determinado diagnóstico¹¹. Como reúnem resultados de vários estudos clínicos independentes, permitem melhor avaliação das intervenções disponíveis e evidenciam as questões de pesquisa mais urgentes ou que precisam ser melhores estudadas, além de evidenciar intervenções benéficas, evitando novas pesquisas desnecessárias sobre o assunto¹¹.

Esse agrupamento de vários estudos a respeito de determinado assunto, em uma RS, tornou-se ainda mais relevante mediante o crescimento exponencial de pesquisas e publicações científicas, ao final da Segunda Guerra Mundial¹⁶. Como os profissionais de saúde apresentam uma alta carga horária de trabalho e afazeres diários, reunir dados de diferentes estudos em RS e MA, tornou-se essencial para auxiliá-los na tomada de decisões baseadas no mais alto nível de evidência científica e para que eles se mantenham atualizados em relação às intervenções clínicas¹. Porém, esse grande aumento no número dessas publicações, fez com que muitos autores não se preocupassem em seguir uma metodologia rigorosa na confecção das mesmas, o que tem prejudicado a qualidade dessas revisões e aumentado muito o risco de incorporação de vieses^{22,37}. Logo, embora RS com ou sem MA tenham facilitado a tomada de decisões baseada em evidências científicas, também têm gerado críticas e controvérsias²⁰.

A elaboração de uma RS é bem complexa, envolve vários passos³⁵. Com a criação da declaração PRISMA, em 1996, houve uma melhora na redação das RS e, conseqüentemente, na qualidade das mesmas, visto que funcionava como uma ferramenta de checklist em que os autores verificavam se estavam presentes todos os itens necessários para a redação de uma RS com qualidade³⁵. Porém, ainda assim sentiu-se a necessidade de desenvolver ferramentas que avaliassem, especificamente, a qualidade metodológica e o risco de viés das RS⁹. Isso porque as RS devem ser confiáveis e válidas, uma vez que serão utilizadas, posteriormente, para tomada de decisões⁹. Além disso, as RS estão muito sujeitas a vieses, logo,

aceitar os resultados delas sem uma avaliação crítica, pode acarretar em tomada de decisões incorretas²⁷.

Com o objetivo de melhorar a qualidade das RS, mais de 24 instrumentos foram desenvolvidos, porém, um número relativamente pequeno avalia todas as etapas importantes na construção das mesmas². A maioria não é utilizada, por serem extensos ou não apresentarem instruções claras sobre seu uso². O primeiro instrumento para avaliar a qualidade das RS foi desenvolvido por Oxman e Guyatt, 1991²¹ e Sacks, em 1987¹⁰. O questionário denominado de OQAQ apresentava boa confiabilidade e validade para avaliar a qualidade das revisões, mas, a principal limitação era a necessidade de julgamento na aplicação do questionário²¹. Anos mais tarde, em 2001, foi desenvolvida a declaração QUOROM, uma lista de verificação para avaliar a qualidade de MA de ECRs²².

Atualmente, uma das ferramentas mais utilizadas para auxiliar os leitores de RS, especificamente, de ensaios clínicos randomizados, a analisar criticamente a qualidade metodológica delas, é o AMSTAR (Avaliação de Revisões Sistemáticas Múltiplas, do inglês Assessment of Multiple Systematic Reviews)⁴. Esse instrumento foi desenvolvido em 2007 e é constituído por uma lista de verificação de 11 itens, que avalia a metodologia aplicada em cada etapa de construção de uma RS⁴. Estudos recentes demonstram que o AMSTAR possui boa confiabilidade e validade⁹. Em 2010, o AMSTAR passou por uma revisão, levando ao desenvolvimento do instrumento R- AMSTAR^{5,9}. Posteriormente, em 2017, foi desenvolvido o AMSTAR 2, uma nova atualização do AMSTAR⁵. Essa lista de verificação contém agora 16 itens, mantendo 10 da ferramenta original, e faz uma avaliação mais ampla em comparação com o AMSTAR, visto que abrange estudos randomizados e não randomizados de intervenções em saúde², porém, leva o dobro do tempo para ser preenchida, em relação ao AMSTAR³⁴. O AMSTAR 2 possui CIE razoável entre metodologistas recentes e experientes na avaliação de RS³⁴. Porém, algumas questões específicas exigem mais treinamento, para melhorar a CIE³⁴. Além disso, a capacidade discriminativa do AMSTAR 2 ainda precisa ser melhorada, necessitando de mais investigações a respeito³². Apesar da necessidade de aprimoramento, o uso dessas ferramentas para avaliar a qualidade metodológica das RS, tem sido importante para alertar sobre o fato de que muitas dessas revisões não estão sendo conduzidas de maneira adequada⁵.

Outra importante ferramenta desenvolvida recentemente avalia, especificamente, o risco de viés nas RS (ROBIS, do inglês Risk Of Bias In Systematic Reviews)²⁵. O uso dessa ferramenta ainda possui algumas limitações²⁵. Deve ser utilizado por revisores de diferentes áreas, embora seja relevante algum conhecimento metodológico e / ou de conteúdo²⁴. O AMSTAR é considerado mais simples que o ROBIS, cuja execução é mais demorada e requer maior conhecimento metodológico e estatístico e experiência em atuação clínica⁴. O AMSTAR avalia mais a metodologia da revisão em questão, enquanto o ROBIS se volta mais para os resultados⁴. Alguns estudos mostram melhora da qualidade metodológica das revisões após introdução do AMSTAR. Essas duas ferramentas podem ser utilizadas como complementares para avaliar as RS⁴.

Quanto à validade, há uma associação positiva entre AMSTAR e AMSTAR 2 e entre ROBIS e AMSTAR 2³¹. A CIE da AMSTAR 2 é um pouco menor do que o AMSTAR, e maior do que o ROBIS³¹. As medições de validade indicam que o AMSTAR 2 está intimamente relacionado ao ROBIS e ao AMSTAR³¹. AMSTAR, AMSTAR 2 e ROBIS exibem boa validade de face e conteúdo²⁹. Com relação à aplicabilidade, ROBIS é a que leva mais tempo para ser concluída²⁹.

Também devemos destacar o desenvolvimento do instrumento MECIR (do inglês Methodological Expectations for Cochrane Intervention Reviews), com o objetivo de checar e guiar, de maneira rigorosa, as buscas na literatura³³. Os instrumentos PRESS (do inglês Peer Review of Electronic Search Strategies) e MECIR fornecem mais detalhes para a realização de buscas do que PRISMA e AMSTAR, além de orientações detalhadas para esta etapa e maior rigor³³. A declaração PRISMA foi criada para servir como um checklist na redação de RS, porém, também pode auxiliar numa análise prévia de qualidade da RS, apesar de não ser indicada avaliar qualidade de MA^{35,36}. A ferramenta AMSTAR 2 é mais indicada e válida para avaliar a qualidade de MA, porém, deve-se utilizar outros instrumentos para complementar essa avaliação³⁶. AMSTAR- PLUS avalia a qualidade metodológica da MA em si e dos estudos incluídos na MA³⁶.

Apesar de existirem várias ferramentas que avaliam a qualidade das RS²⁵, relativamente poucas avaliam todas as etapas importantes de uma RS²⁷ e poucas possuem validade e confiabilidade testada³⁷. Além disso, ainda não há um consenso aceito a respeito de qual ferramenta usar²⁵. As ferramentas atuais de avaliação da

qualidade metodológica de RS precisam ser constantemente revisadas e atualizadas²³.

De forma geral, para melhorar a qualidade metodológica das RS, autores, revisores e editores devem trabalhar em conjunto, planejando melhor a confecção das RS e utilizando de forma correta as ferramentas adequadas para avaliar a qualidade e risco de viés das mesmas^{5,28,33}. Já existem evidências que mostram que a qualidade metodológica das RS aumentou com a aplicação do AMSTAR⁵. São necessárias mais pesquisas e evidências a respeito do AMSTAR 2 e ROBIS, que são ferramentas bem recentes²⁹. Além disso, necessita-se de uma melhor instrução quanto ao uso e interpretação das ferramentas de avaliação mais recentes, principalmente o ROBIS²⁹.

É de extrema importância ressaltar que a elaboração do protocolo bem como a condução da metodologia proposta numa RS exige que a equipe possua habilidades e conhecimentos de metodologia científica e estatística bem fundamentados. Todas as etapas de uma RS são importantes e devem ser bem compreendidas por aqueles que as desenvolvem e aqueles que as utilizam como fonte de evidências científicas. Na graduação, o conhecimento em metodologia científica para avaliar estudos, seria muito relevante, pois, quando entende-se como é construída uma RS, principalmente do ponto de vista metodológico, fica mais fácil interpretar os dados dos estudos primários e das RS, entender a qualidade daquele estudo ou revisão, verificar a existência de viés e basear o conhecimento nisto, realizando uma prática clínica baseada em evidências. Na pós-graduação, saber a respeito de metodologia científica tem uma relevância ainda maior, visto que os estudantes estão envolvidos na área de pesquisa e, portanto, devem estar preparados para interpretar e conduzir pesquisas e RS com a melhor qualidade possível, para ajudar na propagação das decisões em saúde baseadas em evidências. Para isso, é extremamente importante que estudantes e professores entendam as etapas de construção de uma boa RS e saibam aplicar as ferramentas para avaliar as revisões.

7 CONCLUSÃO

O desenvolvimento e adoção da declaração PRISMA contribuiu para melhorias na redação e, conseqüentemente, na elaboração e qualidade das RS. Entre as ferramentas específicas para análise de qualidade das RS publicadas, destacam-se o AMSTAR, AMSTAR 2 e AMSTAR-PLUS. Já a ferramenta ROBIS é a primeira desenvolvida com rigor, para avaliar, especificamente, o risco de viés em RS. AMSTAR e ROBIS podem ser usadas como ferramentas complementares para avaliar RS e tais instrumentos de avaliação devem ser entendidos como um campo em constante evolução.

REFERÊNCIAS*

1. Silva V, Grande AJ, Martimbianco AL, Riera R, Carvalho AP. Overview of systematic reviews: a new type of study: part I: why and for whom? Sao Paulo Med J. 2012; 130(6): 398-404.
2. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007; 7: 10.
3. Cordeiro AM, Oliveira GMd, Rentería JM, Guimarães CAJRdCBdC. Revisão sistemática: uma revisão narrativa. 2007; 34(6): 428-31.
4. Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi M, Minozzi S. Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. J Clin Epidemiol. 2018; 99: 24-32.
5. Dosenovic S, Jelicic Kadic A, Vucic K, Markovina N, Pieper D, Puljak L. Comparison of methodological quality rating of systematic reviews on neuropathic pain using AMSTAR and R-AMSTAR. BMC Med Res Methodol. 2018; 18(1): 37.
6. Clarke M, Chalmers I. Reflections on the history of systematic reviews. BMJ Evid Based Med. 2018; 23(4): 121-2.
7. Osborn JF, Bulman JS, Petrie A. Further statistics in dentistry. Part 10: Sherlock Holmes, evidence and evidence-based dentistry. Br Dent J. 2003; 194(4): 189-95.
8. Bassani R, Pereira GKR, Page MJ, Tricco AC, Moher D, Sarkis-Onofre R. Systematic reviews in dentistry: current status, epidemiological and reporting characteristics. J Dent. 2019; 82: 71-84.
9. Pieper D, Jacobs A, Weikert B, Fishta A, Wegewitz U. Inter-rater reliability of AMSTAR is dependent on the pair of reviewers. BMC Med Res Methodol. 2017; 17(1): 98.
10. Sacks HS, Berrier J, Reitman D, Ancona-Berk V, Chalmers TCJNEJoM. Meta-analyses of randomized controlled trials. 1987; 316(8): 450-5.
11. Egger M, Smith GD. Meta-Analysis: potentials and promise. BMJ. 1997; 315(7119): 1371-4.
12. Oxman AD. Checklists for review articles. BMJ. 1994; 309(6955): 648-51.
13. Atallah AN. Systematic reviews of medical literature and metanalyses. Sao Paulo Med J. 1996; 114(5): 1285-7.

14. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ*. 2000; 320(7234): 537-40.
15. CASTRO AAJ. Revisão sistemática e meta-análise. 2001.
16. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof*. 2002; 25(1): 12-37.
17. Thomson D, Russell K, Becker L, Klassen T, Hartling L. The evolution of a new publication type: steps and challenges of producing overviews of reviews. *Res Synth Methods*. 2010; 1(3-4): 198-211.
18. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010; 7(9): e1000326.
19. Bown MJ, Sutton AJ. Quality control in systematic reviews and meta-analyses. *Eur J Vasc Endovasc Surg*. 2010; 40(5): 669-77.
20. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nature*. 2018; 555(7695): 175-82.
21. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991; 44(11): 1271-8.
22. Shea B, Dubé C, Moher DJ, Srinivasan S, Altman DG. Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. 2001; 323(7306): 122-39.
23. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009; 62(10): 1013-20.
24. Higgins JP, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. *Res Synth Methods*. 2013; 4(4): 351-66.
25. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016; 69: 225-34.
26. Buhn S, Mathes T, Prengel P, Wegewitz U, Ostermann T, Robens S, et al. The risk of bias in systematic reviews tool showed fair reliability and good construct validity. *J Clin Epidemiol*. 2017; 91: 121-8.
27. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017; 358: j4008.
28. Pieper D, Koensgen N, Breuing J, Ge L, Wegewitz U. How is AMSTAR applied by authors - a call for better reporting. *BMC Med Res Methodol*. 2018; 18(1): 56.

29. Gates A, Gates M, Duarte G, Cary M, Becker M, Prediger B, et al. Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: protocol for a descriptive analytic study. *Syst Rev*. 2018; 7(1): 85.
30. Pieper D, Puljak L, Gonzalez-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *J Clin Epidemiol*. 2019; 108: 26-33.
31. Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol*. 2019; 114: 133-40.
32. Leclercq V, Beaudart C, Tirelli E, Bruyere O. Psychometric measurements of AMSTAR 2 in a sample of meta-analyses indexed in PsycINFO. *J Clin Epidemiol*. 2020; 119: 144-5.
33. Puga M, Atallah AN. What editors, reviewers, researchers and librarians need to know about the PRESS, MECIR, PRISMA and AMSTAR instruments with regard to improving the methodological quality of searches for information for articles. *Sao Paulo Med J*. 2020; 138(6): 459-64.
34. Martinez-Zapata MJ NdGE, Canelo C, Vasquez-Mejia A, Merchan-Galvis A, Madera-Anaya M V-GA, Muñoz P, Zaror C. Agreement in the assessment of systematic reviews with the AMSTAR 2 tool: novice versus expert methodologists. *Abstracts of the 26th Cochrane Colloquium*. 2020.
35. Chapman JR. Editorial Perspective: time for another grading system-from PRISMA to AMSTAR 2. *Global Spine J*. 2020; 10(5): 674-5.
36. Luchini C, Veronese N, Nottegar A, Shin JI, Gentile G, Granzio U, et al. Assessing the quality of studies in meta-research: review/guidelines on the most important quality assessment tools. *Pharm Stat*. 2021; 20(1): 185-95.
37. Pieper D, Lorenz RC, Rombey T, Jacobs A, Rissling O, Freitag S, et al. Authors should clearly report how they derived the overall rating when applying AMSTAR 2-a cross-sectional study. *J Clin Epidemiol*. 2021; 129: 97-103.
38. Buhn S, Ober P, Mathes T, Wegewitz U, Jacobs A, Pieper D. Measuring test-retest reliability (TRR) of AMSTAR provides moderate to perfect agreement - a contribution to the discussion of the importance of TRR in relation to the psychometric properties of assessment tools. *BMC Med Res Methodol*. 2021; 21(1): 51.
39. Franco JVA, Meza N. Authors should also report the support for judgment when applying AMSTAR 2. *J Clin Epidemiol*. 2021; 138: 240.
40. Lorenz RC, Pieper D, Rombey T, Jacobs A, Rissling O, Freitag S, et al. Reply to letter to the editor by Franco et al. AMSTAR 2 overall confidence rating: a call for even more transparency. *J Clin Epidemiol*. 2021; 138: 241-2.