

Revisão sistemática e meta-análise de estudos observacionais em Psicologia

Revisión sistemática y meta-análisis de estudios observacionales en Psicología

Systematic Review and Meta-Analysis of Observational Studies in Psychology

Gabriela Andrade da Silva

Emma Otta

Universidade Federal do Sul da Bahia, Brasil

Universidade de São Paulo, Brasil

Resumo

O objetivo do presente trabalho foi levantar e sumarizar informações sobre metodologias para a realização de revisões sistemáticas e meta-análises de estudos observacionais, com considerações sobre o seu emprego e relevância na área da Psicologia. Revisão sistemática é uma revisão de literatura com base em uma pergunta de pesquisa, que usa métodos sistemáticos para a busca, seleção e avaliação crítica de estudos, resultando em uma sumarização. Meta-análise é um conjunto de técnicas estatísticas para combinar e sintetizar os resultados de dois ou mais estudos realizados de forma independente e selecionados em uma revisão sistemática. Essas metodologias têm elevada relevância na construção do conhecimento, por sintetizar informações de estudos publicados em diferentes fontes; diminuir o atraso entre as descobertas científicas e suas aplicações à prática profissional; e permitir testar novas hipóteses a partir dos dados já coletados. Foram descritas as etapas para realização de revisão sistemática. Os desafios específicos relacionados com a inclusão de estudos observacionais, bem como a aplicação desta técnica para o campo da Psicologia, foram abordados a fim de orientar os pesquisadores interessados em realizar revisões sistemáticas e meta-análises, bem como psicólogos, outros profissionais de saúde e gestores que buscam informações em evidências científicas para apoiar as suas práticas.

Palavras-chave: Prática clínica baseada em evidências, revisão, meta-análise, estudo observacional, metodologia, Psicologia

Resumen

El objetivo de este estudio fue identificar y resumir la información sobre las metodologías para la realización de las revisiones sistemáticas y metanálisis de estudios observacionales con consideraciones sobre su uso y relevancia en el campo de la Psicología. Una revisión sistemática es una revisión de la literatura basada en una pregunta de investigación que utiliza métodos sistemáticos para la búsqueda, selección y evaluación crítica de los estudios incluidos, lo que resulta en una sumarización. Metanálisis es un conjunto de técnicas estadísticas para combinar y sintetizar los resultados de dos o más estudios realizados de forma independiente y seleccionados en una revisión sistemática. Estas metodologías tienen gran relevancia en la construcción del conocimiento, pues sintetizan la información de los estudios publicados en diferentes fuentes, disminuyen la demora entre los descubrimientos científicos y sus aplicaciones a la práctica profesional y permiten la prueba de nuevas hipótesis a partir de datos ya recogidos. Se describen los pasos para llevar a cabo una revisión sistemática. Han sido abordados los desafíos específicos relacionados con la inclusión de los estudios de observación y la aplicación de esta técnica en el campo de la Psicología, con el fin de guiar a los investigadores interesados en la realización de revisiones sistemáticas y metanálisis, así como psicólogos, otros profesionales de la salud y gestores que buscan información en la evidencia científica para apoyar sus prácticas.

Palabras clave: Práctica clínica basada en la evidencia, revisión, metanálisis, estudio observacional, metodología, Psicología

Gabriela Andrade da Silva, Instituto de Humanidades, Artes e Ciências Jorge Amado, Universidade Federal do Sul da Bahia, Itabuna, BA, Brasil.

Emma Otta, Instituto de Psicologia da Universidade de São Paulo, São Paulo, SP, Brasil.

Nota: Partes deste texto foram extraídas da tese de doutorado da primeira autora.

Agradecemos à Dra. Maria Imaculada Cardoso Sampaio pela colaboração no presente trabalho, sobretudo no tópico relativo às estratégias de busca; à Dra. Tamara Melnik, pelas importantes observações sobre metodologia para realização de revisões sistemáticas; e ao Dr. José de Oliveira Siqueira, pelas contribuições referentes às análises estatísticas.

A correspondência relacionada com este artigo dirige-se a Gabriela Andrade da Silva, direção eletrônica: gabriela.silva@ufsb.edu.br



Abstract

The objective of this study was to identify and summarize information on methodologies for conducting systematic reviews and meta-analysis of observational studies, with considerations about their use and relevance in the field of psychology. A systematic review is a literature review based on a research question that uses systematic methods for search, selection and critical appraisal of the included studies, resulting in a summarization. Meta-analysis is a set of statistical techniques to combine and synthesize the results of two or more independent studies, which were included in a systematic review. These methodologies have high relevance in the construction of scientific knowledge, because they synthesize information from studies published in different sources; decrease the delay between scientific discoveries and their applications to professional practice; and allow the testing of new hypotheses from data already collected. The steps to perform a systematic review are described. Specific challenges related to the inclusion of observational studies and the application of this technique to the field of psychology are addressed, in order to guide researchers interested in conducting systematic reviews and meta-analyses, as well as psychologists, other health professionals and managers who seek information on scientific evidence to support their practices.

Keywords: Evidence-based Practice, Review, Meta-analysis, Observational study, Methodology, Psychology

O presente artigo apresenta considerações sobre o uso de Revisões Sistemáticas (RS) e meta-análise no campo da Psicologia Baseada em Evidências, com foco na aplicação dessa técnica a estudos observacionais. O campo da Psicologia apresenta particularidades em relação a outras áreas de saúde, pois lida diretamente com a subjetividade, tanto dos indivíduos participantes da pesquisa, quanto do entrevistador, que pode exercer influência sobre a interpretação e mensuração dos fenômenos psicológicos. A observação já foi considerada uma técnica de menor valor que a experimentação, devido à interferência do pesquisador no meio pesquisado, que acarretaria perda da objetividade e rigor científico, mas atualmente, sua contribuição à Psicologia é considerada igualmente relevante, pois permite a compreensão dos comportamentos e seus determinantes, bem como das interações humanas (Cano & Sampaio, 2007).

Estudos observacionais são definidos, no presente artigo, como aqueles em que os investigadores não realizam uma intervenção, podendo incluir delineamentos descritivos, como estudos epidemiológicos, séries de casos e relatos de caso; analíticos, como estudos de coorte, caso-controle ou transversais; e de diagnóstico, prognóstico e acurácia de instrumentos clínicos ou de pesquisa (Holly, 2012). Dado o uso frequente de delineamentos observacionais em psicologia, contemplando diversos temas de interesse (Cano & Sampaio, 2007), a Psicologia Baseada em Evidências deve apropriar-se de métodos para sintetizar os resultados dessas evidências por meio de RS, facilitando a sua incorporação à prática profissional dos psicólogos e às políticas públicas.

Entende-se por revisão sistemática: “Uma revisão de uma pergunta claramente formulada que usa métodos sistemáticos e explícitos para identificar, selecionar e avaliar criticamente pesquisas relevantes; e coletar e analisar dados dos estudos incluídos na revisão” (*The Cochrane Collaboration*, 2005, p. 45).

Meta-análise, também conhecida como revisão sistemática quantitativa, é um conjunto de técnicas estatísticas que permitem sintetizar os resultados de dois ou mais estudos realizados de forma independente e incluídos em uma RS. O emprego de meta-análise leva a novos resultados e a conclusões, que podem ser semelhantes ou diferentes das apontadas pelos estudos primários incluídos (Glass, 1976; *The Cochrane Collaboration*, 2005). Isso acontece porque, ao revisar a literatura de forma convencional, cada estudo conta como um voto favorável ou desfavorável a uma determinada hipótese, conforme o seu resultado positivo ou negativo. No entanto, o resultado negativo não é evidência da ausência de

efeito, mas sim ausência de evidência do efeito, ou seja, significa que o efeito, caso existente, não pôde ser detectado dentro dos limites da pesquisa. Um estudo pode falhar em alcançar resultados estatisticamente significativos por outras razões que não a ausência de efeito, como por exemplo o tamanho de amostra insuficiente. A meta-análise, porém, combina dados dos diferentes estudos, aumentando o poder estatístico do teste (O'Sullivan, 2010).

Diversas razões determinam a importância de se realizar RSs e meta-análises em Psicologia: (a) fornecer subsídio para as decisões de profissionais e usuários dos serviços, ampliando a eficiência das buscas por resultados de pesquisas científicas, pois sintetizam informações de estudos publicados em diferentes fontes; (b) diminuir o atraso entre as descobertas científicas e suas aplicações à prática profissional (Greenhalgh, 2008); (c) possibilitar a sumarização de dados de um grande número de estudos, que não poderiam ser abordados em uma revisão narrativa (O'Sullivan, 2010); (d) minimizar vieses nos resultados das revisões, ao adotar critérios explícitos para a busca e seleção de estudos primários, diferentemente de revisões narrativas; (e) estabelecer a possibilidade de generalização das descobertas científicas, por meio da comparação de resultados de diferentes estudos e verificação de sua consistência (Greenhalgh, 2008); (f) verificar o quanto uma nova iniciativa de pesquisa acrescentaria ao conhecimento científico, por meio do conhecimento abrangente da área que é proporcionado por uma RS (O'Sullivan, 2010); (g) testar hipóteses concretas por meio da combinação de estudos, com aplicação de meta-análise.

Organizações voltadas para a Prática Baseada em Evidências têm se empenhado em ampliar o número de RSs e meta-análises e elevar a sua qualidade metodológica. A Colaboração Cochrane, que atua principalmente na área médica, objetiva preparar, manter e assegurar o acesso a RSs sobre efeitos de intervenções em saúde. Com características semelhantes, a Colaboração Campbell objetiva oferecer apoio a políticos, gestores e demais profissionais na tomada de decisões em relação a programas de intervenção. Suas atividades são preparar, manter e promover acesso a RSs nas áreas de educação, justiça criminal, política, serviço social, psicologia e economia (Sánchez-Meca, Boruch, Petrosino & Rosa-Alcázar, 2002).

O foco da Colaboração Cochrane são as RSs e meta-análises sobre efetividade ou eficácia de intervenções na área de saúde, para as quais elaborou diretrizes consistentes, considerando o ensaio clínico aleatorizado ou padrão-ouro para a avaliação de diferentes formas de prevenção ou tratamento (Higgins & Green, 2011). Por exemplo, as psicoterapias têm sido avaliadas por RSs da Colaboração Cochrane quanto à sua eficácia e efetividade no tratamento de transtornos mentais (Melnik & Attalah, 2011). Paralelamente, a Colaboração Cochrane está elaborando diretrizes para RSs sobre a acurácia de testes diagnósticos (*Diagnostic Test Accuracy Working Group*, 2013). De acordo com a Colaboração Cochrane, ao avaliar o efeito de intervenções em saúde, RSs de estudos observacionais são recomendadas apenas em situações em que a realização de ensaios clínicos aleatorizados é difícil ou inadequada por razões éticas (Higgins & Green, 2011).

A concepção de que estudos observacionais produzem evidências de nível inferior às geradas por ECRs é bastante difundida, conforme mostram as pirâmides de evidências mais divulgadas na área de saúde. Pirâmides de evidências são representações gráficas de ordenações das pesquisas em níveis hierárquicos, conforme o campo ampliado correspondente à pergunta de pesquisa, à natureza e ao delineamento do estudo. O topo da pirâmide é ocupado por pesquisas que oferecem maior confiabilidade e a base ocupada por estudos metodologicamente menos confiáveis. Embora didáticas, essas representações devem ser relativizadas. Em razão da complexidade do campo da saúde – e, em especial, da Psi-

ciologia – o estabelecimento de uma hierarquia rígida de níveis de evidência baseada no tipo de pesquisa pode ser considerado demasiadamente simplista.

Estudos observacionais podem ser os mais indicados para responder a perguntas de pesquisa que não estão relacionadas à avaliação de intervenções, como por exemplo: associação entre fatores de risco e ocorrência de uma condição, avaliação de instrumentos psicométricos ou de rastreamento, variáveis relacionadas ao prognóstico e pesquisas exploratórias. Também são adequados ao campo da epidemiologia, para determinar a prevalência ou incidência de uma determinada condição. Até mesmo no campo de pesquisa de avaliação da eficiência ou eficácia de intervenções, em que é praticamente consenso entre pesquisadores considerar o ensaio clínico aleatorizado como uma evidência de nível mais elevado que os estudos observacionais, uma RS que buscou meta-análises baseadas em ambos os delineamentos problematizou essa concepção, tendo concluído que, em cada questão clínica estudada, os resultados de estudos observacionais (de coorte ou caso-controle) foram semelhantes aos de ensaios clínicos aleatorizados (Concato, 2004). Além disso, ensaios clínicos aleatorizados, por seu rigor metodológico e necessidade de padronização, podem ser pouco precisos ao avaliar a eficiência das intervenções, embora apropriados para avaliar sua eficácia. Por exemplo, a efetividade de uma intervenção na comunidade pode ser diferente de sua eficácia em um contexto de ensaio controlado (Holly, 2012, Stroup et al., 2000); e populações minoritárias ou com maior vulnerabilidade, como mulheres e idosos, em geral não atendem aos critérios de inclusão de ensaios clínicos (Holly, 2012). Sugere-se, assim, substituir as hierarquias rígidas que estabelecem os chamados níveis de evidências pela criticidade do pesquisador, que deve identificar o grau em que os delineamentos são apropriados para responder à sua pergunta de pesquisa (Ho, Peterson & Masoudi, 2008).

O *Levels of Evidence Working Group (Oxford Centre for Evidence-Based Medicine, OCEBM, 2011)* construiu classificações de níveis de evidências de acordo com o delineamento para sete campos ampliados de pesquisa em saúde: estudos epidemiológicos, diagnóstico, prognóstico, benefícios de um tratamento, prejuízos de um tratamento e rastreamento. O documento advertiu que os níveis de evidência ocupados por cada delineamento podem variar. Por exemplo, estudos com qualidade baixa, imprecisão ou inconsistência entre eles, ou com tamanho de efeito muito pequeno devem ser graduados em níveis de evidência mais baixos. Da mesma maneira, estudos com tamanho de efeito grande podem ser graduados em níveis de evidência mais elevados. Em todos os campos ampliados de pesquisa, uma RS geralmente ocupa nível mais alto de evidência que um estudo individual. Considera-se a hierarquia de níveis de evidência proposta pela OCEBM ampla, podendo ser usada em diversos campos de pesquisa em saúde, e razoavelmente flexível. A figura 1 mostra um exemplo de pirâmide de níveis de evidência adaptado da hierarquia proposta pelo *Levels of Evidence Working Group* para estudos observacionais de prevalência ou incidência (*Oxford Centre for Evidence-Based Medicine, 2011*).

Estudos observacionais produzem evidências relevantes para a construção do conhecimento científico em Psicologia e a importância de reuni-los em revisões sistemáticas e meta-análises vem sendo reconhecida. No entanto, ainda há necessidade de diretrizes metodológicas sobre RSs de estudos com diferentes delineamentos, populações e contextos, bem como métodos estatísticos para sintetizar os dados de estudos observacionais (Gülmezoglu, Say, Betrán, Villar & Piaggio, 2004). O presente trabalho objetivou levantar e sintetizar informações para a realização de RSs e meta-análises de estudos observacionais com boa qualidade, levantando considerações sobre o emprego dessas metodologias na área da Psicologia.

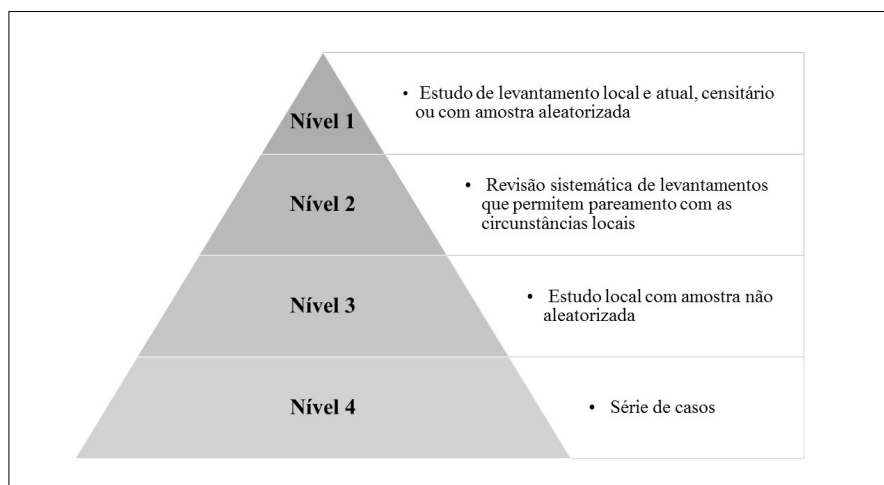


Figura 1. Pirâmide de níveis de evidência para responder à pergunta “Quão comum é o problema?”
(Adaptada de *Levels of Evidence Working Group, Oxford Centre for Evidence-Based Medicine, 2011*)

1. Formulação da pergunta de pesquisa e avaliação da necessidade da Revisões Sistemáticas

O primeiro passo para a realização de uma RS ou meta-análise é formular a pergunta de pesquisa, usando os elementos importantes para o delineamento do estudo. Alguns modelos têm sido propostos para auxiliar na definição das perguntas, em geral conhecidos por mnemônicos, como o PICO: Participantes, Intervenção, Controle e *Outcomes* (desfechos), que se aplica a questões sobre efetividade e/ou eficácia de intervenções (Greenhalgh, 2008; Higgins & Green, 2011; Saimbert, Pierce & Hargwood, 2012). No entanto, esse modelo nem sempre se aplica a estudos observacionais, pois nem todos preveem uma intervenção, de forma que a pergunta não tem um “I”. Dependendo do campo de pesquisa, a pergunta pode não ter um “C” (controle), como nos estudos sobre prognóstico (Saimbert, Pierce & Hargwood, 2012). Estudos sobre fatores de risco ou proteção precisam levar em conta a exposição “E” (Booth, 2006). Assim, outros modelos têm sido propostos para dar conta dos elementos relevantes a cada tipo de pergunta de pesquisa (Booth, 2004; Saimbert, Pierce & Hargwood, 2012). Cabe ao pesquisador identificar aquele que pode orientá-lo de forma adequada na definição da pergunta referente ao seu objeto de estudo, identificando os elementos importantes do delineamento científico.

Para ampliar a abrangência dos modelos, duas pesquisadoras brasileiras propuseram o modelo PVO, ou Participantes, Variáveis e *Outcomes* (desfechos) (Peres Biruel & Rocha Pinto, 2011). Esses três elementos constitutivos da pergunta devem ser determinados da seguinte maneira: **P - Participantes:** as características dos participantes que se deseja contemplar na investigação (ex.: adultos de sexo masculino, gestantes, idosos), bem como a condição ou problema investigado (ex.: depressão, transtorno de estresse pós-traumático, transtorno de ansiedade generalizada). **V - Variáveis:** variáveis que serão usadas como controle ou como subgrupos de comparação de participantes, as quais podem ser usadas como filtro ou identificadas como categorias na construção da estratégia de pesquisa. **O - Outcomes (desfechos):** resultado verificado nos participantes, ou, mais precisamente, o indicador da modificação ou variação das condições dos participantes em relação às variáveis definidas. O tipo de desfecho de interesse deve ser definido antes da busca de estudos, mas não deve ser incluído na pergunta de pesquisa, visto que será definido pelos resultados da mesma.

Após formular a pergunta de pesquisa, é preciso avaliar a necessidade da RS (Higgins & Green, 2011), verificando-se, por meio de consulta às bases de dados – sobretudo as especializadas em RS, como *The Cochrane Library* e a biblioteca da Colaboração Campbell –, a existência de outras RSs que possam responder à pergunta formulada. Caso existentes, estas devem ser avaliadas quanto à sua qualidade e data da última atualização. RSs pré-existentes de baixa qualidade são inadequadas para responder à pergunta de pesquisa, sendo necessária uma nova revisão. RSs antigas e de boa qualidade devem ser atualizadas à luz dos novos estudos.

2. Definição dos critérios de elegibilidade

Os critérios de elegibilidade orientam os revisores sobre qual é ou quais são os tipos de delineamento mais adequados para responder à pergunta de pesquisa, qual é o tipo de participante que o estudo deve incluir; e qual é o desfecho estudado (Higgins & Green, 2011). Ao estabelecer critérios de inclusão (características que fazem com que um determinado estudo seja admitido na revisão) e exclusão (características que levam à rejeição de um determinado estudo, uma vez que este preencha os critérios de inclusão) o revisor estabelece as condições para minimizar as chances de que o processo de seleção de estudos primários produza viés nos resultados. Por esse motivo, esses critérios devem ser definidos *a priori*, isto é, antes do início da coleta de dados, para evitar que a forma de seleção dos estudos produza viés nos resultados.

3. Definição de estratégias de busca

A busca sistemática é orientada pela pergunta de pesquisa e sua estratégia deve ser construída, preferencialmente, em conjunto com um bibliotecário com experiência em RS. Para mapear todo o conhecimento científico disponível na área, deve ser incluído o maior número possível de bases de dados que indexem trabalhos científicos, mesmo gerando sobreposição entre os resultados retornados. O conhecimento das bases de dados da área em que está sendo feita a revisão é de suma importância para que se maximize o número de estudos localizados sobre o tema pesquisado. No campo da Psicologia, a PsycINFO, Psycodoc e BVS-PSI ULAPSI são referências; bases de dados eletrônicas da Medicina, da Educação ou de ciência em geral também devem ser consultadas, como por exemplo: ISI Web of Knowledge, SCOPUS, Lilacs, PubMed, EMBASE e ERIC.

Além do uso das bases de dados eletrônicas e das listas de referências bibliográficas dos estudos encontrados, recomenda-se a busca na chamada “literatura cinzenta” (em inglês *gray literature*), ou seja: teses e dissertações (podendo-se usar acervos eletrônicos como a Biblioteca Digital de Teses e Dissertações ou o *Dissertation Abstracts*), anais de eventos e contato com os principais especialistas da área, para verificar se conhecem outros estudos em andamento ou ainda não publicados. Estudos em andamento, embora não tenham seus resultados incluídos na RS, devem ser referenciados para contribuir com futuras atualizações (Higgins & Green, 2011). A literatura cinzenta deve ser analisada com cuidado, visto que não é submetida à revisão rigorosa realizada por pareceristas de periódicos científicos.

Ao relatar a busca sistemática, o revisor deve indicar as fontes consultadas, a estratégia de busca em cada fonte, as datas das consultas, se houve limitação idiomática e o marco temporal. Recomenda-se evitar restrições na busca, como a limitação idiomática ou de ano de publicação, pois pode produzir viés nos resultados (Higgins & Green, 2011). Sugere-se, ainda, empreender esforços para identificar estudos de países em desenvolvimento, que têm maior chance de ser publicados em periódicos não in-

dexados e em idiomas diferentes do inglês (Gülmezoglu et al., 2004), evitando o viés devido ao local em que os estudos foram realizados.

4. Seleção de estudos e coleta de dados

A fase de coleta de dados é realizada concomitantemente com a fase de identificação e seleção de estudos, para que a revisão seja atualizada durante todo o processo. Formulários padronizados devem conter a identificação do revisor, as referências dos estudos coletados e características dos estudos: método, participantes, variáveis, resultados.

Geralmente, o uso de múltiplas bases para a busca sistemática de trabalhos implica na localização de um número excessivo de registros, sobretudo em RS de estudos observacionais. Isso ocorre porque não é possível limitar a busca de acordo com o tipo de estudo, como é o caso de revisões sobre eficácia e efetividade de intervenções, que restringem os retornos a ensaios clínicos controlados. Assim, sugere-se que a seleção de estudos seja iniciada por um rastreamento de títulos e resumos (Gülmezoglu et al., 2004).

A eliminação de registros duplicados, decorrentes da inclusão de mais de uma base de dados, pode ser facilitada pelo uso de um *software* de gerenciamento, como por exemplo, o EndNote® Web¹ (*software* gratuito), cuja ferramenta “*find duplicates*” (encontrar duplicatas) compara os registros por meio dos seguintes campos: autor, ano, título e tipo de referência (ex.: artigo, livro, capítulo etc.). No entanto, essa identificação automática pode ser prejudicada pela diferença de formato nos registros de diferentes bases de dados. Por exemplo, autores com nomes originários da Península Ibérica e América Latina costumam ter dois sobrenomes e nem sempre os registros diferenciam corretamente aquele que deve ser usado nas citações. Portanto, após o uso de ferramenta automática para detecção das duplicatas, segue-se a fase de identificação manual, que muitas vezes se torna um processo exaustivo (Gülmezoglu et al., 2004).

A tarefa de seleção resulta na classificação dos estudos nos seguintes grupos: (a) **Identificados** – todos os localizados pela busca nas bases de dados; (b) **Não selecionados** – dentre os identificados, são aqueles que, conforme os títulos e resumos, claramente não preenchem os critérios de inclusão; (c) **Selecionados** – dentre os identificados, são aqueles cujos títulos e resumos indicam que, aparentemente, preenchem os critérios de inclusão; (d) **Excluídos** – selecionados pelo resumo, mas que não preenchem os critérios de inclusão, conforme avaliação do texto completo; (e) **Incluídos** – atendem a todos os critérios de inclusão, após a avaliação do texto completo (Higgins & Green, 2011).

Ao menos dois revisores devem participar do processo de seleção, realizando todas as etapas de forma independente. Em caso de divergências, pode-se buscar consenso, ou consultar um terceiro revisor. O grau de concordância entre eles deve ser avaliado por meio de testes estatísticos (Higgins & Green, 2011).

Ao relatar o processo de seleção de estudos, o revisor deve apresentar o fluxo de pesquisa, em geral apresentado na forma de um diagrama, como o recomendado pela Declaração PRISMA² (Urrútia & Bonfill, 2010).

¹ <https://www.myendnoteweb.com/EndNoteWeb.html>

² <http://www.prisma-statement.org/2.1.4%20-%20PRISMA%20Flow%202009%20Diagram.pdf>

5. Análise de qualidade dos estudos incluídos

Os resultados e conclusões de uma RS ou meta-análise podem sofrer viés caso os estudos primários incluídos estejam enviesados (Shamliyan, Kane & Dickinson, 2010). Por isso, embora o termo qualidade tenha muitas concepções, para a finalidade de uma RS, pode-se conceber que está relacionado com o uso de procedimentos metodológicos para evitar viés (Sanderson, Tat & Higgins, 2007). Assim, o revisor deve avaliar a qualidade dos estudos primários incluídos na revisão. O método usado para essa avaliação deve ser definido antes do início da coleta de dados e apresentado no protocolo de pesquisa, juntamente com a fundamentação da escolha.

A maior parte dos instrumentos de avaliação da qualidade dos estudos incluídos em RS da área de saúde considera o ensaio clínico aleatorizado como padrão-ouro, por considerar que essa metodologia reduz o potencial de viés e variáveis de confusão. Assim, atribuem pontuações mais baixas a estudos com outros delineamentos, embora possam ser mais apropriados para responder a perguntas de pesquisa que não se referem à eficácia de intervenções (Baker, Young, Potter & Madan, 2010). Portanto, recomenda-se aos revisores buscar um instrumento de avaliação da qualidade adequado ao tipo de estudos incluídos na revisão.

Uma das principais diretrizes que orientam a avaliação de estudos observacionais é o STROBE³, um *checklist* de itens que devem constar dos relatos de estudos de coorte, caso-controle ou transversal (Elm et al., 2007). O objetivo do instrumento é estabelecer diretrizes para o relato de estudos observacionais. Conforme os próprios autores orientam, “embora a clareza do relato seja um pré-requisito para a avaliação, o *checklist* não é um instrumento para avaliar a qualidade de uma pesquisa observacional” (Elm et al., 2007, p. 248). Instrumentos que avaliam o relato buscam verificar se todas as informações importantes estão disponíveis ao leitor, enquanto aqueles que buscam avaliar a qualidade metodológica buscam verificar se foram adotados procedimentos que minimizem o risco de viés. Assim, é importante ter em mente que o revisor busca avaliar a validade do estudo e não necessariamente a qualidade do relato (Sanderson, Tat & Higgins, 2007), embora ambas estejam relacionadas em algum grau. Um exemplo dessa diferença é a amostragem: o STROBE considera correto qualquer artigo em que os critérios usados para compor a amostra estejam claros, mas em uma RS, deve ser atribuída maior qualidade a estudos observacionais que tenham usado amostra aleatória e não de conveniência.

Duas RSs localizaram 86 (Sanderson, Tat & Higgins, 2007) e 96 (Shamliyan, Kane & Dickinson, 2010) instrumentos para avaliar a qualidade metodológica de estudos observacionais. No entanto, constataram que não há um instrumento que tenha sido usado em ampla escala com essa finalidade (Sanderson, Tat & Higgins, 2007) e não há padrão-ouro para avaliar a validade interna ou externa de estudos observacionais (Shamliyan, Kane & Dickinson, 2010). Essas revisões também mostraram lacunas que comprometem o uso da maioria deles em futuras pesquisas, como a ausência de informação sobre como foram desenvolvidos e de dados sobre a sua validade. Além disso, diversos instrumentos foram desenvolvidos para estudos específicos, limitando suas possibilidades de aplicação a outros contextos.

Ao revisor que busca um instrumento adequado para avaliação da qualidade metodológica de estudos observacionais, sugere-se consultar as tabelas 5, 6 e 7 do artigo de Sanderson, Tat e Higgins

³ <http://www.strobe-statement.org/> (STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of STrengthening the Reporting of OBServational studies in Epidemiology).

(2007), que sumarizaram os 86 instrumentos encontrados na RS para avaliação de qualidade de estudos observacionais, 13 deles desenvolvidos especificamente com o objetivo de uso em RSs. Os instrumentos para avaliação da qualidade de pesquisas foram classificados em quantitativos, na forma de escalas que produzem uma pontuação; qualitativos, na forma de um *checklist*, ou qualitativos que podem ser resumidos em uma decisão, ou seja, *checklists* que permitem classificar os estudos em categorias, como “baixa”, “média” e “alta” qualidade.

O emprego de escalas para avaliação quantitativa permite que as pontuações sejam usadas como fator de ponderação na análise estatística dos resultados de uma meta-análise, atribuindo-se maior peso aos trabalhos de melhor qualidade. No entanto, é discutível se essa pontuação é realmente representativa, permitindo comparar a qualidade metodológica dos artigos científicos. A quantificação implica em se estabelecer um peso para cada item da escala, o qual é incerto, por não ser possível a comparação com um padrão-ouro para aferir sua validade. As ponderações usadas nas escalas se mostraram inconsistentes em RS (Sanderson, Tat & Higgins, 2007).

Uma RS (Shamliyan, Kane & Dickinson, 2010) constatou que apenas cinco instrumentos se propuseram a avaliar a qualidade metodológica de estudos de prevalência ou incidência, sendo que apenas um (Loney, Chambers, Bennett, Roberts & Stratford, 1998) era aplicável a qualquer tipo de estudo de incidência ou prevalência em saúde. Para esse propósito, recomenda-se o uso da escala de Gianakopoulos, Rammelsberg, Eberhard e Schmitter (2012), publicada posteriormente à RS. Embora tenha a desvantagem de ser baseada em abordagem quantitativa (pontuação), essa escala de aplicação rápida (média de oito minutos) permite avaliar estudos de incidência ou prevalência quanto à sua amostragem, mensuração e análise e obteve concordância satisfatória entre revisores.

Ao escolher um método para avaliação da qualidade metodológica de estudos observacionais incluídos em uma RS, sugere-se aos revisores, na medida do possível, buscar um instrumento que (a) inclua um número pequeno de domínios avaliados; (b) seja específico em relação ao delineamento e área de estudo; (c) seja um *checklist* qualitativo e não uma escala quantitativa; e (d) tenha tido um desenvolvimento cuidadoso e/ou evidências de sua validade e confiabilidade (Sanderson, Tat & Higgins, 2007). Nas RSs em Psicologia e saúde mental, considera-se fundamental que o instrumento avalie também a forma como o diagnóstico foi realizado, devido à complexidade de se definir transtornos mentais de forma operacional e à existência de diversos instrumentos para identificá-los e mensurá-los, todos com considerável possibilidade de viés decorrente da subjetividade do avaliador e do participante.

Escalas de auto-avaliação (isto é, constituídas de perguntas objetivas respondidas pelo próprio participante) são frequentes em estudos de Psicologia e seu uso deve ser condicionado à existência de informações satisfatórias sobre a sua validade, verificada por estudos que tenham determinado a acurácia (Greenhalgh, 2008; Holly, 2012). Até mesmo instrumentos de auto-avaliação devidamente validados tendem a superestimar a ocorrência de transtornos mentais nas populações de estudo. Isso ocorre porque, em geral, são desenvolvidos para rastreamento em contexto clínico, em que é preferível encaminhar mais pessoas para avaliação profissional a deixar de identificar uma pessoa que pode apresentar um determinado problema. Assim, o ponto de corte costuma ser calculado de forma a evitar falsos negativos, o que aumenta a probabilidade de falsos positivos. De fato, constatou-se que cerca de 50% das mulheres que obtiveram pontuação acima do ponto de corte na principal escala usada para identificar depressão pós-parto não apresentavam esse transtorno (Mathey, 2010) e que o uso de escalas

de auto-avaliação produziu prevalências de depressão pós-parto significativamente maiores que o uso de critérios de diagnóstico baseados em entrevista clínica (Andrade da Silva, 2013). Portanto, estudos em Psicologia e saúde mental poderiam alcançar maior precisão ao avaliar os participantes por meio de entrevista clínica realizada por, no mínimo, dois profissionais de saúde mental treinados, trabalhando de forma independente e cegos para resultados de outros instrumentos, calculada sua taxa de concordância.

6. Avaliação do grau de heterogeneidade dos estudos incluídos

Críticas à meta-análise frequentemente levantaram o problema conhecido como “misturar maçãs e laranjas”, ou seja, combinar estudos sobre desfechos distintos, com metodologias distintas e cuja síntese em valor único, portanto, teria significado duvidoso do ponto de vista prático (Higgins & Green, 2011). Quando os estudos incluídos apresentam metodologias demasiadamente diferentes, não há sentido em sintetizar estatisticamente seus resultados.

Considera-se heterogeneidade qualquer tipo de variabilidade entre os estudos incluídos em uma RS. Heterogeneidade estatística é o termo específico que se refere a uma consequência da variabilidade clínica (variabilidade entre participantes) e/ou metodológica (diferenças quanto ao delineamento e qualidade dos estudos incluídos) (Higgins & Green, 2011).

RSs de estudos observacionais costumam apresentar elevada heterogeneidade, devido às diferenças metodológicas entre os estudos incluídos, de forma que muitas têm a síntese estatística (meta-análise) inviabilizada (Giannakopoulos et al., 2012). Assim, nas meta-análises de estudos observacionais, é fundamental que a heterogeneidade seja avaliada qualitativamente e quantitativamente, por meio de ferramentas estatísticas, como o teste de qui-quadrado, procedimento que testa a hipótese de que os desfechos são iguais em todos os estudos. Outro parâmetro que pode ser usado é o I^2 ($I^2 = [(Q - g)/Q] \times 100\%$, onde Q = valor do qui-quadrado, g = graus de liberdade do teste), que descreve a porcentagem de variabilidade nas estimativas de efeito que se devem à heterogeneidade, e não ao erro de amostragem (acaso). Considera-se que valores acima de 50% indicam heterogeneidade substancial (Higgins & Thompson, 2002; Higgins & Green, 2011) e que, nesses casos, a análise qualitativa é mais indicada, dispensando-se a meta-análise.

Visualmente, a heterogeneidade pode ser observada nos gráficos de floresta (Anzures-Cabrera & Higgins, 2010; Lewis & Clarke, 2001), construídos a partir da estimativa pontual de cada estudo e do intervalo de confiança (95%). Um exemplo de gráfico de floresta pode ser visualizado na figura 2. Esse gráfico, em geral, usa quadrados para representar as estimativas pontuais de cada estudo. Linhas horizontais, que interceptam esses quadrados, representam as estimativas intervalares (intervalo de confiança a 95%). O valor médio, que sintetiza os resultados dos estudos incluídos, é representado por um losango na parte inferior do gráfico. Valores atípicos (outliers) podem ser identificados visualmente nos gráficos de floresta. Recomenda-se que não sejam excluídos em busca de maior homogeneidade, pois o procedimento de excluir estudos com base em seus resultados pode gerar viés (Higgins & Green, 2011). Pode-se avaliar o impacto dos outliers na análise de sensibilidade, conforme explicado na seção 7 do presente artigo.

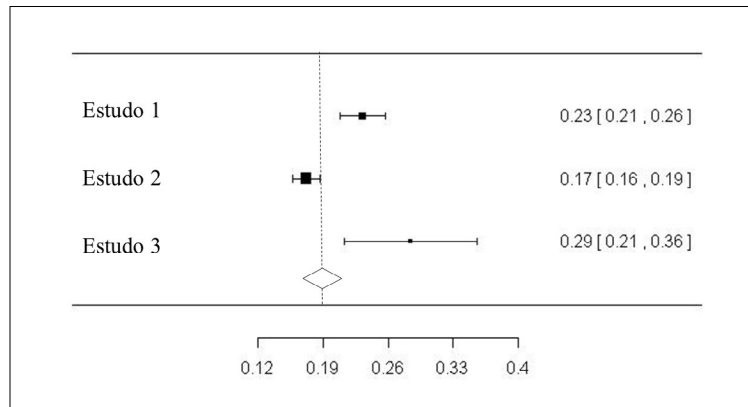


Figura 2. Exemplo fictício de gráfico de floresta. Os quadrados representam as estimativas pontuais dos estudos incluídos, as linhas representam os respectivos intervalos de confiança e o losango “diamante” representa a síntese estatística.

Um dos fatores relacionados à heterogeneidade é o viés de publicação, isto é, desvio nos resultados que ocorre quando os estudos publicados não representam adequadamente o universo de trabalhos realizados sobre um determinado assunto (Delgado-Rodríguez, 2001). Trabalhos sem evidência de efeito ou com efeito em direção oposta à esperada pelos pesquisadores teriam menor chance de ser publicados, porque o pesquisador não submeteu o trabalho a periódicos científicos ou porque foi recusado pelos pareceristas (Csada, James & Espie, 1996; Rosenthal, 1979; Stroup et al., 2000). Provavelmente, esse viés é mais comum em estudos caso-controle ou de coorte, em que há uma direção de efeito esperada pelo pesquisador, que em estudos psicométricos, de prevalência e de incidência. No entanto, revisões de todos os tipos de delineamento estão sujeitas a outro tipo de viés de publicação, relacionado ao local de origem.

A expressão “viés de publicação contra o Sul” tem sido usada para referir-se à desproporção entre a produção científica de países de renda alta (referidos como “Norte”) e os de renda média ou baixa (referidos como “Sul”). Foram apontados quatro motivos para essa diferença: (1) menores investimentos financeiros em pesquisa e consequente “fuga de cérebros” de países “do Sul”; (2) menor pressão em países “do Sul” que em países “do Norte” pela publicação de artigos; (3) os delineamentos escolhidos por autores de países “do Sul” seriam, em geral, os menos valorizados na literatura internacional, como estudos qualitativos e observacionais em detrimento dos experimentais; (4) menor chance de publicação de artigos produzidos em países “do Sul” nos periódicos “do Norte”, devido a preconceito dos editores e pareceristas, chamado “racismo editorial” (Victora & Moreira, 2006).

Dados estatísticos corroboraram a afirmação de que países em desenvolvimento apresentam menor produção científica. Uma revisão bibliométrica verificou que apenas 7,9% dos artigos científicos da área de saúde publicados entre 1992 e 2001 eram de países de renda média e 2,7% de países de renda baixa (Paraje, Sadana & Karam, 2005). Em Psicologia da saúde, constatou-se que países europeus, Estados Unidos da América, Japão e Austrália são responsáveis por 90% das pesquisas (Kern de Castro & Bornholdt, 2004). A menor chance de publicação de artigos de países latino-americanos, quando comparados aos de outros países, aparentemente não se justifica por uma eventual diferença na quali-

dade dos estudos, visto que uma revisão que comparou a qualidade metodológica e da apresentação dos resultados de artigos latino-americanos e de outros países não encontrou diferença significativa entre os dois estratos (Sampaio, 2013). O viés de publicação referente ao local de origem pode comprometer significativamente a elaboração de programas e políticas públicas no campo da Psicologia nos países em desenvolvimento, visto que estes apresentam contexto ambiental, social e cultural diferente de países desenvolvidos, o que limita a generalização dos resultados das pesquisas.

Técnicas estatísticas permitem avaliar o possível impacto do viés de publicação. Uma delas é o *fail-safe number* (número a prova de falhas), ou seja, o cálculo do número de estudos com resultado nulo, não publicados ou não localizados, que seria necessário para alterar o resultado da meta-análise (Orwin, 1983). Outra forma de investigar o viés de publicação é o gráfico de funil (Anzures-Cabrera & Higgins, 2010; Egger, Smith, Schneider & Minder, 1997; Higgins & Green, 2011), que relaciona os resultados dos estudos com o tamanho da amostra ou o erro padrão. Parte-se do pressuposto de que estudos com amostras maiores apresentam maior precisão, enquanto estudos com pequenas amostras tendem a apresentar dispersão aleatória simétrica em relação ao resultado médio encontrado na meta-análise, de forma que os pontos do gráfico assumem o formato de um funil. Formatos assimétricos podem indicar vieses de publicação ou de outros tipos, como por exemplo: baixa qualidade metodológica dos estudos com amostras pequenas, ou variabilidade ao acaso (Higgins & Green, 2011). Um exemplo de gráfico de funil, extraído da Tese de Doutorado de Andrade da Silva (2013), pode ser encontrado na figura 3. Foi investigada a prevalência de depressão pós-parto por meio de revisão sistemática de literatura, mantendo-se registro da cidade e país em que o estudo foi realizado, quando essa informação estava disponível. O gráfico A mostra a prevalência em cada estudo, em relação ao erro padrão, e os pontos se distanciam do formato de funil, indicando heterogeneidade elevada. Nos gráficos B e C, o país e a cidade, respectivamente, foram inseridos como moderadores nos modelos, obtendo-se dispersão que se aproxima do formato de funil, indicando menor heterogeneidade.

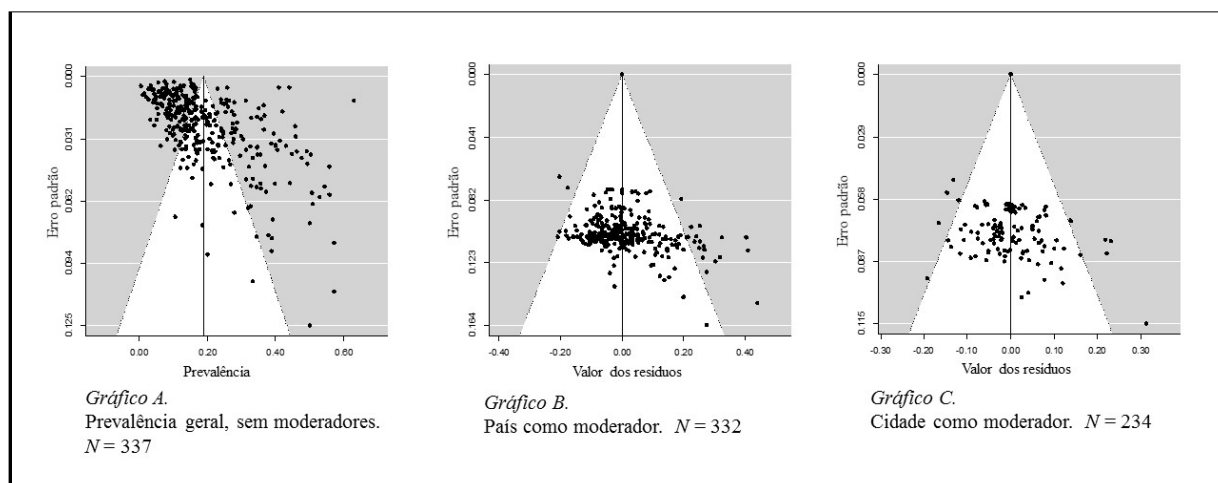


Figura 3. Gráficos de funil para avaliar heterogeneidade dos estudos incluídos em revisão sistemática sobre a prevalência de depressão pós-parto. Gráfico A: prevalências dos estudos incluídos em relação ao erro padrão. Gráfico B: resíduos em relação ao erro padrão, usando país como moderador; Gráfico C: resíduos em relação ao erro padrão, usando cidade como moderador. Figura extraída de Andrade da Silva, 2013.

Para viabilizar a síntese estatística dos resultados de estudos heterogêneos, é possível usar o modelo de efeitos aleatórios, método que combina os tamanhos de efeito dos estudos individuais, incorporando a heterogeneidade à estimativa síntese (construída a partir dos resultados dos estudos incluídos). Essa análise assume que os estudos incluídos na meta-análise são uma amostra tomada de uma população de estudos e não se parte do pressuposto de homogeneidade entre eles (Delgado-Rodríguez, 2001).

7. Síntese estatística dos resultados

A realização de meta-análise, caso prevista, deve ser planejada antes do início da coleta de dados, definindo-se a medida geral e os subgrupos de análise. Informações de cada estudo incluído deverão ser coletadas, como o tamanho do efeito da variável independente sobre as dependentes, o erro padrão e o tamanho da amostra. As análises estatísticas variam de acordo com o tipo de variável em questão (nominal, ordinal ou contínua) e com a pergunta de pesquisa. Os resultados de uma meta-análise são apresentados em três partes: descrição dos estudos, qualidade dos estudos e resultado das análises estatísticas. Além do resultado geral, indicando uma estimativa pontual e um intervalo de confiança para o desfecho, deverão ser apresentadas as análises por subgrupos, quando conveniente, e as análises de sensibilidade (Higgins & Green, 2011).

Análises de sensibilidade são ferramentas para investigar o quanto os resultados encontrados na meta-análise são sensíveis à metodologia que foi usada na RS. São calculados novos resultados, realizando-se mudanças nos critérios de inclusão ou na forma de coletar os dados, de forma a comparar e discutir a diferença entre esses resultados e os resultados originais da pesquisa. Procura-se, dessa maneira, obter maior segurança na aplicação prática dos resultados da meta-análise (Higgins & Green, 2011). Sugere-se realizar análises de sensibilidade levando em conta variáveis metodológicas, qualidade dos estudos e exclusão de *outliers*.

8. Discussão

A discussão deve conter uma síntese dos resultados obtidos, discutidos em relação ao tamanho e direção do efeito (Higgins & Green, 2011) e interpretados no contexto de outras evidências (Urrútia & Bonfill, 2010), indicando-se qual é o potencial de generalização dos resultados. Independentemente da confirmação ou rejeição das hipóteses de pesquisa, devem ser exploradas explicações alternativas aos resultados apresentados (Stroup et al., 2000).

É importante declarar também as limitações do estudo, ou seja, questões relacionadas ao risco de viés, considerando-se a qualidade metodológica dos estudos incluídos, a seleção de estudos e as variáveis de confusão (Stroup et al., 2000; Urrútia & Bonfill, 2010). Recomenda-se, ainda, que as RSs e meta-análises contenham em sua discussão as implicações para a prática profissional e a futura pesquisa na área (Higgins & Green, 2011).

Considerações finais

O presente trabalho levantou e sistematizou informações relevantes para a realização de RSs de estudos observacionais em Psicologia, procurando orientar pesquisadores interessados em realizar esse tipo de pesquisa, bem como psicólogos, outros profissionais de saúde e gestores que busquem informação em evidências científicas para subsidiar suas práticas. Foi abordada a importância de se realizar RSs e meta-análises, tanto para sumarizar as informações de diversas fontes e deixá-las mais acessíveis, quanto para testar novas hipóteses sobre os dados já coletados e publicados.

Considera-se que os esforços empregados por pesquisadores das áreas de medicina e enfermagem – nas quais as RSs são mais comuns – para criar diretrizes, protocolos e instrumentos para RSs e meta-análises trouxeram muitas contribuições para o emprego dessas metodologias em qualquer área de saúde. No entanto, no campo da Psicologia há algumas especificidades que precisam ser observadas pelos revisores, tais como a importância de se levar em consideração a subjetividade do pesquisador e dos participantes da pesquisa; a grande relevância dos estudos observacionais para a compreensão de fenômenos psicológicos; a complexidade de se avaliar a validade de instrumentos psicométricos e comparar técnicas de intervenção; e a necessidade de se incluir as bases de dados específicas da Psicologia na busca sistemática. A maior parte das recomendações metodológicas abordadas pelo presente artigo podem ser aplicadas às diversas áreas de saúde, mas foram abordadas também essas especificidades do campo da Psicologia.

Dada a importância dos estudos observacionais para a construção do conhecimento científico em Psicologia, foram abordados os principais desafios enfrentados por revisores ao realizar de RSs de estudos com esses delineamentos, tais como: (a) estabelecer os níveis de evidência de estudos observacionais para os diferentes campos de pesquisa, visto que as principais pirâmides de evidências se referem aos estudos sobre eficácia de intervenções; (b) identificar um modelo que oriente a formulação de uma pergunta de pesquisa; (c) lidar com o grande número de registros recuperados, que é frequente em RSs de estudos observacionais; (d) encontrar um bom instrumento para avaliação da qualidade metodológica dos estudos observacionais incluídos na revisão; (e) lidar com a alta heterogeneidade frequentemente encontrada entre os estudos incluídos, que pode dificultar a síntese estatística dos dados (meta-análise); e (f) avaliar criticamente a possibilidade de viés de publicação, ou seja, à possibilidade de existirem estudos não publicados por não terem apresentado evidências que dessem subsídio à hipótese de pesquisa, ou devido ao local de origem do estudo – o chamado “viés de publicação contra o sul” (Victora & Moreira, 2006).

É importante que o relato de uma RS documente todos os pontos importantes para a avaliação de sua validade pelos pesquisadores, profissionais e usuários interessados no assunto abordado por ela. Para elevar a qualidade do relato, recomenda-se seguir as diretrizes da Declaração PRISMA, um *checklist* criado por um grupo de pesquisadores com experiência em RSs e meta-análises e que deve ser usado por revisores, editores e pareceristas para elevar a qualidade do relato desses estudos (Urrútia & Bonfill, 2010). Considerando que RSs e meta-análises de estudos observacionais apresentam desafios particulares, relacionados aos vieses inerentes aos delineamentos de pesquisa, em 1997 foi realizado um *workshop* que reuniu especialistas em meta-análises de estudos observacionais para debater uma agenda incluindo questões sobre viés, busca sistemática, heterogeneidade, categorização de estudos e métodos estatísticos. Esse grupo, denominado *Meta-Analysis of Observational Studies in Epidemiology* (MOOSE), gerou recomendações específicas para RSs de estudos observacionais (Stroup et al., 2000).

Ressalta-se que a RS é uma linha de pesquisa à qual o autor deverá dar continuidade. Portanto, não deve ser abandonada depois de concluída: precisa ser continuamente atualizada, acrescentando-se a ela novos estudos publicados (Higgins & Green, 2011). Sugere-se, ainda, buscar incluir a revisão em bases de dados específicas, como por exemplo, as bases de dados da Colaboração Cochrane e da Colaboração Campbell, para evitar a duplicação de esforços por outros autores.

Referências

- Andrade da Silva, G. (2013). *Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise* (Tese de doutorado, Universidade de São Paulo, Brasil). Recuperado de <http://www.teses.usp.br/teses/disponiveis/47/47132/tde-24052013-143728/>
- Anzures-Cabrera, J. & Higgins, J. (2010, Jan-Mar). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods, 1*(1), 66-80. DOI: 10.1002/jrsm.6
- Baker, A., Young, K., Potter, J. & Madan, I. (2010). A review of grading systems for evidence-based guidelines produced by medical specialties. *Clinical Medicine, 10*(4), 358-363. doi: 10.7861/clinmedicine.10-4-358
- Peres Biruel, E. & Rocha Pinto, R. (2011). *Bibliotecário: Um profissional a serviço da pesquisa*. Anais do XXIV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação. Maceió, Alagoas, Brasil.
- Booth, A. (2004). Formulating answerable questions. In A. Booth & A. Brice (Eds). *Evidence-Based Practice for Information Professionals* (pp. 61-70). London, UK: Facet Publishing. Retrieved form <http://ebliptext.pbworks.com/>
- Booth, A. (2006). Clear and present questions: formulating questions for evidence based practice. *Library Hi Tech, 24*(3), 355-368. <http://dx.doi.org/10.1108/07378830610692127>
- Cano, D. S. & Sampaio, I. T. A. (2007). O método de observação na psicologia: considerações sobre a produção científica. *Interação em Psicologia, 11*(2), 199-210. Recuperado de <http://ufpr.br>
- Concato, J. (2004, Jul). Observational versus experimental studies: What's the evidence for a hierarchy? *The Journal of the American Society for Experimental NeuroTherapeutics, 1*(3), 341-347. doi:10.1602/neurorx.1.3.341
- Csada, R. D., James, P. C. & Espie, R. H. M. (1996, Sep). The "file drawer problem" of non-significant results: Does it apply to biological research? *Oikos, 76*(3), 591-593.
- Delgado-Rodríguez, M. (2001). Glossary on meta-analysis. *Journal of Epidemiology and Community Health, 55*, 534-536. doi:10.1136/jech.55.8.534
- Diagnostic Test Accuracy Working Group (2013). *Handbook for DTA Reviews*. Recuperado de <http://srdta.cochrane.org/handbook-dta-reviews>
- Egger, M., Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. doi: <http://dx.doi.org/10.1136/bmj.315.7109.629>
- Elm von, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. [STROBE Initiative] (2007, Oct). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Annals of Internal Medicine, 147*(8), 573-577. doi:10.7326/0003-4819-147-8-200710160-00010
- Giannakopoulos, N., Rammelsberg, P., Eberhard, L. & Schmitter, M. (2012, Jun). A new instrument for assessing the quality of studies on prevalence. *Clinical Oral Investigations, 16*(3), 781-788. doi: 10.1007/s00784-011-0557-4
- Glass, G. (1976, Nov). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.
- Greenhalgh, T. (2008). *Como ler artigos científicos: Fundamentos da medicina baseada em evidências* (Trad. A. P. Fajardo) Porto Alegre, Brasil: Artmed.
- Gülmezoglu, A. M., Say, L., Betrán, A. P., Villar, J., & Piaggio, G. (2004, Jul). WHO systematic review of maternal mortality and morbidity: Methodological issues and challenges. *BMC Medical Research Methodology, 4*(16). doi:10.1186/1471-2288-4-16 Recuperado de <http://www.biomedcentral.com/1471-2288/4/16>

- Higgins, J. P. T. & Green, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions* [Version 5.1.0, atualizado em Março de 2011]. The Cochrane Collaboration, 2011. Recuperado de www.cochrane-handbook.org
- Higgins, J. P. T. & Thompson, S. G. (2002, Jun). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. DOI: 10.1002/sim.1186
- Ho, P. M., Peterson, P. N. & Masoudi, F. A. (2008). Evaluating the evidence: Is there a rigid hierarchy? *Circulation*, 118, 1675-1684. doi: 10.1161/CIRCULATIONAHA.107.721357
- Holly, C. (2012). Systematic review of observational evidence. In C. Holly, S. W. Salmond & M. K. Saimbert (Eds.), *Comprehensive Systematic Review for Advanced Nursing Practice* (pp. 189-208). New York, US: Springer. Retrieved from <http://www.midwiferycollege.org/>
- Kern de Castro, E. K. & Bornholdt, E. (set, 2004). Psicologia da saúde X psicologia hospitalar: definições e possibilidades de inserção profissional. *Psicologia, Ciência e Profissão*, 24(3), 48-57. Recuperado de http://pepsic.bvsalud.org/scielo.php?pid=S1414-98932004000300007&script=sci_arttext
- Lewis, S. & Clarke, M. (2001, Jun). Forest plot: Trying to see the wood and the trees. *British Medical Journal*, 322, 1479-1480. <http://dx.doi.org/10.1136/bmj.322.7300.1479>
- Loney, P., Chambers, L., Bennett, K., Roberts, J. & Stratford, P. (1998). Critical appraisal of the health research literature: Prevalence or incidence of a health problem. *Chronic Diseases in Canada*, 19(4), 170-176. Retrieved from researchgate.net
- Mathey, S. (2010, Jan). Are we overpathologising motherhood? *Journal of Affective Disorders*, 120(1-3), 263-266. doi:10.1016/j.jad.2009.05.004
- Melnik, T. & Atallah, A. (2011). *Psicologia baseada em evidências: Provas científicas da efetividade da psicoterapia*. São Paulo, Brasil: Santos.
- Oxford Centre for Evidence-Based Medicine (2011). *The Oxford 2011 Levels of Evidence*. Author/ Levels of Evidence Working Group. Retrieved from <http://www.cebm.net/index.aspx?o=5653>
- Orwin, R. G. (1983). A Fail-Safe N for Effect Size in Meta-Analysis. *Journal of Educational Statistics*, 8(2), 157-159.
- O'Sullivan, D. (2010). Metanálise. En G. M. Breakwell, S. Hammond, C. Fife-Schaw, J. A. Smith & V. G. Haase (Eds.), *Métodos de pesquisa em psicologia* (Trad. F. R. Elizalde). Porto Alegre, Brasil: Artmed.
- Paraje, G., Sadana, R., & Karam, G. (2005, May). Increasing international gaps in health-related publications. *Science*, 308, 959-960. Retrieved from researchgate.net
- Rosenthal, R. (1979, may). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641. <http://dx.doi.org/10.1037/0033-2909.86.3.638>
- Saimbert, M. K., Pierce, J., & Hargwood, P. (2012). Developing clinical questions for systematic reviews. In C. Holly, S. W. Salmond, & M. K. Saimbert (Eds.), *Comprehensive Systematic Review for Advanced Nursing Practice* (pp. 189-208). New York, US: Springer. Retrieved from <http://www.midwiferycollege.org/>
- Sampaio, M. I. C. (2013). *Qualidade de artigos incluídos em revisão sistemática: comparação entre latino-americanos e de outras regiões* (Tese de doutorado, Universidade de São Paulo, Brasil). Recuperado de <http://www.teses.usp.br/teses/disponiveis/47/47132/tde-11122013-084214/>
- Sánchez-Meca, J., Boruch, R., Petrosino, A. & Rosa-Alcázar, A. (dic, 2002). La Colaboración Campbell y la práctica basada en la evidencia. *Papeles del Psicólogo*, 83, 44-48. Recuperado de <http://www.papelesdelpsicologo.es/vernumero.asp?id=896>
- Sanderson, S., Tat, I. D., & Higgins, J. P. (2007, Apr). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology*, 36(3), 666-676. doi: 10.1093/ije/dym018

- Shamliyan, T., Kane, R., & Dickinson, S. (2010, Oct). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, *63*(10) 1061-1070. doi:10.1016/j.jclinepi.2010.04.014
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., ..., & Thacker, S. B. (2000, Apr). Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*, *283*(15), 2008-2012. doi:10.1001/jama.283.15.2008
- The Cochrane Collaboration. (2005). *Glossary of Terms in The Cochrane Collaboration* [Version 4.2.5]. Recuperado de <http://www.cochrane.org/sites/default/files/uploads/glossary.pdf>
- Urrútia, G. & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica (Barcelona)*, *135*(11), 507-511. Recuperado de <http://www.prisma-statement.org/PRISMA%20Spanish%20Sept%202010.pdf>
- Victora, C. & Moreira, C. (2006). Publicações científicas e as relações Norte-Sul: Racismo editorial? *Revista de Saúde Pública*, *40*(N Esp), 36-42. <http://dx.doi.org/10.1590/S0034-89102006000400006> Recuperado de <http://www.revistas.usp.br/rsp/article/view/32083>

Recebido 17 de julho de 2014
Revisão aceita 08 de janeiro de 2015
Aceitado 20 de janeiro de 2015

Resenha das autoras

Gabriela Andrade da Silva é Doutora em Psicologia Experimental e Professora Adjunta do Instituto de Humanidades, Artes e Ciências Jorge Amado da Universidade Federal do Sul da Bahia. E-mail: gabriela.silva@gmail.com

Emma Otta é Doutora em Psicologia Experimental e Professora Titular do Instituto de Psicologia da Universidade de São Paulo. E-mail: emmaotta@gmail.com

Primeira publicação online: 16 de março de 2015