UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

MARIA IMACULADA CARDOSO SAMPAIO

Qualidade de artigos incluídos em revisão sistemática:
comparação entre latino-americanos e de outras regiões

São Paulo
2013

MARIA IMACULADA CARDOSO SAMPAIO

Qualidade de artigos incluídos em revisão sistemática: comparação
entre latino-americanos e de outras regiões
(Versão corrigida)

Tese apresentada ao Instituto de Psicologia da
Universidade de São Paulo, como parte dos
requisitos para a obtenção do grau de Doutor em
Psicologia.

Área de concentração:
Psicologia Experimental

Orientadora: Profa. Dra. Sonia Beatriz Meyer

São Paulo
2013

Autorizo a divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na publicação
Biblioteca Dante Moreira Leite
Instituto de Psicologia da Universidade de São Paulo

Sampaio, Maria Imaculada Cardoso.

Qualidade de artigos incluídos em revisão sistemática: comparação entre latino-americanos e de outras regiões / Maria Imaculada Cardoso Sampaio; orientadora Sonia Beatriz Meyer. - São Paulo, 2013.

223 f.

Tese (Doutorado – Programa de Pós-Graduação em Psicologia. Área de Concentração: Psicologia Experimental) – Instituto de Psicologia da Universidade de São Paulo.

1. Trabalho científico 2. Psicologia baseada em evidências 3. Revisão sistemática 4. Depressão pós-parto 5. Prevalência 6. América Latina I. Título.

Q180

Nome: Sampaio, Maria Imaculada Cardoso

Título: Qualidade de artigos incluídos em revisão sistemática: comparação entre latino-americanos e de outras regiões

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo como parte dos requisitos para a obtenção do grau de Doutor em Psicologia

Aprovado em: 4 de outubro de 2013.

Banca examinadora

Prof^a. Dr^a Sonia Beatriz Meyer

Instituição: IP-USP

Prof^a. Dr^a. Briseida Dogo de Resende

Instituição: IP-USP

Prof. Dr. Piotr Trzesniak

Instituição: UNIFEI

Prof^a. Dr^a. Tamara Melnik

Instituição: UNIFESP

Prof. Dr. Adrés Eduardo Aguirre Antúñez

Instituição: IP-USP

In Memoriam

In Memoriam de Caio Luiz Cardoso Sampaio, meu professor, meu marido, pai dos meus filhos, meu único e verdadeiro amor.

Nos primeiros dias do nosso namoro, no ano de 1975, o Caio me trouxe à USP e me explicou o que era uma Universidade.

Então Caio, capitão estelar, ainda estou aqui na Universidade e ofereço este trabalho a você!

"Atenção senhor Scotty, velocidade de transcorder, preparar para deixar o planeta Terra, temos uma longa viagem pela frente."

"Para casa - respondeu Kirk.

Para o Universo Alternativo?

O senhor entende isso?

Sim, capitão. E concordo com sua partida. Só peço que me coloque desacordado com um disparo de *feiser* em tonteio antes de partir. Meus agentes confirmarão qualquer história que eu contar depois."

(Star Trek – Episódios da Série Clássica)

Nave Star Trek

Dedicatória

Dedico esta tese ao Professor Doutor Arrigo Leonardo Angelini que, dentre tantas coisas maravilhosas que fez pela Psicologia brasileira, no ano de 1954, realizou uma revisão da literatura sobre “O emprego do ácido glutâmico na terapia da debilidade mental”. O estudo teve como objetivo

...rever e comparar os resultados de tais pesquisas, no sentido de verificar até que ponto são eles concordantes e autorizariam o emprego dessa droga como terapêutica da deficiência mental. (Angelini, 1954, p.130)

Muito além do seu tempo, mesmo sem utilizar o termo revisão sistemática, o estudo reuniu e “sistematizou” 36 artigos sobre o tema e concluiu que:

Sómente quando um maior número de pesquisas, feitas com o maior rigor científico, evidenciarem resultados positivos definidos, e se ficar bem também provado que o uso da droga não constitui perigo para a integridade do organismo, então poderemos empregar o ácido glutâmico, sem receio, na terapêutica da deficiência mental. (p.131)

Penso que a Psicologia Baseada em Evidências deu seus primeiros passos no Brasil com este artigo.

Agradecimento(s)

Muitas pessoas contribuíram para que essa etapa dos meus estudos fosse concluída, agradeço a cada uma, sensibilizada. Morrendo de medo de ser injusta, vou agradecer especialmente a algumas pessoas e algumas instituições aqui. Quem se sentir injustiçado, por favor, desculpe-me, foi falha mesmo, nada intencional.

À Profa. Dra. Sonia Beatriz Meyer, minha orientadora, que me adotou e foi competente, envolvida e carinhosa, meu muito obrigada, do fundo do coração.

À Gabriela Andrade da Silva, autora da tese irmã, como disse a Sonia. Obrigada por surgir e remodelar meu estudo. Já sinto saudades dos nossos cafés, reunião por Skype à noite, e-mails... Ainda bem que temos projetos futuros juntas. Ainda bem!

À Aparecida Angélica Paulovic Z. Sabadini, amiga, parceira de tantos projetos, bibliotecária exemplar, obrigada pela revisão cuidadosa das referências, leitura atenta de parte do texto e por ficar comigo até o último minuto.

Aos membros da banca de defesa, que compartilharam comigo um momento único e especial da minha vida acadêmica e promoveram um debate rico e instrutivo, além de fornecerem contribuições excelentes para a versão revisada deste trabalho.

À Helina, muito obrigada por arrumar minhas tabelas, figuras, formatar o texto e cuidar de mim com tanto carinho e dos passarinhos também.

Obrigada aos colegas da Biblioteca Dante Moreira Leite: Aline; Ana Rita; Carla; Cristiane; Elaine; Fernanda; Flávio; Lilian; Lúcia; Marcos; Marta; Nilza; Renato; Roseni; Silvana; Tatiana; Teresa; Vanessa e Wanderley, por todo o apoio e paciência nos últimos tempos, quando minha cabeça estava na tese e na tese e na tese.

À Profa. Dra. Emma Otta por ter me apoiado na decisão de fazer meus estudos no Programa de Psicologia Experimental.

À Eneida Cardoso Sampaio (*In Memoriam*) por ter me apresentado à Biblioteconomia e à Elza Correa Granja, por me apresentar à Psicologia e por terem me ensinado tantas coisas, agradeço emocionada.

À Sonia Maria Caetano de Souza, obrigada pelas orientações seguras e atenção.

À minha mãe, Maria do Nascimento Leite e meu pai, José Leite da Silva (*In Memoriam*) por terem me ensinado a nunca desistir e lutar pelo que eu quero. Mãe, amo você, vamos ter mais tempo juntas agora!

Aos meus filhos, Cassio Luiz Cardoso Sampaio e Glauco Luiz Cardoso Sampaio (na ordem de chegada) simplesmente por existirem, por serem meu orgulho maior, por fazerem minha vida mais feliz e por terem me trazido duas filhas: Maria Soledad Zabalza Sampaio e Janaína Cipriano. Duas moças lindas e queridas.

À minha irmã e meus irmãos, cunhado, cunhadas e sobrinhos pelo carinho e por estarem sempre presentes em minha vida.

À USP, ao IPUSP e aos seus funcionários, meus colegas, professores e alunos.

Ao Departamento de Psicologia Experimental do Instituto de Psicologia da Universidade de São Paulo por me receber com tanto carinho e pelo apoio financeiro para obtenção de artigos por comutação bibliográfica.

Aos editores e autores que me trouxeram a inquietação para fazer este estudo.

Epígrafe

“Em princípio, a comunidade científica é uma república democrática.
Qualquer pessoa com ideias novas ou com críticas válidas acerca
das ideias correntes pode publicar seu trabalho.”

(John Ziman, A força do conhecimento: a dimensão científica da sociedade, 1981)

RESUMO

Sampaio, M. I. C. (2013). *Qualidade de artigos incluídos em revisão sistemática: comparação entre latino-americanos e de outras regiões*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Introdução: A qualidade de um artigo científico é diretamente proporcional à qualidade metodológica e à apresentação dos resultados da pesquisa que o originou. Perguntas, ou hipóteses, muito bem explicitadas, objetivos claramente definidos, metodologia adequada que conduza aos objetivos propostos, com base em procedimentos metodológicos rigorosos e suficientemente descritos e justificados, são elementos essenciais no relato de pesquisas. **Objetivo:** Avaliar a qualidade metodológica e da apresentação de resultados de artigos sobre prevalência de depressão pós-parto, incluídos em revisão sistemática, comparando artigos latino-americanos e de outras regiões. **Metodo:** O delineamento do estudo foi de pesquisa documental, com análise crítica da literatura. As etapas de definição das bases de dados, elaboração da estratégia de busca, classificação dos estudos de acordo com os critérios de inclusão e análise dos estudos foram efetuadas por duas revisoras independentes. A qualidade dos artigos foi mensurada de acordo com a Escala de Loney, comparados o desempenho dos artigos latino-americanos e não latino-americanos e a procedência das revistas que os publicaram. Títulos, resumos e palavras-chave também foram avaliados, em função de normas internacionais e terminologias da área. Técnicas estatísticas foram aplicadas e as hipóteses testadas com o uso de testes não paramétricos. **Resultados:** Foram recuperados 1.894 registros de estudos originais e, após o processo de seleção, 337 foram incluídos na revisão sistemática. Desses, 34 eram estudos latino-americanos que foram avaliados e comparados com uma amostra aleatória pareada de 34 artigos não latino-americanos. A qualidade da apresentação da metodologia e dos resultados dos dois grupos de artigos não apresentou diferença significativa, sendo considerada fraca em ambos os casos. Em relação à procedência das revistas, também, não foram encontradas diferenças significativas. **Discussão:** Artigos científicos são evidências que podem apoiar a tomada de decisão, tanto na prática clínica como na gestão pública. Para que possam ser utilizados com essa finalidade, a qualidade da metodologia e os resultados devem ser primorosos. **Conclusão:** Revisar o vasto conhecimento psicológico, gerado nos últimos anos, em revisões sistemáticas, narrativas e exploratórias é um caminho para o registro e análise constante dos procedimentos utilizados e das intervenções que podem promover melhorias na qualidade de vida das pessoas. No entanto, a capacidade do pesquisador em desenvolver estudos que produzam resultados confiáveis e generalizáveis demanda investimentos, tanto na América Latina, quanto em outras regiões.

Palavras-chave: Trabalho científico. Revisão sistemática. Psicologia baseada em evidências. Prevalência. Depressão pós-parto. América Latina.

ABSTRACT

Sampaio, M. I. C. (2013). *Quality articles included in the systematic review: Comparison between Latin-american and others regions*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Introduction: The quality of an article is directly proportional to the methodological quality and to the presentation of its associated research results. Questions or hypothesis well explained, clearly defined objectives, proposed objectives supported by an adequate methodology – based on sufficiently described and justified rigorous methodological procedures – are all essential elements in the correct presentation of research findings. **Objective:** Assess the quality of both the methodology and results presentation on articles about postpartum depression, included in systematic review, contrasting Latin-American with foreign articles. **Method:** The study is of the documentary research type with critical literary analysis. Two independent reviewers performed the definition of the data sources, definition of the search criteria, classification of the studies according to the selection criteria and data analysis. The quality of the articles was measured according to the Loney Scale in comparing the performance of Latin-American and foreign articles and the location of the respective publishing journals. Titles, summaries and keywords were also evaluated in terms of international normative references and field specific terminologies. Technical statistics were applied and the hypothesis tested through the usage of non-parametric tests. **Results:** 1,894 original study records were retrieved and 337 included in the systematic review after passing through the selection process. Amongst those, 34 were identified as Latin American studies and which were evaluated and compared to a random sample of 34 foreign articles. The quality of both the methodology presentation and results of the two groups or articles did not show a significant difference. In both cases, quality was found to be low. The same results in regards to regional differences could be found when comparing the country of origin of the journals. **Discussion:** Scientific articles are evidences that can support decision making both in the clinical practice and in public policymaking. That can only be achieved if the quality of methodological quality and results presentation are high. **Conclusion:** The review of the vast sources of psychological knowledge (created over the last few years) in systematic reviews, narratives and exploratory is a path to record the constant analysis of the utilized procedures and of the intervention methods that could promote quality of life improvements. However, the ability of researchers to develop studies that generate reliable results that can be generalized (both in Latin American and worldwide) requires significant investments.

Keywords: Scientific work. Systematic review. Psychology evidence based. Postpartum depression. Prevalence. Latin America.

RESUMEN

Sampaio, M. I. C. (2013). *Calidad de artículos incluidos en revisión sistemática: comparación entre latinoamericanos y otras regiones*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo.

Introducción: La calidad de un artículo científico es directamente proporcional a la calidad metodológica y a la presentación de los resultados de la investigación que lo originó. Preguntas o hipótesis muy bien explicitadas, objetivos claramente definidos, metodología adecuada que conduzca a los objetivos propuestos con base en procedimientos metodológicos rigurosos y suficientemente descriptos y justificados, son elementos esenciales en el relato de las investigaciones. **Objetivo:** Evaluar la calidad metodológica y la presentación de resultados de artículos sobre la prevalencia de depresión post parto. Estos artículos han sido sujetos a una revisión sistemática, comparando artículos latinoamericanos y de otras regiones. **Método:** El delineamiento del estudio se ha constituido en una investigación documental con un análisis crítico de la literatura pertinente. Las etapas de definición de las bases de datos, elaboración de la estrategia de búsqueda, clasificación de los estudios de acuerdo con los criterios de inclusión, y análisis de los mismos, fueron efectuadas por dos revisoras independientes. La calidad de los artículos fue medida de acuerdo con la Escala de Loney comparando el desempeño de los artículos latinoamericanos y no latinoamericanos y la procedencia de las revistas que los publicaron. Títulos, resúmenes y palabras clave también fueron evaluados en función de las normas internacionales y terminologías del área. Técnicas estadísticas fueron aplicadas y las hipótesis evaluadas con el uso de pruebas no paramétricas. **Resultados:** Fueron recuperados 1.894 registros de estudios originales y, después del proceso de selección, 337 fueron incluidos en la revisión sistemática. De éstos 34 estudios latinoamericanos fueron evaluados y comparados con la muestra aleatoria pareada de 34 artículos no latinoamericanos. La calidad de la presentación de la metodología y de los resultados de los dos grupos de artículos no presentó diferencia significativa siendo considerada débil en ambos casos. En relación a la procedencia de las revistas, tampoco fueron encontradas diferencias significativas. **Discusión:** Los artículos científicos son evidencias que pueden apoyar la toma de decisión, tanto en la práctica clínica como en la gestión pública. Para que puedan ser utilizados con esta finalidad, la calidad de la metodología y los resultados deben ser primorosos. **Conclusión:** Revisar el vasto conocimiento psicológico, generado en los últimos años en revisiones sistemáticas, narrativas y exploratorias es un camino para el registro y análisis constante de los procedimientos utilizados y de las intervenciones que pueden promover mejoras en la calidad de vida de las personas. No obstante, la capacidad del investigador de desenvolver estudios que produzcan resultados confiables y generables demandan inversiones, tanto en América Latina, como en otras regiones.

Palabras clave: Trabajo científico. Revisión sistemática. Psicología basada en evidencias. Prevalencia. Depresión post parto. América Latina.

LISTA DE ILUSTRAÇÕES

Figura 1 – Métodos de Pesquisa	52
Figura 2 – Pirâmide das Evidências Científicas.....	55
Figura 3 – Um autor latino-americano diante de editor internacional?.....	79
Figura 4 – Fluxograma da seleção de estudos para a revisão sistemática.....	108
Figura 5 - Universo do estudo comparativo.....	144
Figura 6 – Frequência total das pontuações aos artigos (68), em função da Escala de Loney et al. (1998)	144
Figura 7 Comparação das pontuações totais obtidas na Escala de Loney et al. (1998) pelos artigos latino-americanos e de outras regiões (em números de artigos).....	151
Figura 8 – Comparação das pontuações totais obtidas na Escala de Loney et al. (1998) pelas revistas latino-americanas e não latino-americanas (em número de artigos).....	152
Figura 9 – Pontuações dos artigos em relação ao objetivo do estudo.....	153

LISTA DE TABELAS

Tabela 1 – Escala de Qualidade para Avaliação de Artigos Científicos (ARR).....	32
Tabela 2 – Diferenças entre a Revisão Sistemática Clássica e a Revisão Sistemática Exploratória.....	64
Tabela 3 – Tipos de Pesquisa Definidos na Força Tarefa APA.....	64
Tabela 4 – Escala de Loney et al. (1998).....	96
Tabela 5 – Bases de dados, data de consulta e de registros recuperados	106
Tabela 6 – Estratégias de busca para as bases de dados consultadas.....	107
Tabela 7 – Concordância entre avaliadoras quanto à inclusão de estudos pelos títulos, resumos e textos completos.....	109
Tabela 8 – Títulos das revistas e numero dos artigos incluídos na revisão sistemática	111
Tabela 9 – Países em que a coleta de dados dos trabalhos incluídos na revisão sistemática foi realizada.....	119
Tabela 10 – Data de publicação e frequência dos estudos incluídos na revisão sistemática.....	122
Tabela 11 – Distribuição dos países em que a coleta de dados foi realizada: latino-americanos e de outras regiões.....	125
Tabela 12 – Distribuição dos artigos conforme o ano de publicação e a origem da revista.....	127
Tabela 13 – Título e país e procedência da revista, frequência de artigos por revista e índices de impacto no JCR (FI), índice H no SCImago e FI no SciELO.....	129
Tabela 14 – Título da revista que publicou os estudos latino-americanos, país e frequência de publicação.....	133
Tabela 15 – Título do estudo latino-americano, revista estrangeira que publicou o artigo, país de origem do estudo, número de citações no Google Acadêmico, Scopus e WOK e ano de publicação.....	135

Tabela 16 – Título do artigo e da revista, país de origem, número de citações no Google Acadêmico, WOK e Scopus e ano de publicação.....	137
Tabela 17 – Concordância entre avaliadoras para os itens da Escala de Loney et al. (1998).....	143
Tabela 18 - Número e frequência de artigos que obtiveram pontuações em cada item da Escala de Loney et al. (1998).....	146
Tabela 19 – Frequência de artigos que realizaram a coleta de dados em países latino-americanos e não latino-americanos que obtiveram pontuação em cada item da Escala de Loney et al. (1998).....	154
Tabela 20 - Frequência de artigos publicados em revistas latino-americanas e não latino-americanas que obtiveram pontuação nos itens da Escala de Loney et al. (1998)	157
Tabela 21 - Posicionamento dos estudos latino-americanos e não latino-americanos em função da Escala Loney et al. (1998).....	161
Tabela 22 - Posicionamento dos artigos publicados em revistas latino-americanas e não latino-americanas em função da Escala de Loney et al. (1998)....	161
Tabela 23 - Países latino-americanos e número de registros LILACS e em títulos de revistas em SciELO.....	176

LISTA DE ABREVIATURAS E SIGLAS

ARR - Acceptable, Revise, Reject (escala de avaliação de artigos)

ARRR - Accepted (without revision), Revision (minor), Revision (enough), Rejected (idem ARR, com revisão e ampliação)

APA – American Psychological Association

BIREME – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde

BVS-PSI ULAPSI – Biblioteca Virtual em Saúde - Psicologia - União Latino-americana de Psicologia

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CINAHL – Cumulative Index to Nursing & Allied Health Literature

CLASE – Citas Latinoamericanas en Ciencias Sociales y Humanidades

Comut – Programa de Comutação Bibliográfica

DECS – Descritores em Ciências da Saúde

DPN – Depressão pós-natal

DPP – Depressão pós-parto (em inglês, PPD – postpartum depression)

DSM-IV – Manual Diagnóstico e Estatístico de Transtornos Mentais. 4ª Edição

EMBASE – Excerpta Medica Database

ERIC –Education Resources Information Center

EVIPNet – Rede de Políticas Informadas por Evidências

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

FI – Fator de Impacto

IC – Intervalo de Confiança

IDH – Índice de Desenvolvimento Humano

IPUSP – Instituto de Psicologia da Universidade de São Paulo

Lilacs – Literatura Latino-Americana e do Caribe em Ciências da Saúde (base de dados)

MEDLINE – Medical Literature Analysis and Retrieval System Online (Sistema Online de Busca e Análise de Literatura Médica)

MESh – Medical Subjects Headings

NLM – National Library of Medicine's

OMS – Organização Mundial de Saúde

PBE - Práticas Baseadas em Evidências

PEP – Psychanalytic Electronics Publishing

PePSIC – Periódicos Eletrônicos em Psicologia

PICO – Participantes, Intervenção, Controle e *Outcomes* (desfechos)

PNUD – Programa da Nações Unidas para o Desenvolvimento

PPBE – Prática da Psicologia Baseada em Evidências

ProQuest – Base de dados de dissertações, teses e outros tipos de documentos científicos.

Psicodoc – Base de dados do Colégio de Psicólogos de Madrid

PsycINFO – Psychological Information (APA)

PsycLIT – Psychological Literatura (APA)

PsycNet – Rede de Fonte de Informações da APA

PubMed – Base de dados publicada pelo National Center for Biotechnology Information (NCBI) na National Library of Medicine (NLM)

PVO – Participantes, Variáveis e *Outcomes* (desfechos)

RedAlyc – Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

SCAD – Serviço de Acesso ao Documento

SciELO – Scientific Electronic Library Online (Biblioteca Científica Eletrônica Online)

SciMago - SCImago Journal & Country Rank – Portal de indicadores bibliométricos

SCOPUS – Base de dados multidisciplinar, produzida pela editora Elsevier

Sociofile – Sociological Abstracts (base de dados)

UNIFESP – Universidade Federal do Estado de São Paulo

WOK - Web of Knowledge

WOS - Web of Science

SUMÁRIO

1 APRESENTAÇÃO	20
2 INTRODUÇÃO	25
2.1 O processo de avaliação de artigos científicos	27
2.2 Avaliação de artigos na área de Psicologia	37
2.3 Pergunta de pesquisa	44
2.4 Tipos de pesquisas utilizados pela ciência psicológica	47
2.5 Revisão Sistemática: padrão-ouro das evidências científicas?	53
2.5.1 Aprendendo com a Medicina Baseada em Evidências	55
2.5.2 Evidências científicas: a tomada de decisão nos processos psicológicos	59
2.5.2.1 Psicologia Baseada em Evidências	68
2.6 Artigos publicados em revistas latino-americanas contêm boas evidências?	77
2.7 Por quê Prevalência de DPP?	80
2.7.1 Estudos de Prevalência	80
2.8 Depressão Pós-Parto: um grave transtorno do humor	84
3 OBJETIVOS E HIPÓTESES	87
3.1 Objetivo geral	87
3.2 Objetivos específicos	87
3.3 Hipóteses	87
4 MÉTODO	88
4.1 Etapas da revisão sistemática	89
4.1.1 Formulação da pergunta da revisão sistemática	89
4.1.2 Critérios de inclusão dos trabalhos	89
4.1.3 Bases de dados selecionadas	91
4.1.4 Definição da estratégia de busca	92
4.1.5 Confronto de duplicidades	93
4.1.6 Armazenamento e gerenciamento dos dados dos registros	93
4.1.7 Triagem pelo título do trabalho	94
4.1.8 Triagem pelo resumo	94
4.1.9 Obtenção dos textos completos	94
4.1.10 Avaliação do texto completo	95
4.1.11 Inclusão de outras variáveis para avaliação metodológica	101
4.1.12 Verificação do nível de concordância entre as revisoras	102
4.2 Procedimentos específicos deste estudo comparativo	102
4.2.1 Universo do estudo	103
4.2.2 Instrumentos	103
4.2.3 Critérios de seleção	103
4.3 Avaliação da qualidade de forma	103
4.4 Análise dos dados	104
5 RESULTADOS	105
5.1 Resultados gerais da revisão sistemática	105

5.2 Seleção dos estudos	108
5.3 Análises descritivas da Revisão Sistemática	110
5.4 Resultados da qualidade de artigos incluídos na revisão sistemática	124
5.5 Qualidade metodológica e da apresentação dos resultados dos estudos	140
5.6 Testes de hipóteses	150
5.7 Análise da qualidade dos títulos, resumos e palavras-chave dos artigos	162
5.7.1 Avaliação dos títulos dos artigos	163
5.7.2 Avaliação dos resumos dos artigos	164
5.7.3 Avaliação das palavras-chave dos artigos	167
6 DISCUSSÃO	168
6.1 Buscando a literatura para a revisão sistemática	170
6.2 Diminuindo os vieses de seleção: concordância entre avaliadoras	173
6.3 Presença de estudos latino-americanos na revisão	175
6.4 Publicação em inglês e impacto dos artigos latino-americanos	177
6.5 Avaliação de conteúdo e busca por evidência científica	180
6.6 Qualidade de forma: títulos, resumos palavras-chave	185
7 CONCLUSÃO	186
7.1 Direções Futuras	189
REFERÊNCIAS	191
APÊNDICE A - Instruções para seleção e análise dos estudos	204
APÊNDICE B - Referências dos estudos avaliados	217

1 APRESENTAÇÃO

Esta tese é um conjunto de textos organizados de acordo com normas e procedimentos, encerra um ciclo da minha vida, um importante ciclo, por isso vou me permitir falar na primeira pessoa do singular nesta seção do trabalho. Foi realizada no Programa de Pós-Graduação em Psicologia Experimental, do Instituto de Psicologia da Universidade de São Paulo, sete anos depois da conclusão do meu mestrado, na Escola de Comunicação e Artes da mesma Universidade, na qual também conclui a graduação, no ano de 1989. Portanto, minha formação é na área de Comunicação, precisamente, em Ciência da Informação ou Biblioteconomia. Fazer o doutorado em Psicologia foi uma forma de me aproximar ainda mais da área que atuo há 23 anos, desde que me formei bibliotecária, e realiza dois desejos: 1) colocar em prática o que temos tentado mostrar nos cursos que oferecemos sobre escrita científica; 2) aprender mais sobre uma área que vem despertando muito meu interesse: o uso de evidências científicas na tomada de decisão em saúde. Claro que preenche uma lacuna que considero importante na minha formação: operar no universo da pesquisa e usar o método científico, efetivamente.

Afinal, o que é uma tese? Tem origem na palavra grega *Thesis* e quer dizer proposição. Consiste numa asserção discutida e defendida por alguém, com base em hipóteses ou pressupostos e, atualmente, encerra um ciclo de estudos acadêmicos em uma determinada universidade. Entre o início dos meus estudos pós-graduados e a conclusão desta tese, muitas coisas aconteceram.

Desde o projeto inicial para participação do processo de seleção, denominado: *Psicologia Latino-Americana Baseada em Evidências: Diretrizes para uma Prática em Prol de Melhores Condições de Vida* à finalização deste trabalho, algumas coisas mudaram e outras permaneceram inalteradas. Das coisas que mudaram e que gostaria de destacar foi a chegada à minha vida acadêmica e pessoal de uma colega de turma que se tornou peça essencial deste estudo e da qual jamais esquecerei. Na verdade, Gabriela Andrade da Silva, foi a minha co-orientadora e me ensinou muitas coisas. Interferiu na transformação do meu projeto, dos meus objetivos, auxiliou-me a redefinir minhas hipóteses, e, principalmente, apresentou-me a uma área nunca antes apreendida: a prevalência da depressão pós-parto. Do projeto inicial ao estudo aqui apresentado as mudanças foram profundas e o aprendizado imenso.

Outra mudança a ser comemorada foi a chegada da professora Sonia Beatriz Meyer no meu projeto. Do pouco contato que eu tinha com a Sonia inicialmente, até as interferências que ela fez no meu texto, sugerindo, questionando e participando ativamente da construção do trabalho, houve uma trajetória de surpresas felizes. Tenho a plena convicção que a providência a colocou no meu caminho para colaborar muitíssimo na remodelação do meu projeto. Sonia Meyer, sua presença é forte nesta tese!

Dentre as mudanças, duas convicções se mantiveram firmes e inabaláveis: 1) a de querer aprender e falar da Psicologia Baseada em Evidências, e 2) a de verificar se estudos latino-americanos apresentam diferenças de qualidade em relação aos não latino-americanos.

Buscar caminhos para incrementar o interesse de pesquisadores e autores da área de Psicologia pela Psicologia Baseada em Evidências, que é uma área importante e necessária para a região, foi uma das motivações deste estudo. Para tanto, era necessário avaliar artigos científicos, pois são essas as evidências consideradas mais válidas na atualidade.

A oportunidade de participar de uma revisão sistemática, discutindo cada etapa com a Gabriela, definindo critérios, buscando nas bases de dados, avaliando os artigos por título, pelo resumo e, finalmente, pelo texto completo, foram lições que levarei para o resto da minha vida. Enquanto esse percurso era percorrido, eu ia aprendendo sobre prevalência, depressão pós-parto, taxas de concordância entre avaliadores, técnicas de amostragem, taxas de perdas, sensibilidade, especificidade, intervalo de confiança e outras variáveis que faziam parte do estudo. Assim, quando me perguntam: “Por que prevalência de depressão pós-parto?”, eu respondo: “O motivo foi o de participar de uma revisão sistemática e a prevalência de depressão pós-parto foi a oportunidade que eu tive de estudar uma área nova e que integra a Psicologia com a Ciência da Informação”. Participar de todas as etapas de uma revisão sistemática foi uma oportunidade ímpar de aprendizado. Mesmo que não houvesse a tese, o conhecimento adquirido já teria valido o esforço.

Os objetivos de “fortalecer a pesquisa de processos comportamentais básicos em nosso país, com o sentido de se obter independência na solução de problemas, através da produção do conhecimento científico” e “a partir das oportunidades criadas por esse fortalecimento, formar novos pesquisadores e docentes, qualificados para difundir e multiplicar esse conhecimento em outros centros do país

e derivar desse conhecimento a solução prática de problemas sociais e regionais”, declarados na página web do Programa de Pós-graduação em Psicologia Experimental, do Instituto de Psicologia da USP,¹ também chamou minha atenção e despertou-me um interesse em realizar meus estudos no programa.

A interação com profissionais da informação, pesquisadores, professores, editores e autores responsáveis pela geração do conhecimento científico na área de Biblioteconomia e Psicologia, promovida pela minha atuação como bibliotecária e pela coordenação da Biblioteca Virtual em Saúde – Psicologia (BVS-Psi), do Brasil e da América Latina, além da participação na Diretoria da Associação Brasileira de Editores Científicos de Psicologia (ABECiPsi), desde seu nascimento, em 2006, permitiu-me assistir, fascinada, a realização de sonhos e acompanhar a eterna luta pela melhoria da qualidade das revistas científicas latino-americanas. A defesa da visibilidade do nosso conhecimento também é uma batalha a qual acompanho com muito interesse. Tudo isso me encorajou a avançar na empreitada deste trabalho.

Assim, ainda que modestos e parciais, entrego os resultados dos meus estudos, esperando ter contribuído, minimamente, para ampliar a discussão sobre a qualidade de artigos científicos e, também, em uma área que tende a se fortalecer e se expandir: a Psicologia Baseada em Evidências. A Ciência da Informação também será impactada pelo avanço dessa nova área, afinal, estamos falando de informação na tomada de decisão.

No segundo Capítulo, o assunto é introduzido e as características dos artigos científicos são apresentadas. O sistema de revisão por pares, algumas escalas para revisão de manuscritos e um breve relato do processo de avaliação das revistas de Psicologia, ilustram parte do capítulo. O problema da retratação de artigos e a má conduta de autores são brevemente introduzidos. Em seguida, as partes que compõem o artigo científico, com ênfase para métodos utilizados pela Psicologia esclarecem o leitor. As evidências científicas na tomada de decisão nos processos psicológicos e como instrumento da gestão de políticas públicas ganham ênfase e relevância no texto. É lançada uma pergunta sobre as revistas latino-americanas publicarem boas evidências e uma breve explicação sobre prevalência e depressão pós-parto encerram o capítulo. Em meio às definições e às fundamentações, a literatura é revisada.

¹ http://www.ip.usp.br/portal/index.php?option=com_content&view=article&id=975%3Aobjetivos-do-programa&catid=191&Itemid=114&lang=pt

O objetivo geral e os específicos que guiaram o desenvolvimento do estudo com as três hipóteses testadas no trabalho formam o Capítulo 3.

O método de pesquisa utilizado foi descrito em detalhes no Capítulo 4. Estudos transversais de análise crítica da literatura, que podem, também, ser chamados de delineamentos de pesquisa documental, são aqueles cujos objetivos e hipóteses verificam-se pela análise de documentos bibliográficos ou não bibliográficos, demandando uso de metodologia adequada. A coleta, seleção, organização e análise dos dados foram feitas de forma sistematizada para se chegar aos resultados esperados e descritas nesse capítulo.

Os resultados compõem o Capítulo 5, quando dados gerais da revisão sistemática, que embasou o trabalho de Gabriela, e os que compararam artigos latino-americanos e de outras regiões, são apresentados. As buscas bibliográficas, a seleção dos textos para avaliação, os critérios de inclusão, a concordância entre as revisoras, as estatísticas descritivas e os testes de hipóteses são apresentados. A análise dos títulos, resumos e palavras-chave encerram esta parte da tese.

As discussões dos principais resultados e o diálogo com autores que desenvolveram estudos de natureza semelhante conformam o Capítulo 6. A presença de estudos latino-americanos na revisão sistemática e a importância da qualidade metodológica e da apresentação dos resultados em artigos científicos, que os habilitam a serem usados como evidências na tomada de decisão em gestão e políticas públicas, são debatidas.

O Capítulo 7 conclui a tese e reforça a necessidade de a Psicologia latino-americana se apoderar das técnicas de revisão sistemática para aplicação da Psicologia Baseada em Evidências como ferramenta de apoio ao avanço das pesquisas e na solução de problemas comuns da população da região. Para finalizar são apresentadas algumas direções de estudos futuros que apontam para a importância em se desenvolver protocolos de verificação de qualidade metodológica e da apresentação de resultados em todas as áreas, para que sejam utilizadas por pesquisadores ainda na fase de definição da metodologia de estudos em andamento. Sugestão para que a revisão integrativa seja adotada em estudos na área também foi oferecida. Parcerias entre pesquisadores de Psicologia e profissionais da informação são sugestões para ampliação da construção de estudos acadêmicos de interesse para as áreas de atuação desses profissionais. Finalmente, é feito um apelo para que a comunidade científica, as agências de fomento e os

programas de pós-graduação revejam alguns critérios de avaliação da produção científica dos cientistas, pesquisadores e docentes, considerando a qualidade dos estudos e desestimulando o produtivismo exagerado, que valoriza os aspectos quantitativos em prol dos qualitativos.

2 INTRODUÇÃO

A qualidade de um artigo científico é diretamente proporcional à qualidade metodológica e à apresentação dos resultados da pesquisa que o originou. “A qualidade científica de uma pesquisa e do artigo que a divulga resulta, fundamentalmente, da sua validade externa (o poder de generalizar os resultados) e da sua validade interna (se efetivamente está sendo medido o que se deseja mensurar)” (Hoppen, 1998, p. 155). A relevância do tema de pesquisa, apresentado em um objetivo muito bem definido, as teorias e os estudos que sustentam a importância da pesquisa, embasados em uma revisão da literatura atualizada e julgada com critérios rigorosos de seleção, metodologia consistente, se possível testada em pesquisas anteriores, coleta e análise de dados com o mínimo de viés e profundidade, além do poder de síntese dos resultados, são características que permitem avaliar a qualidade da apresentação de um relato de pesquisa. A habilidade do pesquisador (autor) em sintetizar e redigir corretamente, de preferência baseada em diretrizes especializadas, cada uma dessas partes da pesquisa no bom artigo científico, faz a diferença do relato.

O artigo científico é o meio por excelência para publicação de resultados de pesquisas, relatos de experiência profissional, revisão de teorias/temas e outros tipos de contribuições que pesquisadores e profissionais entregam à sociedade. É um texto técnico-científico, escrito por um ou mais autores, de acordo com as normas editoriais de uma determinada revista. Pode ser o resultado de uma pesquisa científica, da revisão ou da reflexão sobre um determinado tema ou teoria, que são apresentados em forma de artigos empíricos originais, de revisão ou teóricos (Sabadini, Sampaio, & Koller, 2009a). O *Dicionário de Biblioteconomia e Arquivologia* resume: artigo científico é um “texto escrito que foi aprovado para publicação ou publicado num periódico científico” (Cunha & Cavalcanti, 2008, p. 32). A breve definição deixa claro que o artigo científico passa por um processo de avaliação, pois necessita ser aprovado. A avaliação antes da aprovação é a grande diferença do artigo científico em relação aos outros tipos de publicação científica. Especialistas avaliam o texto e emitem pareceres antes da publicação, validando o conhecimento ali registrado e contribuindo com sua experiência para a melhoria do texto.

Escrever um artigo científico é muito diferente de escrever um poema, conto, romance ou ficção. Não depende tanto de talento, embora esse seja um facilitador para autores de qualquer tipo de escrita. A criatividade também é um fator que auxilia na redação de bons artigos científicos. A vantagem é que autores podem ser formados e treinados na tarefa de redação de textos de qualidade. Um conjunto de regras, normas e procedimentos conduz a redação dos artigos e serve de orientação para os autores menos experientes. De qualquer forma, depende de habilidades, mas o interessante é que essas podem ser desenvolvidas e melhoradas com a prática, não dependendo tanto de vocação.

Quando a velocidade de geração do conhecimento científico já não permitia a demora para publicar o livro “nasceu, por sua vez, a conscienciosa invenção do artigo científico, meio de transmitir e preservar os conhecimentos que passavam a crescer com rapidez maior do que permitiria transpô-los para a forma de livros” (Price, 1976, p. 101). A ciência é um empreendimento humano que busca compreender e explicar os fenômenos e descobrir a ordem das coisas. “A matéria prima que alimenta a ciência são todas as outras contribuições já dadas por seus contemporâneos e predecessores” (Price, 1976, p.117). Se o principal resultado do trabalho de pesquisa de um homem é o conhecimento, podemos dizemos que esse homem está dedicado à ciência, logo é um cientista. Os trabalhos em andamento, ou aqueles que o pesquisador (não havia mulheres na ciência, ainda...) decidia compartilhar, eram lidos nas reuniões das sociedades científicas. Para ampliar a audiência, as Sociedades passaram a publicar as atas das reuniões. Por isso, muitas revistas tinham o nome, e têm até hoje, de Atas...(contribuição direta de Piotr Trzesniak).

Como vimos, o artigo se diferencia dos outros tipos de trabalhos científicos pelo processo de avaliação ao qual é submetido. Esse tipo de avaliação recebe o nome de *peer-review* (em inglês), que significa revisão por pares, ou *referee system*, sistema de arbitragem, que é a revisão feita por especialistas antes da aceitação ou da rejeição de um manuscrito para publicação em uma revista científica (Pessanha, 1998). No entanto, outros processos de avaliação estão ganhando força atualmente, o que podemos chamar de avaliação de conteúdo pós-publicação. Deste trataremos mais à frente, agora vamos discutir sobre como teve início o processo de avaliação por pares, considerado o meio por excelência de avaliação de artigos científicos.

2.1 O processo de avaliação de artigos científicos

“Na produção científica, a quantidade e a qualidade são aspectos centrais em qualquer área do conhecimento, pois são indicadores do seu vigor, constituindo-se em parâmetro para sua mensuração e caminho para o seu desenvolvimento crescente” (Soares, Victoria, Cavalieri, & Bottino, 2006). Em relação à quantidade, Price (1976) analisou matematicamente o crescimento exponencial das revistas científicas e criticou: “Até agora nada se disse da qualidade da pesquisa, em oposição à sua quantidade. Qualidade, naturalmente, é algo muito mais difícil, naturalmente, de determinar e exigirá investigação muito mais séria do que tem merecido” (p. 163). Price gostaria de saber que, no ano de 2010, Haslam e Laham exploraram sistematicamente a relação entre quantidade *versus* qualidade *versus* impacto de publicações acadêmicas. Os autores analisaram o estilo da redação e o prestígio institucional dos autores, por meio dos índices bibliométricos, embora alertassem que os fatores de impacto não capturam as complexidades multidimensionais da qualidade dos artigos. Elegeram uma amostra intencional de 86 psicólogos sociais com grau de doutorado e analisaram suas publicações na *Web of Knowledge* (WOK), antes e depois do doutoramento. A quantidade de artigos publicados foi um forte preditor do impacto no grupo de pesquisadores, embora os autores argumentem que a qualidade também é limitada, pois uma revista com fator de impacto alto pode apresentar artigos com impacto modesto. Essa descoberta implica que a qualidade e a quantidade refletem maior realização científica: pesquisadores mais realizados tendem a publicar mais artigos e a fazê-lo em revistas de maior prestígio. Concluem que qualidade e quantidade de publicações não são fatores antagônicos, mas alertam que as descobertas não significam que pesquisadores devem apontar para a quantidade acima de tudo. Artigos que aparecem em revistas “melhores” são, em média, mais impactantes, mesmo sabendo que a qualidade das pesquisas não pode ser reduzida aos índices bibliométricos. Price, provavelmente, ficaria confuso em relação à validade desses índices, mas certamente concluiria que ainda não é com o fator de impacto que o problema da avaliação da qualidade está resolvido. O problema da qualidade do artigo não pode ser resolvido por nada feito depois que ele é publicado. FI é um indicador post-mortem, não influi na qualidade. É *apenas* uma tentativa (meio fracassada) de avaliá-lo (contribuição de Piotr Trzesniak).

Vale esclarecer que o Fator de Impacto (FI) (de dois anos) de uma revista “é um indicador bibliométrico utilizado para medir as citações a um determinado periódico. É calculado dividindo-se o número de citações recebidas por um periódico em um determinado ano pelo número de artigos publicados por ele nos dois anos anteriores” (Sampaio & Sabadini, 2009, p. 108). O impacto de um artigo pode ser medido pelo número de vezes que o mesmo é citado por outros trabalhos. Antes restritas às bases de dados WOK e Scopus, atualmente, o Google Acadêmico faz essas medições também, com a (des)vantagem de considerar todas as publicações que ele localiza na Internet, independente de indexação nas bases de dados. Vale lembrar que não considera todas as disponíveis, só as que ele encontra e consegue medir.

O primeiro modo para transmissão de ideias entre os cientistas foi a correspondência pessoal. As cartas eram trocadas entre os homens da ciência e a comunicação com os amigos era utilizada para se fazer o relato e a discussão das descobertas científicas, o que ficava restrito a pequenos grupos de interessados que discutiam essas ideias e, quase sempre, concordavam entre si (Stumpf, 1996). Essa troca informal de informação, a respeito das novas experiências, foi muito bem recebida pela comunidade científica e permitiu maior divulgação a respeito do que se fazia na Europa, naquele tempo, em termos de ciência. Como eram grupos de amigos, as informações eram disseminadas de maneira dirigida, uma vez que seus autores nunca as enviavam para aqueles que pudessem refutar suas teorias ou rejeitar seus experimentos. As descobertas científicas eram “...limitadas a um pequeno círculo de pessoas, essas ‘dissertações epistolares’, não se constituíram no método ideal para a comunicação do fato científico e das teorias” (Stumpf, 1996, p. 383). “...a correspondência através de trocas de cartas entre cientistas e estudiosos foi criada e mantida por trinta anos, graças a Père Mersenne (1588-1648), o que deu início à primeira rede de cientistas” (Beaulieu, 1995; Maury, 2003 citados por Bégault, 2009, p. 92). Estava formado o colégio invisível, tão bem aceito e popular no meio acadêmico e científico.

Embora diferente da troca das cartas pessoais do início do século XII, podemos questionar se não acontece coisa semelhante com algumas revistas na atualidade, pois, até hoje, algumas não enviam os artigos para avaliação externa e publicam artigos de acordo com os interesses do editor. O envio dos manuscritos

para que os pares avaliem e deem pareceres é uma forma de minimizar o problema da “amizade científica” e dar maior credibilidade às descobertas da ciência.

O processo de avaliação da produção científica surgiu de forma embrionária, juntamente com a publicação das primeiras revistas científicas, no século XVIII. Em 1731, a *Royal Society of Edinburgh* publicou no prefácio do primeiro volume da revista *Medical Essays and Observations* um procedimento para seleção de artigos baseado na revisão por especialistas de fora da instituição. A identidade do revisor não era informada, somente o parecer era enviado ao autor. Vinte anos depois, a *Royal Society of London* estabelecia o processo de revisão dos textos por juízes externos. Nascia um processo de avaliação de manuscritos que ganharia forma e se tornaria o modo mais utilizado de seleção de artigos para publicação de todos os tempos.

Os revisores são juízes encarregados de avaliar a qualidade do produto científico e são parte do controle social, pois avaliam seus pares em todos os âmbitos institucionais e distribuem as recompensas (Davyt & Velho, 2000). No caso do artigo científico, a recompensa é a aceitação para publicação. A revisão por pares é definida, segundo Davyt e Velho (2000), como a “análise por iguais”, ou ainda como “mecanismo auto-regulador da ciência moderna”. É, também,

um método organizado para avaliar o trabalho científico, que é usado pelos cientistas para garantir que os procedimentos estejam corretos, estabelecer a plausibilidade dos resultados e distribuir recursos escassos – como o espaço em revistas, fundos de pesquisa, reconhecimento e reputação. (Chubin & Hackett, 1990, p. 2)

O sistema de assessores, ou revisores, não apareceu tal como é conhecido hoje, mas tem se modificado ao longo da história, em resposta ao desenvolvimento da própria instituição científica e dos seus contextos (Davyt & Velho, 2000; Zuckerman & Merton, 1971). O processo objetiva verificar se o tema é relevante, original, e se o texto fornece informações, argumentos e interpretações que constituam apoio suficiente para as conclusões. A revisão visa a contribuir com a construção do conhecimento, mais do que a mera decisão de publicar ou não. O editor, a seu juízo, tem o poder de concordar ou discordar do parecer, sejam eles positivos ou negativos. O parecerista pode, por exemplo, estar defendendo ou protegendo um autor ou bloqueando uma nova ideia. Pode ainda estar acontecendo um “Conflito de interesses entre o autor e o parecerista, seja de ordem pessoal,

econômica, ideológica e cultural. Nesse caso, editores enfrentam a difícil tarefa de detectá-los e contorná-los a fim de garantir a qualidade e a credibilidade da revista” (Reiners et al., 2002, p. 4).

A avaliação final é sempre do editor e o reconhecimento e respeito por seu trabalho vai depender de sua postura ética e conhecimento da área. Em relação à questão ética na publicação e o conflito entre autores, editores e pareceristas, muito se tem debatido, porém, o consenso está distante de ser estabelecido (Buela-Casal, 2003; Goldenberg, 2001; Greene, 1998; Rodrigues, Crespo, & Miranda, 2006; Werneck, 2006). O que permeava as cartas aos amigos do início das publicações da ciência, hoje é chamado de conflito de interesse, e muitas revistas exigem que os autores declarem não haver tal conflito na redação do seu trabalho. Rodrigues et al. (2006) demonstram a preocupação com o assunto e enfatizam a

necessidade de verificar como ocorre o jogo de interesses, quando maus profissionais utilizam-se de associações de classe para promoção mútua e aproveitam-se desse respaldo, ou mesmo de sua inclusão em corpos editoriais de periódicos, para restringir a publicação de “inimigos”, de posições teóricas diferentes das suas. (p. 38)

Outra preocupação dos editores reside na situação em que o revisor compete com o autor em um mesmo tema e pode dirigir o parecer contrário à publicação do manuscrito. Embora a revisão seja às cegas, ou duplo-cego, dificultando ao revisor identificar o autor, muitas vezes, pelo assunto, pelas referências citadas no trabalho ou apenas buscando o título ou alguma frase peculiar na *web*, é possível identificar a autoria (Martínez, 2012).

Desde que surgiu, o processo de revisão de manuscritos antes da publicação tem sido aprimorado e diversas revistas publicam roteiros para auxiliar os revisores na árdua tarefa. Trata-se de preocupação legítima, que interfere diretamente na qualidade do conhecimento científico. Os aspectos a serem analisados em um manuscrito diferem de área para área, mas alguns pontos são comuns e não podem ser desconsiderados.

No ano de 1956, Dvorak publicou um guia para analisar relatos de pesquisas. O autor argumentava que os estudantes de doutorado tinham dificuldade para selecionar um problema de pesquisa, preparar a metodologia, coletar, processar e interpretar dados e escrever o relatório, denominado por ele de tese. Sugeria que os estudantes tivessem aulas para adquirir experiência como pesquisadores e aulas de

redação de textos científicos. Dvorak se surpreenderia se soubesse que, passados quase 60 anos da criação de sua escala, a problemática continua, acentuada por uma necessidade imensa de se produzir muito para ser bem avaliado.

Johnson (1957) estudou a habilidade de estudantes do curso de métodos de pesquisa em educação para avaliar e criticar artigos de pesquisa; concluiu que eles necessitavam de treinamento para desenvolverem a atividade. Com o objetivo de facilitar a capacitação dos alunos na tarefa, ele desenvolveu um método para avaliar artigos de pesquisas científicas no campo da Educação, como forma de preparar os alunos na elaboração do texto científico.

O “Report of the Committee on Evaluation of Research”, de autoria de Wandt e colaboradores (1967), relata o procedimento desenvolvido para avaliar a qualidade das pesquisas publicadas em revistas da área educacional. Os artigos foram avaliados mediante uma escala, com cinco características, que pontuava a qualidade da publicação. A tabela 1, adaptada do “Report” apresenta a escala.

Tabela 1 - Escala de Qualidade para Avaliação de Artigos Científicos (ARR)

Nível de qualidade	Descrição
5	Excelente - um modelo de boas práticas
4	Bom - poucos defeitos
3	Medíocre - nem bom, nem ruim
2	Pobre - alguns defeitos graves
1	Completamente incompetente - um exemplo horrível

Nota. Tabela traduzida e adaptada de “An evaluation of educational research published in journals. Report of the Committee on Evaluation of Research”, de Wandt et al., 1967, *American Educational Research Association* (Unpublished), p. 3. Recuperado de <http://www.indiana.edu/~educy520/readings/wandt65.pdf>

Da amostra de 125 artigos, publicados no ano de 1962, em revistas da área de Educação, os especialistas encontraram 19% aceitáveis (aceitável para publicação sem revisão), 41% com revisão (aceitável para publicação depois de revisão) e 40% rejeitado (não aceitável para publicação). Das iniciais das palavras “Acceptable”, “Revise” e “Reject” surgiu a sigla ARR, que deu nome à escala. Foram identificadas 25 características nos 125 artigos publicados. Os procedimentos organizados e detalhados se transformaram em um relevante guia para avaliação do relato de pesquisas na área.

Em 1971, Ward, Hall e Schramm replicaram o estudo conduzido em 1962 pelo *Committee of the American Educational Research Association* (AERA) e avaliaram a qualidade das pesquisas de Educação por meio de artigos publicados em revistas científicas específicas da área. Foram avaliados 121 artigos que cumpriam os requisitos definidos para a seleção da amostra. Considerando novas demandas da comunidade científica, a Escala ARR foi aplicada com uma ampliação no número de itens a serem avaliados, passando de 25 na primeira escala para 33 no estudo de 1971. Com os oito itens incluídos pelos pesquisadores, a escala avaliou: título, problema da pesquisa, revisão da literatura, procedimentos, análise de dados, conclusão, forma e estilo. A nova escala gerou um item a mais também na classificação dos artigos: 1) aceito para publicação sem revisão (A); 2) aceito para publicação com pequena revisão (R); 3) aceito para publicação com bastante

revisão (R); e 4) rejeitado (R), e passou a ser denominada Escala ARRR (Ward et al., 1975).

Hall, Ward e Comer (1988) retomaram a questão da avaliação da qualidade do relato das pesquisas na área educacional e explicaram que numerosos guias haviam sido desenvolvidos nos últimos 20 anos, enfatizando o trabalho do *Committee on Evaluation of Research, American Educational Research Association* (AERA). Os autores compararam os resultados encontrados aos de 1971, e publicados em 1975, por Ward et al. A amostra foi composta por 128 artigos submetidos à Escala ARRR e analisados segundo os mesmo critérios, sem nenhuma inovação. Os autores concluíram que houve melhoria na qualidade das pesquisas publicadas entre 1971 e 1983, data de coleta dos dados da última análise, e comentaram que 75% dos juízes, consultados no ano de 1983, eram revisores ou editores de revistas, o que pode ter favorecido os resultados da avaliação.

Objetivando avaliar o “estado da arte da área de conhecimento de Sistemas de Informação, no Brasil”, Hoppen (1998, p. 151) analisou 163 artigos publicados em revistas científicas de Administração, nos anos de 1990. Foram estudados os temas abordados, as estratégias de busca e as metodologias utilizadas nas pesquisas empíricas que resultaram em artigos. Dentre diversas variáveis, o autor constatou que “a qualidade científica dos artigos ainda é baixa, em razão da não explicitação de como as teorias de base são operacionalizadas, da descrição incompleta dos procedimentos metodológicos e da não validação dos instrumentos de pesquisa” (p. 151).

Na metade dos anos de 1990, foi constituído um grupo internacional de pesquisadores de ensaios clínicos, estatísticos, epidemiologistas e editores da área de biomédicas que, em 1996, apresentou as normas conhecidas como *Consolidated Standards of Reporting Trials* (CONSORT) ou Padronização Consolidada dos Ensaios Clínicos. O objetivo era o de criar normas para melhorias da comunicação e análises de ensaios clínicos. Sua última atualização foi no ano de 2011. O CONSORT é apoiado por diversas revistas da área médica, que exigem que os ensaios clínicos sejam registrados em algum site ou órgão regulamentador para serem publicados, e tem a adesão de grupos editoriais, incluindo o Comitê Internacional de Editores de Revistas Médicas, o (ICMJE, O Grupo de Vancouver), o Conselho de Editores Científicos (CSE), e a Associação Mundial de Editores Médicos (WAME) (Candeiro et al., 2011). A utilidade de CONSORT é reforçada por

um acompanhamento contínuo da literatura biomédica, que lhe permite ser modificada dependendo dos méritos para manutenção dos indicadores ou agregando novos itens, quando necessário. A abordagem, baseada em evidências, utilizada para desenvolver o CONSORT também tem sido utilizada para desenvolver padrões para a comunicação de metanálises de ensaios clínicos aleatorizados e de estudos observacionais (Altman et al., 2001). As normas do CONSORT foram estendidas para cobrir, também, os desenhos de estudos não farmacológicos. Intervenções como cirurgias, procedimentos técnicos, dispositivos de reabilitação, psicoterapia comportamental, terapia à base de plantas e outros procedimentos receberam diretrizes objetivando a melhoria dos seus relatos de pesquisas (Boutron, Moher, Douglas, Schult, & Ravaud, 2008).

A qualidade científica de artigos que utilizaram a pesquisa empírica tipo *survey* foi avaliada por Froemming et al. (2000) por meio da análise de 124 trabalhos, publicados em duas revistas e anais de um congresso, considerados os “veículos brasileiros mais representativos daquilo que se produz e publica em termos científicos dentro da área de marketing” (p. 202). A conclusão da análise, “em termos gerais, sinaliza pouca preocupação com aspectos relevantes da qualidade metodológica da investigação” (p. 216).

A qualidade metodológica dos ensaios clínicos controlados randomizados, publicados em revistas de ortopedia e de traumatologia do esporte indexadas no ISI (Institute for Scientific Information), foi comparada com a das revistas indexadas apenas no MEDLINE. Foram analisados 266 ensaios clínicos controlados randomizados, sendo 225 publicados em revistas indexadas no ISI e 41 em revistas MEDLINE. Os ensaios clínicos publicados em revistas ISI apresentaram maiores pontuações metodológicas, segundo os critérios definidos no estudo, em relação aos estudos publicados nas revistas MEDLINE (Peccin, 2005).

A avaliação da qualidade de relatos de pesquisas publicados em artigos científicos vem ampliando seu alcance como área de estudo. Na área de otorrinolaringologia, uma dissertação de mestrado classificou os objetivos e os desenhos de estudo dos artigos publicados em quatro revistas da área de fonoaudiologia, com o objetivo de analisar a qualidade metodológica das publicações sobre voz e laringe. As revistas selecionadas para a avaliação apresentaram “adequação metodológica insuficiente para responder às perguntas

clínicas, ou seja, os desenhos de estudo não estão adequados às respectivas perguntas” (Vieira, 2006, p. 45).

Cursos de redação científica têm sido oferecidos nas principais universidades do Brasil e livros que orientam sobre a escrita científica são publicados em grande escala. Gilson Volpato é um dos chamados “gurus” dessa modalidade de capacitação para autores e seus livros são consumidos pela comunidade científica em busca de aprovação para publicação em revistas internacionais e nacionais com fator de impacto alto. Interessa aos autores publicarem em revistas com FI alto, pois são revistas reconhecidas e valorizadas, além de muito consultadas. Os livros: *Bases Teóricas para Redação Científica* (2007), *Administração da Vida Científica* (2009), *Dicas para Redação Científica* (2010a), *Pérolas da Redação Científica* (2010b), *Método Lógico para Redação Científica* (2011) e, o mais atual, *Ciência: da Filosofia à Publicação* (2013) são leituras obrigatórias para iniciantes e, também, para experientes na matéria da redação científica com qualidade. Especificamente, na área de Psicologia, outro texto que vem tendo uma excelente aceitação é o *Publicar em Psicologia*, publicado pela Associação Brasileira de Editores Científicos de Psicologia (ABECiP), em parceria com a Instituto de Psicologia da USP. O livro pode ser baixado gratuitamente da internet, a partir do link <http://www.ip.usp.br/portal/images/stories/biblioteca/Publicarempsicologiversao2012.pdf>.

Um maior rigor no processo de avaliação antes da publicação poderia, também, evitar a situação que vem sendo apontada por especialistas como um grave problema na ciência: o elevado número de retratações em artigos científicos. Retratação é quando uma inconsistência grave é apontada em um artigo e o autor envia sua defesa para publicação. A frequência da retratação é um importante indicador da saúde do empreendimento científico, pois revela os erros e os equívocos publicados nos trabalhos. Artigos retratados representam falhas, evidências equivocadas e fracasso do projeto, independente da causa (Foo, 2011).

O número de retratações vem despertando o interesse de estudiosos e levando à análise desse tipo de ocorrência nas publicações científicas (Fang, Steen, & Casadevall, 2012; Foo, 2011). Os autores analisaram os artigos indicados como retratados no PubMed e classificaram as retratações de acordo com a fraude identificada: falsificação ou fabricação de dados, plágio, publicação duplicada, ou por outras razões (por exemplo, erro da revista ou disputa de autoria). Em maio de 2012,

o PubMed reunia 25 milhões de artigos, indexados desde o ano de 1940, e foram localizados 2.074 artigos retratados, sendo o mais antigo publicado em 1973 e retratado em 1977. A maioria dos artigos foi retratada por má conduta dos autores e os casos mais notórios são problemas de ética na pesquisa. Os autores concluem que falta maior vigilância de revisores, editores e leitores na avaliação dos artigos fraudulentos. As instituições governamentais, que patrocinam grande parte das pesquisas, também precisam estar mais vigilantes, assim como agências de fomento e os jornalistas científicos. Maiores investimentos em capacitações de pesquisadores e autores também podem resultar na diminuição do número de retratações de artigos científicos, uma vez que a má conduta científica tende a diminuir, concluem Fang et al (2012).

Visando reunir o maior número possível de trabalhos que discutem a questão Manes e Flach (2012) realizaram uma revisão sistemática “da produção científica sobre o tema fraude, a partir de artigos publicados em periódicos internacionais de contabilidade indexados às bases de dados ISI e Scopus, no período entre 2000 e 2011” (p. 166). A partir da revisão de 61 artigos os autores concluíram que o tema “fraudes” vem ocupando vários espaços em publicações, tanto nacionais, quanto internacionais, e que os autores prefeririam discutir sobre a responsabilidade dos pareceristas e as punições que lhes dizem respeito, caso seja necessário. A punição se dá, principalmente, quando existem escândalos financeiros que envolvem as publicações.

- Dentre as muitas formas de má conduta de pesquisadores, a publicação de dados fraudulentos é uma das mais graves, pois a confiança e a validade da pesquisa é seriamente prejudicada.
- Ter um artigo retratado pode prejudicar injustamente a reputação de um autor que cometeu um erro sem intenção e passível de explicação.
- Além disso, um artigo pode conter informação valiosa, porém descartá-lo, porque não há clareza em sua redação, não o invalida.
- A máxima “publicar ou morrer” é um fator que influi demasiadamente na qualidade, pois o FI e o índice h estão aí como uma espada na cabeça de pesquisadores e professores em todos os países.

O índice h ou *h-index*, em inglês, é uma ferramenta para combinar quantidade e qualidade de produção científica, com base nos trabalhos publicados (Marques, 2013).

2.2 Avaliação de artigos na área de Psicologia

A Psicologia é uma das áreas que, há muitos anos, preocupa-se com a qualidade dos relatos de pesquisas. No ano de 1929 foi publicado o primeiro *Manual de Publicação*, um documento de sete páginas com um conjunto de procedimentos para aumentar a compreensão da leitura dos textos científicos. Criado sob a responsabilidade do *United States National Research Council*, seus elaboradores incluíam psicólogos, antropólogos e profissionais de publicação (American Psychological Association [APA], 2010, p. xiii).

Em 1952, o manual foi ampliado e publicado como um suplemento de 55 páginas no *Psychological Bulletin*. Revisões foram publicadas em 1957 e 1967 (APA, 1952, 1957, 1967). A primeira edição apresentava normas para a escolha de palavras adequadas, gramática, pontuação e formatação, políticas de publicação de periódicos, além de orientações para “acondicionamento e transporte” das revistas (APA, Council of Editors, 1952, p. 442).

Em resposta às crescentes complexidades da comunicação científica, as edições subsequentes foram lançadas em 1974, 1983, 1994 e 2001. Conhecido principalmente pela simplicidade e consistência de seu estilo de citação e normalização das referências, o *Manual de Publicação da American Psychological Association* (*Publication Manual of the American Psychological Association*) também estabeleceu normas para o uso da linguagem que tiveram efeitos de longo alcance. Particularmente influentes foram as *Diretrizes para a língua não-sexista em Revistas da APA*, publicado pela primeira vez como uma modificação para a edição de 1974, o que proporcionou alternativas práticas para a linguagem sexista então de uso comum (APA Task Force on Sexual Bias in Graduate Education, 1975). As diretrizes para reduzir o viés na linguagem foram atualizadas ao longo dos anos e, atualmente, fornecem orientações práticas para escrever sobre raça, etnia, sexo, idade, orientação sexual, deficiência e *status* (APA, 2010, pp. 70-77).

Antes mesmo de lançar a sexta edição, no ano de 2009, antecipando a revisão do Manual, a APA formou o *Working Group on Journal Article Reporting Standards* (JARS) para preparar recomendações para a submissão de manuscritos às revistas publicadas pela Associação. As recomendações tratam de a) normas para apresentação de todos os tipos de artigos científicos; b) normas específicas

para relatos de pesquisas com manipulação experimental ou avaliação de intervenções, usando desenhos de pesquisa que envolvam aleatorização e não aleatorização; e c) normas para apresentação de estudos de metanálise. Os informes também encorajam os autores a providenciarem mais detalhes nos seus relatos de pesquisas e examinam os obstáculos e os benefícios do desenvolvimento de relatórios de normas editoriais (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008).

O *Manual de Publicação da American Psychological Association* encontra-se na sexta edição e orienta os autores em relação ao estilo que deve ser seguido para a redação de artigos científicos a serem submetidos às revistas da Associação (APA, 2010). A última versão incluiu procedimentos para elaboração de revisões sistemáticas, com ou sem metanálise, além de inovações em relação ao ambiente digital, com ênfase para o papel do *Digital Object Identifier* (DOI), como uma maneira confiável de localizar informações. Foram introduzidas, também, novas orientações sobre a questão ética da determinação da autoria principal e colaboração; duplicações de uma mesma publicação; plágio e autoplágio, mascaramento dos participantes; validade dos instrumentos de pesquisas; disponibilização de dados para verificação por outros; procedimentos para verificação da clareza e da precisão de pesquisas empíricas; simplificação do título para torná-lo mais propício em publicações eletrônicas; novas diretrizes para redução do viés de linguagem e adequação aos padrões atuais da língua; novas diretrizes para apresentação das estatísticas e abreviaturas. Vale destacar a discussão ampliada do processo de publicação, incluindo a função e o processo de revisão por pares, ética, requisitos legais e políticas em publicação, e orientações sobre como interagir com o editor enquanto o artigo está no prelo. Sem dúvidas, é o mais importante manual de estilo para publicação da área de Psicologia. As orientações do manual regem a Psicologia em diversos países do mundo. Mesmo no Brasil, que conta com a Associação Brasileira de Normas Técnicas (ABNT) para padronizar a normalização de publicações, o *Manual de Publicação* da APA é consenso entre os autores da área de Psicologia e disciplinas correlatas.

Sabadini, Nascimento e Cadidé (2007) analisaram as normas editoriais de 69 títulos de revistas brasileiras de Psicologia com o objetivo de observar as que adotam o estilo publicado pela APA na normalização das citações e das referências. Concluíram que 35 dos 69 títulos analisados (51%) utilizavam as normas

internacionais, o que demonstrava uma tendência de uso do estilo APA de normalizar. Ao acompanhar as revistas, por ser atributo das nossas atividades, podemos afirmar que esse panorama sofreu bastante alteração e muitas outras revistas de Psicologia migraram para o padrão APA.

A preocupação com a qualidade dos artigos científicos em Psicologia levou Witter (2006) a buscar instrumentos para a avaliação da produção científica brasileira. A autora pesquisou e analisou diversas publicações que relatavam a preocupação com a sistematização e criação de instrumentos para análise da produção científica. Dentre as escalas mais difundidas e pesquisadas, ela encontrou a Escala ARRR, desenvolvida por Ward et al. (1975), apresentada anteriormente. Witter (2006) informa que utilizou amplamente a Escala nos cursos de pós-graduação das áreas de Psicologia, Educação e Linguística, para que os alunos avaliassem artigos de pesquisa. Segundo a autora, o instrumento se mostrou muito útil, não só como ferramenta de avaliação, mas também como meio de ensino de redação científica. A autora passou a utilizar a Escala também para que os alunos avaliassem seus próprios relatórios de pesquisa, pois notou que isso ajudava a melhorar a criticidade na avaliação de pesquisas por parte dos estudantes.

A Escala ARRR avalia os artigos de acordo com o número de pontos que o trabalho alcançou na avaliação. Não deve ser utilizada para avaliar resenhas, editoriais, informativos e outros tipos de contribuição que não sejam textos teóricos e relatos de pesquisa. Quando o trabalho é teórico são computados os pontos de acordo com uma tabela específica (Witter, 2006). Ao adaptar a Escala ARRR, a autora inseriu itens para avaliação das referências dos artigos, em relação a seguir a norma adotada pela revista e a temporalidade das referências e tipologia do suporte documental. Graças ao trabalho de Witter (2006), a área da Psicologia tem avaliado seus artigos com um instrumento validado, internacional e nacionalmente, e pode se dedicar à avaliação de sua produção com o rigor necessário à ciência.

No final da década de 1990, a preocupação com a qualificação não foi preocupação com a qualidade, mas com a qualificação, isto é, *onde* estava sendo publicado, e não *o quê* estava sendo publicado nos programas de pós-graduação que, consequentemente, levou à avaliação das revistas científicas. Cada área do conhecimento se responsabilizou por desenvolver os critérios para análise da sua produção, de acordo com eventos da Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), denominado QUALIS.

As demandas para a realização deste empreendimento, que teve a sua versão inicial realizada em 1999, são as seguintes: (a) qualificar os veículos científicos nos quais publicam pesquisadores brasileiros vinculados aos programas de pós-graduação em psicologia; (b) criar mecanismos para viabilizar a manutenção dos periódicos com melhor avaliação; (c) estabelecer parâmetros para o incremento da qualidade dos periódicos da área (Yamamoto et al., 2002, p. 163).

A necessidade de distinguir entre a informação de qualidade e o conhecimento reproduzido em larga escala desqualificou uma parte considerável do material apresentado aos programas de pós-graduação e centros de pesquisa. “Entre os artigos publicados, produções apresentadas em eventos, orientações de trabalhos científicos, livros e outras publicações, é imprescindível sinalizar para a comunidade acadêmica o mérito científico de tais produtos” (Costa & Yamamoto, 2008, p. 22). O Sistema de Avaliação da Pós-graduação foi implantado no ano de 1998, com o objetivo de avaliar os programas de pós-graduação no Brasil.

Costa e Yamamoto (2008) analisaram o sistema de avaliação de periódicos Qualis da área de Psicologia com o objetivo de fazer um estudo do processo de avaliação, que chamaram de meta-avaliação (avaliação da avaliação). A análise demonstrou que há divergência entre os principais atores envolvidos no processo editorial das revistas em relação aos critérios adotados pelas comissões avaliadoras da CAPES. No entanto, há o reconhecimento das melhorias das publicações após a implantação do processo. A recomendação dos autores é que, “uma vez considerado esse processo de avaliação como contribuição para o progresso científico, há que construir alternativas para um modelo capaz de fornecer respostas adequadas às demandas da comunidade científica” (p. 13). Concluíram ainda que

...a Comissão de Avaliação CAPES/ANPEPP atendeu aos objetivos estipulados: os periódicos científicos estão sendo classificados continuamente, desde a primeira edição; os editores, agora, são capazes de visualizar os parâmetros exigidos para os títulos da área, e, apesar de não haver um mecanismo formal voltado exclusivamente para a manutenção dos periódicos melhor avaliados, este aspecto pode ser considerado operacionalizado pelo próprio consumo de tais periódicos por parte da comunidade científica (Costa & Yamamoto, 2008, p. 23).

O processo de avaliação de periódicos Qualis analisa as revistas e não os artigos. Claro, seria humanamente impossível avaliar os artigos em decorrência do grande número de revistas nas quais a área de Psicologia, aqui no Brasil, publica.

No ano de 2012, foram avaliados 2058 títulos de periódicos que receberam a contribuição de autores brasileiros dos diversos programas de pós-graduação. A alternativa encontrada para ampliar o processo foi considerar a indexação em bases de dados, que desde sua implantação provocou a reação dos editores.

Esse novo modelo de avaliação passou, desde logo, a receber críticas de diferentes setores da comunidade científica brasileira. Marques (2009²), no texto “A escala da discórdia”, informa que pesquisadores de áreas diferentes fazem críticas que vão desde apontar como perigoso o peso exagerado dado, nessa escala de avaliação, ao fator de impacto (FI) até a falta de critério na política de avaliação da Capes. (Barone, 2011, p. 88)

As críticas ao novo modelo de avaliação, focado no FI, foram feitas porque a indexação em bases de dados passa a ter uma valorização excessiva da visibilidade e da inserção internacional, já que as bases de dados mais valorizadas, ou seja, aquelas que classificam a revista no melhor extrato da avaliação, A1, são duas norte-americanas: WOK e PsycInfo. A importância da avaliação é reconhecida como elemento essencial para a elevação da qualidade do ensino, mas declara a editora do *Jornal de Psicanálise* (Barone, 2011, p. 89) “a Capes deveria desenvolver critérios de avaliação mais sofisticados e complexos, já que fator de impacto não constitui, efetivamente, uma medida absoluta de qualidade”.

Com o objetivo de oferecer um instrumento de apoio aos processos de avaliação utilizados pelas agências de ciência e de tecnologia e pelas próprias bases de dados, Castro (2011) desenvolveu um protocolo para avaliar as bases de dados em relação ao âmbito, cobertura e público-alvo, mérito científico, qualidade editorial, qualidade gráfica, tipo de conteúdo, idioma e cobertura geográfica. Com a aplicação dos critérios de avaliação definidos no protocolo é possível conhecer os graus de complexidade e as restrições das bases de dados e categorizá-las, de acordo com esses quesitos, considerando a qualidade do conteúdo indexado (Castro, 2011). Quanto mais exigentes forem os critérios de seleção das revistas indexadas em uma determinada base de dados, maior será a qualidade dos artigos publicados. Avaliar as bases de dados, com critérios igualitários e validados por especialistas, é um excelente caminho para enquadrá-las no âmbito temático, cobertura do tipo de documentos e público-alvo ao qual a base é dirigida.

² Marques, F. (2009). A escala da discórdia. *Pesquisa Fapesp Online*, (160). Recuperado de <http://revistapesquisa.fapesp.br/2009/06/01/a-escala-da-discordia/>

A produção científica ganhou uma condição de suma importância na avaliação de docentes, alunos, profissionais de universidade e centros de pesquisa. Dessa maneira, é fundamental a criação de estratégias para avaliação da qualidade dos artigos publicados por essa comunidade. Incentivar autores a revisitarem os artigos científicos em busca de evidências e, neste processo, incluir a avaliação pode ser uma alternativa válida, ainda que o caminho seja árduo e longo. Afinal, estamos falando de um volume imenso de artigos publicados em diversas revistas científicas que necessitam ser sistematizados por alguns requisitos para, então, serem oferecidos como provas para a comunidade.

Uma alternativa para avaliar a qualidade dos artigos e das revistas, denominada *Índice de Qualidade*, foi apresentada por Buela-Casal (2003). Esta justifica a criação dos itens de avaliação por achar lógico que a qualidade dos artigos é muito mais que o número de citações que ele recebe. Em colaboração com Robert Sternberg, presidente da APA em 2003, o autor aponta quinze motivos para dar mais importância ao artigo, propriamente dito, que à revista que o publica; ele faz duras críticas ao processo de revisão por pares e ao conservadorismo de algumas revistas prestigiosas, ou seja, com fator de impacto alto. Seu *Índice de Qualidade* relaciona uma série de critérios para avaliar o conteúdo de artigos em Psicologia e em outras áreas do conhecimento, da qual traduzimos e destacamos uma síntese em forma de perguntas.

- 1) Os resultados são novos, inéditos, relevantes e fazem sentido em um contexto teórico e prático?
- 2) As ideias interessantes e trazem um novo enfoque para tratar velhos problemas?
- 3) Os dados são interpretados sem equívocos?
- 4) É apresentado um marco novo e mais simples para casos mais confusos e complexos?
- 5) Ideias que pareciam irrefutáveis são questionadas?
- 6) A pesquisa tem validade interna e utiliza um desenho metodológico adequado?
- 7) A pesquisa tem validade externa suficiente para que os resultados sejam generalizados?
- 8) O método e os procedimentos são bem descritos e passíveis de replicação?
- 9) Os resultados teóricos e práticos têm um alto grau de aplicação?
- 10) Os resultados teóricos e práticos são úteis à sociedade?

- 11) O tipo de estudo é nomeado com clareza?
- 12) As hipóteses possuem formulações e interpretações adequadas?
- 13) Os instrumentos de avaliação utilizados têm qualidade e garantia científica?
- 14) Os estudos seguem normas adequadas para a apresentação de casos?
- 15) Os estudos para construções e validação de instrumentos seguem a metodologia adequada?

O *Índice de Qualidade* deve ser utilizado por revisores profissionais devidamente formados, que precisam seguir o manual estabelecimento para aplicar os critérios de forma adequada (Buela-Casal, 2003).

Sendo a “Qualidade de Conteúdo o Grande Desafio para os Editores Científicos”, como Trzesniak, Plata-Caviedes e Córdoba-Salgado (2012) apresentam no título de um artigo publicado na *Revista Colombiana de Psicología*, o que fazer para colaborar na mudança do panorama atual? Respondendo a tal pergunta, os autores apresentam uma taxonomia, cujo objetivo é avaliar a qualidade dos artigos com base em três grandes desafios.

1) Desafio para autoras e autores a não ter como objetivo publicar apenas mais um artigo, mas a colaborarem efetivamente com o avanço do conhecimento. Para tanto, devem estar certos de que o conteúdo que estão submetendo para publicação é construtivo e consequente. Trabalhos chamados de aventureiros (incompletos e mal-acabados) são fortes candidatos à recusa por parte dos editores. Na verdade, autores consequentes deveriam se preocupar com o precioso tempo do editor e dos pareceristas antes de tentar passar qualquer trabalho, o que é uma grave infração ética. Manuscritos que reaproveitam partes de outros, já publicados, são vistos por Trzesniak et al. (2012) como muito mais que uma questão de autoplágio, “mas integridade pessoal” (p. 74).

2) Desafios para os pareceristas: antes de iniciar o esforço de analisar o trabalho criteriosamente, verificar se não se trata de um artigo “aventureiro”, ou “mosaico” (feito de partes de outros já publicados); hoje, os portais de revistas facilitam muito essa tarefa. Verificada a “pertinência e validade científica do trabalho” emitir pareceres de modo consciente, objetivo e úteis, que auxiliem o editor e orientem os autores (p. 75).

3) Desafios aos órgãos de fomento: ao avaliar as revistas considerarem não apenas o número de artigos publicados, mas o total, bruto e líquido, de submissões. A

supervalorização da quantidade leva à aceitação de trabalhos de qualidade duvidosa.

Os editores devem desencorajar com firmeza e severidade (porém, sempre de modo cortês) a “submissão aventureira”, aquela que deixa evidente que o autor não se empenhou suficientemente na elaboração do manuscrito, transferindo sua responsabilidade para revisores e editores. Tais submissões devem ser retornadas na revisão de conformidade. Se apresentá-los é falta de ética do autor, tramitá-los, onerando os revisores, é falta de ética do editor. (Trzesniak et al., 2012, p. 76)

Trabalhar com a avaliação da qualidade intrínseca dos artigos não é tarefa fácil, entretanto, já é chegada a hora de autores e editores buscarem caminhos para enfrentar esse desafio. Há que se buscar um “padrão-ouro” nos artigos. Para tanto, forma e conteúdo devem ter apresentação esmerada e rigorosa.

2.3 Artigo científico: forma e conteúdo

A questão da forma, ou seja, a maneira como a informação deve estar organizada nas respectivas partes do artigo é de fundamental relevância. Entretanto, as etapas anteriores à redação do artigo, ou seja, a definição da pergunta de pesquisa, a escolha da metodologia adequada, a aplicação de métodos estatísticos validados e a análise cuidadosa dos dados, são determinantes da qualidade do artigo. Na verdade, são essas etapas que determinam a qualidade do conteúdo do artigo. Assim, vamos iniciar discorrendo sobre o primeiro passo que rege o caminho das pesquisas científicas de modo geral, a pergunta de pesquisa.

2.3 Pergunta de pesquisa

Em toda pesquisa científica, a elaboração da pergunta de pesquisa é o momento decisivo, que determinará o andamento do trabalho e orientará os resultados. O sucesso da pesquisa científica e a posterior publicação dos seus resultados em revistas especializadas começam com uma ideia, que levaremos adiante após a identificação da pergunta de pesquisa que queremos responder.

Para a definição desse momento tão especial da pesquisa podemos recorrer à técnica do PICO – Participantes, Intervenção, Controle e Resultados (em inglês

Outcomes) - (Castro, Saconato, Guidugli, & Clark, 2002; Greenhalgh, 2008; Higgins & Green, Eds., 2006). Os quatro elementos, que devem ser definidos corretamente, uma vez respondidos de forma adequada, culminarão com a pergunta de pesquisa. É neste momento que o conhecimento de técnicas na elaboração de estratégias de buscas em bases de dados pode auxiliar. O pesquisador que não se sentir seguro para a exploração desses elementos pode e deve consultar um profissional de informação. Nessa etapa do trabalho o pesquisador reflete sobre cada um dos seus elementos de pesquisa. Agora falaremos sobre o que cada letra desta técnica indica:

P deve fornecer informações relevantes sobre os participantes da pesquisa. Quem são os participantes da pesquisa? Sobre quem é a questão? Qual é o contexto? São pessoas enfermas? Qual é a enfermidade? Enfim, qual é a situação problema (Greenhalgh, 2008). Exemplificando o **P** temos: *mulheres residentes no Estado de São Paulo com depressão pós-parto (DPP)*.

I nos informa sobre o tipo de tratamento/ação/intervenção que se está estudando, analisando. Por exemplo: *psicoterapia*.

C define a intervenção de comparação, que pode ser um tratamento medicamentoso, exercício físico ou mesmo nenhum tratamento. Por exemplo, no caso das *mulheres residentes no Estado de São Paulo com depressão pós-parto (DPP)* podemos ter como intervenção a psicoterapia. O grupo controle de comparação (C) poderia ser composto por mulheres que estão na fila de espera e que ainda não estão recebendo nenhum tipo de tratamento.

O nos remete aos resultados que esperamos encontrar com o nosso estudo. Podemos dizer que o **O** é o indicador clínico da modificação do estado do paciente após a intervenção. Nesse caso, os resultados esperados podem ser formulados da seguinte maneira: Verificar se a Psicoterapia é eficiente no tratamento de mulheres que apresentam DPP e que residem no Estado de São Paulo.

Organizando os itens de cada um dos elementos do PICO, explicados anteriormente, em uma pergunta de pesquisa, temos: a psicoterapia é eficiente no

tratamento de mulheres, residentes no Estado de São Paulo, que apresentam depressão pós-parto?

O **O**, bem compreendido e detalhado, fornece-nos o objetivo principal do estudo. No caso do exemplo anterior, o objetivo da pesquisa poderia ser: verificar se a psicoterapia é eficiente no tratamento da DPP em mulheres residentes no Estado de São Paulo.

O modelo PICO não pode ser aplicado a todas as perguntas de pesquisa, pois existem estudos que não possuem intervenção, outros não indicam o grupo controle. Por exemplo, um estudo buscou responder à seguinte pergunta: qual é a atuação do psicólogo em relação ao cuidado do paciente na assistência em saúde mental, atendido nos serviços públicos de saúde? Esta questão de pesquisa trata do papel do psicólogo e não aplica nenhuma intervenção. Para este caso, Biruel e Pinto (2011) adaptaram o modelo PICO para o PVO: **P**: Situação problema, participantes, contexto. O **V** se refere às variáveis do estudo e o **O** é o próprio desfecho ou os resultados. Ao organizar a pergunta de pesquisa anterior no PVO, temos:

P - assistência em saúde mental/ serviços públicos em saúde.

V - cuidado do paciente.

O - identificar a atuação do psicólogo no cuidado do paciente atendido nos serviços públicos de saúde.

O **O** continua representando o objetivo do estudo, aquilo que se espera responder com a pesquisa.

Biruel e Pinto (2011) esclarecem que “Perguntas de pesquisa do tipo exploratória que tem por objetivo gerar hipóteses, não necessariamente, propõem a comprovação de intervenções específicas”. Nesse caso, torna-se possível adaptar a técnica sem considerar o **I** (intervenção) e/ou o **C** (controle/comparação).

Além de organizar os elementos da pesquisa, o objetivo de estruturar a pergunta em PICO ou PVO é o de preparar a estratégia de busca a ser submetida às bases de dados. Uma estratégia de pesquisa bem elaborada, utilizando, quando possível, termos hierarquizados em um vocabulário controlado, conectados corretamente pelos operadores booleanos, de truncamento e de proximidade,

garante a recuperação eficiente da informação e o acesso aos documentos relevantes do tema pesquisado.

Construída a estratégia de busca, com base na pergunta de pesquisa bem delineada, é necessário consultar as bases de dados em busca dos artigos científicos de qualidade. As buscas em bases de dados podem resultar em uma intensa variedade de material, desde ensaios controlados aleatorizados multicêntricos e revisões sistemáticas, elegantemente delineadas, até estudos de caso único, revisões analíticas e opiniões pessoais. Todos esses tipos de trabalhos podem ser evidências em algum sentido. O grau de credibilidade em um ou outro trabalho é que vai variar (Greenhalgh, 2008). Por isso, conhecer as fontes de informação qualificadas na área e saber usá-las, tirando vantagem de seus recursos e filtrando suas melhores contribuições, são habilidades que pesquisadores, profissionais e estudiosos precisam desenvolver e aperfeiçoar. A outra vertente da questão é que os artigos devem estar escritos corretamente, com todas as etapas da pesquisa apresentadas e formalizadas no texto.

Sabadini et al. (2009a) discutiram exaustivamente os requisitos para construção de um artigo científico, por isso não vamos tratar desses e nos preocuparemos mais com a questão da metodologia de pesquisa. Essa parte do artigo descreve, em sequência lógica, o método utilizado e as etapas do estudo. Explica o que foi feito para se atingir os objetivos do estudo. Aqui, deve ser declarado o caminho percorrido, os materiais (instrumentos e aparatos), o local de realização da pesquisa, as técnicas e os procedimentos aplicados. O método é fundamental e é a primeira informação que deve constar nesta parte do artigo. Vale destacar os diversos delineamentos de pesquisa utilizados pela ciência psicológica, pois é desejado que tais delineamentos sejam devidamente nomeados nesta parte do trabalho. As bases de dados, como PsycINFO, agrupam os artigos pela metodologia declarada nos artigos. Assim, não explicitá-la fará com que o estudo não seja classificado e perca uma janela de visibilidade.

2.4 Tipos de pesquisas utilizados pela ciência psicológica

Granja (1995) que, dentre outros indicadores, analisou os delineamentos de estudos nas dissertações e teses defendidas no Instituto de Psicologia da Universidade de São Paulo, no período de 1980 a 1989, esclarece que, de acordo

com a natureza, as pesquisas na área de Psicologia podem ser divididas em: Estudo Teórico, Pesquisa de Campo/Laboratório e Pesquisa Documental.

Estudo Teórico

Trabalhos teóricos ocorrem em nível de análise, síntese ou avaliação na taxonomia do Domínio Cognitivo de Bloom, que é estruturada em níveis de complexidade crescente – do mais simples ao mais complexo, exigindo vivência de pesquisa e conhecimento que permitam neles trabalhar. Os trabalhos teóricos elaboram a estrutura da área, constroem conceitos, estabelecem a arquitetura e as interrelações entre áreas e subáreas. São trabalhos que dispensam o uso de questionário/estatística/correlação para conclusão e apresentação de seus resultados, que são fundamentados em estudos profundos de teorias consolidadas (contribuição de Piotr Trzesniak).

Pesquisa Empírica

Também chamada de pesquisa de campo, são pesquisas baseadas na busca pela resposta a uma questão específica, ou a uma hipótese a ser comprovada ou negada, com base em dados relevantes e convenientemente obtidos da experiência e vivência do pesquisador. Identificam novos fenômenos ou buscam descobrir a relação existente entre eles. Podem ser de natureza descritiva, estudam a relação entre duas ou mais grandezas de um fenômeno, ou de natureza experimental, quando o fenômeno sofre manipulação das variáveis, resultando em dados e fatos mensuráveis (Granja, 1995).

Pesquisa Documental

É aquela cujos objetivos ou hipótese foram verificadas por meio da análise de documentos. Neste tipo de pesquisa, os objetivos ou hipóteses são verificados pela análise de documentos, primários ou secundários. Documento é toda a forma de ideias e fatos registrados, independente do suporte material. É suscetível de consultas, pesquisas e estudos e pode ser utilizado como evidência, ou prova.³ De

³ Fonte: <http://pt.wikipedia.org/wiki/Arquiv%C3%ADstica>

acordo com o suporte, pode ser: textual (ex. artigo científico), cartográfico (ex. mapa), iconográfico (ex. fotografia), filmográfico (ex. filmes), sonoro (ex. discos) e informático (ex. arquivos de disco rígido). Documento primário é o que se apresenta tal como produziu seu autor, é a fonte original da informação, como um artigo científico que apresenta um relato de pesquisa (Cunha & Cavalcanti, 2008). Secundário é o documento gerado a partir da análise do documento primário, exemplo: bibliografias, bases de dados, revisões narrativas ou sistemáticas. Em virtude destas nos interessar diretamente, à frente, deteremo-nos um pouco mais nas revisões sistemáticas.

As ciências são caracterizadas pela utilização de métodos científicos, “em contrapartida, nem todos os ramos de estudo que empregam esses métodos são ciências” (Marconi & Lakatos, 2010, p. 65). Ainda segundo os autores, “o método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo”, traçando o caminho a ser seguido para testar as hipóteses e chegar aos resultados. Conforme Granja (1995), e outros autores que serão identificados de acordo com a contribuição, a Psicologia traça os caminhos de suas pesquisas com base em diferentes métodos.

1) *Estudo de caso*: trabalhos que relatam casos de forma convencional ou utilizando rigorosas técnicas de controle na coleta de dados. A estratégia do estudo de caso é analisar e descrever afirmações de uma pessoa, ou grupo de pessoas, com base em um número correspondente de observações (Stiles, 2007). As categorias e análise devem ser definidas a priori.

2) *Levantamento*: estudos que descrevem uma situação ou realidade, valendo-se de estatísticas descritivas para a apresentação dos resultados. Nesse caso, as categorias de análise podem ser definidas posteriormente.

3) *Estudo correlacional*: pesquisas cujos delineamentos fazem uma análise de correlação entre as variáveis selecionadas como objeto de investigação. Exploram o grau em que dois, ou mais, fenômenos se relacionam ou variam, como, por exemplo, idade e peso. Nesse tipo de estudo é necessário que o pesquisador relacione as variáveis com base nos sujeitos da amostra. De posse de pelo menos

duas medidas, o pesquisador consegue analisar os dados e fazer o *coeficiente de correlação*.

4) *Estudo longitudinais de coorte*: estudos observacionais, nos quais os sujeitos são classificados ou selecionados de acordo com critérios pré-estabelecidos. Neste tipo de estudo, grupos de indivíduos são acompanhados ao longo de um determinado tempo e, periodicamente, são investigados por um pesquisador, que vai agrupando os indivíduos em categorias de sujeitos. É um recurso muito utilizado em pesquisas sobre doenças, acidentes, infecções e, mortes, de modo geral. Pode ser chamado também de estudos de incidência (*incidence*) ou de seguimento (*follow-up*).

5) *Estudo transversal*: são estudos que se assemelham aos estudos longitudinais de coorte. No entanto, nos estudos transversais todas as medições são feitas em um único momento e os sujeitos não são seguidos. Este tipo de estudo é utilizado para descrever características das populações em relação às variáveis definidas para a pesquisa.

6) *Estudo experimental*: nesta categoria de estudos, o pesquisador manipula as variáveis de pesquisa. A categoria estabelece subdivisões, como veremos a seguir.

6.1) *Comparação entre grupos*: neste tipo de pesquisa, um ou mais grupos são submetidos a uma só variável, comparando-se o desempenho entre si ou com o grupo controle. O grupo controle é composto por sujeitos que não sofreram nenhum tipo de intervenção durante a fase da pesquisa.

6.2) *Delineamento experimental de caso único*: neste modelo de delineamento, os sujeitos são expostos a uma série de condições, mensurando-se, repetidamente, o desempenho do organismo e verificando-se se há uma relação ordenada entre as condições manipuladas no experimento e as alterações nessas medidas (Sampaio et al., 2008).

6.3) *Estudo misto*: estudos desenvolvidos utilizando-se mais de um tipo de método.

6.4) *Ensaio clínico controlado aleatório*: são estudos considerados evidências com o maior grau de força. Também conhecido por ensaio clínico randomizado (ECR), consiste em um tipo de estudo experimental, desenvolvido em seres humanos, e que visa ao conhecimento do efeito de intervenções em saúde. A simplicidade do seu desenho, quando comparado a outros tipos de estudos, faz que seja fortemente utilizado na pesquisa clínica. Neste estudo, os sujeitos são recrutados de um grupo comum, a partir de uma população de interesse. Em seguida, são decididas quais intervenções os participantes receberão pela randomização. Tecnicamente, esse processo envolve determinar a alocação por meio de números obtidos por sorteio, em vez de características da amostra ou preferências do pesquisador e dos participantes. Após a aplicação das intervenções, realiza-se a leitura de uma ou mais variáveis de desfecho. É o tipo de estudo foco das revisões sistemáticas e é considerado o padrão-ouro para se avaliar intervenções.

6.5) *Ensaio clínico controlado não aleatório*: segue o modelo do ensaio clínico controlado, com a diferença que, neste tipo de estudo, os sujeitos não são selecionados ao acaso, sem a preocupação com métodos para aleatorização.

6.6) *Revisões sistemáticas, com e sem metanálise*: diante do interesse que o método de pesquisa vem despertando e por ter este trabalho um envolvimento direto com uma revisão sistemática, estenderemos a explanação sobre o tema logo adiante.

Conceitos básicos empregados em metodologias de pesquisas devem ser dominados, para facilitar o entendimento de estudos já realizados e preparar o pesquisador para o seu projeto. Reis, Ciconelli e Faloppa (2002) resumem informações claras e objetivas sobre metodologias de pesquisas e explicam o processo de seleção da amostra, a delimitação e demarcação do assunto ou material, a realização do planejamento da pesquisa, bem como da avaliação, interpretação e registro dos resultados, com o uso das estatísticas corretas na análise dos dados.

A pesquisa psicológica depende da capacidade do psicólogo em acomodar novos fenômenos ou problemas ao conhecimento já existente. Para estudos desses fenômenos, diversos métodos de pesquisa estão disponíveis. Chow (2002) divide esses métodos em três categorias: não experimental, experimental e quase

experimental, de acordo com o grau de controle exercido pelo investigador. O autor observa que tais controles não significam constranger os participantes da pesquisa. Em vez disso, um controle é um meio para excluir uma explicação alternativa da pesquisa, dando maior rigor aos dados. O ponto importante, é que a criação de controles no experimento não significa moldar o comportamento dos sujeitos. Se os psicólogos não aplicam nenhum controle no estudo empírico, estão conduzindo um método não experimental. Quando todos os mecanismos de controle, teoricamente baseados, estão presentes, os psicólogos estão conduzindo um experimento. Se por qualquer razão, ao menos um nível de controle não seja devidamente utilizado, os psicólogos estão conduzindo um meio experimento, ou seja, aplicando um método quase experimental (Chow, 2002).

A figura 1 reproduz o fluxograma de Chow (2002) para organização dos tipos de pesquisas.

Figura 1. Métodos de Pesquisa. Reproduzida e traduzida de “Methods in psychological research”, de S. L. Chow, 2002, *Encyclopedia of Life Support Systems (EOLSS)*, (p.10). Recuperada de <http://cogprints.org/2643/1/EOLSSrm.pdf>

O desenho de Chow está distante de esgotar os métodos de pesquisas na área da Psicologia. Os estudos de caso único, que vem crescendo em relevância nas pesquisas psicológicas, não foram apontados no esquema.

O problema de se classificar os delineamentos de pesquisa é que nem todos conhecem a nomenclatura e, atualmente, com a possibilidade de publicações científicas em acesso aberto os artigos podem ser lidos por não especialistas na

área. Volpato (2011, p. 196) recomenda “é desejável facilitar uma conversa mais universal”, encorajando os cientistas a expandirem o alcance de seus trabalhos para além dos muros das universidades e dos centros de pesquisas. O autor orienta que, para se evitar o uso de classificações que nem todos conhecem, é melhor adotar um critério mais universal, com os tipos lógicos de pesquisa que devem mostrar a lógica desta e a dinâmica da investigação. Explicar com detalhes cada uma das fases da pesquisa em linguagem simples, direta e acessível, é um dos caminhos para se democratizar o acesso ao conhecimento. O importante é entender que todas as formas de pesquisa são igualmente importantes, dependendo do contexto e do tipo de pesquisa. O essencial é que o pesquisador analise e avalie cada um dos métodos, decidindo pelo mais adequado à sua pesquisa.

Artigos de qualidade apresentam consistência em relação à descrição de cada uma de suas partes, mas a metodologia e os resultados da pesquisa vem ganhando maior destaque nos relatos. Observamos, atualmente, artigos com introduções curtas e sem sessões dedicadas às conclusões, mas com as partes do método e resultados extensamente descritas. A ciência vem se tornando cada vez mais exigente com a forma de apresentação dos artigos científicos. O desconhecimento de normas internacionais para a elaboração do trabalho científico é inadmissível nos tempos atuais. Enquanto ciência que busca respeito entre a comunidade científica internacional e deseja ser reconhecida por seus pares no mundo, a Psicologia latino-americana não pode prescindir desses requisitos de qualidade nos artigos publicados. Ignorar as exigências em relação a esses quesitos é fechar os olhos para o desenvolvimento e seguir às margens do lugar onde a ciência reside.

2.5 Revisão Sistemática: padrão-ouro das evidências científicas?

Já está se tornando comum ouvir afirmações como: “este estudo é nível 1” ou “tem grau de recomendação B”, como referência aos estudos científicos. Nos consultórios médicos também escutamos frases como “as evidências mostram que o efeito colateral deste medicamento é...” ou “Há evidências que comprovam a eficácia deste medicamento”. São as referências aos resultados de pesquisas, quase sempre publicados em revistas científicas. Algumas bulas de remédios trazem referências a estudos científicos. Precisamos tomar cuidado para não confundirmos

escala de evidência com grau de recomendação. O grau de recomendação refere-se ao tratamento e não ao estudo. A escala de evidência hierarquiza o estudo científico com base na metodologia aplicada para sua geração. A Medicina tem sua escala de evidências hierarquizadas em uma pirâmide, chamada *Pirâmide de Evidências*, como pode ser observado na Figura 2. Tal pirâmide representa graficamente os níveis hierárquicos das evidências, considerando “pergunta de pesquisa, a natureza e o delineamento do estudo” (Silva, 2013, p. 14).

Figura 2. Pirâmide das Evidências Científicas. Reproduzida de *Psicologia Baseada em Evidências: Provas Científicas da Efetividade da Psicoterapia*, de T. Melnik e A. N. Atallah, 2011 (capa), São Paulo: Santos. Ilustrada por L. C. C. Guzman.

A hierarquização da força da evidência que a pirâmide apresenta atende à área da Medicina e pode ser utilizada por outras disciplinas pela análise que fornece do valor dos diversos tipos de métodos de pesquisa. A Psicologia poderá ser a próxima área a desenhar sua pirâmide para acomodação dos seus métodos de pesquisa.

2.5.1 Aprendendo com a Medicina Baseada em Evidências

Cordeiro, Oliveira, Rentería e Guimarães (2007) orientam que a Medicina Baseada em Evidências “pode ser definida como o uso consciencioso, explícito e judicioso das melhores evidências na tomada de decisões relativas à prática com pacientes individuais, unindo evidências produzidas em pesquisa com experiência

clínica”. Os autores esclarecem que a prática clínica “compreende cinco componentes específicos: acesso à informação, avaliação crítica da literatura, principais desenhos de pesquisa clínica, métodos estatísticos e planejamento da pesquisa clínica”. Segundo os pesquisadores é necessário “saber acessar a informação, avaliar a literatura, entender as limitações (vantagens e desvantagens) de cada desenho de estudo, os métodos estatísticos envolvidos com a situação clínica e como planejar pesquisas para que seus resultados sejam válidos” (p. 428).

Há muito, a Medicina Baseada em Evidências vem auxiliando os profissionais da área de saúde no tratamento de pacientes e na tomada de decisão sobre o procedimento mais adequado. Galvão, Sawada e Rossi (2002) lembram que a prática baseada em evidências se originou do trabalho do epidemiologista britânico Archie Cochrane. A publicação da obra *Effectiveness and Efficiency: Random Reflections on Health Services*, no ano de 1972, é uma marco para a Medicina Baseada em Evidências. O médico pesquisador também encorajou e apoiou a criação do primeiro Centro Cochrane, em Oxford, Reino Unido, em 1992 e, em 1993, fundou a *The Cochrane Collaboration* (<http://www.cochrane.org/docs/archieco.htm>). O tema é de tanta relevância, que a área conta com fontes de informação especialmente dedicadas às revisões sistemáticas, como é o caso da Biblioteca Cochrane. A Colaboração Cochrane oferece uma coleção

de fontes de informação de boa evidência em atenção à saúde, em inglês. Inclui as Revisões Sistemáticas da Colaboração Cochrane, em texto completo, além de ensaios clínicos, estudos de avaliação econômica em saúde, informes de avaliação de tecnologias de saúde e revisões sistemáticas resumidas criticamente.⁴

O objetivo da *Cochrane Collaboration* é preparar, manter e disseminar revisões sistemáticas na área da Saúde. Atualmente, existem sete centros Cochrane na Europa (França, Alemanha, Grã-Bretanha, Espanha, Itália, Holanda e Dinamarca), além de centros no Canadá, China, Austrália, Nova Zelândia, África do Sul e Brasil.

A BIREME (Rede Latino-Americana e do Caribe de Informação em Ciências da Saúde) também vem investindo na organização e disseminação das evidências em Saúde. O Portal de Evidências (<http://evidences.bvsalud.org/php/index.php>) oferece:

⁴ Recuperado de <http://cochrane.bvsalud.org/portal/php/index.php?lang=pt>

- Revisões Sistemáticas, Revisões Cochrane (CDSR), Revisões avaliadas (DARE), LILACS;
- Ensaios clínicos CENTRAL, LILACS, estudos sobre dor nas costas;
- Sumários de evidência, atualização na prática ambulatorial, Bandolier, gestão clínica e sanitária;
- Avaliações Econômicas em Saúde – NHS-EED, LILACS;
- Avaliações de Tecnologias em Saúde – HTA-INAHTA, Agências Ibero-Americanas;
- Diretrizes para Prática Clínica – National Guideline Clearinghouse.

Sites da Internet avaliados e eventos relacionados à metodologia Medicina Baseada em Evidências também são apontados no Portal. Além do acesso às revisões sistemáticas, ensaios clínicos, sumário de evidência, avaliações econômicas em saúde, avaliação de tecnologias em saúde e diretrizes para a prática clínica, o Portal fornece também uma metodologia para a aplicação da Medicina Baseada em Evidências.

A Medicina Baseada em Evidências fundamenta suas decisões em evidências obtidas de maneira cientificamente provadas. Uma das maiores dificuldades que a Medicina encontra é identificar a qualidade das informações recuperadas nas publicações. Para tanto, existem métodos específicos de avaliação crítica que amparam a sistematização do conhecimento. As revisões sistemáticas têm a vantagem de seguir métodos científicos rigorosos, que podem ser reproduzidos e criticados. A aplicação da Técnica evita a duplicação de esforços, já que, uma vez prontas, as revisões sistemáticas poderão ser lidas, divulgadas e utilizadas mundialmente, além de atualizadas, lembrando que temos os meios virtuais que possibilitam extrema agilidade na comunicação científica. Assim, a Medicina Baseada em Evidências é a aplicação das melhores evidências no tratamento clínico. Entretanto, para que essas melhores evidências possam ser recuperadas, de forma rápida e segura, pesquisadores precisam se debruçar sobre a literatura publicada e indexada nas bases de dados nacionais e internacionais, selecionando as melhores práticas a partir da sistematização das informações ali publicadas.

De acordo com El Dib (2007), a Medicina Baseada em Evidências é o elo entre a boa pesquisa científica e a prática clínica. Em outras palavras, explica que “a

MBE utiliza provas científicas existentes e disponíveis no momento, com boa validade interna e externa, para a aplicação de seus resultados na prática clínica” (p. 1). A boa evidência está diretamente ligada à efetividade, eficiência, eficácia e segurança da intervenção. O processo inicia-se pela formulação de uma questão clínica, assim, uma pergunta bem formulada é a chave para o sucesso da pesquisa, pois diminui as possibilidades de erros sistemáticos durante o planejamento, a análise estatística e a conclusão do projeto. A autora orienta sobre os quatro itens fundamentais que compõem uma boa pergunta científica: 1) situação clínica (qual é a doença), 2) intervenção (qual é o tratamento de interesse a ser testado), 3) grupo controle (placebo, *sham*, nenhuma intervenção ou outra intervenção) e 4) o desfecho clínico (El Dib, 2007).

Lopes (2000) orienta que a Medicina Baseada em Evidências é a aplicação da prática médica em um contexto no qual a experiência clínica se entrega à capacidade de analisar as evidências científicas de maneira crítica e aplica-lás de forma racional com o objetivo de melhorar a qualidade da assistência médica. O autor afirma que um especialista está apto para a prática quando souber identificar os problemas relevantes do paciente, conseguir converter os problemas em questões que levem às respostas necessárias, conhecer e estar capacitado para consultar as fontes de informação da área, ser capaz de avaliar a qualidade da informação e a força da evidência, certificando ou negando o valor de uma determinada conduta. Chegar a uma conclusão correta em relação ao significado da informação e ter capacidade para aplicar as conclusões dessa avaliação na melhoria dos cuidados prestados aos pacientes também são habilidades que o revisor deve possuir.

A expressão “nível de evidência” ocorre com frequência e se refere apenas a isto: quanto posso confiar nesse artigo? Em geral, o nível de evidência corresponde ao delineamento do estudo e à natureza do artigo, estando as revisões sistemáticas de ensaios clínicos controlados randomizados e ensaios clínicos controlados randomizados bem delineados no topo da pirâmide das evidências, sendo seguidos por estudos observacionais de coorte ou estudos de caso-controle, a seguir os estudos de caso, então estudos de autoridades e oficiais próximos à base. (Greenhalgh, 2008, p. 34)

Assim, podemos resumir que a Medicina Baseada em Evidências é a aplicação das melhores práticas publicadas e analisadas sistematicamente com o objetivo de oferecer o melhor tratamento aos pacientes. Para que a Medicina

Baseada em Evidências seja possível, é necessário que os especialistas elaborem perguntas de pesquisas bem estruturadas, busquem as evidências em artigos publicados em veículos reconhecidos e bem avaliados pela comunidade científica, analisem a relevância e a pertinência da prática clínica ali registrada e sintetizem as melhores evidências, publicando os resultados da revisão em artigo de periódico ou livro. Para que tudo isso seja possível, a qualidade metodológica e o relato das pesquisas que geraram o artigo científico é de suma importância.

Do mesmo modo como vem acontecendo com a Medicina, a Psicologia desenvolveu técnicas e métodos de intervenção eficazes e que precisam ser avaliados e sintetizados para uso do psicólogo na sua prática clínica e do gestor na definição de políticas públicas. Nesse sentido, parece oportuna uma discussão que traga ferramentas para avaliação e sistematização do conhecimento gerado pela área e que se introduza a Psicologia Baseada em Evidências (PBE) como apoio ao fazer da Psicologia no Brasil e em outros países da América Latina, uma vez que, como veremos a seguir, em alguns países desenvolvidos essa prática está bem estruturada.

2.5.2 Evidências científicas: a tomada de decisão nos processos psicológicos

Enquanto profissão que completou 50 anos de regulamentação no Brasil, no ano de 2012, a Psicologia avançou muito em relação à geração e disseminação do conhecimento produzido por pesquisadores e estudiosos da disciplina. O número de fontes de informação que divulgam a produção de Psicologia da América Latina, que serão apresentadas no capítulo quatro desta tese, demonstra que esforços estão sendo efetuados para dar a visibilidade necessária ao conhecimento gerado na região.

São diversos e prestigiosos recursos que fornecem informações atualizadas logo após a publicação. Entretanto, sabemos que a falta de tempo para se buscar em todas as bases de dados, ler os resumos, selecionar os documentos e depois analisá-los é fator determinante para que o profissional desista de buscar os melhores resultados para aplicá-los na solução de problemas. Não recorrer às melhores práticas, tratando-se de Psicologia, uma das ciências que trabalha com a vida humana, pode significar a aplicação de procedimentos antigos e superados em detrimento do uso de técnicas recém-descobertas ou mesmo atualizadas.

O conhecimento de procedimentos exitosos, quase sempre registrados em boas publicações, pode levar o profissional à solução de problemas de forma rápida e efetiva, amenizando o sofrimento ao qual o paciente está exposto. A Psicologia construiu um volume de conhecimento que precisa ser revisitado por especialistas para uma prática apoiada em conhecimento científico.

Excesso de publicações para consulta não é privilégio da Psicologia e nem um problema da era atual, “... por volta de 1830, que o processo havia alcançado um ponto absurdo: cientista algum podia ler todos os periódicos ou manter-se suficientemente a par de todos os trabalhos publicados que tivesse relevância para seus próprios estudos” (Price, 1976, p. 146). O autor relata que “com um conjunto de aproximadamente 300 periódicos sendo publicados, fazia-se necessária uma solução nova” (p. 146). A solução foi a organização dos resumos dos artigos em uma revista especialmente criada para este objetivo: a revista de *abstract*, criada no final do século XIX e início do século XX que, segundo Price (1976, p. 146), era tão “deliberada e controvertida como a própria revista científica”. Em 1976, Prince previa que, de acordo com a tendência histórica, a nova solução apontada – bases de dados em formato eletrônico – seria apenas um paliativo e não a solução radical e somente retardaria em poucas décadas a crise inevitável das revistas científicas. Alertava também que se não “encontrarmos alguma forma de sumarizar os sumários, poderá muito bem acontecer que o artigo impresso desapareça, embora seja difícil nos libertarmos da obsessão de que ele é vital para a ciência” (p. 147). Prince não imaginava como essa “obsessão”, chamada artigo científico, geraria polêmica e discussão de cientistas nos anos posteriores aos seus escritos.

A Psicologia sempre demonstrou estar alerta ao problema de excesso de artigos e no mês de março, do ano de 1895, lançou uma publicação para sumarizar as revistas, considerada o primeiro indexador da área. Chamado *The Psychological Index*, teve como primeiro editor Howard Crosby Warren, da *Princeton University and Livingston*; resumia 1.312 revistas e foi publicado até o ano de 1936. Após a Segunda Guerra mundial, a literatura psicológica aumentou muito e o *Index* não conseguiu se manter atualizado (Benjamin Jr. & VandenBos, 2006).

Fundado em 1904, o *Psychological Bulletin* apresentava cerca de 40 artigos de revisão por ano. Um relatório, de autoria de Fulcher, sugeria que o *Psychological Bulletin* fosse transformado em boletim de *abstracts*, mas a ideia não vingou

totalmente. Entretanto, o periódico publicava uma parte de seu conteúdo com os *abstracts* das principais revistas da área.

No ano de 1921, surge a discussão sobre a possibilidade de se ter um *psychological abstracts journal*, de responsabilidade da APA e, em janeiro de 1927, aparece o primeiro fascículo do *Psychological Abstracts* (PA). No ano de 1973, surge a primeira base de dados, com 75.000 registros oriundos do PA, inaugurando uma nova etapa em relação à sumarização e acesso à informação psicológica. Em 1977, PsycINFO passa a ser acessível via *online*, pelo sistema DIALOG. As novas tecnologias, emergentes nos anos de 1980, impulsionaram a construção de fontes de informação em CD-ROM, e a SilverPlatter Inc, empresa contratada pela APA, imprimiu, no ano de 1986, o CD-ROM PsycLit, com a base de dados originária do PA. Durante um período, subsistiram três fontes de informação com conteúdos muito similares: PA (publicação impressa), PsycINFO (versão *online*) e PsycLit (versão em CD-ROM). No ano 2000, o PsycLit deixou de ser gravado e, em 2005, o PA encerrou a publicação impressa, restando apenas a versão *online* PsycINFO, considerada a mais importante base de dados da área de Psicologia do mundo (Benjamin Jr. & VandenBos, 2006).

É importante destacar tal informação para mostrar que Price estava certo quando afirmou que se não surgisse uma forma de organizar os artigos científicos em fontes de informação, que possibilitassem ao pesquisador e estudioso o acesso de forma rápida e eficiente aos artigos, a revista científica teria sucumbido, mesmo sendo a “coqueluche” da ciência. Buscando contribuir com o uso racional da informação, uma interessante estratégia surgiu para organizar os milhares de artigos em um único estudo: as revisões sistemáticas da literatura.

Segundo o Curso de Revisão Sistemática e Metanálise da Universidade Federal de São Paulo (UNIFESP) - (www.virtual.epm.br/cursos), a primeira revisão sistemática sobre um cenário clínico foi publicada no *Journal of American Medical Association*, em 1955. Anteriormente, algumas publicações abordavam os métodos estatísticos, chamados metanálise, para combinar resultados de pesquisas. Trata-se de um método quantitativo que permite combinar resultados de pesquisas, realizadas de forma independente, e sintetizar suas conclusões, ou mesmo extrair nova conclusão pela aplicação de uma ou mais técnicas estatísticas (Luiz, 2002). O termo metanálise foi adotado por G. V. Glass, psicólogo, no ano de 1976, em um artigo intitulado “Primary, Secondary and Meta-Analysis of Research”, publicado na

revista *Educational Research*, volume 5, páginas 3 a 8. “Antes disso, em diversos trabalhos foram utilizadas técnicas estatísticas para combinar ou reunir dados já publicados ou utilizados em outros estudos sem, entretanto, ter-se cunhado um termo específico” (Luiz, 2002, p. 409). Para que a metanálise seja realizada é necessário que a revisão sistemática reúna dois, ou mais, estudos que respondam à mesma pergunta de pesquisa.

Em 1995, um grupo de cientistas reunidos em Potsdam (Alemanha) definiu como revisão sistemática “a aplicação das estratégias científicas que limitam o viés de seleção e avaliam com espírito científico os artigos e sintetizam todos os estudos relevantes em tópicos específicos.” (Cordeiro et al., 2007, p. 429)

Intensamente utilizada na área da saúde, a revisão sistemática é um tipo de investigação científica. Seu objetivo é levantar, reunir, avaliar criticamente e sintetizar os resultados de diversos estudos primários. A revisão sistemática busca responder a uma pergunta claramente formulada, utilizando métodos sistemáticos e explícitos para recuperar, selecionar e avaliar as pesquisas relevantes, reunindo e analisando os dados dos estudos. Na revisão sistemática, os sujeitos são os estudos primários publicados (unidades de análises), selecionados para compor o grupo dos estudos que darão as respostas à pergunta formulada. As revisões sistemáticas são consideradas estudos observacionais retrospectivos por alguns autores e por outros são situadas como estudos experimentais (Cordeiro et al., 2007).

As revisões sistemáticas têm sido utilizadas para testar hipóteses concretas. No caso de estudos quantitativos, é fácil combinar seus resultados para obter dados padronizados, podendo, inclusive, aplicar-se a metanálise para a integração desses dados. Porém, essa metodologia é insuficiente quando se busca aplicá-la aos estudos qualitativos. Neste caso, pode ser interessante realizar-se uma revisão integrativa, que nos dará uma ideia de como os estudos foram realizados gerando as hipóteses e identificando os aspectos que podem ser estudados com mais detalhes. Revisão integrativa é um método de estudo que inclui a análise de resultados de pesquisas relevantes e que dão suporte à tomada de decisão, tanto na prática clínica, quanto na gestão pública. A partir da síntese do conhecimento sobre um determinado assunto, esse tipo de revisão pode apontar lacunas sobre áreas de estudos descobertas e que necessitam ser preenchidas com a realização de novos trabalhos. A revisão integrativa difere da sistemática pelo método de elaboração, que

aceita estudos com metodologias diferenciadas e com abordagens qualitativas e quantitativas (Mendes, Silveira, & Galvão, 2008).

A maioria dos elementos pertencente a metodologia da revisão sistemática tradicional é válida para a revisão sistemática exploratória, embora não se possa desprezar o desenho diferenciado dos estudos dessa natureza. Essas revisões, em geral, são publicadas em forma de informes técnicos e são utilizadas como ferramentas na tomada de decisão no âmbito da saúde no trabalho (Gabarito et al., 2009). Os mesmos autores explicam que os objetivos da revisão sistemática são:

- avaliar a qualidade e metodologia empregadas nas pesquisas científicas;
- sintetizar a evidência científica;
- oferecer subsídios na tomada de decisão com base em conhecimento científico. Ampliando o leque de possibilidades de revisões (Gabarito et al., 2009), enumeram os estudos em três categorias:

1) *revisão sistemática*: trabalhos que resumem de forma sistemática a evidência científica para analisar a relação entre elas, respondendo a uma pergunta concreta de pesquisa.

2) *Revisão sistemática com metanálise*: além de fazer a relação entre as evidências, aplica uma técnica estatística para sintetizar quantitativamente os resultados dos estudos primários. O objeto da metanálise é proporcionar estimativas mais precisas e integradas.

3) *Revisão sistemática exploratória*: sintetiza as evidências sobre um tema e descreve o conhecimento existente. Serve para gerar hipótese, estabelecer linhas de pesquisas, ou como base para Informes Técnicos. Informes Técnicos são documentos selecionados segundo um determinado critério, que responde a uma pergunta determinada. A Tabela 2 reafirma as diferenças entre os tipos de revisão.

Tabela 2 - Diferenças entre a Revisão Sistemática Clássica e a Revisão Sistemática Exploratória

Tipo de estudo	RS clássica	RS exploratória
Hipóteses	Responde a uma pergunta concreta	Descreve o conhecimento existente
Objetivos	Uma intervenção	Mais de uma intervenção
Estudos incluídos	Qualitativo e quantitativo	Todos os tipos de estudos
Resultados	Uma medida de impacto	Diferentes medidas de impacto
Crítérios de inclusão	Seleção por qualidade	Seleção por relevância

Nota. Tabela reproduzida e traduzida de “Revisiones sistemáticas exploratorias”, de R. M. Gabarito et al., 2009, *Medicina y Seguridad del Trabajo*, 55(216), p. 15. Recuperada de <http://scielo.isciii.es/pdf/mesetra/v55n216/especial.pdf>

Não apenas explorar o que foi feito sobre um tema, mas também analisar onde e como foi feito para gerar hipóteses para futuras perguntas de pesquisas e propor novos estudos que ainda não foram desenvolvidos. O desenho de pesquisa de uma revisão é visto como um momento crítico durante o processo, por isso, assim como todo projeto de pesquisa, necessita de um delineamento cuidadoso e bem estruturado.

A força tarefa da APA, definida para a prática baseada em evidências (Levant, 2005), indicou os tipos de pesquisas mais utilizados pela Psicologia e estas indicações guardam semelhanças com os tipos de pesquisa citados anteriormente.

Tabela 3 - Tipos de pesquisa definidos na Força Tarefa APA

Força Tarefa APA (2005)
Observação clínica (incluindo estudos de casos individuais) e pesquisa básica em Psicologia são valiosas fontes de inovações e hipóteses (no contexto da descoberta científica).
Pesquisa qualitativa pode ser usada para descrever a experiência subjetiva vivida pelas pessoas, incluindo os participantes na psicoterapia.

Força Tarefa APA (2005)

Estudos de caso sistemáticos são particularmente úteis quando agregados, sob a forma de redes de pesquisa, para comparar pacientes individuais a outros com características semelhantes.

Delineamentos experimentais de sujeito único são particularmente úteis para o estabelecimento de relações causais no contexto de um indivíduo.

Pesquisas de saúde pública e etnográficas são especialmente úteis para monitorar a disponibilidade, utilização e aceitação de tratamentos de saúde mental, bem como sugerir formas de alterá-los e maximizar a sua utilidade em um determinado contexto social.

Estudos de processos-resultados são especialmente valiosos para a identificação de mecanismos de mudança no indivíduo.

Estudos de intervenções, tal como aplicados em ambientes naturais (investigação de eficiência), são adequados para avaliar a validade ecológica dos tratamentos.

Os ensaios clínicos randomizados e seus equivalentes lógicos (pesquisa eficácia) são o padrão para fazer inferências causais sobre os efeitos das intervenções (verificação do contexto científico).

Metanálise é um meio sistemático para sintetizar os resultados de vários estudos, testar hipóteses e, quantitativamente, estimar o tamanho dos efeitos.

Nota. Tabela traduzida e adaptada de *Report of the 2005 Task Force on Evidence-based Practice* de Levant, 2005, pp. 7-8. Recuperado de <http://www.apa.org/practice/resources/evidence/evidence-based-report.pdf>

Parece um processo natural que os profissionais nas suas mais diversas linhas de atuação recorram ao conhecimento científico gerado nas universidades e nos centros de pesquisas para a atualização continuada. Gestores também encontram nas evidências suporte qualificado para a tomada de decisão, os políticos para a gestão pública. Antes restrito às bibliotecas das instituições geradoras das novidades científicas, o conhecimento, hoje, está acessível via Internet e todos podem ter acesso a ele, de forma rápida e eficiente, desde que saibam onde está e consigam avaliar a qualidade da informação disponibilizada. Entretanto, avaliar a qualidade desse conhecimento é uma tarefa árdua, que antecede o uso efetivo das provas científicas.

Estamos falando em reaproveitamento do conhecimento a partir da utilização dos resultados de pesquisas científicas na prática profissional e no desenvolvimento de políticas de saúde pública. Há muito se discute qual seria o benefício direto para

a população, advindo dos investimentos efetuados pelo Estado, da produção da ciência, principalmente nos países em desenvolvimento.

No Brasil, por exemplo, grande parte da população arca com os custos da geração do conhecimento nas universidades públicas e nem sequer sabe o que ali é produzido, muito menos como são utilizados os recursos retirados dos impostos pagos pelos contribuintes. As universidades públicas no Brasil são totalmente gratuitas e os alunos de pós-graduação recebem, muitas vezes, bolsas de estudo para desenvolverem suas pesquisas. Nessa situação, nada mais lógico que a população se valer do seu investimento para o próprio benefício. Isso pode ser feito a partir do alcance aos resultados de pesquisas, quer seja pelo profissional para aplicação desses resultados no tratamento dos pacientes, quer seja pelo próprio cidadão, que receberá do profissional o melhor tratamento com a ajuda das evidências científicas de qualidade.

Além de financiar o alto custo da geração do conhecimento desenvolvido nas universidades públicas e institutos de pesquisas, é comum o cidadão ser incluído como participante de pesquisas, das quais quase nunca toma consciência dos resultados obtidos. Saber como o dinheiro dos seus impostos é utilizado em ciência e tecnologia e como suas queixas, respostas e problemas contribuem para o avanço da ciência é uma forma justa de retribuir o investimento aplicado em pesquisas e na pós-graduação das melhores universidades do país.

Notamos a crescente participação do Estado como formulador, financiador e, muitas vezes, executor de atividades voltadas à ciência e à tecnologia. Sendo ele o grande responsável pela promoção da saúde da população, nada mais justo que esperar da comunidade científica retorno pelos investimentos aplicados na geração do novo conhecimento (Barreto, 2004). Aos poucos, a academia vai saindo do isolamento que, muitas vezes, tornava-a semelhante a um templo do saber, para se aproximar do cidadão e devolver parte dos investimentos empregados em pesquisas e estudos.

Tanto no processo de utilização das evidências científicas na prática profissional quanto na tomada de decisão em políticas públicas, os atores envolvidos nessas atividades precisam extrair do “estoque” de conhecimentos gerados e organizados em bases de dados e outras fontes de informação o que lhe será útil na solução de seus problemas.

O processo de transformar o conhecimento em evidência alimentadora de uma decisão é complexo e nunca totalmente científico, sendo permeado por diferentes interpretações e valorações do conhecimento existente, da mesma forma que todo o processo de produzir e disseminar conhecimentos está mediado por relações que se estabelecem no interior da comunidade científica e desta com os vários setores e interesses da sociedade. Com relação a uma mesma questão, no “estoque” podem existir conhecimentos diversificados, produzidos por diferentes disciplinas, muitas vezes com diferentes abordagens teóricas e em diferentes contextos históricos e geográficos. Extrair e utilizar o conhecimento em “estoque” é um processo complexo e não existe uma forma única de fazê-lo, podendo variar desde formas não sistemáticas até formas mais sistemáticas e com metodologias estruturadas. (Barreto, 2004, p. 332)

O enfoque da prática, baseada em evidência, sustenta-se sobre dois princípios fundamentais: 1) o convencimento de que as melhores evidências científicas sobre a eficácia dos programas, serviços e ações procedem das revisões sistemáticas dos estudos avaliativos; 2) que é preciso articular canais de comunicação que permitam levar tais evidências de forma rápida aos políticos, gestores, profissionais que atendem aos pacientes e receptores dos programas, assim como ao público em geral (Sánchez-Meca, Boruch, Petrosino, & Rosa, 2002).

Tão importante quanto o apoio na definição de políticas públicas, as práticas com base em evidências são essenciais no estabelecimento de diagnóstico, prognóstico e na seleção de “intervenções mais eficazes frente aos desfechos clínicos de interesse, especialmente, ao verificarem o impacto das doenças ou condições de saúde sobre a vida das pessoas...” (Dias & Dias, 2006, p. 11). Assim, a formação profissional e a prática atual requerem que os profissionais sejam capacitados para trabalhar com base nesse recurso, garantindo rapidez e eficiência que podem culminar com a melhoria da qualidade de vida das pessoas, ou mesmo no próprio salvamento de uma vida humana.

Um exemplo da relevância do tema e de que a discussão vai ganhar corpo nos próximos anos é a EVIPNet (Evidence-informed Policy Network), rede colaborativa para elaboração e implementação de políticas baseadas em evidências científicas. O Brasil aderiu à EVIPNet, mediante o projeto elaborado pela Secretaria de Ciência, Tecnologia e Insumos Estratégicos, do Ministério da Saúde, e por meio do Departamento de Ciência e Tecnologia (Decit), com a participação da Organização Pan-Americana da Saúde (OPAS).

A EVIPNet Brasil tem por objetivo apoiar o desenvolvimento de políticas baseadas em evidências científicas, permitindo, também, o acesso permanente a elas. Essa iniciativa busca promover o uso sistemático de evidências de pesquisa no processo de elaboração e definição de políticas de saúde, por meio de parcerias entre gestores, representantes do controle social e pesquisadores. (Evidências Científicas, 2009, p. 1084)

O Ministério da Saúde do Brasil, a OPAS/Organização Mundial da Saúde (OMS), o Centro Latino-americano e do Caribe de Informação em Ciências da Saúde (Bireme), a Fundação Oswaldo Cruz (Fiocruz), a Associação Brasileira de Pós-graduação em Saúde Coletiva (Abrasco), a Comissão Intersetorial de Ciência e Tecnologia do Conselho Nacional de Saúde, o Conselho Nacional de Secretários de Saúde (Conass) e o Conselho Nacional de Secretarias Municipais de Saúde (Conasems) são os parceiros desse projeto em busca de colaboradores para a difusão das iniciativas em âmbito nacional. Como área estratégica da saúde no país, a Medicina lançou os pilares do uso de evidências na tomada de decisão.

2.5.2.1 Psicologia Baseada em Evidências

A pluralidade da Psicologia permite que coexistam muitas teorias que competem entre si para explicar um mesmo fenômeno. Em um mesmo dia, um estudante pode aprender sobre uma teoria computacional das emoções e, em seguida, receber explicações sobre o modelo psicanalítico dessas mesmas emoções (Vázquez & Moreno, 2003). Os autores entendem que esta pluralidade não é um valor agregado, mas reflete um estado imaturo de crescimento de uma disciplina. Completam que a fragmentação de conhecimentos e a proliferação de enfoques, mais no âmbito profissional que no acadêmico, resulta de difícil aceitação do ponto de vista epistemológico.

A complexidade do objeto de conhecimento da Psicologia justifica essa pluralidade e possibilita um volume de informação de difícil manejo, seleção e avaliação, tornando a vida do estudioso, pesquisador e profissional ainda mais complicada. O acúmulo e a oferta dessa diáspora teórica dificultam ainda mais o filtrar de modo adequado, com critérios de qualidade, ou de exigências científicas, àquilo que é publicado (Vázquez & Moreno, 2003). Diante dessa explosão informacional, com tantas opções, torna-se complicado acessar, avaliar e analisar

toda informação existente, principalmente, em se tratando de artigos científicos. Podemos voltar e recordar o número de bases de dados que indexam a produção da área e veremos, ao longo deste trabalho, as dificuldades encontradas por um pesquisador para sistematizar um extrato especializado desse conhecimento. A solução não é fácil, mas a alternativa aponta para o desenvolvimento de ferramentas para a seleção, avaliação e classificação do conhecimento publicado como, por exemplo, as utilizadas pela revisão sistemática.

Nos surpreendemos ao localizar um artigo de autoria de Angelini (1954) cujo objetivo foi comparar os resultados de pesquisas sobre o emprego do ácido glutâmico na terapia da *debilidade mental* para verificar se eram concordantes e autorizavam o uso da droga como terapêutica da deficiência mental. O autor reuniu, analisou e sistematizou 36 artigos de acordo com a data, país do estudo, sujeitos, grupo controle, tempo da administração da droga, provas utilizadas, resultados e observações sobre os estudos. No final, concluiu que os resultados eram fracos, por isso, uma generalização dos efeitos positivos do ácido não podia ser feita. O artigo parece ser o embrião da Psicologia Baseada em Evidências no Brasil e não fora citado em nenhum dos trabalhos sobre o tema, dentre tantos que consultamos. Não estar indexado em bases de dados e nem disponível em acesso aberto pode ser o motivo dessa invisibilidade.

A outro psicólogo é atribuído o pioneirismo em reunir e sintetizar estudos em uma metanálise (Glass, 1976). Desde o estudo de Angelini (1954) e o de Glass (1976), apenas recentemente os estudiosos começaram a se preocupar com o problema do excesso de informação e passaram a chamar a atenção para a responsabilidade de todos em analisar esse acúmulo de conhecimento de forma eficiente e com o mesmo empenho, rigor e sistematização da pesquisa primária. Podemos crer que, em um futuro bem próximo, teremos dois tipos de pesquisadores: aqueles que seguirão fazendo as pesquisas clássicas (estudos primários) e os que se dedicarão ao trabalho de integrar essas pesquisas a partir de suas análises (estudos secundários). Afinal, as revisões nos dizem o que está sendo estudado em uma determinada área, assim como suas zonas descobertas e lacunas a serem preenchidas.

O termo *Prática da Psicologia Baseada em Evidências* foi utilizado pela APA apenas no ano de 2005, embora desde a década de 1970 possamos perceber um

interesse pelo uso de evidências científicas nos tratamentos psicoterápicos, a partir da reunião e síntese estatísticas de estudos publicados. Em 1977, Smith e Glass sistematizaram estudos controlados sobre psicoterapia. Dos 375 trabalhos considerados adequados, foi possível observar que 75% dos pacientes tiveram melhoras com o tratamento. Também foi possível identificar que as diferenças teóricas entre as escolas de psicoterapia influenciaram os resultados do tratamento. Concluem que os estudiosos e profissionais da saúde estão em uma posição difícil, considerando o volume e a dificuldade de acesso ao conhecimento publicado em livros e revistas científicas, e que era necessário buscar formas de localizar e extrair conhecimentos de estudos acumulados, e que esta era uma questão metodológica que merecia atenção. Smith e Glass ficariam felizes se soubessem que, no ano de 2013, quando pesquisamos pela expressão “Meta-analysis of psychotherapy”, no Google Acadêmico, encontramos 2.920 referências, sendo o estudo citado por 1.786 trabalhos, o que demonstra o reconhecimento da obra pioneira.

É necessário que os cientistas comecem a pensar em uma ciência sustentável, na qual se reutilizam os resultados já levantados, em vez de consumirem novos recursos para repetir o que já foi feito. Antes de se empreender uma pesquisa sobre um determinado tema, há que se buscar o que foi pesquisado sobre o assunto e seguir, a partir de então, aproveitando as conclusões e os resultados obtidos nos estudos anteriores para avançar rapidamente com a ciência. Os erros e os enganos também devem ser controlados, evitando-se repetir pesquisas fracassadas, que somente oneram os custos da ciência. Tudo deve ser publicado com a qualidade dos bons artigos, para que possamos caminhar com a velocidade que o nosso tempo exige. A esse respeito, não produzir os artigos de qualidade discutível e respeitar as recomendações de Trzesniak et al (2012) são alternativas interessantes a considerar.

Analisar sistematicamente as evidências e aplicar técnicas estatísticas para a medição dos resultados demanda o conhecimento de fontes de informação e domínio de ferramentas para análise estatística. Essa atividade é fortemente desenvolvida no meio dos pesquisadores e docentes de universidades. Entretanto, o psicólogo clínico, uma vez orientado e devidamente preparado, também pode se aventurar nessa área. Instruções claras e de fácil compreensão se apresentam como um bom começo para consolidar a prática da revisão sistemática, que permitirá encontrar as boas evidências científicas entre os profissionais da

Psicologia. Enquanto o psicólogo não se apoderar das técnicas para elaboração de revisões, saber que essas contribuições existem, onde estão e que elas podem economizar muito de seu tempo com pesquisa e leitura de documentos, o progresso da Psicologia será retardado e os pacientes continuarão perdendo muito. Os gestores de políticas públicas também devem ser orientados e incentivados a usar o bom conhecimento disponível em fontes de informação na tomada de decisão.

Segundo Botella e Gambara (2006), a ciência é cumulativa, porém os pesquisadores da Psicologia acreditavam que esse acúmulo era feito naturalmente. Durante décadas, as revisões foram consideradas como um trabalho de segunda categoria, uma espécie de parasita que se alimenta dos trabalhos de outros e que nem eram dignas de serem registradas nos currículos. Claro, o acúmulo se faz por si, mas buscar, analisar e discutir esse estoque de conhecimento é diferente de apenas juntá-lo. Trata-se de gerar conhecimento a partir daquele já publicado, reduzindo o tempo de consulta a vários trabalhos e retirando o que cada texto tem de melhor para a tomada de decisão.

A habilidade para se avaliar textos publicados e classificá-los em um estudo sistemático, que permita detectar as melhores evidências, ainda é recente e não faz parte das atividades de rotina dos pesquisadores e profissionais da Psicologia. Melnik (2006, p. 15) acredita que pelo volume de contribuições publicado nas revistas de Psicologia e o “alcance dos recursos atuais em comunicação e informática, é apenas uma questão de tempo até que essa habilidade seja incorporada como parte integral da prática psicológica” (p.15).

O *Report of the 2005 Presidential Task Force on Evidence-Based Practice*, da *American Psychological Association* (APA), define Prática Baseada em Evidências como a integração da melhor pesquisa científica disponível com a experiência clínica no contexto das características, cultura e preferências do paciente (Levant, 2005). Em um contexto mais amplo, podemos definir a Psicologia Baseada em Evidências (PBE) como a integração de sólidas teorias psicológicas com a melhor evidência de pesquisa disponível, resultante de pesquisas de alta qualidade metodológica, aliada à experiência profissional na tomada de decisão em relação ao tratamento clínico, elaboração de guias e procedimentos e definição de políticas públicas na área de Psicologia (Levant, 2005).

No contexto das ciências sociais, educativas e do comportamento e com o objetivo de alcançar os princípios da prática baseada em evidências foi fundada a

Colaboração Campbell, uma organização internacional criada seguindo o modelo da Colaboração Cochrane. Sánchez-Meca et al. (2002) explicam que a Colaboração Campbell, fundada em 2000, na Filadélfia (Pennsylvania), foi criada à imagem e semelhança da Colaboração Cochrane, e é uma organização internacional que reúne investigadores, profissionais e gestores de diversos segmentos das ciências sociais (Educação, Serviços Sociais, Psicologia, Criminologia, Economia) interessados em ajudar aos políticos, gestores e demais profissionais a tomarem decisões bem informadas em relação aos programas, serviços e ações a serem implementados. A ideia é que, a exemplo da Colaboração Cochrane, a Colaboração Campbell tenha seus centros regionais, dotados de infraestrutura para o desenvolvimento dos trabalhos dos Grupos de Revisão.

Confirmando as afirmações de Sánchez-Meca et al. (2002) quanto à aplicação das melhores evidências em outras áreas do conhecimento, temos o trabalho de Cantisano, Depolo e Dominguez (2007), que apresenta uma revisão sistemática sobre o assédio moral. Os autores pesquisaram intensamente o tema em 86 estudos empíricos, com 93 amostras submetidas a uma metanálise. A matriz da metanálise foi utilizada para comprovar o modelo de avaliação estrutural. Os resultados confirmam a hipótese relativa aos fatores do ambiente organizacional como preceptores do assédio. A revisão pode ser utilizada pelos gestores como caminho na tomada de decisão quanto ao ambiente organizacional e na aplicação de medidas efetivas de controle das situações operacionais, economizando o tempo do gestor com consultas a diversos artigos.

Sánchez-Meca et al. (2002) orientam que o uso das melhores evidências científicas se aplica a qualquer contexto, tratamentos, intervenções, serviços e ações de modo geral. Os autores explicam que a corrente é denominada *prática baseada em evidência*, ou *política baseada em evidência* e nasceu no cerne das ciências da saúde, originada da medicina baseada em evidências, e está se expandindo para as ciências sociais, gerando correntes nas áreas da Educação, Psicologia, Trabalho Social, Criminologia e Economia.

No campo da Psicologia, Pheula e Isolan (2007, p. 74) utilizam o termo “tratamentos baseados em evidências” para se referirem às “intervenções que possuem evidência de eficácia em pesquisas empíricas”. Explicam que os tratamentos psicológicos com base em evidências têm sido aplicados nos Estados Unidos e, “atualmente, há uma clara orientação em direção à sua indicação para os

transtornos psiquiátricos mais prevalentes”. Buscando colaborar com a área ainda incipiente no Brasil, os autores revisam as intervenções psicossociais para as principais formas clínicas dos transtornos mentais na infância e na adolescência, e os desafios para a pesquisa em tratamentos baseados em evidência. A partir de uma revisão na base de dados MedLine, de 1985 a 2005, os autores analisaram as evidências e concluíram que a psicoterapia vem sendo utilizada como suporte empírico em depressão, ansiedade, transtornos do comportamento disruptivo e transtorno de déficit de atenção/hiperatividade. Concluíram que quando diversos artigos são avaliados e sintetizados, as certezas sobre as intervenções passam a ter base em conhecimento e isso lhes atribui maior valor.

A primeira tese que efetuou revisão sistemática com metanálise na área de Psicologia é de autoria de Melnik (2006) e teve como objetivo avaliar a eficácia das intervenções psicossociais para o tratamento da disfunção erétil, em comparação aos medicamentos via oral, injeção local e outras intervenções. O estudo demonstrou que há evidências de que a psicoterapia de grupo pode melhorar a função erétil. A revisão foi atualizada no ano 2007 e publicada na Biblioteca Cochrane, contribuindo com os profissionais que necessitam tratar pacientes com esse distúrbio. A consulta ao estudo de Melnik permite grande economia de tempo na leitura dos vários artigos que compõem a revisão, revertendo em ganho real, tanto para o profissional quanto para o paciente.

Objetivando analisar sistematicamente a produção científica relacionada à interação entre mães e crianças nascidas pré-termo e seu desenvolvimento nas fases pré-escolar e escolar, do período de 1998 a 2004, Klein e Linhares (2006) realizaram um levantamento bibliográfico nas bases de dados Medline, PsycInfo e LILACS e identificaram 33 artigos de revistas científicas, dos quais 26 foram selecionados e 22 efetivamente analisados. O estudo identificou diferenças na interação entre díades mãe-criança: prematura e a termo nas fases pré-escolar, escolar e adolescente, e apontou que comportamentos interativos maternos exerceram efeito moderador diferencial no desenvolvimento de crianças vulneráveis com antecedentes de risco biológico.

Ao resenhar o livro de Goodheart, Kazdin, e Sternberg (2006), Witter (2007) esclarece que os cuidados com a saúde vêm se baseando nas evidências resultantes de pesquisas e a Psicoterapia também vem recebendo o mesmo tipo de atenção. Complementa que “Há um forte apoio científico para que as políticas de

saúde e os governos adotem na saúde uma prática baseada em evidências” (p. 523). Atualmente, espera-se que a Psicoterapia se desenvolva mediante a aplicação de resultados de pesquisas com o “máximo de controle possível, o que certamente pede avaliações sistemáticas...” (p. 524). Witter entende que é necessário eliminar a separação entre a pesquisa e a prática clínica, pois se a prática não consome dados da pesquisa, o atendimento pode ser prejudicado. Para uma prática cientificamente informada, o especialista em Psicoterapia precisa:

(a) conhecer a literatura científica de psicologia, dos processos e intervenções psicoterápicas e sua aplicabilidade a casos específicos; (b) incorporar os sólidos princípios terapêuticos da teoria e da pesquisa; (c) ter amplas e profundas habilidades terapêuticas; (d) um esforço terapêutico genuíno para compreender o problema da pessoa, o contexto e as circunstâncias; (e) consultar outros especialistas e pares quando necessário; (f) ser crítico quanto ao processo, as reações do paciente, progresso e resultados da psicoterapia e (g) ver a aprendizagem como um processo ao longo da vida. (Witter, 2007, p. 523)

O primeiro livro publicado no Brasil com o título *Psicologia Baseada em Evidências* foi organizado por Melnik e Atallah (2011), sendo o seu subtítulo *Provas Científicas da Efetividade da Psicoterapia*. A obra reuniu 11 temas de extrema relevância para os estudos de Psicologia e nos seus 22 capítulos apresentou um panorama das evidências atuais em relação à efetividade da Psicoterapia no tratamento dos distúrbios psiquiátricos. Segundo os organizadores, na apresentação do livro, cada capítulo apresenta uma revisão sistemática publicada na Cochrane Library sobre a efetividade da Psicoterapia no tratamento dos transtornos mentais comentado por um ou mais especialistas na área. Melnik e Atallah indicam a necessidade que as outras áreas da Psicologia entendam que utilizar o conhecimento na tomada de decisão é de suma importância para uma boa prática profissional, assim como para a gestão das políticas públicas.

A tese de doutorado de Melo Junior (2011), defendida na Faculdade de Ciências Médicas da Universidade Estadual de Campinas (Unicamp), apresenta uma revisão sistemática de literatura recente e atualizada sobre prevalência de Depressão Pré-Natal e DPP, com aplicação de metanálise. Os artigos foram recuperados nas bases de dados MedLine, EMBASE, WOK, CINAHL e SciELO, sem restrição de idioma, limitando o período de publicação dos artigos de 2000 a junho de 2010. Não foram consultadas as bases de dados específicas da Psicologia brasileira, hospedadas na Biblioteca Virtual de Psicologia e nem a Literatura Latino-

Americana em Ciências da Saúde (LILACS), o que pode ter gerado viés nos resultados pela ausência de importantes estudos publicados nos países latino-americanos cujas revistas não estão indexadas nas bases de dados consultadas. No total, foram localizados 1.265 artigos no PubMed, 660 no EMBASE, 38 no Scielo, 1.896 na WOS e 50 no CINAHL. Eliminando-se as sobreposições e os artigos que, pelo título, foram considerados fora dos critérios de inclusão, restaram 487 estudos. Com base nos resumos, foram selecionados 169. Outros 20 trabalhos foram levantados nas referências dos artigos analisados. O conjunto dos 189 artigos compôs a amostra da revisão para sistematização da prevalência de depressão pós-parto, mas não foram fornecidas as informações detalhadas sobre o método de avaliação de qualidade dos estudos, o que deixa uma lacuna na revisão sistemática.

A tese de Silva (2013), da qual participamos como segunda revisora independente nas etapas de definição das bases de dados para consulta, elaboração das estratégias de busca, seleção dos estudos para inclusão e análise da qualidade metodológica, usando-se técnicas estatísticas para determinar o grau de concordância entre as duas avaliadoras, teve como objetivo geral “sintetizar dados de prevalência de DPP por meio de revisão sistemática da literatura seguida de metanálise e identificar as variáveis regionais e metodológicas que contribuíram para as diferenças de resultados entre estudos” (p. 62). O trabalho não conseguiu estabelecer a prevalência de DPP a partir dos trabalhos sistematizados em decorrência da baixa qualidade metodológica dos estudos revisados. A autora apresenta como um dos itens da conclusão que a revisão sistemática e a metanálise, que são ferramentas da Psicologia Baseada em Evidências, trouxeram novas informações e conclusões a respeito dos dados de prevalências de DPP, que provavelmente, não seriam obtidas através de uma revisão narrativa comum (Silva, 2013). Além de estabelecer comparações entre os dados de prevalência de países desenvolvidos e em desenvolvimento, a discussão sobre o tema da PBE, introduzida no trabalho, é um importante elemento para a ampliação do uso de revisão sistemática e metanálise na Psicologia. O roteiro, detalhadamente estruturado e testado para a elaboração da revisão sistemática apresentado na tese, é um interessante manual que pode ser destacado e utilizado como guia para o desenvolvimento de trabalhos dessa natureza.

A Psicologia ainda não construiu sua pirâmide de evidências, ou seja, não graduou a importância de seus estudos, mas já começa a olhar para esse horizonte

como promissor e facilitador na ordenação da conduta psicológica, como vimos nos estudos apresentados. Os exemplos aqui apontados servem para embasar a discussão de que há um grande campo para a PBE. É urgente iniciarmos ações que busquem sistematizar o conhecimento gerado no país e na região em prol de uma Psicologia mais econômica e sustentável, que não repita esforços já desenvolvidos, cujos resultados já foram publicados e estão disponíveis para consulta nas fontes de informação da área. A urgência de soluções, característica da era atual, e o volume de informação disponível nas bases de dados justificam uma nova modalidade de uso e aplicação do conhecimento.

A atuação do psicólogo e do psicanalista, juntamente com os médicos, exige da Psicologia a aplicação de práticas que dialoguem diretamente com a área da saúde. Dessa forma, buscar as melhores evidências, sistematizá-las, publicar os resultados, para que outros profissionais possam delas se valer na solução de problemas psicológicos, aparece como um novo campo do qual o psicólogo deve se apoderar. Destacamos o contexto da América Latina, pois nos países desenvolvidos, como, por exemplo, os Estados Unidos da América, a Psicologia Baseada em Evidências apresenta destaque enquanto área de atuação profissional e lá os profissionais já estão pondo em prática essa técnica (Levant, 2005).

Para que os psicólogos sejam competentes em avaliação, com amplo domínio de metodologia científica para uma interface produtiva entre ciência e prática, é necessário que a formação inclua competências para lidar com problemas de amostragem, observação, registro, interpretação de dados e lógica, escolhas e análises estatísticas e, também, no pensamento crítico (Witter, 2007). Além da habilidade para desenvolver pesquisas, os alunos e os profissionais devem se especializar em redação científica de qualidade. Registrar o caminho da pesquisa, com a apresentação formal de todos os passos, dando ênfase na descrição dos resultados, é uma competência que vai além dos itens citados por Witter e que dará à Psicologia brasileira e latino-americana o sonhado *status* de ciência consolidada.

Transformar o conhecimento científico em evidência não é tarefa fácil e requer o aperfeiçoamento da capacidade nacional e regional, tratando-se de América Latina, em produzir e utilizar a ciência já produzida e publicada em revistas e outros tipos de publicações científicas. Um bom início pode ser a conscientização da comunidade acadêmica de que o conhecimento por ela gerado pode ser utilizado por um colega, na solução de um problema de saúde, ou por um gestor, na definição

de políticas públicas de saúde para a população. A continuidade está vinculada a construir uma nova consciência quanto ao uso do conhecimento, que deve ser voltado à racionalização. A conclusão é que é necessário gerar conhecimento de qualidade e de fácil compreensão, para isso devemos avançar rumo às publicações com mais consistência e profissionalismo, deixando de lado o discurso científico rebuscado, de difícil entendimento e com ares de erudito. Para que possa ser utilizado deve ser transformado em publicações de fácil leitura e compreensão, que permitam ao leitor conhecer o que foi feito, como foi feito e, principalmente, quais foram os resultados obtidos. Revisitar o saber, selecionar de acordo com critérios claros e definidos *a priori*, analisar, classificar e publicar em veículos de amplo acesso são práticas que precisam emergir urgentemente na comunidade científica latino-americana. O volume de informação disponível aponta para a urgência em racionalizar vários artigos em revisões que sintetizem a massa do conhecimento publicado.

É fundamental que a tomada de decisão, tanto na prática psicológica quanto na definição de políticas públicas, seja norteadas pelo juízo ético, rigor e cuidado do profissional e do gestor na busca, seleção e aplicação das evidências. Para que os tomadores de decisão possam se valer da evidência, é necessário aperfeiçoar a capacidade nacional e regional, tratando-se de América Latina, em produzir e avaliar boas evidências. A discussão suscita uma pergunta importante que aparece como um subitem a seguir.

2.6 Artigos publicados em revistas latino-americanas contêm boas evidências?

A pergunta anterior inquieta autores e editores no Brasil e em outros países latino-americanos. Atitudes preconceituosas são declaradas por editores que debatem a questão da publicação de países ricos e pobres em editoriais e artigos publicados.

A discussão sobre a situação das publicações em países desenvolvidos e em desenvolvimento não é recente. Gibbs (1995) escreveu um artigo sobre a ciência perdida nos países em desenvolvimento pelo fato de que as revistas publicadas nesses países não estarem disponíveis nas bases de dados. O autor explicava que era vital que os países em desenvolvimento pudessem comunicar os resultados de pesquisas com os países ricos, mas demonstrava que muitos editores ainda

acreditavam que a baixa representação das revistas dos países pobres era devida à má qualidade da ciência gerada nessas nações e por isso os manuscritos dos autores desses países eram rejeitados. No ano de 2012, a questão da visibilidade das revistas de Psicologia publicadas na América Latina parecia estar resolvida, como demonstra o número de publicações indexadas nas principais bases de dados internacionais (Sampaio & Sabadini, 2012).

Um artigo publicado pelo editor da *British Journal of Psychiatry* trata do problema: uma forma de racismo em nossa prática editorial, e é importante reconhecer que existem muitas outras, é encontrada em nossas atitudes em relação aos artigos recebidos de países com menor renda média. Estes artigos tendem a ser rejeitados pelas revistas mais bem classificadas. Tais atitudes podem ser por motivos racistas, mas a maioria dos editores não tem consciência disso e fica chocado com tal alegação. Então utilizam disfarces eficazes para a rejeição dos manuscritos, como, por exemplo, a questão do rigor metodológico ou o estilo de escrita da revista (Tyrer, 2005).

No Brasil, Victora e Moreira (2006) usaram a expressão “viés de publicação contra o Sul” fazendo referência à desproporção entre a produção científica de países de renda alta (Norte) e os de média/baixa (Sul) para discutir o preconceito editorial nas publicações. Os autores fizeram uma enquete, perguntando para 351 primeiros autores de artigos publicados na *Revista de Saúde Pública*, *Cadernos de Saúde Pública* e *Revista Brasileira de Epidemiologia* se acreditavam haver preconceito dos editores de revistas produzidas em países ricos contra artigos submetidos por autores de países menos desenvolvidos, como o Brasil. Dos autores que responderam, 25% concordaram totalmente, 50% concordaram parcialmente e apenas 8% discordaram.

Na vida acadêmica, poucas coisas são mais frustrantes do que receber uma carta com a expressão: “pensamos que seu artigo seria mais adequado para uma publicação local ou regional”. Ou então, “recebemos um número de manuscritos muito superior ao que podemos publicar, e seu artigo não recebeu prioridade suficiente para ser incluído... O autor muitas vezes se sente como mostra a Figura 3.” (Victora & Moreira, 2006, p. 37)

Figura 3. Um autor latino-americano diante de editor internacional?. Reproduzida de “Publicações científicas e as relações Norte-Sul: racismo editorial?”, de C. G. Victora e C. B. Moreira, 2006, *Revista de Saúde Pública*, 40, p. 38. Recuperado de <http://dx.doi.org/10.1590/S0034-89102006000400006>

Victora e Moreira (2006) apontam três itens que precisam ser considerados pelos autores dos países do Sul que, certamente, serão notados pelos editores dos países do Norte: 1) qualidade da redação: com ênfase na redação em inglês para não nativos da língua; 2) falta de empatia com o leitor: temáticas típicas de pesquisas locais que não têm alcance internacional e que, muitas vezes, possuem questões específicas que demandam explicações detalhadas; 3) forma do artigo: muitas vezes o manuscrito não atende às normas editoriais da revista, são muito longos, há falta de cuidado na descrição de cada uma das seções. O editor sempre pergunta se vale a pena enviar um trabalho para revisão com grandes chances de ser reprovado.

Atualmente, na maior parte das revistas internacionais de qualidade, 90% ou mais dos artigos são rejeitados imediatamente sem revisão externa. Os editores tomam essa decisão rapidamente, sem ler todo o artigo. Os critérios utilizados incluem a aparente qualidade do artigo (em parte baseada na redação e forma, mas também na metodologia e sua extensão), seu presumido interesse para os leitores da revista, e sem dúvida a autoria; autores conhecidos ou filiados a instituições de renome indiscutivelmente apresentam maiores chances de passar pelo crivo inicial... (Victora & Moreira, 2006, p. 40)

Se a questão da forma nos artigos científicos está solucionada, ou seja, os aspectos normativos são um desafio superado, o conteúdo ainda gera muito embate e clama por métodos de avaliação que definam sua qualidade (Trzesniak et al., 2012). O mote da questão é: resultados de pesquisas publicados em artigos são considerados as melhores evidências científicas que existem e, aliados à prática profissional e às teorias reconhecidas, são elementos de extrema importância na tomada de decisão? Recorrer à melhor evidência, ou seja, ao melhor resultado de pesquisas publicado e disponível é o que o profissional deve fazer para apoiar a sua tomada de decisão, para tanto, é necessário conhecer como avaliar o conteúdo desses documentos.

Claro que, considerando o volume do conhecimento publicado e disponível atualmente, é humanamente impossível avaliar a produção de uma área sem fazer um recorte estreito dos seus trabalhos. No nosso caso, o recorte foi feito nos estudos sobre prevalência de DPP.

2.7 Por quê Prevalência de DPP?

No capítulo que apresenta este trabalho, explicamos que participar como segunda revisora de uma revisão sistemática foi uma oportunidade para entrar em uma área nova de estudo, inquietante e desafiadora. Assim, não repetiremos as justificativas pela opção do estudo com o tema prevalência de DPP. Entretanto, desmembraremos cada um dos elementos desse tema para situar nosso estudo.

2.7.1 Estudos de Prevalência

A prevalência quantifica uma dada população que atende a uma condição particular, por exemplo, mulheres com depressão pós-parto em um dado ano, foco deste estudo. Loney e Stratford (1999) definem prevalência como um número de pessoas em uma determinada população que tem uma doença específica ou uma condição momentânea. Explicam ainda que, na verdade, prevalência é o número de pessoas com um problema de saúde, em um determinado período, dividido pelo total da população num mesmo momento e local. Produzem um retrato da situação de uma comunidade, com base em avaliações individuais do estado de saúde de cada um dos membros de um dado grupo. Estudos de prevalência são estudos

transversais que medem a prevalência de uma característica, como o percentual de mulheres deprimidas em uma determinada população de parturientes.⁵

Reis et al. (2002, p. 54) apresenta a fórmula para se encontrar prevalência:

“Prevalência

Número de pessoas que têm a doença num determinado ponto do tempo

Número de pessoas que estão expostas ao risco num determinado ponto do tempo”.

Um estudo transversal é um estudo descritivo, no qual estados de doenças e exposição são medidos simultaneamente em uma determinada população para se obter um “instantâneo” da frequência e características de uma doença. Os dados coletados como parte de um estudo transversal, podem ser usados para avaliar a prevalência de condições agudas ou crônicas da população (Boyle, 1998). Os estudos transversais têm vantagens em relação a outros métodos de pesquisa pois não requerem acompanhamento, logo são menos custosos e podem ser representativos de uma população. Ainda que com custos mais baixos, o desenho desse tipo de estudo não é o mais indicado para pesquisas de prevalência de DPP, como demonstrou Silva (2013).

Estudos de prevalência podem se estender a uma situação restrita, por exemplo, pacientes de uma determinada clínica, enfoque restrito, ou para uma nação inteira, enfoque amplo. Nos últimos anos, a preocupação com as necessidades em saúde mental em contextos clínicos tem promovido a ampliação dos estudos de populações inteiras, ou seja, cidades, países. Essa perspectiva da saúde tem estimulado o desenvolvimento de diversos estudos de prevalência de transtornos mentais que se destinam ao planejamento de políticas públicas, programas de saúde, avaliação e tratamentos com melhor alocação de recursos. Nessa questão, as diretrizes para avaliar a qualidade dos estudos de prevalência surgem para desenvolver critérios de análise e validação dos inquéritos aplicados à comunidade (Boyle, 1998).

A validade dos estudos de prevalência é determinada em função da amostragem, medição e resultados. Estudaremos cada um desses critérios separadamente, como recomenda Boyle (1998).

⁵ <http://www.cochrane.org/sites/default/files/uploads/glossary.pdf>

1) *Amostragem*: a validade de um estudo de prevalência exige uma amostra precisa e representativa da população alvo (Loney & Stratford, 1999). Representatividade é uma qualidade associada ao método utilizado para definir a amostra estatística e a avaliação cuidadosa das características dos respondentes. A amostra é um recorte da população alvo, que permite obter informações sobre essa totalidade. Algumas características dessa população, que podem defini-la com precisão, são: idade, sexo, língua, etnia, renda e residência; são características sociodemográficas. Quando uma parte da população é excluída do estudo – por dificuldade com a língua ou residência em áreas remotas –, os motivos devem ser informados no relato de pesquisa. A amostra com maior qualidade para estudos de prevalência é a aleatorização ou probabilística, pois garante que cada respondente elegível tenha a mesma chance de ser escolhido (Boyle, 1998). O tamanho da amostra também deve ser planejado de acordo com características desejadas para o estudo, tais como: nível de confiança, margem de erro e estimativa de prevalência na população estudada. A técnica de amostragem (censo, amostra aleatória, não aleatória), o tamanho da amostra, os critérios de inclusão/exclusão dos participantes, a taxa de perdas, a apresentação das características das participantes perdidas e do perfil sociodemográfico da população estudada, são fatores essenciais na definição da metodologia de pesquisa e devem ser relatados no informe científico.

2) *Medição*: um estudo certificado utiliza instrumentos que fornecem medições confiáveis e válidas. Artigos com o rigor científico desejável em estudos de prevalência também devem primar pelo instrumento de coleta de dados, que deve ser confiável e validado para a população a ser estudada. A utilização de um instrumento cuja validade e confiabilidade já foram testadas em outros estudos é sempre preferível à criação de um novo recurso para medição (Boyle, 1998). A qualidade de estudos de prevalência é potencializada pelo uso de métodos padronizados de coleta de dados, que são confirmados empiricamente por estudos de avaliação de medição. Normalização do método refere-se não só à obtenção de informações dos entrevistados, mas também à formação do entrevistador, à supervisão, ao uso de dois avaliadores para evitar, ao máximo, o viés da subjetividade. O tipo de instrumento para avaliação de DPP (autoavaliação/entrevista clínica), a apresentação das medidas de validade e de confiabilidade do instrumento e o cuidado para que o desfecho de saúde seja

medido de forma não enviesada garantem a boa qualidade da medição dos estudos (Boyle, 1998; Giannakopoulos, Rammelsberg, Eberhard, & Schmitter, 2012; Loney & Stratford, 1999).

3) *Resultados*: o principal objetivo dos estudos de prevalência é produzir estimativas da frequência de transtornos. Essas estimativas devem ter abrangência global e serem detalhadas em subgrupos populacionais. Assim, estudos de maior qualidade devem determinar se as taxas de respostas foram adequadas e detalhar as análises por subgrupos. Estimativas em estudos de prevalência devem ser acompanhadas por intervalos de confiança ou a informação necessária para calculá-los. Um intervalo de confiança para avaliar o desfecho e a descrição das participantes elegíveis que se recusaram a participar, assim como, as análises de sensibilidade são ferramentas indispensáveis para avaliarmos o quanto os resultados encontrados nos estudos são sensíveis à metodologia utilizada (Boyle, 1998; Giannakopoulos et al., 2012; Loney & Stratford, 1999).

Boyle (1998) esclarece que as seguintes perguntas devem ter respostas positivas, quando indagadas aos relatos de estudos de prevalência.

Amostragem

O questionário aborda uma amostra representativa dos participantes da população alvo?

A população alvo está definida claramente?

Amostragem probabilística foi utilizada para identificar potenciais participantes?

A característica dos participantes se enquadra na população alvo?

Medição

O questionário obtém medidas confiáveis e válidas sobre problemas psiquiátricos e outros conceitos-chave?

Os métodos de coleta de dados são padronizados?

Os instrumentos de pesquisa/questionamento são confiáveis?

Os instrumentos de pesquisa/questionamento são válidos?

Resultados

Características especiais na definição da amostragem foram consideradas na análise?

Os relatórios incluem intervalos de confiança nas estimativas estatísticas?⁶

Diretrizes básicas para avaliação da qualidade de estudos de prevalência podem auxiliar no julgamento informado sobre a validade do artigo científico, resultante da pesquisa, além de demonstrar preocupação com a área. Localizamos diretrizes, rigorosamente elaboradas, para a avaliação crítica de artigos sobre prevalência ou incidência de transtornos da saúde, o que sugere uma preocupação e dedicação ao assunto. As fundamentações das grandezas identificadas nessas diretrizes embasaram nossa discussão anterior (Boyle, 1998; Giannakopoulos et al., 2012; Loney, Chambers, Bennett, Roberts, & Stratford, 1998; Loney & Stratford, 1999).

Considerando que as diretrizes para avaliação de relatos sobre prevalência podem pontuar o grau de qualidade do conteúdo desse tipo de estudos, utilizar um determinado transtorno, como a DPP, para fazer um recorte do universo das prevalências é um caminho interessante e oportuno, para a aplicação de tais diretrizes. Vamos, então, introduzir a DPP no nosso trabalho.

2.8 Depressão Pós-Parto: um grave transtorno do humor

A depressão tem recebido atenção dos profissionais da saúde, pois vem se configurando como a “doença do século XXI, dada a sua alta prevalência e potencial para prejudicar a capacidade de funcionamento dos indivíduos, bem como sua relação com outras doenças” (Silva, 2013, p. 36). Em um trabalho que revisou as diretrizes para a tomada de decisão com base em evidências sobre depressão, Fleck et al. (2009) esclarecem que a prevalência da depressão em países ocidentais varia de 3 a 11% na população em geral. Em 1996, a Organização Mundial de Saúde (OMS) estimou que a depressão foi a quarta maior causa de mortalidade no ano de 1990 e que, no ano 2020, poderá ser a segunda maior causa de mortalidade no mundo e a primeira em países em desenvolvimento (Murray & Lopez, 1997).

⁶ Tradução livre de Boyle (1998).

As classificações mais utilizadas para pontuar a depressão são: a *Classificação Internacional de Doenças* (CID 10) e o *Manual Diagnóstico e Estatístico de Transtornos Mentais*, da *American Psychiatry Association* (DSM IV). Está classificada como um *Transtorno Afetivo*, que são aqueles que apresentam uma perturbação fundamental, alterando o humor ou o afeto, como uma depressão ou uma euforia (CID 10). Essa alteração do humor é, geralmente, acompanhada de modificação no nível global de atividades e a maioria dos episódios do transtorno tende a ser recorrente e pode estar relacionada com situações ou fatos estressantes. O DSM IV classifica a *depressão* dentro dos *Transtornos do Humor* e também baseia a classificação nos episódios depressivos.

O termo depressão pós-parto é utilizado para designar qualquer episódio depressivo que transcorra nos meses seguintes ao nascimento do bebê, com variações entre três, seis meses e até um ano. A DPP é caracterizada por uma série de sintomas, tais como o humor deprimido persistente, a diminuição do prazer e menor interesse nas atividades habituais, distúrbios do apetite e do sono, a função psicomotora fica comprometida e há perda do nível de energia, da concentração e da auto-estima (Cantilino et al., 2010). O DSM IV caracteriza a depressão pós-parto como um transtorno que afeta a mulher no período que vai de duas semanas até os 12 meses após o parto. A mulher deve experimentar quatro sintomas adicionais de uma lista extensa, que inclui: alterações no apetite ou peso, sono e atividade psicomotora; diminuição da energia; sentimentos de desvalorização ou culpa; dificuldades em pensar, concentrar-se ou tomar decisões, ou pensamentos recorrentes sobre morte, ideação suicida, planos ou tentativas de suicídio. Deve haver, ainda, um prejuízo clinicamente significativo no funcionamento social, profissional ou de outras áreas importantes da vida da mãe (Gonçalves, 2010).

A depressão pós-parto é uma patologia que, muitas vezes, não é diagnosticada. Provavelmente,

porque as mulheres com depressão são relutantes em discutir os seus sintomas com o seu médico assistente, pediatra ou obstetra. As razões incluem o receio de serem vistas como más mães, medo de serem vistas como “malucas” e o estigma associado com o facto de terem um problema psiquiátrico. (Gonçalves, 2010, p. 21)

O tratamento da DPP, geralmente, apresenta bons resultados, quando se tem o diagnóstico e procedimentos clínicos precoces. Se não for tratada em tempo hábil,

pode evoluir para a cronicidade e a baixa adesão ao tratamento, podendo se estender para o segundo/terceiro ano pós-parto e causando enfraquecimento significativo, morbidade e, às vezes, mortalidade da mulher. Uma providência importante é o rastreio da DPP entre a quarta e sexta semanas do pós-parto, objetivando detectar sintomas positivos da patologia. Nessas avaliações, “o rastreio sistemático para depressão pós-parto usando questionários, ajuda a melhorar a identificação dos sintomas e a acelerar o tratamento” (Gonçalves, 2010, p. 21).

A breve apresentação da DPP não teve a intenção de explicar com detalhes a questão, mas, sim, introduzir o tema para justificar a importância em se avaliar estudos de prevalência dessa patologia que pode causar sérios problemas de saúde à população.

Sendo a DPP um grave problema na população mundial, como demonstrado por Silva (2013), e a prevalência um dos elementos na definição de políticas públicas de saúde, analisar a qualidade da apresentação de estudos dessa natureza incluídas em uma revisão sistemática, considerada padrão-ouro de evidência científica, pode ser um bom caminho para demonstrar a relevância da Psicologia Baseada em Evidências na tomada de decisão de gestores públicos.

Visando contribuir para a ampliação dessa discussão, nosso estudo busca avaliar a qualidade de artigos sobre prevalência de DPP incluídos em uma revisão sistemática com metanálise, comparando a origem do estudo e a procedência das revistas latino-americanas e de outras regiões, como forma de apresentar uma nova opção de avaliação de artigos e expandir e fortalecer a PBE em nossa região.

3 OBJETIVOS E HIPÓTESES

Neste capítulo apresentamos as proposições estabelecidas para o nosso estudo.

3.1 Objetivo geral

Comparar a qualidade metodológica dos estudos e da apresentação dos resultados em artigos científicos latino-americanos e de outras regiões incluídos em revisão sistemática sobre prevalência de DPP.

3.2 Objetivos específicos

- 1) Avaliar a qualidade metodológica e da apresentação dos resultados de um universo de artigos incluídos em uma revisão sistemática sobre prevalência de DPP, de acordo com a Escala de Loney et al. (1998).
- 2) Comparar a qualidade metodológica e da apresentação dos resultados dos artigos que tiveram a coleta de dados em países latino-americanos e em outras regiões.
- 3) Comparar a qualidade metodológica e da apresentação dos resultados dos artigos publicados em revistas latino-americanas e de outras regiões.
- 4) Analisar se artigos que declararam ter como objetivo avaliar a prevalência de DPP tiveram qualidade diferente dos que abordaram o tema e informaram ter outro objetivo.
- 5) Analisar a qualidade dos títulos, resumos e palavras-chave de acordo com o manual de publicações da American Psychological Association (2012) e com dois vocabulários controlados da área de saúde.

3.3 Hipóteses

H1: Artigos científicos que tiveram a coleta de dados em países latino-americanos e de outras regiões possuem a mesma qualidade, de acordo com a Escala de Loney et al. (1998).

H2: Artigos científicos publicados em revistas latino-americanas e de outras regiões possuem a mesma qualidade, de acordo com a Escala de Loney et al. (1998).

H3: Artigos que declararam ter como objetivo avaliar a prevalência de DPP têm a mesma qualidade, em relação aos que não informaram ter esse objetivo.

4 MÉTODO

O método de pesquisa utilizado foi de estudo transversal de análise crítica da literatura. Pode ser chamado, também, de delineamento de pesquisa documental. Pesquisa documental “É aquela cujos objetivos ou hipóteses podem ser verificados através da análise de documentos bibliográficos, ou não bibliográficos, requerendo metodologia (coleta, organização, análise de dados) compatíveis com os mesmos.” (Witter, 1990, p.22). Neste tipo de pesquisa os documentos são denominados unidades de análise, sejam eles primários ou secundários. Assim, este é um estudo secundário, cujo universo são artigos científicos publicados em revistas científicas e que foram incluídos em revisão sistemática sobre DPP. Ao conjunto de todos os artigos que estimaram a prevalência de DPP em um dado grupo de participantes, chamaremos de população. O objeto de nossa pesquisa é o conjunto dos artigos latino-americanos e a amostra aleatória pareada de artigos incluídos na mesma revisão sistemática.

Para chegarmos aos artigos objetos do presente estudo utilizamos dados obtidos pelo trabalho de Silva (2013), do qual participamos como segundo revisora. Na avaliação de qualidade dos trabalhos participantes de uma revisão sistemática a presença de um segundo revisor é fundamental para diminuir a possibilidade de viés na análise dos estudos. Ward, Hall e Schramm (1975), criadores da Escala ARRR para avaliação de artigos de pesquisa enfatizam essa necessidade nas conclusões do seu trabalho, explicando que dois avaliadores podem tornar o critério de excelência dos artigos mais explícito. A primeira tese de doutorado produto da revisão sistemática, relatando os resultados relativos à prevalência de DPP, foi defendida em março de 2013, por Gabriela Andrade da Silva, estudiosa do tema. Nossa tese explorou os aspectos qualitativos dos artigos latino-americanos e comparou com uma amostra aleatória de não latino-americanos. Assim, iniciamos este capítulo relatando as etapas da revisão sistemática, mediante nossa participação, para finalizar mostrando o caminho que nos levou aos resultados.

4.1 Etapas da revisão sistemática

Analisar os artigos incluídos na revisão sistemática foi uma oportunidade de testar a qualidade metodológica e da apresentação dos resultados de artigos sistematicamente revisados. A revisão sistemática é uma metodologia de estudo científico que possui diversas estratégias para a prevenção de vieses. O processo sistemático é planejado, a priori, com critérios estabelecidos para cada etapa. A seguir, as etapas da revisão sistemáticas são apresentadas.

4.1.1 Formulação da pergunta da revisão sistemática

Como em toda pesquisa científica, a definição da pergunta de pesquisa foi um momento crítico da revisão sistemática, pois a partir dessa pergunta os objetivos foram traçados e os resultados obtidos. Seleccionados os elementos importantes para o delineamento da pesquisa, e com base neles, foi elaborada a seguinte pergunta: *É possível sintetizar em uma revisão sistemática, com metanálise, dados de prevalência de DPP, e comparar as variáveis regionais e metodológicas, que contribuíram para as diferenças de resultados entre os estudos?*

4.1.2 Critérios de inclusão dos trabalhos

Como recomendado pelo manual da Cochrane (Higgins & Green, 2006) todos os tipos de trabalhos, considerados científicos e que se enquadrem nos critérios, devem ser incluídos na revisão sistemática. Nove critérios nortearam as avaliadoras no processo de análise dos estudos:

- **Critério A:** O estudo é primário e quantitativo, com delineamento transversal ou de coorte longitudinal.
- **Critério B:** Foi avaliada a prevalência de depressão pós-parto na amostra estudada.

- **Critério C:** Os participantes do estudo são mães humanas.
- **Critério D:** A amostra estudada não foi submetida a nenhuma intervenção com o objetivo de prevenir DPP.
- **Critério E:** O método de diagnóstico é mencionado claramente e baseia-se em instrumento objetivo de medida, ou em avaliação feita por profissional de saúde mental.
- **Critério F:** O período do puerpério, em que o diagnóstico foi feito, está claramente mencionado. Pelo menos uma das avaliações de DPP ocorreu de 8 dias (inclusive) a 365 dias (inclusive) após o parto.
- **Critério G:** O estudo foi realizado na população geral e não declara incluir população exclusivamente com características específicas, como: doenças crônicas (ex.: diabetes, depressão, hipertensão), outras doenças graves (ex.: câncer), complicações na gestação (ex.: eclampsia), bebês com condições de saúde alteradas (ex.: má formação fetal, baixo peso, prematuridade), homossexualidade, adoção, uso de drogas na gestação (álcool, drogas ilícitas).
- **Critério H:** Em pelo menos uma das avaliações para identificar DPP, o resultado de prevalência está relatado de forma clara, contendo todos os seguintes dados: período após o parto, instrumento, tamanho de amostra e número absoluto ou porcentagem de participantes consideradas deprimidas.
- **Critério I:** O texto completo do estudo foi publicado em inglês, português ou espanhol.

Os critérios foram definidos considerando as recomendações das diretrizes para estudos de prevalência e as características da DPP, de acordo com as orientações

da *American Psychiatric Association* (1994) e da literatura especializada. O próximo passo foi a seleção das fontes de informação para busca dos estudos.

4.1.3 Bases de dados selecionadas

Uma recomendação muito importante, tratando-se de revisão sistemática, é que as fontes de informação consultadas sejam as mais abrangentes e representativas da área, e que a consulta permita esgotar o maior número possível de bases de dados. Assim, a busca contemplou 14 fontes de informação eletrônicas que indexam a literatura da área. Em Psicologia conhecemos três fontes de informação especializada.

- BVS-PSI ULAPSI – Reúne informações sobre periódicos, teses, livros, filmes e outros materiais publicados na área de Psicologia, tem como objetivo dar visibilidade ao conhecimento científico gerado nos países da América Latina. Contempla o Periódicos Eletrônicos em Psicologia - PePsic.
- Psycodoc – Desenvolvida pelo Colégio de Psicólogos de Madrid, inclui periódicos, trabalhos publicados em congressos e livros na área, editados na Espanha, Portugal e América Latina, desde 1975 até a atualidade.
- PsycInfo (Psychological Information), criada e mantida pela APA, reúne, organiza e divulga a literatura relevante publicada internacionalmente na área da Psicologia e disciplinas correlatas, desde 1989. É considerada a mais importante base de dados na área da Psicologia.

Lembrando o critério de exaustão exigido para revisão sistemática, outras bases de dados, não específicas, que possuem material de interesse para a Psicologia também foram consultadas:

- Biblioteca Virtual de Dissertações e Teses do Instituto Brasileiro de Informação Técnico-Científica (IBICT).
- CINAHL – Criada em 1981, a base de dados da EBSCO contém mais de 1.000.000 de registros e oferece cobertura completa aos periódicos de enfermagem, em inglês, e publicações da National League for Nursing e da American Nurses' Association. Abrange a enfermagem, a biomedicina, a biblioteconomia, a medicina alternativa/complementar, a saúde do consumidor e 17 disciplinas ligadas à saúde.

- Lilacs (Literatura Latino-Americana e do Caribe em Ciências da Saúde) – Base de dados cooperativa do Sistema BIREME. Compreende revistas científicas, teses, livros, capítulos de livros, anais de congressos ou conferências, relatórios técnico-científicos e publicações governamentais em Ciências da Saúde, publicadas nos países da região, a partir de 1982.
- EMBASE – ou *Excerpta Medica Database*, é uma base de dados produzida pela Elsevier, que indexa a literatura na área de biomedicina e farmacologia.
- ERIC – Considerada a maior fonte de informação existente na área de educação, reúne informações de artigos de periódicos científicos de prática e de pesquisa. É produzida pelo *Education Resources Information Center*.
- Proquest – Fonte de informação de acesso restrito que indexa diversos tipos de produtos científicos, como revistas, teses e outros documentos.
- PubMed – é uma das mais importantes bases de dados desenvolvidas no mundo. Disponível gratuitamente na *web*, indexa periódicos e outros materiais na área de ciências da saúde.
- RedAlyc - Revistas de diversas áreas do conhecimento em texto completo disponíveis em acesso aberto.
- SciELO - Portal de revistas científicas de todas as áreas do conhecimento que publica artigos científicos em texto completo.
- SCOPUS – Base de dados multidisciplinar, produzida pela editora Elsevier desde 2004, com cobertura aos periódicos desde 1960. Indexa artigos de diversas áreas.
- WOK – Uma das mais importantes ferramentas para recuperação da informação. Reúne diversas bases de dados, inclusive a Web of Science (WOS).

Após a definição das bases de dados, foram desenvolvidas estratégias de busca bem delineadas, muitas vezes, com a ajuda de terminologias específicas da área.

4.1.4 Definição da estratégia de busca

Um elemento que interfere diretamente na recuperação dos trabalhos é a estratégia de busca. As fontes de informação possuem particularidades que demandam a elaboração de estratégias diferentes e específicas. O uso dos

operadores booleanos e dos elementos de truncamento são importantes e podem facilitar a recuperação dos trabalhos para composição da amostra geral da revisão. Assim, foram elaboradas diferentes estratégias de busca que contemplavam esses elementos e outras particularidades de cada fonte de informação, como, por exemplo, o idioma principal e a presença, ou ausência, de vocabulário controlado.

4.1.5 Confronto de duplicidades

Após a recuperação dos trabalhos candidatos à inclusão na revisão sistemática, foi necessário excluir as duplicações, pois um mesmo registro aparecia em várias bases de dados. Para a verificação da duplicidade foi usada a ferramenta *Find duplicates* do software Endnote®Web, ferramenta associada à WOK, que tem um mecanismo de identificação de duplicatas pelo título do trabalho, nome do primeiro autor, ano e periódico de publicação. No entanto, como não há padrão para armazenamento dessas informações nas bases de dados, foi necessário eliminar diversas duplicatas manualmente. Um dos principais motivos foi a diferença no formato do nome do primeiro autor, pois enquanto algumas bases usam formato abreviado (ex.: Silva, G. A), outras usam formato completo (Ex.: Silva, Gabriela Andrade) e outras, ainda, colocam o primeiro nome antes do sobrenome (ex.: Gabriela Andrade da Silva). Essas diferenças impediram que diversas duplicatas fossem identificadas automaticamente. Algumas bases de dados, como as da BVS-Psi ULAPSI, não interagiam com a ferramenta do Endnote®Web e, nesse caso, a entrada dos dados e a exclusão de duplicações também foram feitas de forma manual.

4.1.6 Armazenamento e gerenciamento dos dados dos registros

Excluídas as duplicidades, os registros foram exportados do Endnote®Web em arquivo formato RIS (extensão do próprio Endnote), que permite a edição de textos. Um software especialmente desenvolvido para o estudo, em linguagem Ruby, transformou os metadados dos registros em dados legíveis pelo Microsoft Excel 2010. Essa planilha continha os campos: autor, título, e informações sobre o periódico, dissertação, tese ou evento no qual o trabalho foi publicado. Além dos dados essenciais para identificação dos estudos, a planilha continha três campos

adicionais, nos quais era marcado se o trabalho viria a ser incluído/excluído pelo título, pelo resumo ou pelo texto completo.

4.1.7 Triagem pelo título do trabalho

Após o armazenamento dos dados na planilha, passamos à etapa de triagem pelo título, objetivando, quando possível, eliminar trabalhos que não atendiam aos critérios A a G, estabelecidos na seção 4.1.2. Quando havia dúvidas em relação ao título, o trabalho não era excluído, e ficava reservado para a etapa de avaliação do resumo. Se não havia consenso entre as avaliadoras, os registros seguiam marcados como aprovados para análise na próxima etapa.

4.1.8 Triagem pelo resumo

Os estudos aprovados com base no título, assim como os que não atingiram o consenso entre as avaliadoras, foram, então, submetidos à avaliação com base nos critérios. Novamente, em caso de dúvidas, os trabalhos eram encaminhados à próxima etapa, no caso, a análise do texto na íntegra.

4.1.9 Obtenção dos textos completos

Grande parte dos textos completos foi obtida via internet, pelo Portal de Periódicos da Capes, que possibilita o acesso aos textos completos de mais de 31 mil publicações periódicas, internacionais e nacionais. O SciELO também assegurou a recuperação de diversos artigos e os repositórios de teses e dissertações possibilitaram o acesso a grande parte dos documentos.

Os trabalhos que não puderam ser obtidos pelos portais apontados foram solicitados aos serviços de comutação: Programa de Comutação Bibliográfica (Comut), mantido pelo Instituto Brasileiro de Informação Científica (IBICT) e pelo Serviço de Acesso ao Documento (SCAD), desenvolvido pela BIREME. Esse serviço permite a recuperação e envio de textos que não estão disponíveis em formato digital mediante o pagamento dos textos recebidos. Todos os textos para avaliação foram salvos e compartilhados no serviço web de armazenamento

Dropbox, que funciona com o conceito de “computação em nuvens”, para que as duas avaliadoras tivessem acesso a todos os trabalhos, ao mesmo tempo.

4.1.10 Avaliação do texto completo

Na avaliação do texto completo, os estudos foram novamente analisados relativamente ao preenchimento dos critérios de inclusão, agora de A até O, pois foram levados em conta também os itens referentes à clareza do relato (H) e ao idioma do texto (O). Também foi efetuada uma nova verificação de duplicidades, mas dessa vez em relação ao uso dos mesmos dados em trabalhos diferentes. Estudos com dados duplicados, identificados comparando-se, na planilha, os nomes dos autores, locais e datas das coletas de dados, número de participantes e prevalência encontrada, foram incluídos apenas uma vez na revisão.

Somente os artigos aprovados na análise dos textos completos foram avaliados de acordo com uma tradução e adaptação da Escala de Loney et al. (1998). A opção por essa Escala se deu porque ela avalia pesquisas sobre prevalência, é de fácil aplicação e permite análises quantitativas, por fornecer uma pontuação que pode ser usada para comparar estatisticamente os artigos incluídos na revisão quanto à qualidade. A Escala, apresentada na Tabela 4, é dividida em três seções, a primeira com seis quesitos e a segunda e a terceira com um cada. Cada quesito atendido recebe um ponto, de modo que o trabalho que atende a todos eles totaliza oito pontos.

A Escala de Loney et al. (1998) é utilizada para avaliar qualquer estudo de prevalência ou incidência, demandando adaptações relativas ao tipo de desfecho que está sendo esperado.

Tabela 4 - Escala de Loney et al. (1998)

A. Os métodos do estudo são válidos?

1. O delineamento do estudo e os métodos de amostragem são apropriados para responder à pergunta de pesquisa?
2. A base amostral é adequada?
3. O tamanho da amostra é adequado?
4. São usados critérios objetivos, adequados e padronizados para medir o desfecho de saúde?
5. O desfecho de saúde é medido de uma forma não enviesada?
6. A taxa de resposta é adequada? As pessoas elegíveis que recusaram participar são descritas?

B. Qual é a interpretação dos resultados?

7. As estimativas de prevalência ou incidência são dadas com intervalos de confiança e detalhadas por subgrupo, quando adequado?

C. Qual é a aplicabilidade dos resultados?

8. Os participantes e o contexto são descritos em detalhe e podem ser generalizados para outras situações similares?
-

Nota. Tabela traduzida e adaptada de “Critical appraisal of the health research literature: Prevalence or incidence of a health problem”, de P. L. Loney et al., 1998, *Chronic Diseases in Canada*, 19(4).

Os critérios de avaliação de cada quesito da escala, apresentados a seguir, foram desenvolvidos e aplicados em conjunto pela doutoranda e pela autora da tese-irmã, Gabriela Andrade da Silva, a todos os trabalhos incluídos na revisão sistemática. Um subconjunto destes, com a respectiva pontuação, corresponde às unidades de análise deste estudo.

Item 1: O delineamento do estudo e os métodos de amostragem foram apropriados para responder à pergunta de pesquisa?

Não foi necessário avaliar o delineamento, visto que um dos critérios de seleção excluiu os estudos que não eram adequados aos propósitos da revisão, restando apenas os transversais ou de coorte longitudinal. Assim, a avaliação do item 1 dependeu apenas do tipo de amostra usada no estudo, que poderia ser adequada ou inadequada. Para decidir se esse item seria pontuado a amostra foi, primeiramente, classificada nas categorias a seguir.

- *Censo*: foi considerada censitária a pesquisa que avaliou todas as mães de uma determinada localidade geográfica (município, vila, estado), que deram à luz a bebês vivos no período de tempo estipulado pelo trabalho avaliado. Levamos em consideração o quanto o universo em que o censo foi feito era representativo da população na localidade geográfica. Por exemplo, se o estudo buscou determinar a prevalência de DPP em determinada cidade, todas as mães que deram à luz nessa cidade, em um determinado período, deveriam ter sido entrevistadas para que o estudo se caracterizasse como censo. Uma pesquisa realizada em uma única instalação de saúde, de uma cidade em que existiam ou poderiam existir outras instalações, não foi considerada censitária, mesmo que todas as mães que deram à luz naquela unidade em determinado período tenham sido avaliadas quanto à presença de DPP.
- *Amostra aleatória*: foi considerada amostra aleatória ou probabilística, aquela em que cada participante teria igual oportunidade de ser selecionada, ou seja, deveria ser feita seleção ao acaso de puérperas pertencentes à população estudada (Lunsford & Lunsford, 1995). O método de aleatorização foi transcrito para campo aberto do formulário de coleta de dados, tal como declarado na publicação. Nos casos em que o trabalho declarou que a amostra era aleatória, mas não especificou como foi feita a seleção das participantes, foi marcada a opção “não está claro”. Não foram consideradas probabilísticas as amostras sistemáticas ou consecutivas.
- *Amostra não aleatória*: qualquer método que não correspondeu à definição de amostra aleatória, e que não configurou censo, foi classificado como amostra não aleatória.

Quando a amostra foi classificada como não aleatória, o procedimento de amostragem foi transcrito em campo aberto do formulário, tal qual relatado na publicação, e codificado em uma das categorias descritas a seguir.

- *Conveniência*: método de seleção não baseado em regras estabelecidas *a priori*, em que as possíveis participantes foram recrutadas aproveitando-se a sua disponibilidade e a disponibilidade dos entrevistadores (Lunsford & Lunsford, 1995).
- *Consecutiva*: todas as mulheres que visitaram uma determinada instituição (geralmente, centro de acompanhamento pré-natal ou maternidade), durante um determinado período de tempo, e que foram convidadas a participar da pesquisa, consecutivamente.
- *Amostra sistemática*: configurou-se quando a amostra foi selecionada por um método sistemático, como: entrevistar todas as participantes que tiverem bebês em determinados dias da semana ou do mês, entrevistar mulheres cujos prontuários tinham números pares ou ímpares, usar um intervalo numérico fixo para selecionar as participantes (ex.: entrevistar a 5ª, a 10ª, a 15ª, etc, de uma sequência qualquer).
- *Outra*: qualquer tipo de amostra que não se enquadrava nas definições anteriores. Foram pontuados os estudos que declararam ter realizado censo ou ter usado amostra aleatória ou probabilística, e que forneceram detalhes sobre o procedimento de amostragem. No caso de amostra probabilística, o trabalho deveria ter relatado uso de sorteio, tábua de números aleatórios ou outro método que permitisse selecionar participantes ao acaso. Esse rigor na avaliação da técnica de amostragem foi necessário em decorrência da frequente má compreensão do conceito de amostra aleatória por pesquisadores, constatada por estudos que claramente o usaram de forma equivocada.

Item 2: A base amostral foi adequada?

Consideramos que a base amostral foi adequada quando as participantes foram selecionadas de um conjunto de dados apropriado para se evitar vieses, como por exemplo, dados do censo ou de pesquisa porta-a-porta. Inadequada foi a seleção de participantes por meios que apresentam maior chance de gerar vieses na amostra,

como: lista telefônica (exclui mulheres sem telefone), prontuários de maternidade (excluem partos caseiros e os realizados em outros hospitais), participantes que decidiram se voluntariar pela Internet.

Item 3: O tamanho da amostra foi adequado?

O tamanho da amostra é inversamente proporcional à amplitude do intervalo de confiança, de forma que os resultados são mais precisos em estudos com amostras maiores (Loney et al., 1998). Assim, antes de darmos início à coleta de dados de um estudo de prevalência, foi preciso calcular o tamanho de amostra adequado, usando análise de poder, e levando em consideração o grau de precisão desejado, a taxa de erro aceitável, além de uma estimativa de prevalência da condição na população. Considerando a prevalência estimada de DPP de 12,6% (Melo Junior, 2011) e precisão da estimativa absoluta de 3%, com nível de significância de 5%, a amostra adequada para um estudo epidemiológico sobre DPP seria de 470 participantes. Portanto, atribuiu-se um ponto no item 3 da Escala de Loney et al. (1998) a estudos que contaram com 470 participantes ou mais. Esse critério é um tanto arbitrário, pois os parâmetros para estabelecer a amostra podem variar entre os estudos. O ideal seria que cada pesquisador relatasse como o tamanho amostral foi planejado nas suas condições e contexto de pesquisa.

Item 4: Foram usados critérios objetivos, adequados e padronizados para medir o desfecho de saúde?

A medida de desfecho deve ser reproduzível e válida com alta sensibilidade e especificidade (Loney et al., 1998). Medidas de confiabilidade se referem à capacidade do instrumento de discriminar entre indivíduos, separando as diferenças individuais reais daquelas que são provenientes do processo de medida (Boyle, 1998). Consideramos que esse dado estava presente quando o estudo mencionou alguma medida de confiabilidade como, por exemplo: consistência interna (alfa de *cronbach*, *split-half*) ou concordância teste-reteste do instrumento.

Medidas de validade, por sua vez, estabelecem quanto um instrumento pode fazer discriminações significativas e úteis entre os indivíduos (Boyle, 1998). Considerando que escalas de auto avaliação eram válidas quando foram mencionados dados de sensibilidade, e/ou especificidade, e/ou valor preditivo positivo, e/ou valor preditivo negativo. Instrumentos de avaliação clínica foram considerados válidos quando foi

usado critério padronizado, e não entrevista semiestruturada ou aberta; e quando foi mencionada a concordância entre avaliadores, para toda a amostra estudada ou parte dela. O item foi pontuado positivamente nos estudos que apresentaram informações de ambos os parâmetros, ou seja, confiabilidade e validade, não sendo suficiente mencionar somente um deles. Estudos que não explicitaram valores numéricos para a confiabilidade e a validade, mas indicaram que estas foram avaliadas em trabalhos anteriores, apresentando as referências, foram pontuados.

Item 5: O desfecho de saúde foi medido de forma não enviesada?

Esse item foi avaliado positivamente quando todas as condições a seguir foram satisfeitas:

- a) o estudo usou mais de um avaliador;
- b) foi mencionado grau de concordância entre eles;
- c) foi declarado que os entrevistadores estavam cegos quanto à condição da participante e aos resultados obtidos em outras avaliações, quando existentes.

Item 6: A taxa de resposta foi adequada? As características das pessoas elegíveis que se recusaram a participar foram descritas?

A validade da estimativa de prevalência é prejudicada quando há um número grande de recusas, ou de participantes selecionados que não estiveram disponíveis para mensuração. Os participantes podem ser perdidos por duas razões: 1) recusaram-se a participar do estudo para o qual foram convidados; 2) deixaram de participar de uma ou mais etapas de seguimento longitudinal. Loney et al. (1998) sugerem que uma taxa de resposta maior ou igual a 70% é aceitável. O item 6 foi avaliado positivamente quando ambas as condições descritas a seguir, foram satisfeitas:

- a) A taxa de perda foi menor que 30%, considerando-se a soma das puérperas convidadas que recusaram participar (*dropouts*) e aquelas que foram perdidas ao longo do tempo, no caso dos estudos longitudinais (*attrition*).
- b) Se a perda foi maior que zero, foram descritas as características de perfil das mulheres elegíveis que deixaram de participar, e/ou que foram perdidas ao longo do estudo.

Item 7: As estimativas de prevalência ou incidência foram apresentadas com intervalos de confiança e detalhadas por subgrupo, se adequado?

Esse item foi pontuado quando a prevalência foi relatada com intervalos de confiança e o trabalho apresentou análises de subgrupos. A apresentação de intervalos de confiança para outros parâmetros, como razão de prevalência, razão de chance (*odds ratio*) ou risco relativo, não foi considerada para atribuição de pontuação. Este item avaliou a qualidade da apresentação dos resultados e mede o poder de generalização do estudo.

Item 8: Os participantes e o contexto são descritos em detalhes e podem ser generalizados para outras situações similares?

O item 8 foi pontuado quando foi apresentado o perfil sociodemográfico da amostra estudada, permitindo aos leitores avaliar a aplicabilidade para o seu próprio contexto. Esse, também, é um elemento de avaliação de resultados, segundo Loney et al. (1998).

4.1.11 Inclusão de outras variáveis para avaliação metodológica

Considerando as vantagens de recorrermos a outras diretrizes para avaliação da qualidade metodológica de estudos epidemiológicos (Alhabib, Nur, & Jones, 2010; Bennet, Einarson, Taddio, Koren, & Einarson, 2004; Prins, Blanker, Bohenen, Thomas, & Bosch, 2002; Shamliyan, Kane, & Dickinson, 2010; Shamu, Abrahams, Temmerman, Musekiwa, & Zarowsky, 2011), foram incluídos outros elementos à Escala:

1. tipo de instrumento para identificar DPP: consideramos o uso de entrevista clínica e a aplicação de questionário de auto-avaliação como técnicas mais adequadas para avaliação de DPP.
2. Planejamento do tamanho amostral: observamos se os autores do trabalho indicaram ter calculado o tamanho de amostra mínima para o estudo de prevalência.
3. Critérios de inclusão e exclusão da população estudada, constituindo uma variável que pode ter implicações sobre o resultado de prevalência. (Silva, 2013).

Todas as variáveis foram analisadas após o preenchimento de um formulário eletrônico denominado “Revisão Sistemática de Literatura sobre Prevalência de

DPP”, disponibilizado no sistema Lime Survey (<http://www.limesurvey.org/>) e desenvolvido pelas revisoras⁷.

4.1.12 Verificação do nível de concordância entre as revisoras

Análises de concordância compararam as avaliações independentes de ambas as revisoras quanto à seleção dos artigos por título, resumo e pela Escala de Loney et al. (1998). Para isso, foi adotado o seguinte procedimento:

1. ambas as revisoras classificaram os títulos e os resumos dos trabalhos localizados, conforme critérios de inclusão de A a I definidos previamente. A concordância entre as revisoras foi estabelecida para essas etapas.
2. Os textos completos selecionados pelo resumo foram lidos totalmente pela primeira revisora, que aplicou os critérios de inclusão A a I. O papel da segunda revisora foi fazer a leitura dos textos completo de todos os estudos latino-americanos (68) e da amostra pareada aleatorizada de estudos não latino-americanos (68).

Quando não havia concordância em todos os itens, era efetuado o consenso para o estabelecimento do nível de concordância entre as avaliadoras nesta etapa também.

4.2 Procedimentos específicos deste estudo comparativo

Após a finalização da nossa participação na revisão sistemática, que foi concluída com a avaliação dos estudos latino-americanos e da amostra pareada aleatorizada de estudos de outras regiões, partimos para a definição do universo deste estudo.

⁷O formulário eletrônico de avaliação do texto completo pode ser consultado em: <http://revsist.limequery.com/62197/lang-pt-BR> (acessado em 17/01/2013).

4.2.1 Universo do estudo

O universo desta pesquisa foi composto por estudos latino-americanos publicados em revistas científicas incluídos na revisão. Assim, da população de artigos incluídos na revisão sistemática, isolamos os artigos que tiveram coleta de dados em países latino-americanos e comparamos a qualidade com os de outras regiões. Comparamos, também, os artigos publicados em revistas latino-americanas com os publicados em revistas de outras regiões.

4.2.2 Instrumentos

Por ser o universo do nosso estudo um extrato dos trabalhos incluídos na revisão sistemática, os estudos latino-americanos e de outras regiões foram avaliados de acordo com a Escala de Loney et al. (1998), seguindo os mesmos critérios de inclusão e de avaliação da qualidade metodológica e da apresentação dos resultados descritos anteriormente.

4.2.3 Critérios de seleção

A seleção dos artigos pelo título, resumos e texto completo foi efetuada pelas duas revisoras de forma independente. Na última etapa (avaliação do texto completo) as discordâncias foram solucionadas com discussões para consenso. Roteiros com instruções (Apêndice A) e treinamentos para o desenvolvimento de todas as etapas da seleção dos registros e aplicação da Escala de Loney et al. (1998) foram efetuados visando à equiparação das informações entre as revisoras.

4.3 Avaliação da qualidade de forma

O título, o resumo e as palavras-chave atribuídas aos artigos foram de extrema relevância na recuperação e na avaliação para inclusão/exclusão dos artigos na revisão sistemática, assim, efetuamos a avaliação da qualidade desses elementos no universo de artigos deste trabalho. Os títulos e resumos foram analisados com base nas orientações do manual de publicação da American Psychological Association (2012) e as palavras-chave foram verificadas em função do controle da

Terminologia em Psicologia, disponível na Biblioteca Virtual em Saúde – Psicologia (BVS-Psi, www.bvs-psi.org.br), que tem como base o Thesaurus of Psychological Index Terms (2007), e do vocabulário controlado “Descritores em Ciências da Saúde” (DECS), desenvolvido a partir do MeSH - Medical Subjects Headings, da National Library of Medicine, em Washington, DC.

4.4 Análise dos dados

A concordância entre avaliadoras quanto à pontuação total (isto é, a soma das pontuações nos oito itens) da Escala de Loney et al. (1998), considerada variável de contagem, foi mensurada pelo coeficiente de correlação intraclasse, usando modelo de duas vias com efeitos mistos (*two-way mixed effects model*) (Shrout & Fleiss, 1979; McGraw & Wong, 1996). Um valor de coeficiente de correlação intraclasse igual a um indica que as avaliações são idênticas nas comparações realizadas, e foram considerados aceitáveis valores acima de 0,7.

Foi descartada a mensuração de confiabilidade das avaliações das duas revisoras pelo índice de kappa, pois ele não é indicado em situações em que se espera que a ocorrência das categorias do objeto avaliado seja desproporcional. Esse é o caso, por exemplo, da avaliação de alguns dos critérios de inclusão, pois esperávamos maior número de estudos classificados como “verdadeiro” do que como “falso”. Assim, seria esperado alto nível de concordância pelo acaso, resultando num valor máximo de k (valor que corresponde à concordância perfeita entre avaliadores) mais baixo (Eugenio, 2000). Optou-se por avaliar a concordância entre as avaliadoras usando os dados referentes à porcentagem de estudos em que as avaliações apresentaram respostas idênticas.

A análise dos dados foi efetuada de forma quantitativa, sendo utilizados medidas de frequências para apresentação da distribuição dos dados com relação às categorias e testes não paramétricos, ou de livre-distribuição, como também são chamados. Como a distribuição dos dados não foi normal, conforme o teste de distribuição de Kolmogorov-Smirnov (KS), utilizamos o teste de Mann Whitney para testar as distribuições das duas categorias de artigos e testamos as hipóteses aplicando o teste qui-quadrado (X^2). As hipóteses foram testadas através da comparação dos quatro subgrupos: estudos latino-americanos versus estudos de outras regiões e revistas latino-americanas versus revistas de outras regiões.

Utilizamos testes de qui-quadrado para variáveis categóricas e testes não paramétricos para as pontuações na Escala de Loney et al. (1998). Consideramos significativos os testes que resultaram em p-valor menor que 0,05.

No próximo Capítulo apresentamos as estatísticas descritivas dos dados gerais da revisão sistemática, bem como as comparações do universo de artigos avaliados em cada item da Escala de Loney et al. (1998). Os dados referentes à avaliação de forma também são ilustrados.

5 RESULTADOS

Desde a definição da pergunta de pesquisa até a obtenção dos resultados apresentados a seguir, diretrizes e a literatura científica foram consultadas em busca de instrumentos que permitissem cumprir o objetivo de comparar a qualidade metodológica dos estudos e da apresentação dos resultados em artigos científicos latino-americanos e de outras regiões, estes incluídos em revisão sistemática sobre prevalência de DPP.

O presente capítulo apresenta os resultados de todas as etapas percorridas nesta pesquisa, desde o levantamento da literatura, passando pela seleção dos trabalhos, até as estatísticas dos dados levantados. Na sequência, foram apresentados os resultados gerais da revisão sistemática, comparada a qualidade metodológica e da apresentação dos resultados dos artigos que tiveram a coleta de dados em países latino-americanos e de outras regiões e dos artigos publicados em revistas latino-americanas e de outras regiões. Após as análises descritivas foram apresentados os testes de hipóteses da presente Tese. A análise descritiva dos títulos, resumos e palavras-chave encerra o capítulo.

5.1 Resultados gerais da revisão sistemática

A aplicação das estratégias de busca resultou na recuperação de 3.909 registros, nas 14 bases de dados consultadas. A Tabela a seguir ilustra as fontes de informação e o número de registros localizados, nas datas em que as duas avaliadoras efetuaram a consulta, de forma independente.

Tabela 5 - Bases de dados, data de consulta e registros recuperados

Base	Data	Número de registros
Biblioteca Digital de Teses e Dissertações	09/10/2011	17
BVS-Psi ULAPSI	19/10/2011	405
CINAHL	17/10/2011	178
EMBASE	30/09/2011	582
ERIC	03/10/2011	5
WOK	29/09/2011	1117
LILACS	01/10/2011	35
ProQuest	09/10/2011	24
Psicodoc	03/10/2011	61
PsycInfo	29/09/2011	403
PubMed	29/09/2011	229
Redalyc	06/10/2011	6
SciELO	06/10/2011	17
Scopus	29/09/2011	830
Total		3909

A próxima tabela, de número seis apresenta as bases de dados consultadas e as estratégias de busca utilizadas na recuperação dos estudos.

Tabela 6 – Estratégias de busca para as bases de dados consultadas

Base	Estratégia
Biblioteca Digital de Teses e Dissertações	Depressão pós-parto AND prevalência
BVS-PSI ULAPSI	Depressão pós-parto
CINAHL	"postpartum depression" in Mesh AND prevalence in Mesh
EMBASE	Postpartum AND depression/de AND prevalence/de ⁸
ERIC	(postpartum depression or postnatal depression or puerperal depression) AND preval*
WOK	(postpartum depression OR postnatal depression OR puerperal depression) AND preval*
LILACS	(postpartum depression OR depresion postparto OR depressão pos-parto OR depressao puerperal OR puerperal depression OR depressão pos-natal OR postnatal depression OR depresion postnatal) AND preval\$
ProQuest	(postpartum depression OR postnatal depression OR puerperal depression) AND (prevalence)
Psicodoc	Depresión post-parto
PsycINFO	((postpartum depression OR postnatal depression OR puerperal depression) AND preval*)
PubMed	postpartum depression" in Mesh AND prevalence in Mesh ⁹
Redalyc	Depression post-parto
SciELO	Prevalência depressão pós-parto
Scopus	("postpartum depression" OR "postnatal depression" OR "puerperal depression") AND preval*

Nota. Tabela adaptada de Silva (2013) para o nosso estudo.

⁸ O sinalizador "/de" significa que o termo foi procurado em descritores, para que a busca resultasse mais específica e mais qualificada. Usando os descritores, o trabalho não é localizado caso o termo procurado esteja apenas no título, resumo ou corpo do texto.

⁹ MESH significa *Medical Subjects Heading*, ou Cabeçalhos de Assuntos Médicos, em tradução livre. Trata-se do *thesaurus* da área de saúde, ou seja, o termo foi procurado no campo de descritores e não livremente no título, resumo e outros campos. É o equivalente ao "/de" da EMBASE.

Na definição das estratégias de busca, recorremos ao uso dos operadores booleanos OR e AND para combinação dos termos dentro e fora dos parênteses, aos operadores de truncamento para facilitar as buscas, evitando a repetição de estratégias similares, e das aspas para restrição à expressão exata.

5.2 Seleção dos estudos

A seleção dos estudos para obtenção dos resultados obedeceu a três etapas: avaliação de títulos, avaliação de resumos e avaliação dos textos completos, sempre de acordo com os critérios previamente estabelecidos. O fluxograma, apresentado na sequência, mostra o número de trabalhos selecionados e eliminados em cada etapa para elegibilidade dos estudos.

Figura 4. Fluxograma da seleção de estudos para a revisão sistemática.

No processo de seleção dos trabalhos foi necessário fazer a exclusão dos estudos repetidos. Utilizamos o software EndNote® Web (acessado em 04/09/2012) para agilizar o processo de exclusão das duplicatas. Ao final do processo foram excluídos 2.015 trabalhos, dos 3.909 iniciais, restando 1.894 registros para avaliação

pelo título. Ao aplicar os critérios de exclusão nos títulos dos trabalhos eliminamos 580 registros, restando 1.314 trabalhos para avaliação pelos resumos. Essa etapa possibilitou a exclusão de 421 trabalhos, sendo reservados 893 para avaliação pelo texto completo. Não foi possível fazer a leitura de 115 trabalhos, uma vez que vários não puderam ser recuperados pelos sistemas de comutação e outros por não terem o texto completo disponível nas línguas definidas para a revisão sistemática: inglês, espanhol e português. Após a leitura dos textos completos a primeira revisora excluiu 441 estudos, incluindo 337 na revisão sistemática. Como uma das revisoras, passamos por todas as etapas da revisão, até a leitura dos textos, quando fizemos a leitura de 136 textos completos, distribuídos da seguinte maneira: 68 estudos latino-americanos e uma amostra aleatória pareada de 68 estudos de outras regiões para o estabelecimento da concordância entre as revisoras.

O estudo de Silva (2013) considerou aceitável a taxa de concordância entre as revisoras na avaliação dos títulos, resumos e textos completos, igual ou superior a 70%. A Tabela 7 apresenta as taxas de concordância entre as avaliadoras, realizada a partir de avaliações cegas dos títulos, resumos e dos 68 textos completos que coletaram dados em países latino-americanos e dos 68 que compuseram o grupo controle.

Tabela 7 - Concordância entre avaliadoras quanto à inclusão de estudos pelos títulos, resumos e textos completos.

Critério de inclusão	Títulos	Resumos	Textos completos	Textos completos
	(N=1.894)	(N=1.314)	da América Latina	de outros países
	%	%	(N=68)	(N=68)
			%	%
Concordância na decisão final pela seleção dos trabalhos	78,6	80,8	79,7	73,2
Critério A: estudo primário e quantitativo, delineamento transversal ou de coorte longitudinal	95,4	78,5	91,3	93,0
Critério B: avaliou prevalência de DPP	72,5	65,8	81,2	81,7
Critério C: participantes são mães humanas	89,0	70,9	89,9	88,7

Critério de inclusão	Títulos	Resumos	Textos completos	Textos completos
	(N=1.894)	(N=1.314)	da América Latina	de outros países
	%	%	(N=68)	(N=68)
			%	%
Critério D: amostra não foi submetida à intervenção com objetivo de prevenir DPP	99,1	71,5	89,9	90,1
Critério E: método de diagnóstico mencionado claramente e baseado em instrumento objetivo de medida, ou em avaliação feita por profissional de saúde mental	100,0	69,4	91,3	84,5
Critério F: pelo menos uma das avaliações de DPP ocorreu de 8 dias (inclusive) a 365 dias (inclusive) após o parto	90,9	64,2	78,3	76,1
Critério G: estudo realizado na população geral	93,5	70,7	84,1	83,1
Critério H: resultado de prevalência relatado de forma clara	-	-	79,7	71,8

Nota. Tabela adaptada de Silva (2013) para o nosso estudo.

A concordância final pela seleção dos trabalhos para inclusão de estudos latino-americanos e não latino-americanos (grupo controle) foi adequada, lembrando que a concordância aceitável era de 70%. Na seleção pelo título houve 78,6% de concordância, para os resumos 80,8%, no texto completo houve 79,7% para os estudos latino-americanos e 73,2% para os não latino-americanos.

5.3 Análises descritivas da Revisão Sistemática

A revisão sistemática da literatura objetiva avaliar e sintetizar todos os estudos relevantes em um tópico único, com metodologia específica e claramente definida para controlar e minimizar os erros da conclusão. Não foi possível cumprir esta exigência metodológica em nosso estudo. Ao utilizar método científico reproduzível, para sintetizar com rigor e de forma criteriosa os resultados de outros estudos, fornece a avaliação de um conjunto de dados que pode ser utilizado com confiança

na tomada de decisão. Os resultados completos da revisão sistemática sobre prevalência de DPP, que possibilitaram o sequenciamento dos trabalhos aqui analisados, estão disponíveis na tese de Silva (2013). Por ser a prevalência da DPP a condição que norteou todo o processo de busca e avaliação dos estudos objetos deste trabalho, voltamos à tese de Silva (2013) para apresentar as prevalências de DPP obtidas na revisão sistemática:

As prevalências de DPP encontradas nos estudos incluídos variaram de 0,5% (IC 95% = 0,00% a 1,48%) em Singapura a 62,8% (IC 95% = 60,44% a 65,23%) nos Estados Unidos da América. A prevalência média dos 337 trabalhos, ponderada pelos tamanhos amostrais, foi de 16,63% (IC 95% = 16,50%-16,76%). (p. 143)

A seguir as análises descritivas referentes à revisão sistemática, e que nos interessam diretamente, são apresentadas, uma vez que tratam de artigos e revistas científicas. Acompanhamos todas as etapas da revisão sistemática com o objetivo de observar alguns aspectos das revistas científicas que aguçavam nossa curiosidade e que buscamos analisar neste processo.

Dos 337 estudos incluídos na revisão sistemática, excluindo as dissertações e teses, extraímos 307 artigos incluídos que foram publicados em 147 revistas científicas. A Tabela 8 apresenta os títulos organizados de acordo com a frequência e o percentual em que os trabalhos apareceram.

Tabela 8 - Títulos das revistas e número dos artigos incluídos na revisão sistemática

Título da revista		F (N.A.)	(%)
1)	Journal of Affective Disorders	23	7,5
2)	Archives of Women's Mental Health	20	6,5
3)	British Journal of Psychiatry	11	3,6
4)	Journal of Psychosomatic Obstetrics & Gynecology	10	3,3
5)	Journal of Reproductive and Infant Psychology	8	2,6
6)	Journal of Women's Health	8	2,6
7)	BJOG: an International Journal of Obstetrics and Gynaecology	7	2,3

	Título da revista	F (N.A.)	(%)
8)	Maternal and Child Health Journal	7	2,3
9)	Australian and New Zealand Journal of Psychiatry	5	1,6
10)	Pediatrics	5	1,6
11)	Psychological Medicine	5	1,6
12)	Revista Chilena de Obstetricia y Ginecología	5	1,6
13)	Social Psychiatry and Psychiatric Epidemiology	5	1,6
14)	Acta Psychiatrica Scandinavica	4	1,3
15)	American Journal of Obstetrics and Gynecology	4	1,3
16)	Comprehensive Psychiatry	4	1,3
17)	Journal of Obstetric Gynecologic and Neonatal Nursing	4	1,3
18)	American Journal of Psychiatry	3	1,0
19)	BMC Psychiatry	3	1,0
20)	Journal of Clinical Nursing	3	1,0
21)	Journal of Maternal-Fetal and Neonatal Medicine	3	1,0
22)	Journal of Midwifery & Women's Health	3	1,0
23)	Revista Médica de Chile	3	1,0
24)	Acta Obstetricia et Gynecologica Scandinavica	2	0,7
25)	Australian and New Zealand Journal of Public Health	2	0,7
26)	Birth-Issues in Perinatal Care	2	0,7
27)	BMC Pregnancy and Childbirth	2	0,7
28)	BMC Public Health	2	0,7
29)	British Journal of Clinical Psychology	2	0,7
30)	Clinical Practice and Epidemiology in Mental Health	2	0,7
31)	International Journal of Gynecology and Obstetrics	2	0,7
32)	International Journal of Nursing Studies	2	0,7
33)	International Urogynecology Journal	2	0,7
34)	Irish Journal of Psychological Medicine	2	0,7
35)	Journal of Abnormal Psychology	2	0,7

	Título da revista	F (N.A.)	(%)
36)	Journal of Consulting and Clinical Psychology	2	0,7
37)	Journal of Epidemiology and Community Health	2	0,7
38)	Journal of Obstetrics and Gynaecology Research	2	0,7
39)	Journal of Psychosomatic Research	2	0,7
40)	Journal of Reproductive Medicine	2	0,7
41)	Midwifery	2	0,7
42)	Obstetrics and Gynecology	2	0,7
43)	Pakistan Journal of Medical Sciences	2	0,7
44)	Psychiatry and Clinical Neurosciences	2	0,7
45)	Psychiatry Research	2	0,7
46)	Psychology, Health & Medicine	2	0,7
47)	Research in Nursing & Health	2	0,7
48)	Revista de Psiquiatria Clínica	2	0,7
49)	Revista de Saúde Pública	2	0,7
50)	Scandinavian Journal of Caring Sciences	2	0,7
51)	SLEEP	2	0,7
52)	Social Science & Medicine	2	0,7
53)	Women & Health	2	0,7
54)	Acta Obstetricia et Gynecologica Scandinavica	1	0,3
55)	Acta Pædiatrica	1	0,3
56)	Actas Luso-Españolas de Neurologia Psiquiatria y Ciencias Afines	1	0,3
57)	African Health Sciences	1	0,3
58)	American College of Obstetricians and Gynecologists	1	0,3
59)	Anales de la Facultad de Medicina (Perú)	1	0,3
60)	Anxiety Disorders	1	0,3
61)	Archives of Disease in Childhood	1	0,3
62)	Archives of Gynecology and Obstetrics	1	0,3

	Título da revista	F (N.A.)	(%)
63)	Archives of Psychiatric Nursing	1	0,3
64)	Asian Journal of Psychiatry	1	0,3
65)	Atención Primaria	1	0,3
66)	Australian & New Zealand Journal of Obstetrics & Gynaecology	1	0,3
67)	Australian Journal of Public Health	1	0,3
68)	Bipolar Disorders	1	0,3
69)	BMC Family Practice	1	0,3
70)	British Journal of Community Nursing	1	0,3
71)	British Journal of Midwifery	1	0,3
72)	British Journal of Psychiatry	1	0,3
73)	British Medical Journal	1	0,3
74)	Bulletin of the World Health Organization	1	0,3
75)	Cadernos de Saúde Pública	1	0,3
76)	Canadian Family Physician	1	0,3
77)	Canadian Journal of Nursing Research	1	0,3
78)	Child Abuse & Neglect	1	0,3
79)	Child: Care, Health and Development	1	0,3
80)	Ciência, Cuidado e Saúde	1	0,3
81)	Colômbia Médica	1	0,3
82)	Community Mental Health Journal	1	0,3
83)	Contemporary Nurse	1	0,3
84)	Developmental Psychology	1	0,3
85)	Dreaming: Journal of Association for the Study of Dreaming	1	0,3
86)	European Psychiatry	1	0,3
87)	Family Practice Research Journal	1	0,3
88)	Gaceta Médica de México	1	0,3
89)	Health & Social Work	1	0,3

	Título da revista	F (N.A.)	(%)
90)	Health and Social Care in the Community	1	0,3
91)	Health Care for Women International	1	0,3
92)	Infant Behavior and Development	1	0,3
93)	Infant Mental Health Journal	1	0,3
94)	Infanto - Revista de Neuropsiquiatria da Infância e Adolescência	1	0,3
95)	International Journal of Epidemiology	1	0,3
96)	International Journal of Nursing Practice	1	0,3
97)	International Journal of Psychiatry in Clinical Practice	1	0,3
98)	International Journal of Public Health	1	0,3
99)	International Medical Journal	1	0,3
100)	Iranian Journal of Public Health	1	0,3
101)	Israel Medical Association Journal	1	0,3
102)	Jornal Brasileiro de Psiquiatria	1	0,3
103)	Journal of Adolescent Health	1	0,3
104)	Journal of Behavioral Health Services & Research	1	0,3
105)	Journal of Clinical Endocrinology & Metabolism	1	0,3
106)	Journal of Community Health Nursing	1	0,3
107)	Journal of Epidemiology	1	0,3
108)	Journal of Mental Health	1	0,3
109)	Journal of Nervous and Mental Disease	1	0,3
110)	Journal of Nursing Research	1	0,3
111)	Journal of Psychiatric Research	1	0,3
112)	Journal of Reproductive Medicine for the Obstetrician and Gynecologist	1	0,3
113)	Journal of Research in Medical Sciences	1	0,3
114)	Journal of Social and Clinical Psychology	1	0,3
115)	Journal of the American Board of Family Medicine	1	0,3

	Título da revista	F (N.A.)	(%)
116)	Journal of the American Osteopathic Association	1	0,3
117)	Journal of the College of Physicians and Surgeons Pakistan	1	0,3
118)	Journal of the Egyptian Public Health Association	1	0,3
119)	Journal of the Oklahoma State Medical Association	1	0,3
120)	Kaohsiung Journal of Medical Sciences	1	0,3
121)	Revue Canadienne de Psychiatrie	1	0,3
122)	Lancet	1	0,3
123)	Medicina (Kaunas)	1	0,3
124)	Morbidity and Mortality Weekly Report	1	0,3
125)	Nursing Research	1	0,3
126)	Pacific Health Dialog	1	0,3
127)	Pain	1	0,3
128)	Pediatrics International	1	0,3
129)	Perspectives in Psychiatric Care	1	0,3
130)	PlosOne	1	0,3
131)	Psicología Conductual	1	0,3
132)	Psico-USF	1	0,3
133)	Psychiatric Services	1	0,3
134)	Psychologica	1	0,3
135)	Psychosomatic Medicine	1	0,3
136)	Public Health Nutrition	1	0,3
137)	Revista Brasileira de Ginecologia e Obstetrícia	1	0,3
138)	Revista Brasileira de Psiquiatria	1	0,3
139)	Revista Brasileira de Saúde Materno-Infantil	1	0,3
140)	Revista Colombiana de Obstetricia y Ginecología	1	0,3
141)	Revista de Psicopatología y Psicología Clínica	1	0,3
142)	Revista Peruana de Epidemiología	1	0,3
143)	Rural and Remote Health	1	0,3

Título da revista		F (N.A.)	(%)
144)	Singapore Medical Journal	1	0,3
145)	Transcultural Psychiatry	1	0,3
146)	Tropical Medicine and International Health	1	0,3
147)	Women's Health Issues	1	0,3
Total		307	100

Nota. F = Frequência. N.A. = Número Absoluto.

Optamos por transcrever os títulos das revistas tal como apareceram na legenda dos artigos, excluindo a partícula *The*, tão comum nas revistas com títulos em inglês. Os artigos incluídos na revisão sistemática foram publicados de maneira bastante diluída entre as 147 revistas. Três dessas merecem destaque por terem mais de 10 artigos incluídos e concentrarem 18% dos trabalhos. O Journal of Affective Disorders, responsável por 23 artigos da amostra, é um órgão da International Society for Affective Disorders e publicada pela editora holandesa Elsevier. Declara em sua página publicar trabalhos sobre depressão, mania, ansiedade e pânico. Com um fator de impacto de 3,91 nos últimos cinco anos, de acordo com o Journal Citation Reports (JCR), não está indexada na base de dados Scopus e, por isso, não apresenta o Índice H (H Index). Archives of Women's Mental Health contribuiu com 20 artigos na revisão sistemática. É um órgão oficial da International Association for Women Health, publicada pela Springer Wien. Seu escopo abrange a psicodinâmica, aspectos sociais e biológicos da psiquiatria e dos distúrbios psicossomáticos. Seu Índice H é de 33, medido pelo SCImago Journal & Country Rank. Embora declare ser indexada pelo JCR, seu fator de impacto não está disponível. The British Journal of Psychiatry participou com 12 trabalhos. É publicada pelo Royal College of Psychiatrists de Londres, e foi fundada em 1853. Abrange todos os ramos da psiquiatria, com ênfase nos aspectos clínicos. Em sua página principal declara o fator de impacto de 6,6 no JCR e não possui Índice H por não estar indexada em Scopus. O número elevado de revistas aqui representado

demonstra que as buscas foram exaustivas e abrangentes e reforça a importância das revistas na comunicação da ciência.

A representatividade de diversos países em revisões sistemáticas é um atributo de valor para trabalhos dessa natureza, uma vez que assegura um panorama globalizado dos estudos. Dessa maneira, alcançar os estudos gerados na maior parte dos países do mundo foi uma questão perseguida nas buscas por estudos. A Tabela 9 ilustra os países de origem dos estudos.

Para classificação dos países, quanto ao nível de desenvolvimento socioeconômico, utilizamos a tabela de Índice de Desenvolvimento Humano (IDH), elaborada pelo Programa das Nações Unidas para Desenvolvimento (PNUD)¹⁰. Para o PNUD, o conceito de desenvolvimento humano é um processo de ampliação de escolhas das pessoas para que elas tenham capacidades e oportunidades de serem aquilo que desejam ser. Vincula o bem-estar da sociedade ao desenvolvimento do ser humano, retirando o foco apenas do crescimento econômico, ou da renda. O IDH mede o desenvolvimento humano por três dimensões: renda, saúde e educação, contrapondo-se a outro indicador muito utilizado, o Produto Interno Bruto (PIB), que considera apenas a dimensão econômica do desenvolvimento.

No Relatório do Desenvolvimento Humano, de 2011, foi possível analisar o IDH dos 187 países classificados. Neste estudo, consideramos desenvolvidos os países do primeiro bloco, classificados como nível de “desenvolvimento humano muito elevado”. A Noruega aparece como o país mais desenvolvido, seguido pela Austrália, Holanda, EUA e Nova Zelândia nas cinco primeiras posições da lista. Dois países latino-americanos apareceram na lista dos mais desenvolvidos: Chile (44º lugar) e Argentina (45º).

No segundo bloco, denominado países com “desenvolvimento humano elevado”, estão vários países latino-americanos. O Uruguai apareceu no primeiro lugar deste bloco (48º), Cuba (51º), México (57º), Panamá (58º), Costa Rica (69º), Venezuela (73º), Peru (80º), Equador (83º), Brasil (84º), Colômbia (87º) e Belize (94º).

O termo América Latina foi estabelecido pelo imperador francês Napoleão III para designar os países com predomínio das línguas romanas, ou latinas, que são idiomas provenientes da família das línguas que se originaram do Latim,

¹⁰ www.pnud.org.br/IDH/DH.aspx

especialmente do Latim vulgar. No final da Segunda Guerra Mundial, a Comissão Econômica para América Latina e Caribe (CEPAL) consolidou o uso da expressão para designar os países subdesenvolvidos, ou seja, aqueles que viviam abaixo da linha do desenvolvimento, e o termo se converteu em uma divisão econômica e social (Sampaio & Sabadini, 2012). Classificamos os países da América Latina de acordo com a Biblioteca Virtual da América Latina (www.bvmemorial.fapesp.br), desenvolvida pela Fundação Memorial da América Latina. Assim, para o nosso estudo são 24 países os latino-americanos: Argentina, Belize, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, El Salvador, Equador, Guatemala, Guiana, Guiana Francesa, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, Porto Rico, República Dominicana, Suriname, Uruguai e Venezuela.

A seguir, são relacionados os países e as frequências dos trabalhos incluídos na revisão sistemática.

Tabela 9 - Países em que a coleta de dados dos trabalhos incluídos na revisão sistemática foi realizada

País	F (N.A.)	(%)
Estados Unidos da América (EUA)	76	22,6
Brasil	26	7,7
Austrália	24	7,1
Reino Unido (UK)	23	6,8
Canadá	16	4,7
China	13	3,9
Turquia	11	3,3
Suécia	9	2,7
Chile	8	2,4
Japão	8	2,4
Israel	7	2,1
Itália	7	2,1
Espanha	7	2,1
Noruega	6	1,8
Portugal	6	1,8
Irã	5	1,5

País	F (N.A.)	(%)
Hong Kong	5	1,5
Vários países	5	1,5
Grécia	4	1,2
Irlanda	4	1,2
Holanda	4	1,2
Paquistão	4	1,2
Alemanha	3	0,9
Nova Zelândia	3	0,9
Nigéria	3	0,9
África do Sul	3	0,9
Suíça	3	0,9
Emirados Árabes Unidos	3	0,9
Colômbia	2	0,6
Hungria	2	0,6
Índia	2	0,6
Malásia	2	0,6
México	2	0,6
Nepal	2	0,6
Peru	2	0,6
Cingapura	2	0,6
Uganda	2	0,6
Vietnã	2	0,6
Armênia	1	0,3
Bangladesh	1	0,3
Dinamarca	1	0,3
Egito	1	0,3
Finlândia	1	0,3
França	1	0,3
Islândia	1	0,3
Jamaica	1	0,3
Jordão	1	0,3
Líbano	1	0,3
Lituânia	1	0,3
Malta	1	0,3
Mongólia	1	0,3
Marrocos	1	0,3
Polônia	1	0,3

País	F (N.A.)	(%)
Coreia do Sul	1	0,3
Arábia Saudita	1	0,3
Tailândia	1	0,3
Venezuela	1	0,3
Zâmbia	1	0,3
Zimbábue	1	0,3
Total	337	100,0

Os 337 estudos incluídos na revisão sistemática foram realizados em 58 países, sendo cinco com coleta de dados em vários países, estudos interculturais. Desses, duzentos e quarenta (240) fazem parte do bloco denominado países desenvolvidos; noventa e quatro (94) dos países em desenvolvimento e três considerados mistos (o mesmo estudo foi realizado em países desenvolvidos e em desenvolvimento). Recorremos à Silva (2013) para apresentar a prevalência de DPP encontrada nos estudos incluídos na revisão sistemática, de acordo com o nível socioeconômico de desenvolvimento dos países em que os dados foram coletados.

A prevalência média foi de 21,31% (IC 95% 20,87%-21,85%) nos 94 estudos realizados em países em desenvolvimento e 16,02% (IC 95% 15,89%-16,16%) nos 240 estudos realizados em países desenvolvidos. Três estudos realizaram coleta de dados em países desenvolvidos e em desenvolvimento, e tiveram prevalência média de 29,55% (IC 95% 26,96%-32,15%). (Silva, 2013, p.128).

Vale destacar a participação dos EUA, com 76 estudos incluídos, o que parece denotar preocupação dos pesquisadores com a questão da DPP, mas também pode ser justificada pela alta produtividade científica daquele país. O Pesquisa Mundi, site dedicado à discussão de Bases de Dados/Informações, Bibliotecas Digitais/Virtuais, Arquivos de Acesso Aberto, no dia 22 de abril de 2013, apresentou o *ranking* dos países mais produtivos na quantidade de trabalhos científicos, de acordo com a base de dados SCImago, alimentada pela plataforma Scopus, da Elsevier. Os EUA apareceram em primeiro lugar, com 519.573 trabalhos

publicados no período de 2001 a 2011, o que confirma a primazia do país encontrada no nosso estudo¹¹.

Nos países em desenvolvimento, a prevalência detectada nos estudos foi mais alta (21,31%), em relação aos desenvolvidos (16,02%), embora tenhamos levantado mais estudos na segunda categoria de países (240). O Brasil aparece na segunda posição, com 26 trabalhos aprovados nos critérios para inclusão na revisão sistemática, demonstrando que o assunto também vem chamando a atenção de pesquisadores no país, ou então que os artigos científicos brasileiros estão mais visíveis nas diversas bases de dados consultadas. No mesmo *ranking* do Pesquisa Mundi, citado no parágrafo anterior, o Brasil está em 13º lugar, com 49.664 trabalhos publicados no mesmo período de 2001 a 2011, o que demonstra que temos galgado posições respeitáveis nas listas mundiais da produção científica.

Bases de dados locais e internacionais foram consultadas, inclusive o Scopus, base que gera o *ranking* da análise do Pesquisa Mundi, o que, certamente, interferiu na quantidade de trabalhos recuperados. No próximo capítulo retomaremos essa discussão em função do número de bases de dados consultadas e do esforço que vem sendo feito para que o conhecimento gerado nos países latino-americanos esteja visível e disponível na Internet.

Não houve restrição ao período de publicação quando efetuadas as buscas nas bases de dados, uma vez que observamos que o descritor *postpartum depression* foi criado em 1973, no *Thesaurus of Psychological Index Terms* (2007), e uma revisão sistemática deve localizar e analisar todos os estudos sobre o tema de pesquisa, independente da data em que a pesquisa foi realizada. A Tabela 10 ilustra os anos dos estudos localizados, a frequência e o percentual de trabalhos incluídos na revisão sistemática.

Tabela 10 - Data de publicação e frequência dos estudos incluídos na revisão sistemática

Ano de Publicação	F (N.A.)	(%)
1980	1	0,3
1984	1	0,3
1986	1	0,3

¹¹ www.pesquisamundi.org/2013/04/producao-cientifica-do-brasil-aumenta.html)

Ano de Publicação	F (N.A.)	(%)
1989	1	0,3
1990	1	0,3
1991	1	0,3
1992	1	0,3
1993	5	1,5
1994	5	1,5
1995	4	1,2
1996	5	1,5
1997	4	1,2
1998	8	2,4
1999	7	2,1
2000	15	4,5
2001	12	3,6
2002	14	4,2
2003	11	3,3
2004	18	5,3
2005	24	7,1
2006	27	8,0
2007	17	5,0
2008	43	12,8
2009	41	12,2
2010	33	9,8
2011	35	10,4
2012	2	0,6
Total	337	100,0

Embora as buscas tenham sido efetuadas no ano de 2011, tivemos acesso a dois estudos publicados em formato eletrônico, na modalidade *ahead of print*, que antecipa a publicação dos artigos antes da conclusão do fascículo.

Observamos que, na década de 1980, o assunto não gerava muitas pesquisas publicadas em revistas científicas ou dissertações e teses, pois foram localizados apenas quatro estudos nesse período. Nos anos de 1990, houve um crescimento do número de pesquisas sobre o tema. A partir de 2000, temos um aumento expressivo de trabalhos que foram incluídos na revisão sistemática.

Chamou a nossa atenção os anos de 2008 e 2009, com 26% dos trabalhos localizados.

Voltaremos a discutir esses dados em função da criação de veículos para armazenamento e disseminação do conhecimento na área de ciências da saúde, na qual para efeito de organização das áreas, a Psicologia está inserida.

5.4 Resultados da qualidade de artigos incluídos na revisão sistemática

Após a análise descritiva dos dados gerais da revisão sistemática, apresentamos os resultados referentes aos trabalhos que coletaram dados em países latino-americanos e não latino-americanos. Comparar a qualidade dos artigos publicados em revistas latino-americanas e não latino-americanas encerrou esta etapa da análise.

O critério: “ter coletado dados em países latino-americanos” orientou a seleção dos trabalhos, pois nosso objetivo está diretamente ligado à qualidade dos estudos e das revistas latino-americanos. A literatura nos mostrou que países latino-americanos estiveram pouco representados nas revisões sistemáticas internacionais voltadas ao problema de DPP. Encontramos diversos estudos que tratavam do tema, mas não faziam parte das revisões sistemáticas identificadas (ex.: Coutinho, Baptista, & Moraes, 2002; Cruz, Simões, & Fasal-Cury, 2005; Hasbún et al., 1999; Moraes et al, 2006; Urdaneta et al., 2011). Silva (2013, p. 55) explica que dois argumentos justificam a não presença de estudos latino-americanos nas revisões levantadas:

...data de publicação das revisões, pois a mais recente é de 2004 e maior parte dos estudos latino-americanos sobre DPP é posterior a essa data; e restrição das estratégias de busca, que geralmente, se limitaram a publicações em inglês e não exploram bases de dados como BVS-Psi, SciELO, Lilacs, Psycodoc e Redalyc.

O desconhecimento de bases de dados é um problema que pode gerar viés em revisão sistemática e precisa ser superado, para obtermos estudos reais que possam apoiar decisões efetivas na área de saúde.

Foram encontradas prevalências de DPP de 00,92% a 57,14% nos estudos latino-americanos. Os dados corroboram a alta variação da prevalência encontrada

nas pesquisas em populações ao redor do mundo que foram incluídas na revisão sistemática, mínima de 0,5% e máxima de 65,32%.

Para que o trabalho fosse localizado, avaliado e incluído na revisão, deveria estar indexado em uma das 14 bases de dados rastreadas. Do total de artigos recuperados, vencidas as etapas de seleção pelo título, resumo e texto completo, analisamos 68 estudos latino-americanos e 68 não latino-americanos, para ser estabelecida a concordância entre revisores na revisão sistemática. Dos 68 latino-americanos avaliados, 24 foram excluídos pelos critérios e 10 foram eliminados por serem publicados em dissertações e teses, restando 34 artigos publicados em revistas científicas, base da comparação. Para o estabelecimento do grupo comparativo foi aleatorizada, utilizando o SPSS, uma amostra de 34 artigos cuja coleta de dados não foi feita em país latino-americano.

Dessa maneira, coincidentemente, ficamos com um universo de 68 artigos científicos para análise, o mesmo número inicial de estudos latino-americanos, antes da exclusão pelos critérios da revisão e dos trabalhos de mestrado e de doutorado. Por isso, mantivemos a identificação 68 para o conjunto inicial, e usamos 34+34 para o final. Os países e frequências dos estudos latino-americanos e não latino-americanos são discriminados na Tabela 11.

Tabela 11 - Distribuição dos países em que a coleta de dados foi realizada: latino-americanos e de outras regiões

Países	F (N.A.)	(%)
Latino-americanos		
Brasil	18	26,5
Chile	8	11,8
Colômbia	2	2,9
México	2	2,9
Peru	2	2,9
Venezuela	1	1,5
Caribe (país não identificado)	1	1,5
Subtotal	34	50,0
Não latino-americanos		
EUA	5	7,4
Suécia	4	5,9
China	3	4,4
Austrália	2	2,9
Canadá	2	2,9
Espanha	2	2,9
Israel	2	2,9
Reino Unido	2	2,9

Países	F (N.A.)	(%)
Coreia do Sul	1	1,5
Dinamarca	1	1,5
Itália	1	1,5
Jamaica	1	1,5
Japão	1	1,5
Jordão	1	1,5
Holanda	1	1,5
Lituânia	1	1,5
Noruega	1	1,5
Nova Zelândia	1	1,5
Cingapura	1	1,5
Turquia	1	1,5
Subtotal	34	50,0
Total	68	100,0

Com exceção de um estudo com população brasileira, os primeiros autores de todos os estudos tiveram o mesmo país de origem da população analisada. No caso do estudo brasileiro, o primeiro autor era de origem australiana e os demais tinham afiliação no Brasil. Consideramos o primeiro autor o responsável pela correspondência com os leitores, do qual obtivemos o endereço completo para identificação do país. Assim, 98,5% dos estudos latino-americanos são de autoria de latino-americanos.

Foram 24 países latino-americanos aceitos para nosso estudo, de acordo com a classificação da Biblioteca Virtual da América Latina, e temos 25% desses representados na amostra. Brasil e México são os países mais populosos da região, estando o Brasil na 5ª posição mundial, com 198.360.943 habitantes e o México com 116.146.768, em 11º lugar, de acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE)¹². O fato de o Brasil aparecer com 26,5% dos trabalhos e o México com 2,9% pode ser explicado pelos esforços que se tem feito para dar visibilidade ao conhecimento científico gerado nos dois países. Em se tratando da área de Psicologia, no Brasil temos fontes de informação consolidadas, no México temos projetos incipientes e difíceis de avançarem. No próximo capítulo discutiremos a questão das bases de dados desenvolvidas na região e a participação dos países latino-americanos nessa empreitada.

O Brasil e o Chile apareceram nas duas primeiras posições em relação aos estudos latino-americanos, com 26,5% e 11,8% dos trabalhos, respectivamente. A Colômbia, México e Peru contribuíram com dois artigos cada e a Venezuela com um

¹² www.ibge.gov.br/paisesat

estudo. Consideramos um estudo do Caribe como latino-americano, embora o país não estivesse identificado.

Nos dados gerais da revisão sistemática, observamos que os EUA estiveram na primeira posição, com 76 estudos incluídos, seguidos pelo Brasil, com 26 trabalhos aprovados nos critérios, incluindo as dissertações e as teses. Com esse panorama, é natural que quando extraíssemos apenas os estudos latino-americanos o Brasil se posicionasse no topo da lista dos países latino-americanos e os USA dos não latinos. Por ser este um estudo que se preocupa com a relevância das fontes de informação na geração de revisões sistemáticas, voltaremos a discutir os dados na próxima seção desta tese.

Diferentemente da amostra geral da revisão sistemática, que incluiu artigos dos anos de 1980, nossa seleção situa o ano de início das publicações na década de 1990, como observamos da Tabela 12.

Tabela 12 - Distribuição dos artigos conforme o ano de publicação e a origem da revista

Ano de publicação	Revistas latino-americanas	(%)	Revistas não latino-americanas	%	F (N.A)	%
1993	1	4,3	1	2,2	2	2,9
1995	1	4,3	1	2,2	2	2,9
1998	0	0	1	2,2	1	1,5
1999	2	8,7	1	2,2	3	4,4
2000	1	4,3	3	6,7	4	5,9
2001	0	0	3	6,7	3	4,4
2002	1	4,3	1	2,2	2	2,9
2003	1	4,3	2	4,4	3	4,4
2004	1	4,3	5	11,1	6	8,8
2005	1	4,3	3	6,7	4	5,9
2006	2	8,7	4	8,9	6	8,8
2007	1	4,3	2	4,4	3	4,4
2008	4	17,4	7	15,5	10	14,7
2009	4	17,4	5	11,1	9	13,2
2010	3	13,0	2	4,4	5	7,4
2011	0	0	5	11,1	5	7,4
Total	23	100,0	45	100,0	68	100,0

Em consonância com os resultados da amostra geral da revisão sistemática, os anos de 2008 e 2009, também, foram os mais produtivos para os estudos de prevalência de DPP. A amostra geral apresentou 84 trabalhos, de 2008 e 2009. Nossa amostra traz 8 estudos nos mesmos anos.

As revistas científicas se consagraram como o meio, por excelência, para publicações dos resultados de pesquisas. Nesse sentido, conhecer as revistas que publicaram os estudos latino-americanos incluídos na revisão sistemática e compará-las com um grupo de outras regiões é uma forma de conhecer os instrumentos nos quais a produção sobre prevalência de DPP foi escoada no período.

Atualmente, indicadores de uso e impacto são importantes na medição da qualidade das revistas científicas. Desses indicadores, destacamos três, que vêm sendo utilizados pelas agências de fomento e organismos que avaliam a ciência no Brasil, na análise das revistas do nosso estudo, como se segue. O Fator de Impacto (FI), do Journal Citation Reports (JCR), é medido com base nas citações aos periódicos indexados na WOK, nos últimos dois anos. Nosso levantamento foi feito em junho de 2013, no JCR para o FI, consideradas as citações às revistas no ano de 2012. Na mesma data, consultamos o Índice H (H Index) do SCImago Journal & Country Rank, que é gerado pelas citações aos periódicos indexados por Scopus, desde 1996. O FI SciELO é um índice bibliométrico medido por meio das citações dos periódicos publicados no próprio Portal, ou seja: SciELO Brasil considera as revistas publicadas no Portal do Brasil; SciELO Chile se baseia nas revistas do Portal do Chile e assim sucessivamente. Consultamos o SciELO em abril de 2013. Apenas os portais certificados por SciELO possuem o módulo de bibliometria, por isso as revistas publicadas no Peru não possuem FI, já que o país está em busca da certificação. O FI medido neste estudo considerou os últimos dois anos das citações em SciELO.

A Tabela 13 mostra os títulos das revistas em três subgrupos: 1) revistas latino-americanas que publicaram estudos latino-americanos; 2) revistas não latino-americanas que publicaram estudos latino-americanos; 3) revistas não latino-americanas que publicaram estudos não latino-americanos; seguidos do número dos artigos por revista e dos fatores de impacto nos três indexadores: JCR, Scopus e SciELO. Consideramos o país de publicação de acordo com informações do SCImago, ou do Research Gate (www.researchgate.net), rede profissional para

cientistas e pesquisadores que fornece dados detalhados sobre as revistas científicas.

Tabela 13 - Título e país de procedência da revista, frequência de artigos por revista e índices de impacto no JCR (FI), Índice H no SCImago e FI no SciELO

Nome da revista	País	F (N.A.)	JCR FI	SCImago Índice H	SciELO FI
Revistas latino-americanas/estudos latino-americanos					
Revista Chilena de Obstetricia y Ginecología	Chile	5	-	5	0,1
Revista Médica de Chile	Chile	3	0,360	23	0,2
Revista de Saúde Pública	Brasil	2	1,066	38	0,9
Anales de la Facultad de Medicina	Peru	1	-	-	-
Ciência, Cuidado e Saúde	Brasil	1	-	-	-
Colômbia Médica	Colômbia	1	0,094	6	0,1
Gaceta Médica de México	México	1	0,128	11	0
Infanto: Revista de Neuropsiquiatria da Infância e Adolescência	Brasil	1	-	-	-
Jornal Brasileiro de Psiquiatria	Brasil	1	-	10	0,2
Psico USF	Brasil	1	-	-	0,2
Revista Brasileira de Ginecologia e Obstetrícia	Brasil	1	-	7	0,3
Revista Brasileira de Psiquiatria	Brasil	1	1,856	24	0,7
Revista Brasileira de Saúde Materno Infantil	Brasil	1	-	10	0,3
Revista Colombiana de Obstetricia y Ginecología	Colômbia	1	-	2	0,1
Revista de Psiquiatria Clínica	Brasil	1	0,633	10	0,7
Revista Peruana de Epidemiologia	Peru	1	-	-	-
Subtotal		23			

Nome da revista	País	F (N.A.)	JCR FI	SCImago Índice H	SciELO FI
Revistas não latino-americanas/Estudos latino-americanos					
Archives of the Women`s Mental Health	Austria	2	2,009	33	-
Clinical Practice and Epidemiology in Mental Health	EUA	2	-	-	-
Journal of Psychosomatic Obstetrics & Gynecology	Reino Unido	2	1,590	35	-
BMC Psychiatry	Reino Unido	1	2,233	35	-
Health and Social Care in the Community	Reino Unido	1	1,185	32	-
Infant Behavior & Development	Reino Unido	1	1,752	44	-
Lancet	Reino Unido	1	38,278	453	-
Transcultural Psychiatry	Reino Unido	1	0,662	22	-
Subtotal		11			
Revistas não latino-americanas/Estudos não latino-americanos					
BJOG: an International Journal of Obstetrics and Gynaecology	Reino Unido	4	3,760	99	-
Journal of Affective Disorders	Holanda	2	3,295	99	-
Journal of Psychosomatic Research	Reino Unido	2	3,268	90	-
Acta Obstetrica et Gynecologica Scandinavica	EUA	1	1,850	62	-
American Journal of Psychiatry	EUA	1	14,721	224	-
Archives of Womens Mental Health	Áustria	1	2,009	33	-
Atención Primaria	Espanha	1	0,957	24	-
Australian and New Zealand	Nova Zelândia	1	3,293	60	-

Nome da revista	País	F (N.A.)	JCR FI	SCImago Índice H	SciELO FI
Journal of Psychiatry					
BMC Pregnancy and Childbirth	Reino Unido	1	2,516	24	-
British Journal of Clinical Psychology	EUA	1	2,333	55	-
Canadian Family Physician	Canada	1	1,808	-	-
Dreaming: Journal of the Association for the study Of Dreaming	EUA	1	0,789	20	-
International Urogynecology Journal	Reino Unido	1	2,169	53	-
Irish Journal of Psychological Medicine	Irlanda	1	-	10	-
Israel Medical Association Journal	Israel	1	0,978	33	-
Psychiatry and Clinical Neurosciences	Reino Unido	1	2,040	39	-
Journal of Epidemiology & Community Health	Reino Unido	1	3,392	96	-
Journal of Obstetric, Gynecologic and Neonatal Nursing	Reino Unido	1	1,033	34	-
Journal of Reproductive and Infant Psychology	Reino Unido	1	0,606	23	-
Journal of Reproductive Medicine for the Obstetrician and Gynecologist	Reino Unido	1	-	-	-
Maternal and Child Health Journal	Alemanha	1	1,875	39	-
Medicina (Kaunas)	Lituânia	1	0,550	15	-
Midwifery	Reino Unido	1	1,116	30	-
Pacific Health Dialog	Nova Zelândia	1	-	10	-
Pain	Holanda	1	5,644	156	-
Pediatrics	EUA	1	5,119	194	-
Scandinavian Journal of Caring Sciences	Noruega	1	0,961	35	-
Social Psychiatry and Psychiatric	Alemanha	1	2,861	66	-

Nome da revista	País	F (N.A.)	JCR FI	SCImago Índice H	SciELO FI
Epidemiology					
Social Science & Medicine	Reino Unido	1	2,733	129	-
Subtotal			34		
Total de artigos			68		

Dos 34 artigos que tiveram a coleta de dados em países latino-americanos, 23 foram publicados em revistas latino-americanas e 11 em revistas de países de outras regiões.

O Brasil é o responsável pelo maior número de contribuições em relação aos países latino-americanos, com 10 artigos publicados em nove revistas brasileiras e oito publicados em revistas não latino-americanas. Destacamos a Revista de Saúde Pública, com dois artigos, que é a única a apresentar indicadores de impacto nos três índices: FI de 1,066; Índice H de 38 e FI SciELO Brasil de 0,9. Vale mencionar a única revista especializada em Psicologia da lista, a Psico USF, com 0,2 de FI no SciELO. O tema estudado na revisão sistemática não circulou intensamente pelas revistas de Psicologia como esperávamos, uma vez que as revistas da área estão visíveis e acessíveis abertamente em portais como SciELO, PePSIC, RedAllyc e indexadas em muitas bases de dados, como podemos confirmar no estudo de Sampaio e Sabadini (2012).

Dos países latino-americanos, o Chile também se destaca com oito artigos publicados, dos 68 avaliados. Na primeira posição encontramos a Revista Chilena de Obstetricia y Ginecología, que publicou cinco artigos do próprio país. A Revista Médica do Chile, com FI de 0,36, Índice H de 23 e FI de 0,2 no SciELO, aparece em segundo lugar, dentre as revistas publicadas em países da América Latina, com três artigos.

Colômbia, México e Peru completam o rol dos países latinos que tiveram revistas publicando estudos com população latina. A Colômbia Médica, FI de 0,094, Índice H de 6 e FI de 0,2 no SciELO, e a Revista Colombiana de Obstetricia y

Ginecologia, Índice H de 2 e FI no Scielo de 0,1, publicaram um artigo cada. Anales de la Facultad de Medicina e Revista Peruana de Epidemiologia, as duas sem nenhum indicador de citação no JCR, publicaram um artigo cada. Vale destacar que somente as duas revistas do Peru, e duas brasileiras não são publicadas em acesso aberto no SciELO.

O Reino Unido contribuiu com 47% dos artigos, dos 34 estudos não latino-americanos estabelecidos para comparação. Contribuindo com um artigo na amostra, Lancet, uma das mais importantes revistas do mundo, publicada no Reino Unido, é a que possui o maior FI no *ranking*: 38,278. A mesma publicação lidera em relação ao H Índice: 453. O artigo publicado na Lancet é de autoria de brasileiros. American Journal of Psychiatry, publicada nos EUA e segunda colocada nos indicadores bibliométricos aqui utilizados, surge com uma contribuição, tendo Índice H de 224 e 14,721 de FI.

Visando a conhecer quais países da região estão escoando seus estudos para o exterior, verificamos quais foram os países que publicaram seus artigos em revistas estrangeiras. A Tabela 14 apresenta os títulos dessas revistas e a frequência das publicações.

Tabela 14 - Título da revista estrangeira que publicou os estudos latino-americanos, país e número de publicação

Nome da revista	País do Estudo	F (N.A.)
Archives of the Women's Mental Health	Brasil	2
BMC Psychiatry	Brasil	1
Clinical Practice and Epidemiology in Mental Health	Brasil	1
Clinical Practice and Epidemiology in Mental Health	México	1
Infant Behavior & Development	Brasil	1
Health and Social Care in the Community	Caribe	1
Journal of Psychosomatic Obstetrics & Gynecology	Brasil	1
Journal of Psychosomatic Obstetrics & Gynecology	Chile	1
Lancet	Brasil	1
Transcultural Psychiatry	Brasil	1
Total		11

O Brasil foi o país latino-americano com mais estudos publicados em revistas não nacionais: oito artigos. Chile, México e a região do Caribe tiveram um artigo publicado em revista estrangeira. Com 16% dos artigos latinos publicados em

revistas estrangeiras tornou-se interessante conhecer o número de citações recebidas pelos estudos. Oito revistas indexadas na WOK, uma vez que, com exceção da *Clinical Practice and Epidemiology in Mental Health*, todas possuem FI no JCR e todas apareceram registradas na Scopus, já que têm Índice H no SCImago. Analisamos se possuem citações nas duas bases de dados apontadas anteriormente e no Google Scholar (www.scholar.com), ou Google Acadêmico, como é conhecido no Brasil. Nos últimos anos, acompanhamos um forte movimento em prol das publicações de artigos de autores brasileiros em língua inglesa, sejam eles em revistas brasileiras que publicam em inglês, ou em revistas estrangeiras.

Os argumentos para esse movimento, que aponta para a publicação em inglês, giram em torno da limitação da leitura do português por pesquisadores que não entendem nossa língua. O debate sobre publicar em inglês levanta questionamentos sobre a convicção de que uma das funções centrais das revistas é disseminar o conhecimento e que, em busca de maior interlocução com os pesquisadores estrangeiros, publicam matérias em inglês quando os temas interessaram ao público local de um determinado país. (Lopes & Piscitelli, 2004). Na verdade, o inglês se tornou a língua franca da ciência em um mundo que é economicamente, cientificamente e culturalmente dominado pelos países anglo-americanos, explicam os autores. Todo cientista deve dominar o inglês, pelo menos em certa medida, para obter o reconhecimento internacional e para acessar as publicações relevantes e atuais. (Meneghini & Packer, 2007).

O Google Acadêmico lançou suas métricas em 2002 e passou a oferecer um novo sistema para avaliação bibliométrica de revistas científicas. Aprimorando seu sistema de coleta de citações aos trabalhos científicos, vem se tornando um importante instrumento de medição do número de citações aos trabalhos, com a vantagem sobre as ferramentas tradicionais dessa natureza, especificamente a WOK e o Sciverse Scopus, que abrange o Scopus e o SCImago. A limitação das duas bases de dados decorre do fato de fazerem um recorte bem seletivo das revistas publicadas no mundo. Segundo Rodrigues e Oliveira (2012), o Ulrich, lista mais abrangente de revistas com revisão por pares do mundo, reunia 92 mil títulos no ano 2011, quando as bases foram consultadas. Desses, a WOK indexava 10 mil títulos e Scopus 19 mil. Dada a restrição dessas fontes de informação o Google Acadêmico surgiu como uma alternativa “democrática” para medir as citações e vem aparecendo em estudos bibliométricos. Como exemplos de estudos ilustramos o

trabalho de Pinto e Ponte Júnior (2008) que analisaram a visibilidade de autores da área de ciência da informação na WOK e no Google Acadêmico, e o de Delgado López-Cozar e Repiso (2013), cujo objetivo foi comparar as citações das revistas da área de comunicação em três fontes de informação: Scopus, WOK e Google Acadêmico. Assim como os autores do segundo exemplo, recorreremos ao Google Acadêmico, à Scopus e à WOK para analisar as citações aos estudos latino-americanos publicados em revistas não latino-americanas.

Tabela 15 - Título do estudo latino-americano, revista estrangeira que publicou o artigo, país de origem do estudo, número de citações no Google Acadêmico, Scopus e WOK e ano de publicação.

Titulo do artigo e da revista	País do estudo	Citações no Google Acadêmico	Citações Scopus	Citações WOK	Ano de publicação
Validation of Edinburgh Postnatal Depression Scale (EPDS) in Chilean post-partum women. <i>Journal of Psychosomatic Obstetrics & Gynecology</i> .	Chile	85	6	57	1985
Postnatal depression in Southern Brazil: prevalence and its demographic and socioeconomic determinants. <i>BMC Psychiatry</i> .	Brasil	65	26	19	2011
Postpartum depression: in relation to life events and patterns of coping. <i>Archives of Women's Mental Health</i> .	Brasil	64	27	Não consta	2004
Violence against women by their intimate partner during pregnancy and postnatal depression: a prospective cohort study. <i>The Lancet</i> .	Brasil	61	34	30	2010
Perinatal depression among black Caribbean women. <i>Health and Social Care in the Community</i> .	Caribe	49	33	28	2004
Translation, validation and cultural aspects of postpartum depression screening scale in Brazilian Portuguese. <i>Transcultural Psychiatry</i> .	Brasil	29	16	Não consta	2007
Validation of the Edinburgh Postpartum Depression Scale in a population of puerperal women in Mexico. <i>Clinical Practice and Epidemiology in Mental Health</i>	México	21	16	Não consta	2006

Titulo do artigo e da revista	País do estudo	Citações no Google Acadêmico	Citações Scopus	Citações WOK	Ano de publicação
Validation of a screening instrument for postpartum depression in Southern Brazil. <i>Journal of Psychosomatic Obstetrics & Gynecology</i> .	Brasil	12	8	6	2009
Postpartum depression according to time frames and sub-groups: a survey in primary health care settings in Rio de Janeiro, Brazil. <i>Archives of Women's Mental Health</i> .	Brasil	5	2	1	2011
Differentials and income-related inequalities in maternal depression during the first two years after childbirth: birth cohort studies from Brazil and the UK. <i>Clinical Practice and Epidemiology in Mental Health</i> .	Brasil	7	3	Não consta	2009
Chronicity and severity of maternal postpartum depression and infant sleep disorders: A population-based cohort study in southern Brazil. <i>Infant Behavior and Development</i>	Brasil	2	1	1	2011

As consultas aos índices bibliométricos foram feitas em 19 de junho de 2013. Por ser o Google Acadêmico a fonte de informação que apresenta os maiores índices de citações aos artigos, hierarquizamos nossos artigos de acordo com suas ocorrências.

O Chile foi o país que teve o artigo com mais citações no Google Acadêmico (85) e que recebeu 57 citações na WOK e seis no Scopus. O Brasil ficou posicionado no segundo, terceiro e quarto lugar, com os artigos citados entre 65 e 61 vezes. Os mesmos trabalhos tiveram 26, 27 e 34 citações, respectivamente, no Scopus. Destes destacamos o artigo publicado na BMC Psychiatry, que com dois anos de publicação, recebeu 65 citações no Google Acadêmico, 26 no Scopus e 19 na WOK. O artigo publicado na Lancet, revista com os maiores índices de citações no JCR e no Scopus, embora com 61 citações no Google Acadêmico, foi o que apresentou mais citações no Scopus (34) e na WOK (30), observando sua data de publicação, 2010, essas marcas foram obtidas nos últimos dois anos. Os trabalhos assinalados como “não consta” foram publicados em fascículos anteriores à indexação da revista na WOK, por isso não aparecem com dados de citações.

Com o objetivo de avaliar o impacto dos estudos latino-americanos publicados em revistas latino-americanas, analisamos as citações aos artigos nos três índices de citações.

Tabela 16 - Título do artigo e da revista, país de origem, número de citações no Google Acadêmico, WOK e Scopus e ano de publicação

Título do artigo e da revista	País da revista	Número de citações no Google Acadêmico	Citações Scopus	Citações WOK	Ano de publicação
Rastreamento da depressão pós-parto em mulheres atendidas pelo Programa de Saúde da Família. <i>Revista Brasileira de Ginecologia e Obstetrícia</i> .	Brasil	108	Não consta	Não indexada	2005
Prevalência da depressão pós-parto e fatores associados. <i>Revista de Saúde Pública</i> .	Brasil	88	23	Não consta	2006
Depressão pós-parto: prevalência e correlação com o suporte social. <i>Infanto: Revista de Neuropsiquiatria da Infância e Adolescência</i> .	Brasil	21	1	Não indexada	2002
Prevalencia de depresión postparto y factores asociados en Santiago, Chile. <i>Revista Médica de Chile</i> .	Chile	19	18	23	1995
Depresión postparto realidad en el sistema público de atención de salud. <i>Revista Chilena de Obstetrícia y Ginecología</i> .	Chile	18	Não indexada	Não consta	2003
Depresión postparto, apoyo social y calidad de vida en mujeres de Cali, Colombia. <i>Colombia Médica</i> .	Colômbia	17	Não consta	Não consta	2000
Depresión postparto: prevalencia de test de rastreo positivo en puérperas del Hospital Universitario de Brasília, Brasil. <i>Revista Chilena de Obstetrícia y Ginecología</i> .	Chile	14	Não indexada	Não consta	2004

Título do artigo e da revista	País da revista	Número de citações no Google Acadêmico	Citações Scopus	Citações WOK	Ano de publicação
Escala de Depressão Pós-natal de Edimburgo para triagem no sistema público de saúde. <i>Revista de Saúde Pública</i> .	Brasil	15	11	Não consta	2009
Depressão pós-parto em Recife – Brasil: prevalência e associação com fatores biossociodemográficos. <i>Jornal Brasileiro de Psiquiatria</i>	Brasil	13	Não consta	Não indexada	2010
A estrutura das representações sociais de mães puérperas acerca da depressão pós-parto. <i>Psico USF</i>	Brasil	13	Não consta	Não indexada	2007
Aspectos gineco-obstétricos em mujeres que desarrollan una depresión em el postparto. <i>Revista Colombiana de Obstetricia y Ginecología</i> .	Colômbia	10	Não consta	Não indexada	1993
La depresión postparto en madres adolescentes de Bucaramanga, Colombia. <i>Revista Chilena de Obstetricia y Ginecología</i> .	Chile	8	Não consta	Não indexada	2006
Mejorando la pesquisa de depresión posparto a través de un instrumento de tamizaje, la escala de depresión posparto de Edimburgo. <i>Revista Médica de Chile</i> .	Chile	6	5	Não consta	2008
Prevalencia y factores de riesgo asociados a la depresión posparto en puérperas de un Centro de Atención Primaria en el Sur de Chile. <i>Revista Médica de Chile</i> .	Chile	6	3	0	2008
Prevalencia de depresión posnatal en mujeres atendidas en hospitales públicos de Durango, México. <i>Gaceta Médica de México</i> .	México	5	7	Não consta	2010

Título do artigo e da revista	País da revista	Número de citações no Google Acadêmico	Citações Scopus	Citações WOK	Ano de publicação
Qualidade de vida na depressão pós-parto na adolescência. <i>Revista Brasileira de Psiquiatria</i>	Brasil	0	0	Não consta	Não consta
Depresión posparto: ¿se encuentra asociada a la violencia basada en género? <i>Anales de la Facultad de Medicina</i>	Peru	4	Não Indexada	Não indexada	2009
Depresion postparto: prevalencia y factores de riesgo. <i>Revista Chilena de Obstetricia y Ginecología</i> .	Chile	3	Nao consta	Não indexada	1999
Prevalencia de depresión posparto en primigestas y multiparas valoradas por la Escala de Edimburgo. <i>Revista Chilena de Obstetricia y Ginecología</i> .	Chile	2	1	Não indexada	2010
Prevalencia y factores asociados a depresión posparto en mujeres atendidas en establecimientos de salud del primer nivel de atención en Lima Metropolitana, junio 2004. <i>Revista Peruana de Epidemiologia</i> .	Peru	2	Não indexada	Não indexada	2008
Alteração tireoidiana: um fator de risco associado à depressão pós-parto. <i>Revista Brasileira de Saúde Materno Infantil</i> .	Brasil	1	0	Não indexada	2009
Investigação dos fatores indicativos de depressão pós-parto em dois grupos de puérperas. <i>Revista Ciência, Cuidado e Saúde</i> .	Brasil	0	Não indexada	Não indexada	2009
Escalas de auto-avaliação de depressão pós-parto: estudo no Brasil. <i>Revista de Psiquiatria Clínica</i> .	Brasil	0	Não consta	Não consta	1999

O artigo mais citado no Google Acadêmico (108 citações) foi publicado na Revista Brasileira de Ginecologia e Obstetrícia, que não é indexada na WOK e não aparece na busca do Scopus. Em seguida, posicionado com 88 citações na mesma

ferramenta, aparece um artigo da Revista de Saúde, que embora indexada na WOK, a busca retornou como não encontrado. Os artigos são indexados a partir do fascículo avaliado e aprovado para indexação nas bases de dados. Por isso, mesmo a revista sendo indexada, os artigos anteriores à data de aprovação da revista não são incluídos. Os estudos publicados na Revista Médica do Chile foi o único que apresentaram citações nos três índices consultados: 19 no Google Acadêmico, 18 no Scopus e 23 na WOK.

Mesmo sem indexação na WOK e sem presença no Scopus, pois a base foi lançada no ano de 2004 e muitos trabalhos são anteriores a esta data, os artigos publicados em revistas latino-americanas apresentam citações no Google Acadêmico. Não existem parâmetros que permitam comparar os indicadores de citações do Google Acadêmico, com as outras duas ferramentas, mas podemos dizer que a ferramenta está mudando o comportamento da bibliometria no mundo e que, em breve, teremos indicadores importantes e amplos o suficiente para analisarmos o impacto de um texto na comunidade científica.

Indicadores bibliométricos de citações ganharam a adesão de pareceristas na avaliação de projetos de pesquisas, na seleção de alunos para programas de pós-graduação e utilizados em outros sistemas de avaliação de autores. O estudo científico não finaliza com sua publicação, pois, para ser valorizado o trabalho precisa ser lido e citado, retroalimentando a cadeia do conhecimento científico. Apesar da relevância dos indicadores de citações, avaliar a qualidade metodológica e da apresentação dos resultados dos estudos publicados em formato de artigos é uma prática que deve ser incorporada às nossas atividades, se quisermos avançar rumo ao uso das evidências na prática profissional e na tomada de decisão em gestão pública.

5.5 Qualidade metodológica e da apresentação dos resultados dos estudos

A qualidade metodológica e da apresentação dos resultados dos estudos foram avaliadas pela escala de Loney et al. (1998), que analisa pesquisas sobre prevalência ou incidência, fornecendo uma pontuação que varia de zero a oito. A cada questão com resposta afirmativa, foi atribuído um ponto, totalizando oito pontos para o artigo que atingisse todos os quesitos.

Optamos por usar a Escala de Loney et al. (1998) por sua pontuação permitir a comparação estatística dos artigos, além de sua fácil aplicação, uma vez feitos os treinamentos sobre os itens aferidos. Quando definimos o instrumento de avaliação, juntamente com Silva (2013), não tínhamos conhecimento de um instrumento de avaliação de qualidade de estudos de prevalência, elaborado posteriormente, no ano de 2012.

Giannakopoulos et al. (2012), construíram uma escala que parece mais flexível na pontuação dos itens que avaliam estudos de prevalência e que acrescenta um novo item para medir a metodologia e os resultados de pesquisas. Os autores chamaram a Escala de um novo instrumento para avaliar a qualidade dos estudos sobre prevalência. A Escala é baseada, principalmente, em critérios estatísticos e dividida em três blocos: amostragem, medição e análise. Os pontos atribuídos para cada um dos critérios são somados para formar um Índice de Qualidade Total (TQS). A Escala pontua os trabalhos em quatro subgrupos: 0-4 (pobre), 5-9 (moderado), 10-14 (bom) e 15-19 (excelente).

Comparamos a Escala de Loney et al. (1998) com a de Giannakopoulos et al. (2012) e observamos bastante semelhança nos itens de avaliação, tais como: tipo de estudo, representatividade da amostra e método de amostragem, características da população alvo, taxa de resposta e descrição dos participantes perdidos, critérios de inclusão e exclusão dos participantes, aplicação do instrumento para medição do transtorno, procedimentos estatísticos, considerando o intervalo de confiança do resultado de prevalência. A Escala de Giannakopoulos et al. (2012) incluiu um item para verificar se o estudo foi aprovado pela comissão de ética em pesquisa, incentivando os pesquisadores a protegerem os indivíduos em estudos futuros. A Escala foi testada por dois dentistas e obteve boa concordância entre os avaliadores que analisaram 400 artigos sobre transtorno temporomandibular, selecionados aleatoriamente. Por possuir uma escala com mais pontos, o novo instrumento oferece melhores possibilidades de avaliação dos estudos em decorrência de sua maior capacidade de discriminação. Entretanto, encontramos o artigo relatando o teste da nova escala no final do ano de 2012, quando nossos dados já haviam sido coletados.

Importante informar que a análise de concordância foi efetuada com base nos estudos que tiveram coleta de dados em países latino-americanos e na amostra aleatorizada de estudos com coleta de dados em outras regiões. Dessa maneira, a

concordância foi estabelecida com 68 estudos latino-americanos e 68 não latino-americanos. A avaliação das duas revisoras teve índices de concordância aceitáveis, sendo que o item sete (apresentação dos resultados de prevalência com intervalo de confiança e detalhamento de subgrupos) foi o único que ficou abaixo do considerado como aceito para a revisão sistemática, com taxas de concordância de 60,5% para estudos da América Latina e 50,0% os estudos de outros países. O item 5 (Desfecho medido de forma a apresentar baixo risco de viés) teve concordância de 100%, tanto para os estudos latino-americanos, quanto para os não latino-americanos. Os itens um e dois (Delineamento e método de amostragem e Base amostral (*sampling frame*) receberam frequências de 97,2% das duas revisoras. Embora com treinamentos para o uso da Escala, o entendimento dos itens, muitas vezes, não era tão claro e interferiu na análise de algumas das questões, como no item 7. As instruções sobre o preenchimento da Escala (Apêndice A) e as reuniões de consenso foram importantes instrumentos para garantir a diminuição do vieses nessa etapa da pesquisa. A Tabela 17 apresenta o grau de concordância em cada item da Escala. Vale lembrar que os itens de concordância se referem aos 68 estudos latino-americanos e aos 68 não latino-americanos, chamado de grupo controle. Os dados dizem respeito aos estudos aprovados nos itens de inclusão da revisão sistemática. Nesses números estão as 10 dissertações e teses excluídas de nosso estudo de avaliação da qualidade de artigos. Dos 68 estudos não latino-americanos (grupo controle) foram incluídos 36, os quais foram avaliados segundo a Escala de Loney et al. (1998). Assim, a concordância foi feita com base em estudos latino-americanos (N= 44) e não latino-americanos (N= 36).

Tabela 17 - Concordância entre avaliadoras para os itens da Escala de Loney et al. (1998)

Item	Nível de concordância	
	Estudos latino-americanos	Estudos não latino-americanos
	N=44	N=36
1. O delineamento do estudo e os métodos de amostragem são apropriados para responder à pergunta de pesquisa?	83,7	97,2
2. A base amostral é adequada?	86,0	97,2
3. O tamanho da amostra é adequado?	90,7	86,1
4. São usados critérios objetivos, adequados e padronizados para medir o desfecho de saúde?	72,1	77,8
5. O desfecho de saúde é medido de forma não enviesada?	100,0	100,0
6. A taxa de resposta é adequada? As pessoas elegíveis que se recusaram a participar são descritas?	79,1	83,3
7. As estimativas de prevalência, ou incidência, são dadas com intervalos de confiança e detalhadas por subgrupo, quando adequado?	60,5	50,0
8. Os participantes e o contexto são descritos em detalhe e podem ser generalizados para outras situações similares?	93,0	91,7

Nota. Tabela adaptada de Silva (2013) para o nosso estudo.

Os itens de um a seis da Escala se referem à qualidade metodológica dos estudos de prevalência e o sete e oito dizem respeito à interpretação e aplicabilidade dos resultados. Os itens um e dois tiveram concordância maior para os não latino-americanos: 97,2% nos dois quesitos, ao contrário do três que teve vantagem para os estudos latino-americanos (90,7%). O item quatro ficou próximo da concordância estabelecida para o estudo de 70% e o cinco teve 100% de concordância. O item seis recebeu maior concordância para os não latino-americanos e o sete ficou abaixo do limite para os não latino-americanos. A concordância ficou acima de 90% no item oito da Escala.

Para o estudo comparativo foram avaliados 68 estudos, sendo 34 com coleta de dados em países latino-americanos e 34 em outras regiões, que foram publicados em 23 revistas latino-americanas e em 45 revistas de outras regiões, como ilustra a Figura 5.

Figura 5. Universo do estudo comparativo.

O universo total dos 68 artigos eleitos para avaliação da qualidade apresentou frequências concentradas em três itens, tendo os demais baixas pontuações, como pode ser observado na Figura 6.

Figura 6. Frequência total das pontuações aos artigos (68), em função da Escala de Loney et al. (1998).

Assim como a avaliação geral dos 337 trabalhos da revisão sistemática, nenhum dos 68 artigos obteve a pontuação máxima (8 pontos). Um artigo teve nota zero (0), ou seja, não pontuou em nenhum item da Escala. As notas variaram de zero a cinco, com maior concentração de estudos com um, dois e três pontos. Com os escores obtidos, podemos afirmar que nossos artigos tiveram desempenho fraco na Escala de Loney et al. (1998). O mesmo aconteceu com a conclusão da avaliação dos estudos na revisão sistemática (Silva, 2013).

A Tabela 18 apresenta as estatísticas em relação à adequação e inadequação dos itens metodológicos e da apresentação dos resultados. Cada um dos itens da Escala foi analisado a seguir.

Tabela 18 - Número e frequência de artigos que obtiveram pontuações em cada item da Escala de Loney et al. (1998)

Itens da Escala	Adequado		Inadequado	
	N.A.	%	N.A.	%
Item 1	3	4,4	65	95,6
O delineamento do estudo e os métodos de amostragem são apropriados para responder à pergunta de pesquisa?				
Item 2	2	3,0	66	97,0
A base amostral é adequada?				
Item 3*	16	23,5	52	76,5
O tamanho da amostra é adequado?				
Item 4	14	20,6	54	79,4
São usados critérios objetivos, adequados e padronizados para medir o desfecho de saúde?				
Item 5	0	0	68	100,0
O desfecho de saúde é medido de uma forma enviesada?				
Item 6	10	14,7	58	85,3
A taxa de resposta é adequada? As pessoas elegíveis que recusaram participar são descritas?				
Item 7*	12	17,6	56	82,3
As estimativas de prevalência ou incidência são dadas com intervalos de confiança e detalhadas por subgrupo, se adequado?				
Item 8	66	97,0	2	3,0
Os participantes e o contexto são descritos em detalhes e podem ser generalizados para outras situações similares?				

Nota. *Itens com diferença estatisticamente significativa em relação à adequação de cada item da Escala de Loney, et al (1998), ($p < 0,05$, no teste de qui-quadrado).

Os 68 artigos do nosso universo obtiveram escores fracos em relação aos oito critérios estabelecidos na avaliação metodológica e de apresentação dos resultados. Não foram observadas diferenças significativas entre as médias das pontuações na Escala de Loney et al. (1998), sendo marcados os itens três e sete da Escala com algum nível de significância, em relação à adequação de cada item da Escala. A seguir cada item e suas respectivas pontuações são analisados.

Item 1 - O primeiro item da Escala (delineamento do estudo e métodos de amostragem apropriados para responder à pergunta de pesquisa) recebeu 95,6% de inadequação. Para que o estudo fosse considerado adequado deveria ser primário e ter como desenho de pesquisa o delineamento transversal (as participantes foram submetidas a apenas uma avaliação ao longo do tempo) ou de coorte longitudinal (as participantes foram seguidas ao longo do tempo e submetidas a mais de uma avaliação, em diferentes períodos). Alguns estudos transversais avaliavam as participantes dois meses após o parto, outros cinco meses, e outros 10 meses. O estudo longitudinal deveria avaliar a mãe pelo menos duas vezes, sendo que uma das avaliações deveria estar compreendida entre o oitavo, até o dia 365, do pós-parto. O delineamento transversal ou longitudinal era um dos critérios de inclusão dos estudos. Estudos de caso-controle e retrospectivos não foram incluídos na revisão sistemática, por não atenderem ao critério A de inclusão e, portanto, não foram avaliados pela Escala de Loney et al (1998).

Assim, o primeiro item da escala avaliou apenas os métodos de amostragem, recebendo pontuação quando a base amostral era aleatória ou censitária. A amostra foi considerada aleatória quando havia descrição da metodologia usada para aleatorização. O estudo que apenas declarasse que a amostra era aleatória, ou "*random sample*", não era considerado suficientemente adequado para pontuar na Escala e era classificado como "não está claro" em relação ao item. Amostra censitária era aquela que avaliou todas as mães de uma determinada localidade geográfica (município, vila, estado), que deram à luz a filhos vivos em um determinado período. Mesmo declarando entrevistar todas as mães de uma única

unidade, por exemplo, uma unidade de saúde específica, o recorte não era considerado requisito para a amostra censitária.

Item 2 - Em relação à base amostral ser adequada, critério dois da Escala, nossos artigos pontuaram com 97,0% de inadequação. O critério foi bastante rígido e reprovou grande parte dos estudos que limitavam as participantes às maternidades e aos ambulatorios nos quais as mães davam continuidade ao atendimento pós-parto. A prevalência é o número de pessoas com um problema de saúde, em um determinado período, dividido pelo total da população num mesmo momento, assim, a amostra deve ser representativa e mostrar um retrato adequado da população. De que adianta saber a prevalência em um determinado hospital quando as medidas deveriam ser feitas para a população em geral, de uma cidade, por exemplo?

Item 3 - O tamanho da amostra pontuava quando a amostra contava com 470 participantes, ou mais, pelo menos em uma das avaliações da DPP. Nossas unidades de análise tiveram 23,5% de adequação e 76,5% de inadequação, sendo a diferença considerada significativa estatisticamente no teste de qui-quadrado: $p < 0,05$. A amostra mínima de 470 participantes fez que muitos estudos não pontuassem na Escala. O número alto de participantes exigido pelo critério tirou ponto de diversos estudos. O ideal seria que cada estudo relatasse como planejou o tamanho amostral nas condições e contexto do estudo e que a escala avaliasse tal procedimento, sem se fixar apenas no tamanho da amostra.

Item 4 - Em relação ao uso de critérios objetivos, adequados e padronizados para medir o desfecho de saúde, nossos estudos obtiveram 20,6% de adequação contra 79,4% de inadequação. Caso o estudo utilizasse mais de um instrumento de avaliação, deveria declarar a validade e a confiabilidade de pelo menos um deles. A Escala valorizou a medida de desfecho por entender que devia ser reproduzível e válida com alta sensibilidade e especificidade. Como era necessário cumprir os dois quesitos: validade e confiabilidade, diversos estudos que apresentaram apenas uma medida perderam ponto no quesito.

Item 5 - Ainda em relação à avaliação da qualidade metodológica dos artigos, o Item cinco da Escala perguntava se o desfecho de saúde era medido de uma forma não enviesada e nenhum trabalho obteve pontuação, ficando o item com 100% de inadequação na avaliação. Nesse quesito, a Escala teve um grau de exigência difícil de ser alcançado pelos artigos. Dos 337 trabalhos da revisão sistemática apenas seis pontuaram no item, confirmando ser um critério de muita exigência metodológica (Silvia, 2013).

Item 6 – O último item para avaliar a qualidade metodológica dos artigos se referia à taxa de resposta das participantes e, se quando elegíveis, e se recusaram a participar do estudo, era descrito o motivo da não participação. Foi de 85,3% o percentual de inadequação do item metodológico na amostra, contra 14,7% de adequação. A pontuação do item demandava o cumprimento de dois requisitos: a taxa de perda deveria ser menor que 30% e deveria ser feita a comparação entre respondentes e não respondentes. O item foi pontuado mesmo quando a comparação indicava diferença significativa entre os dois grupos (respondentes e não respondentes). A revisão sistemática de Silva (2013, p. 140) apontou 15,5% de adequação na amostra de 337 trabalhos, dados muito próximos dos obtidos por nossos artigos.

Item 7 – Quando verificado se as estimativas de prevalência ou incidência foram dadas com intervalos de confiança e detalhadas por subgrupo, se adequado ao estudo, obtivemos 82,3% de inadequação, contra 17,3% de artigos adequados. Para o artigo pontuar nesse item, deveria registrar nos resultados da prevalência o intervalo de confiança (IC), e não apenas a estimativa pontual. A leitura dos dados do artigo demandava atenção ao avaliar o item, pois era necessário observar que o IC se referia à prevalência e não a outras medidas, como as *Odds ratio* referentes a fatores de risco e tão comum nos artigos analisados. Era essencial, também, verificar se os subgrupos importantes do estudo foram considerados, de forma que fornecessem resultados de prevalência específicos para esses subgrupos, com seus respectivos IC. O item fornecia um dos dados mais relevantes para a revisão sistemática, pois era aqui que as taxas de prevalência apareciam. Na amostra geral de Silva (2013, p. 141), apenas 11,9% dos estudos pontuaram no item, ficando abaixo da marca dos nossos artigos. Os estudos latino-americanos apresentaram IC

com maior frequência. O item avaliou um fator de suma importância na apresentação dos resultados de estudos de prevalência e nossos artigos tiveram um desempenho muito aquém do desejado.

Item 8 – Este foi o item mais pontuado na Escala e diz respeito ao perfil sociodemográfico da população do estudo. Perguntava se os participantes e o contexto foram descritos em detalhes e se podiam ser generalizados para outras situações similares. Obtivemos 97,0% dos estudos pontuando no item. Dados como idade, estado civil, escolaridade, renda, paridade e outros dados sociodemográficos apareceram em quase todos os 68 artigos, mantendo a proximidade com os resultados da revisão sistemática, que apontou 94,1% de pontuação no quesito.

5.6 Testes de hipóteses

Nosso estudo estabeleceu três hipóteses, que repetimos a seguir e que serão aqui testadas.

H1: Artigos científicos que tiveram a coleta de dados em países latino-americanos e em outras regiões possuem a mesma qualidade, de acordo com a Escala de Loney et al. (2008).

H2: Artigos científicos publicados em revistas não latino-americanas e de outras regiões possuem a mesma qualidade, de acordo com a Escala de Loney et al. (2008).

H3: Artigos que declararam ter como objetivo avaliar a prevalência de DPP têm qualidade significativamente diferente dos que não declararam ter esse objetivo.

A figura a seguir, compara o desempenho dos estudos latino-americanos e não latino-americanos, de acordo com os itens da Escala.

Figura 7. Comparação das pontuações totais obtidas na Escala de Loney et al. (1998) pelos artigos latino-americanos e de outras regiões (em números de artigos)

Os estudos latino-americanos tiveram pontuação em cinco quesitos, contra três dos não latino-americanos no teste de hipótese aplicado aos nossos artigos. Um não latino-americano não pontuou em nenhum item da Escala. A frequência se manteve maior nas pontuações um e dois da Escala.

As diferenças entre a pontuação na Escala das categorias latino-americanos e não latino-americanos foram testadas usando testes não-paramétricos. O teste de medianas foi realizado considerando como hipótese nula que os estudos latino-americanos e não latino-americanos teriam as mesmas medianas. Essa hipótese foi aceita com base no p-valor ($p=0.510$). O teste de distribuição de Mann Whitney (U), também indicou aceitação da hipótese nula ($p=0,822$) de que ambas as categorias teriam a mesma distribuição. Assim, os dados confirmam nossa hipótese, de que os artigos científicos que tiveram a coleta de dados em países latino-americanos e os que coletaram dados em países não latino-americanos tiveram pontuações semelhantes, de acordo com a Escala de Loney et al. (1998). Os dados mostraram que os estudos das duas categorias tiveram desempenho fraco.

Nosso estudo também teve como um dos objetivos comparar a qualidade metodológica e da apresentação dos resultados dos artigos publicados em revistas não latino-americanas e latino-americanas.

Figura 8. Comparação das pontuações totais obtidas na Escala de Loney et al. (1998) pelas revistas latino-americanas e não latino-americanas (em números de artigos)

Nota. RNL = Revistas não latino-americanas. RL = Revistas latino-americanas.

As diferenças entre as pontuações das duas categorias: revistas latino-americanas e não latino-americanas, foram testadas usando testes não paramétricos¹³. O teste de medianas foi realizado considerando como hipótese nula que as duas teriam as mesmas medianas. Essa hipótese foi aceita com base no p-valor ($p=0.396$). O teste de distribuição de Mann Whitney (U), também indicou aceitação da hipótese nula ($p=0,144$), de que ambas as categorias teriam a mesma distribuição. Assim, a segunda hipótese de nosso estudo é confirmada, uma vez que os artigos científicos publicados em revistas não latino-americanas e latino-americanas tiveram o escore equivalente na Escala de Loney et al. (2008)

¹³A opção por testes não paramétricos foi efetuada porque as distribuições não são normais conforme o Teste de distribuição de Kolgomorov-Smirnov (KS) ($p<0,001$).

O objetivo do estudo é um dos elementos mais importantes na determinação da metodologia adequada para responder à pergunta de pesquisa. A Escala de Loney et al. (1998) é específica para avaliar estudos de prevalência. Dessa maneira, trabalhos com objetivos de analisar a prevalência de DPP deveriam ter desempenho diferenciado daqueles com outros objetivos. Assim, nossa terceira hipótese indicou que artigos que declararam ter como objetivo avaliar a prevalência de DPP teriam escore significativamente diferente daqueles que não declararam ter esse objetivo. A Figura 9 ilustra o resultado do teste da hipótese três deste trabalho.

Figura 9. Pontuações dos artigos em relação ao objetivo do estudo.

As diferenças entre as pontuações dos artigos que declararam e que não declararam ter como objetivo avaliar a prevalência de DPP foram testadas usando testes não paramétricos¹⁴. O teste de medianas foi realizado considerando como hipótese nula que não haveria diferença significativa entre as categorias. Tal hipótese foi negada com base no p-valor ($p=0.815$). O teste de distribuição de Mann Whitney (U), também negou a hipótese nula ($p=0,790$), de que ambas as categorias teriam a distribuição diferente. Dessa maneira, nossa hipótese confirmou a hipótese nula em relação ao objetivo interferir na qualidade metodológica e na apresentação de resultados, medidos pela Escala de Loney et al. (1998). Nos dados gerais da revisão sistemática, Silva (2013) encontrou que os objetivos mais frequentes foram

¹⁴A opção por testes não paramétricos foi efetuada porque as distribuições não são normais conforme o Teste de distribuição de Kolgomorov-Smirnov (KS) ($p<0,001$).

investigar fatores de risco para DPP (52,2%); investigar prevalência de DPP ficou com 48,1% dos estudos. Com base nos dados observamos que o objetivo voltado a avaliar a prevalência, não interferiu na qualidade metodológica dos estudos, mesmo sendo a Escala específica para avaliação de estudos dessa natureza.

Para verificarmos a frequência de heterogeneidade que os resultados das avaliações tiveram na amostra nos artigos eleitos para este estudo, aplicamos o teste qui-quadrado (X^2). Ele serviu, também, para reafirmar, ou negar, as hipóteses estabelecidas para nosso estudo. A Tabela 19 apresenta os itens avaliados na Escala de Loney et al. (1998), as frequências e o resultado de teste de hipótese dos estudos latino-americanos e da amostra aleatória de artigos não latino-americanos aprovados na revisão sistemática.

Tabela 19 - Frequência de artigos que realizaram a coleta de dados em países latino-americanos e não latino-americanos que obtiveram pontuação em cada item da Escala de Loney et al. (1998)

Escala de Loney et al. (1998)	Coleta de dados em países latino-americanos		Coleta de dados em países não latino-americanos		Resultado do teste de qui- quadrado
	N.A.	%	N.A.	%	
Item 1 O delineamento do estudo e os métodos de amostragem são apropriados para responder à pergunta de pesquisa?	3	8,8	0	0	$X^2_1=3,138$, $p<0,076$
Item 2 A base amostral é adequada?	2	5,9	0	0	$X^2_1=2,061$, $p<0,151$
Item 3* O tamanho da amostra é adequado?	4	11,8	12	35,3	$X^2_1=5,231$, $p<0,022$
Item 4 São usados critérios objetivos, adequados e	6	17,6	8	23,5	$X^2_1=0,360$, $p<0,549$

Escala de Loney et al. (1998)	Coleta de dados em países latino- americanos		Coleta de dados em países não latino- americanos		Resultado do teste de qui- quadrado
	N.A.	%	N.A.	%	
padronizados para medir o desfecho de saúde?					
Item 5	0	0	0	0	☐
O desfecho de saúde é medido de forma não enviesada?					
Item 6	6	17,6	4	11,8	$X^2_1=0,469$, $p<0,493$
A taxa de resposta é adequada? As pessoas elegíveis que recusaram participar são descritas?					
Item 7*	10	29,4	2	5,9	$X^2_1=6,476$, $p<0,011$
As estimativas de prevalência ou incidência são dadas com intervalos de confiança e detalhadas por subgrupo, se adequado?					
Item 8	34	100,0	32	94,1	$X^2_1=2,061$, $p<0,151$
Os participantes e o contexto são descritos em detalhes e podem ser generalizados para outras situações similares?					
Total	34	100,0	34	100,0	-

Nota. *Itens com diferença estatisticamente significativa entre países latino- americanos e não latino-americanos, ($p<0,05$ no teste de qui-quadrado).

Apenas os itens três e sete da Escala de Loney et al. (1998) apresentaram diferenças significativas no teste de qui-quadrado, os demais mantiveram a equivalência na frequência. Comentaremos cada item em função dos escores

obtidos, visando confirmar a hipótese de que os estudos latino-americanos tiveram desempenho similar aos não latino-americanos.

O item um da Escala (delineamento do estudo e os métodos de amostragem apropriados para responder à pergunta de pesquisa) teve escore de 8,8% para os estudos latino-americanos, contra zero (0) dos não latino-americanos, mostrando maior pontuação para os estudos latinos, com $X^2_1=3,138$, $p<0,076$. O mesmo panorama aparece no item dois da Escala (base amostral adequada) com 5,9% de frequência favorável aos estudos latino-americanos e com zero (0) novamente para os não latino-americanos, $X^2_1=2,061$, $p<0,151$.

Em contraposição aos dois itens anteriores, o item três aparece com diferenças estatisticamente significante com maior proporção de artigos adequados para os não latino-americanos, que obtiveram 35,3% de pontuação, contra 11,8% dos estudos latino-americanos, sendo $X^2_1=5,231$, $p<0,022$. Houve diferença, também, entre os escores dos não latino-americanos (35,3%) em relação aos latino-americanos (23,5%) no item quatro da Escala, $X^2_1=0,360$, $p<0,549$. No item cinco (desfecho de saúde medido de forma não enviesada) nenhum dos dois grupos pontuou, ficando empatada a baixa qualidade metodológica do item avaliado.

Embora não significativa, maior pontuação apareceu para os estudos latino-americanos no item seis da Escala (taxa de resposta adequada. Pessoas elegíveis que se recusaram a participar estão devidamente descritas), com percentuais de 17,6% para estudos latino-americanos e 11,8% para os não latino-americanos, $X^2_1=0,469$, $p<0,493$. O item sete, que avaliou a questão da apresentação dos resultados, apresentou diferença significativa no teste de qui-quadrado. Estimativas de prevalência ou incidência dadas com intervalos de confiança e detalhadas por subgrupo, se adequado, teve percentuais de 29,4% para os latino-americanos e 5,9% para os não latino-americanos, favorecendo os estudos latinos com $X^2_1=6,476$, $p<0,011$.

Finalizando, com a questão que avalia se os resultados tem aplicabilidade, o item oito (participantes e contexto descritos em detalhes passíveis de generalização para outras situações similares) teve o mais alto escore da Escala, com frequência de 100% para os latino-americanos e 94,1% para os não latino-americanos, com $X^2_1=2,061$, $p<0,151$. Nesse quesito os latino-americanos tiveram um desempenho no máximo de pontuação e, novamente, pontuaram mais que os não latino-americanos.

Embora estatisticamente aparecessem níveis de significância apenas nos itens três e sete, sendo no três maior pontuação para os não latino-americanos e no sete para os latino-americanos, confirmamos a hipótese de que os dois grupos de estudos são similares metodologicamente e na apresentação dos resultados.

O mesmo teste de hipótese foi aplicado para analisar a heterogeneidade dos resultados da avaliação dos artigos publicados em revistas latino-americanas e não latino-americanas. A hipótese definida, *a priori*, foi que não haveria diferença nos escores de avaliação entre as duas categorias de artigos. A Tabela 20 apresenta os oito itens da Escala, a frequência e o resultado do teste.

Tabela 20 - Frequência de artigos publicados em revistas latino-americanas e não latino-americanas que obtiveram pontuação nos itens da Escala de Loney et al. (1998)

Escala de Loney et al. (1998)	Revistas latino-americanas		Revistas não latino-americanas		Resultado do teste de qui- quadrado
	N.A.	%	N.A.	%	
Item 1 O delineamento do estudo e os métodos de amostragem são apropriados para responder à pergunta de pesquisa?	0	0	3	6,6	$X^2_1=1,604$, $p<0,205$
Item 2 A base amostral é adequada?	0	0	2	4,4	$X^2_1=1,053$, $p<0,305$
Item 3* O tamanho da amostra é adequado?	1	4,3	15	33,3	$X^2_1=7,107$, $p<0,008$
Item 4 São usados critérios objetivos, adequados e padronizados para medir o desfecho de saúde?	3	13,0	11	24,4	$X^2_1=1,210$, $p<0,271$
Item 5 O desfecho de saúde é medido	23	100	45	100	100

Escala de Loney et al. (1998)	Revistas latino-americanas		Revistas não latino-americanas		Resultado do teste de qui- quadrado
	N.A.	%	N.A.	%	
de uma forma não enviesada?					
Item 6 A taxa de resposta é adequada? As pessoas elegíveis que recusaram participar são descritas?	4	17,4	6	13,3	$X^2_1=0,200$, $p<0,655$
Item 7 As estimativas de prevalência ou incidência são dadas com intervalos de confiança e detalhadas por subgrupo, se adequado?	5	21,7	7	15,5	$X^2_1=0,400$, $p<0,527$
Item 8 Os participantes e o contexto são descritos em detalhes e podem ser generalizados para outras situações similares?	23	100,0	43	95,5	$X^2_1=1,053$, $p<0,305$
Total	23	100,0	45	100,0	-

Nota. * Teste de qui-quadrado significativo (item 3).

Estatisticamente, não houve diferenças significativas entre as médias das pontuações na Escala de Loney et al. (1998), quando comparamos os estudos publicados nas duas categorias de revistas. O resultado reafirma os dados anteriores em relação à aceitação da Hipótese 2 do presente estudo, que previu que a qualidade dos estudos publicados em revistas latino-americanas seria equivalente à qualidade dos estudos publicados em revistas não latino-americanas. Apenas um item apresentou diferença estatisticamente significativa ($X^2_1=7.107$, $p<0,008$) entre as categorias: o item 3 da Escala (tamanho da amostra), pois as revistas não latino-

americanas apresentaram, proporcionalmente, maior taxa de adequação (33,3%) em relação às revistas latino-americanas (4,3%).

O item 3, tamanho da amostra, obteve pontuação com mais frequência em países não latinos. Estudos com amostras grandes são onerosos. Os dados podem ter refletido os benefícios fortes dos financiamentos das pesquisas existentes nos países ricos?

As revistas não latino-americanas pontuaram mais, também, no item quatro (critérios objetivos, adequados e padronizados para medir o desfecho de saúde) com 24,4%, contra 13,0% para as latino-americanas ($X^2_1=1,210$, $p<0,271$). Ainda é a questão da descrição do método que pesa no resultado obtido.

O item cinco apresenta uma situação de empate entre as duas categorias de revistas, ou seja, revistas latino-americanas são equivalentes às não latino-americanas, com 100% de pontuação da medição do desfecho de saúde com controle de viés.

No último item, que avalia a qualidade metodológica das pesquisas que geraram os artigos incluídos na revisão sistemática, o de número seis (taxa de resposta adequada), as revistas latino-americanas pontuaram acima das não latino-americanas, ou seja, as metodologias dos estudos dessas revistas explicaram mais adequadamente sobre as participantes perdidas no estudo. Os percentuais são de 17,4% para as latino-americanas, contra 13,3% para as não latino-americanas ($X^2_1=0,200$, $p<0,655$). Dos cinco itens que avaliaram a qualidade metodológica dos artigos, as revistas latino-americanas tiveram desempenho melhor, em números absolutos, pontuando favoravelmente em quatro itens, contra um item favorável às não latino-americanas. No item em que os dois grupos de revistas pontuaram no máximo da Escala, houve empate entre os dois grupos.

O item sete da Escala, que diz respeito à interpretação dos resultados em termos do registro dos intervalos de confiança dos estudos, foi mais pontuado pelas revistas latino-americanas, com 21,7% do total, contra 15,5% para as não latino-americanas ($X^2_1=0,400$, $p<0,527$). Ainda em relação à apresentação dos resultados, em termos de aplicabilidade, no item 8 da Escala (participantes e contexto descrito em detalhes e que permitam generalizações para outras situações semelhantes), os artigos publicados em revistas latino-americanas obtiveram 100% de pontuação, e as não latino-americanas 95,5% ($X^2_1=1,053$, $p<0,305$), proporções muito próximas uma da outra.

A seguir é apresentado um resumo comparativo do desempenho dos dois grupos estudados: estudos latino-americanos *versus* não latino-americanos e revistas latino-americanas *versus* não latino-americanos.

Tabela 21 - Posicionamento dos estudos latino-americanos e não latino-americanos em função da Escala Loney et al. (1998)

Item da Escala de Loney, et al. (1998)	Estudos latino-americanos (%)	Item da Escala de Loney, et al. (1998)	Estudos não latino-americanos (%)
8	100	8	94,1
7	29,4	3	35,3
4	17,6	4	23,5
6	17,6	6	11,8
3	11,8	7	5,9
1	8,8	1	0
2	5,9	2	0

Nota. O posicionamento se deu pela pontuação na Escala, do maior para o menor, nos dois grupos.

Analizando as proporções do desempenho dos artigos, em função da Escala de Loney et al (1998), observamos que os estudos latino-americanos tiveram escores mais altos do que os não latino-americanos em seis itens. Como pode ser confirmado na Tabela 19, o item no qual os não latino-americanos obtiveram maior pontuação foi o de número três, que apresentou significância estatística. No item cinco nenhum dos grupos pontuou.

Tabela 22 - Posicionamento dos artigos publicados em revistas latino-americanas e não latino-americanas em função da Escala de Loney et al. (1998)

Item da Escala de Loney et al. (1998)	Revistas latino-americanas (%)	Item da Escala de Loney et al. (1998)	Revistas não latino-americanas (%)
5	100	5	100
8	100	8	95,5
7	21,7	3	33,3
6	17,4	4	24,4
4	13	7	15,5
3	4,3	6	13,3
1	0	1	6,6
2	0	2	4,4

Em relação às revistas notamos que a pontuação dos artigos, de acordo com a Escala de Loney et al. (1998), das publicações não latino-americanas foi maior em quatro itens, contra três das latino-americanas.

Em uma análise final podemos dizer que em números absolutos, os estudos latino-americanos tiveram um desempenho melhor, mas as revistas de outras regiões publicaram estudos com maior qualidade.

5.7 Análise da qualidade dos títulos, resumos e palavras-chave dos artigos

As análises anteriores foram referentes ao conteúdo e na sequência analisaremos os itens de forma dos artigos. Nosso último objetivo específico está vinculado à análise de três elementos importantes que constituem parte do aspecto formal dos artigos, mas que foram de fundamental importância na recuperação e seleção dos estudos: os títulos, os resumos e as palavras-chave. Para tanto utilizamos como instrumento de aferição dos títulos e resumos o manual de publicação da *American Psychological Association* (2012) e das palavras-chave duas terminologias das áreas de saúde e Psicologia.

O Manual de Publicação da *American Psychological Association* (2012) encontra-se na sexta edição e orienta os autores em relação ao estilo que deve ser seguido para a redação de artigos científicos a serem submetidos às revistas publicadas pela Associação. Entretanto, o manual vem ditando as orientações para as revistas de Psicologia e áreas afins em diversas regiões do mundo. Por fazermos parte do Comitê de Análise das Revistas da Área de Psicologia para indexação na LILACS, e publicação no Portal de Periódicos de Psicologia (PePSIC), acompanhamos as revistas de diversos países latino-americanos e temos a visão informal desse panorama. Assim, orientadas pela APA iniciaremos nossa análise pelo título.

5.7.1 Avaliação dos títulos dos artigos

O título é o elemento que apresenta o trabalho e é essencial na decisão sobre a leitura, ou não, do texto. Deve ser claro, objetivo, criativo, curto (a APA recomenda até 12 palavras) e apresentar qualidade dos termos. O título é indexador nas fontes de informação, logo, suas palavras são chaves no processo de busca nas bases de dados. Entretanto, não é recomendado incluir no título os mesmos termos utilizados como palavras-chave, pois se para atribuição dessas palavras o autor deve recorrer ao vocabulário controlado da área, na elaboração do título há a liberdade de escolha dos termos e o autor poderá acrescentar sinônimos ou variações das palavras para ampliação das possibilidades de recuperação do texto nas bases de dados. O título tem relação direta com o objetivo do estudo e deverá apresentar a ideia de um todo, podendo ser considerado o primeiro resumo do trabalho do autor. (Ferreira & Abreu, 2007).

Na nossa pesquisa, o título teve um papel bem mais relevante do que simplesmente nos auxiliar a encontrar e incentivar a leitura do artigo. Foi um elemento de inclusão e exclusão dos trabalhos, como pode ser observado no item em que discutimos a avaliação dos trabalhos pelo título. Ao submetermos os títulos aos critérios de inclusão e exclusão dos trabalhos para a revisão sistemática, que embasou nosso estudo, eliminamos 580, do total de 1894 registros recuperado nas bases de dados. Nesse sentido, o título foi um elemento de decisão na seleção dos trabalhos e quando mais adequado estivesse, maior seria seu poder na definição da amostra. Nossos critérios para avaliação dos títulos dos 68 artigos incluídos no estudo foram:

- o estudo ter como objetivo avaliar a prevalência de DPP.
- Tamanho – recomendação da APA (2012) máximo de 12 palavras. As palavras compostas como, por exemplo, pós-parto foram consideradas um único termo para efeito da contagem.
- Palavras essenciais: depressão pós-parto ou depressão puerperal ou depressão pós-natal ou depressão perinatal em português, ou seu equivalente em inglês e espanhol, idiomas eleitos para leitura do texto completo.

Do universo de 68 artigos, 30 declaram ter como objetivo analisar a prevalência de DPP. Após a leitura e a classificação desses, tivemos 19 títulos considerados adequados e 11 inadequados, de acordo com os critérios de avaliação. Os 11 inadequados foram reprovados no quesito tamanho (mais de 12 palavras), porém todos tinham a palavra DPP e seus sinônimos. Destacamos um título considerado adequado e outro inadequado para ilustrar nossa análise.

Título adequado

“Depressão pós-parto: prevalência e correlação com o suporte social”

Extraído de: Coutinho, D., Baptista, M., & Moraes, P. (2002). Depressão pós-parto: Prevalência e correlação com o suporte social. *Infanto Revista de Neuropsiquiatria da Infância e Adolescência*, 10, 63-71.

O título sucinto, claro e objetivo tinha todas as palavras relevantes que nos permitiram selecionar o trabalho para inclusão na revisão sistemática.

Título inadequado

“Prevalencia y factores asociados a depresión posparto en mujeres atendidas en establecimientos de salud del primer nivel de atención en Lima Metropolitana, junio 2004”

Extraído de: Aramburú, P., Arellano, R., Jáuregui, S., Pari, L., Salazar, P., & Sierra, O. (2008). Prevalencia y factores asociados a depresión posparto en mujeres atendidas en establecimientos de salud del primer nivel de atención en Lima Metropolitana, junio 2004. *Revista Peruana de Epidemiología*, 12(3), 1-5.

Excessivamente longo, o título poderia ser reduzido e nomear os fatores associados à DPP, em vez de declarar o lugar e mês em que o estudo foi realizado. Embora tenha sido possível incluir o trabalho na revisão sistemática, o artigo não seria publicado em uma revista da APA sem a revisão desse elemento.

5.7.2 Avaliação dos resumos dos artigos

O resumo deve conter uma representação breve, mas completa do conteúdo do trabalho. Deve ser claro e conter palavras representativas, pois também é indexador em muitas bases de dados. No resumo evitamos o uso de siglas, abreviaturas e não incluímos referências. O leitor deve encontrar todas as partes do

trabalho no resumo: introdução, metodologia, resultados, discussão e conclusão, quando houver. Resumos longos têm seus conteúdos truncados em bases de dados, pois há limites de caracteres nesse campo. Revistas bem elaboradas declaram o número máximo de palavras, ou caracteres, que o resumo deve conter. A APA (2012) recomenda que o resumo seja o mais sucinto possível, não devendo passar de 250 palavras. A APA ampliou o número mínimo de palavras, pois como observamos em Sabadini, Sampaio e Koller (2009a), a versão anterior do manual recomendava o máximo de 120 palavras. Assim como o título, o resumo é um item de suma relevância na seleção de trabalhos a serem incluídos, ou excluídos, em revisões da literatura, como observamos nas etapas de seleção dos trabalhos que compõem o universo deste estudo. Nossos estudos eram 1.314 trabalhos selecionados pelo título e a avaliação pelo resumo possibilitou a exclusão de 421 trabalhos.

Crítérios para avaliação do resumo:

- ter o máximo de 250 palavras, de acordo com a recomendação da APA (2012).
- Apresentar as principais partes do estudo: objetivo, metodologia, resultados e conclusão. Mesmo que o resumo não fosse estruturado, quando as etiquetas dos campos são apresentadas, deveria apresentar as partes do estudo claramente descritas.

Dos 68 resumos analisados, 49 atenderam as recomendações da APA, inclusive em relação ao tamanho, máximo de 250 palavras. Como exemplo, a seguir, apresentamos um resumo considerado adequado e um inadequado, de acordo com nossos critérios de avaliação.

Resumo adequado

“Background: To investigate the prevalence, socio-cultural and psychosocial risk factors for perinatal depression in Singaporean women.

Method: A prospective cohort of 559 women was interviewed antenatally and at six weeks' postpartum at a tertiary hospital. Women were interviewed for diagnosis of depression using a two-stage design, with a screening questionnaire and diagnostic interview.

Results: Postnatally, a negative confinement experience was associated with depression. Other independent factors included poor emotional support, a past history of depression, unplanned pregnancy and perceived potential conflicts with relatives

over childcare antenatally and dissatisfaction, poor instrumental support postnatally. The prevalence of depression antenatally and postnatally was 12.2% and 6.8%, respectively. Limitations: Measures of satisfaction with social support were based on self-report; there were high dropout rates at six weeks' postpartum; and other modulating social factors such as pre-existing interpersonal conflicts were not studied.

Conclusions: Perinatal depression in Singaporean women is common. Contrary to expectations, a negative "confinement" experience is a significant risk factor for postnatal depression, and is not universally welcomed by women. Depression is modulated by dissimilar sets of psychosocial factors antenatally and postnatally."

Extraído de: Chee, C. Y. I., Lee, D. T. S., Chong, Y. S., Tan, L. K., Ng, T. P., & Fones, C. S.L. (2005). Confinement and other psychosocial factors in perinatal depression: a transcultural study in Singapore. *Journal of Affective Disorders*, 89, 157-166.

Com 179 palavras, de acordo com nossa avaliação, o resumo foi considerado adequado, pois apresentou todas as partes do estudo e a partir dele foi possível decidir pela inclusão do trabalho na revisão sistemática. Estruturado, todos os elementos são facilmente identificados e

Resumo inadequado

"La depresión postparto es un problema de salud pública con prevalencias que oscilan entre 10 y 15% en la literatura mundial. Durante el puerperio existen cambios bioquímicos y estrés que pueden desencadenar estos trastornos. No existen trabajos a nivel nacional que enfoquen estos problemas. Realizamos un estudio prospectivo a nivel de atención primaria para evidenciar la real incidencia, y algunos factores de riesgo para esta condición."

Extraído de: Evans M. , G., Vicuña M. , M., & Marín, R. (2003). Depresion postparto realidad en el sistema publico de atencion de salud. *Revista Chilena de Obstetricia y Ginecología*, 68(6), 491-494.

Pelo resumo, não foi impossível saber se o estudo seria, ou não, incluído na revisão sistemática, em decorrência das informações insuficientes. Foi necessária a consulta ao texto completo para aceitação do trabalho. Além de inadequado, os autores afirmaram que não existiam trabalhos no país sobre o tema e localizamos um estudo da mesma natureza publicado em um fascículo anterior da mesma revista, que também fora incluído na revisão sistemática. Essa informação omitida não mudaria a relevância do resumo e protegeria os autores de afirmações equivocadas.

5.7.3 Avaliação das palavras-chave dos artigos

Palavras-chave são palavras, ou grupo de palavras, que representam o conteúdo do trabalho. Revistas de qualidade primam por artigos que tenham palavras-chave atribuídas mediante consulta a um vocabulário controlado, pois facilitam o trabalho do indexador, que nem sempre é um especialista na área, e garantem a eficiência na recuperação dos artigos. Autores devem conhecer e utilizar os vocabulários controlados de sua área para atribuírem palavras-chave consistentes aos seus trabalhos. Também devem apresentar a versão em inglês e em espanhol dos termos para ampliar o acesso e a visibilidade dos artigos.

Vocabulário controlado ou Tesouro é uma lista de termos controlados e ordenados de acordo com uma hierarquia, geralmente a mesma dos sistemas decimais de classificação utilizados em bibliotecas. São utilizados para indexar e recuperar os trabalhos em bases de dados. Por isso é recomendado que os autores consultem essas ferramentas para atribuição de palavras-chave nos seus manuscritos. (Castro, 2001; Sabadini et al., 2009a). Por identificarmos que as revistas que publicaram os estudos da nossa amostra foram majoritariamente da área de ciências da saúde, optamos por avaliar as palavras-chave de acordo com o instrumento que controla os termos nessa área, desenvolvido a partir do MESH (Medical Subjects Headings da U.S National Library, que é usado na indexação e recuperação da informação na MEDLINE (Castro, 2001), o DECS – Descritores em Ciência da Saúde (DECS). Confrontamos as palavras-chave dos artigos também com a Terminologia em Psicologia, desenvolvida pela Biblioteca Dante Moreira Leite do Instituto de Psicologia da Universidade de São Paulo para indexar e recuperar a informação das Bibliotecas Virtuais de Psicologia latino-americanas e do seu próprio acervo. A terminologia em Psicologia tem como base o *Thesaurus da APA* (2007).

Confrontamos os 68 artigos com os dois instrumentos e identificamos apenas um com todos os termos controlados no DECS. Os demais apresentaram uma ou outra palavra não padronizada. O termo depressão pós-parto foi o mais utilizado como indexador nos artigos (70%) e prevalência apareceu mais frequentemente nos artigos em inglês (prevalence) (30%) do que nos artigos em língua portuguesa e espanhola (20%). Isso explica porque ao fazermos as buscas na BVS-Psi ULAPSI, Psycodoc e RedAlyc tivemos que deixar a busca bem ampla, usando apenas “depressão pós-parto”, como pode ser observado na Tabela 6. A falta de

normalização dos termos foi generalizada para as revistas que publicam seus artigos em português, como por exemplo, o termo: “Período gravídico e puerperal”, sendo que no DECS seria: “Gravidez” e “Período pós-parto”. Em espanhol: a palavra “Latinas” foi usada em um artigo, no DECS o correto é: Hispanoamericanos. Outro artigo, em inglês, apresentou a palavra-chave: *prediction of dreams during pregnancy*, que mais se assemelha a uma frase-chave do que à palavra-chave e, segundo o DECS deveria ser desdobrada em: “*dreams*” e “*pregnancy*”.

O não controle das palavras-chave nos artigos dificultou as buscas por informação nas bases de dados. Poderíamos ter chegado aos trabalhos de interesse para a revisão sistemática bem mais rapidamente e com mais eficiência se os estudos tivessem suas palavras-chave controladas pelos vocabulários da área de saúde, ou da Psicologia. Um ponto que nos chamou a atenção foi a ausência de palavras-chave nos artigos publicados pelas editoras internacionais mais conhecidas e que detêm as revistas com maior fator de impacto, o que pode apontar para uma tendência em abolir esse elemento dos artigos.

Temos nos preocupado em orientar os autores em relação à forma desses e outros itens do artigo científico. Tanto no livro *Publicar em Psicologia: um Enfoque para as Revistas Científicas* (Sabadini, Sampaio, & Koller, 2009b), como nos cursos que temos oferecido na Biblioteca do IPUSP e nos eventos de Psicologia no Brasil e em outros países da América Latina, enfatizamos a relevância desses itens no trabalho científico. Dessa maneira, reforçar tal questão nesta tese, analisando e comentado sobre a importância em melhorar a qualidade dessas partes do artigo foi uma oportunidade de reforçar nossas orientações.

6 DISCUSSÃO

A presente investigação avaliou a qualidade metodológica e da apresentação de resultados de artigos incluídos em uma revisão sistemática sobre prevalência de DPP, comparando os estudos latino-americanos e não latino-americanos e a origem das revistas que os publicaram. Os resultados evidenciaram que as duas categorias de estudos, lamentavelmente, apresentaram qualidade fraca em relação à metodologia e à apresentação de resultados. Não observamos diferenças, também, quanto aos artigos publicados em revistas latino-americanas e não latino-americanas. Os dados mostraram que os artigos avaliados são evidências fracas e

que gestores de saúde e políticos teriam provas científicas de baixa qualidade nos estudos avaliados. “Estudos de prevalência são importantes para facilitar o planejamento e diretrizes para alocação de recursos em saúde e para os profissionais de saúde estimarem as chances que seus pacientes possuem de apresentar um determinado evento.” (Reis, Ciconelli, & Faloppa, 2002, p. 54). Se, como afirmam Sánchez-Meca, Boruch, Petrosino e Rosa (2002), é urgente levar evidências de forma rápida aos políticos, gestores e profissionais, nossas evidências devem ser utilizadas com cuidado, pois são fracas metodologicamente e seus resultados podem não ser aproveitados ao máximo na gestão baseada em evidências. Angelini (1954), ao finalizar sua revisão, concluiu que os resultados dos artigos não tinham qualidade em diversos aspectos e que não era possível recomendar o uso do ácido glutâmico com base nos estudos revisados. Quase 60 anos depois chegamos à conclusão semelhante em nossa avaliação de artigos.

A cidade de Piripiri, no Estado de Piauí, é um exemplo de como as evidências científicas podem auxiliar o gestor público a tomar decisões na administração municipal em saúde. O Secretário de Saúde tinha um problema para resolver, a mortalidade infantil da cidade não estava sendo reduzida da mesma forma que em outras cidades do país. As complicações na gravidez em decorrência da hipertensão levavam as mães à eclampsia, o que culminava com a morte dos recém-nascidos. Em busca de informação que auxiliasse na resolução da questão o Secretário chegou à EVIPNet, Rede de Políticas Informadas por Evidências (<http://brasil.evipnet.org/>), e recebeu informações de como conhecimento científico pode apoiar a tomada de decisão em gestão pública. Apoiada pela equipe de profissionais da informação, a equipe de saúde de Piripiri consultou fontes de informação especializadas e encontrou evidências de qualidade que, juntamente com o conhecimento da prática profissional e o contexto da cidade, apoiaram na elaboração de um programa preventivo de hipertensão arterial em gestantes. A experiência exitosa do uso de evidências em gestão pública está registrada no vídeo *Evidências na Gestão Municipal da Saúde* (Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, 2011). O vídeo enfoca como usar fontes de informação provenientes de pesquisas e práticas em saúde de forma que a tomada de decisão seja embasada em informação de qualidade.

Se um gestor público, ou um político, recorrer à revisão sistemática da qual participamos encontrará 26 estudos brasileiros e prevalências que variaram de 4,0%

a 39,4%. Analisando as prevalências médias por cidade observaria 8,9% em Caxias do Sul (RS), 13,15% em Pelotas (RS), 14,48% em Brasília, 14,69% em Juiz de Fora (SP), 20,1% em Recife (PE), 20,66% em Porto Alegre (RS), 24,55% no Rio de Janeiro (RJ), 24,92% em São Paulo (SP), 26,94% em Belo Horizonte (MG), 33,33% em João Pessoa (PB), 34,29% em Uberaba (MG), 39,13% em Embu Guaçu (SP) e 39,98 em Vitória (ES), (Silva, 2013). Em algumas cidades os dados eram sínteses de prevalências encontradas em mais de um estudo. Por exemplo, em Brasília, um estudo encontrou 13% e outro em torno de 22%. Talvez seja difícil para esse gestor tomar decisão, em nível nacional, com dados que variaram nessa ordem. Porém, temos um panorama sobre o transtorno em 13 cidades do país que mediram a DPP e que poderia apoiar gestores públicos e políticos que se preocupassem com o problema enquanto um caso de saúde pública. Mesmo os estudos apresentando baixo desempenho na avaliação pela Escala de Loney, et al. (1998), os dados poderiam ser utilizados para a tomada de decisão em gestão pública no país.

No vídeo *Evidências na Gestão Municipal da Saúde* o Secretário da Saúde da cidade de Piripiri esclarece que não foram muitas as evidências localizadas e utilizadas na tomada de decisão em relação ao problema de saúde pública de sua cidade, mas algumas foram suficientemente fortes para apoiar na elaboração do plano de combate à hipertensão em gestantes.

Não encontramos na literatura nenhum estudo com objetivo de comparar artigos científicos oriundos de estudos latino-americanos e de outras regiões, o que nos incentiva a explorar seus dados em busca de ampliação das discussões sobre o tema. Tendo como pano de fundo que os artigos científicos são as fontes primárias mais valorizadas na Psicologia Baseada em Evidências, como vimos no capítulo que introduz nosso tema de pesquisa, discutiremos os principais resultados encontrados, dando destaque para os pontos fortes e as limitações do nosso estudo.

6.1 Buscando a literatura para a revisão sistemática

Para chegar ao universo de nosso estudo, que avaliou a qualidade de artigos científicos latino-americanos e comparou com uma amostra aleatória de artigos não latino-americanos, percorremos todas as etapas de uma revisão sistemática sobre DPP, com metanálise. A revisão sistemática, assim como outros tipos de estudo de

revisão, é um tipo de pesquisa que obtém seus dados em fontes primárias de informação sobre determinado tema.

Um diferencial das buscas por artigos deste estudo foi a consulta nas 14 bases de dados. Em revisões sistemáticas sobre o nosso tema não observamos o mesmo número de fontes de informação consultadas. A revisão de Sawyer, Ayers e Smith (2010) utilizou MEDLINE, PsycINFO, PsycArticles APA, PubMed, Web of Science, Scopus e African Index Medicus. Vigod, Villegas, Dennis e Ross (2010) revisaram MEDLINE, CINAHL, EMBASE, PsycInfo e The Cochrane Library. Melo Junior (2011) pesquisou MEDLINE, EMBASE, ISI Web of Knowledge, CINAHL e SciELO. Lobato, Moraes e Reichenheim (2012) buscaram apenas em LILACS e SciELO. Outro diferencial foi a consulta às fontes de informação especializadas em Psicologia e desenvolvidas para indexar os estudos latino-americanos. A BVS-Psi ULAPSI recuperou 405 registros, embora a dificuldade de restrição do assunto na estratégia de busca tenha feito com que muitos registros recuperados fossem descartados, nas análises posteriores. A LILACS forneceu 35 registros. Psycodoc e RedAlyc, bases de dados especializadas em literatura em língua portuguesa e espanhola, retornaram com 61 e seis artigos, respectivamente. SciELO forneceu 17 artigos candidatos à inclusão na revisão. Podemos afirmar que a ampliação das buscas a essas fontes de informação possibilitou alcançar estudos desenvolvidos com população latino-americana e publicados em revistas latino-americanas, permitindo nossas comparações e análises.

O Centro Cochrane Iberoamericano (2011) orienta que, além da MEDLINE e da EMBASE, que são paradigmas internacionais na área de saúde, muito países e regiões produzem fontes de informação dedicadas a organizar e divulgar a literatura por seus cientistas e investigadores, que devem ser consultadas nas revisões sistemáticas, que seguem o modelo da Colaboração. O Manual aponta 10 bases de dados complementares às consultadas por nós, além de incluir a LILACS, como fonte de informação regional para a América Latina, o que denota reconhecimento internacional à base de dados.

- 1) Africa: Índice Médico Africano (indexmedicus.afro.who.int)
- 2) Australia: Índice Médico Australoasiático (www.nla.gov.au/ami)
- 3) China: Base de datos de Literatura Médica China
(www.imicams.ac.cn/cbm/index.asp)

- 4) Mediterrâneo oriental: Índice Médico para la Región Oriental del Mediterráneo (www.emro.who.int/his/vhsl)
- 5) Francesa: PASCAL (international.inist.fr/article21.html)
- 6) Índia: IndMED (indmed.nic.in)
- 7) Coreia: KoreaMed (www.koreamed.org/searchBasic.php)
- 8) Sudeste Asiático: Índice Médico para la Región del Sudeste Asiático - IMSEAR - (library.searo.who.int/modules.php?op=modload&name=websis&file=imsear)
- 9) Ucrânia e a Federação Russa: Panteleimon (www.panteleimon.org/main.php3)
- 10) Pacífico Ocidental: Índice Médico de la Región del Pacífico Occidental – WPRIM - (wprim.wpro.who.int/SearchBasic.php)

As orientações do Centro Cochrane Iberoamericano (2011) recomendam, também, a consulta ao Google Acadêmico, como forma de completar as falhas nas bases de dados por falta de indexação de revistas. Quando efetuamos nossas buscas, entre 29/09/2011 e 19/10/2011, consultamos o buscador acadêmico, mas o número de registros recuperados impossibilitou a análise de todas as referências retornadas, em decorrência do grande volume de duplicidades. Um mesmo trabalho chegava a ter seis entradas. É bem provável que nos próximos anos tenhamos mudanças no comportamento do usuário em relação à busca de informação, pois o Google Acadêmico vem aprimorando seu sistema de busca e perseguindo formas de eliminar as sobreposições de referências, que é o maior problema que enfrentamos atualmente. Além disso, clássicas fontes de informação, como é o caso da PsycNet (<http://psycnet.apa.org/journals/abn/93/2/158>) da APA, que disponibiliza as referências dos artigos das revistas publicadas pela Associação, oferece acesso ao artigo mediante pagamento com cartão de crédito. Certamente, essa parte importante da metodologia da revisão sistemática, ou seja, a busca dos textos completos, ficará muito mais fácil e rápida, com maior garantia de acesso aos textos completos. Nosso estudo deixou de avaliar 47 trabalhos aprovados pelo título e resumo que foram solicitados ao Comut porque não foram recebidos até a data limite para coleta de dados. Com os serviços de acesso aos documentos oferecidos diretamente ao interessado, sem intermediação das bibliotecas, certamente ganharemos velocidade e esse tipo de limitação deixará de ocorrer.

Seguindo as recomendações do Centro Cochrane Iberoamericano (2011), o estudo de Sawyer et al. (2010) foi o único, dentre os analisados, que utilizou uma fonte de informação diferenciada para publicações africanas, African Index Medicus, recomendada pela Cochrane. A base de dados, desenvolvida pela World Health Organization, tem como objetivo promover o uso das revistas publicadas nos países africanos e incentivar os autores a publicarem os seus trabalhos nas revistas regionais. Chamou nossa atenção o maior objetivo para criação da fonte de informação africana. “É uma forma de oposição ao que acontece atualmente: os cientistas e investigadores competem por espaços nas publicações internacionais de maior prestígio”, informa a página principal da fonte de informação. (Tradução nossa de: <http://indexmedicus.afro.who.int/>). A ideia é reter parte do conhecimento gerado na região em veículos locais de publicação, pela sua visibilidade na Internet.

Consultar 14 bases de dados foi um ponto forte do nosso estudo. No entanto, seguindo as recomendações do Centro Cochrane, revisões sistemáticas na área de saúde devem ampliar as buscas para distintas fontes de informações, como as apontadas anteriormente. Outra limitação deste trabalho foi não rastrear as referências dos estudos em busca de trabalhos sobre o tema. Também não foram consultados os principais especialistas na área, buscando saber se haveria outros estudos publicados, não publicados ou em andamento sobre a prevalência de DPP. Essa etapa também é recomendada pela Colaboração Cochrane. Estudos de revisão são complexos e dificilmente se esgotam na primeira versão. Recomendamos, também, em se tratando de estudos da área de Psicologia, que seja consultada a base de dados PEP - *Psychoanalytic Electronic Publishing*, que disponibiliza conteúdo da área de Psicanálise, desenvolvida pela *American Psychoanalytic Association* e *Institute of Psychoanalysis* (Londres).

6.2 Diminuindo os vieses de seleção: concordância entre avaliadoras

Vícios de seleção em estudos de revisão sistemática, com ou sem metanálise, precisam ser controlados, uma vez que a inclusão ou a exclusão de um único trabalho tem potencial para alterar a decisão final da revisão. A existência de discordância na seleção de artigos para revisões sistemáticas constitui importante aspecto metodológico. Há poucos estudos na literatura que avaliam a concordância entre avaliadores na seleção inicial de artigos. (Oliveira, Oliveira, & Bergamaschi,

2006). Giannakopoulos et al. (2012) recorreram à avaliação independente de dois revisores para validar a proposta de um novo instrumento para medir a qualidade de estudos de prevalência e concluíram que a concordância foi adequada e que não se pode validar esse tipo de instrumento por uma única opinião.

Nosso estudo contou com duas revisoras: uma especialista na área de Psicologia e uma especialista em Ciência da Informação. As duas revisoras participaram de todas as etapas de avaliação dos estudos, embora, devido ao grande número de trabalhos recuperados, a segunda revisora tenha participado parcialmente da etapa de avaliação dos textos completos. Esse foi um ponto forte da pesquisa, pois, após as discussões para definição de critérios e formulação do protocolo de pesquisa e as reuniões de consenso sobre as discordâncias, podemos afirmar que vieses foram bastante reduzidos. As revisões de Sawyer et al. (2010), Vigod, Villegas, Dennis e Ross (2010) e Lobato et al. (2012) não mencionaram a participação de dois, ou mais, revisores. Melo Junior (2011) informou que dois revisores avaliaram os estudos de sua revisão sistemática e que as divergências foram resolvidas por consenso.

Apesar de as avaliadoras terem sido treinadas, grande parte das discordâncias foi atribuída à subjetividade inerente ao processo de seleção de artigos por diferentes avaliadores e à necessidade de maior compreensão sobre o assunto. Como revisora não especializada no tema de prevalência de DPP, necessitamos de treinamento para a coleta de dados e orientações sobre alguns pontos específicos dos temas abordados nos estudos, como, por exemplo, as medidas de confiabilidade e validade. As instruções do Apêndice A resolveram grande parte das dúvidas, contribuindo para melhorar a concordância.

Dois itens dos oito critérios de inclusão ficaram abaixo do aceitável (70%) definido para a revisão. O critério E (método de diagnóstico mencionado claramente e baseado em instrumento objetivo de medida, ou em avaliação feita por profissional de saúde mental) com 69,4% de concordância e o critério F (pelo menos uma das avaliações de DPP ocorreu de 8 dias (inclusive) a 365 dias (inclusive) após o parto) com 64,2%. Os demais itens ficaram com níveis de concordância considerados satisfatórios.

Em estudo que avaliou a concordância entre avaliadores na seleção de artigos em revisões sistemáticas, Oliveira, Oliveira e Bergamaschi (2006) analisaram os resultados da concordância entre dois revisores na seleção de artigos para inclusão

em revisão sistemática, verificando passo a passo a decisão dos avaliadores. Segundo a análise de elegibilidade, observaram que o avaliador um decidiu pela inclusão de 30 trabalhos e o avaliador dois por 51. Treze trabalhos identificados como “não claros” pelo avaliador um foram considerados elegíveis pelo avaliador dois. Os autores concluem que a participação de dois avaliadores é condição essencial para minimizar os vieses na seleção de trabalhos para qualquer tipo de revisão.

Assim como no estudo de Oliveira et al. (2006), a presença de discordância entre as avaliadoras no nosso estudo confirma a necessidade de participação de, pelo menos, dois juízes nesta fase da revisão, pois seriam selecionados conjuntos de trabalhos diferentes caso houvesse a participação de apenas um avaliador. Ainda que o risco de vieses na aplicação dos critérios de inclusão dos estudos em uma revisão sistemática seja possível, há que se criar estratégias para que os critérios de elegibilidade dos trabalhos estejam de acordo com os objetivos da revisão mais do que com regras arbitrárias dos avaliadores. (Centro Cochrane Iberoamericano, 2011). As reuniões de consenso, realizadas nos casos discordantes, foram úteis para resolver diferenças de interpretação entre as avaliadoras, possibilitando nova leitura, maior reflexão e revisão da decisão final. Colaboraram, portanto, para diminuir as chances de exclusão de um artigo que, na verdade, deveria ser incluído na revisão sistemática, ou, ao contrário, excluído.

6.3 Presença de estudos latino-americanos na revisão

A revisão sistemática incluiu 44 estudos latino-americanos, publicados em revistas científicas e dissertações e teses, o que equivale a 12% do total de trabalhos revisados. Seis países da região estiveram representados, destacando-se o Brasil e o Chile com grande parte das contribuições. Lembrando que consideramos 24 os países latino-americanos, podemos dizer que a representação regional de 12% foi fraca.

Nosso estudo comparativo analisou 34 artigos dos seis países latino-americanos que contribuíram com a amostra da revisão, uma vez que excluímos as dissertações e teses. O Brasil apareceu com 26,5% e o Chile com 11,8% dos estudos latino-americanos incluídos. Os 68 artigos deste estudo foram publicados em 54 revistas, sendo 23 em revistas latino-americanas, ou seja, 45% são da região

e indexadas na LILACS, importante fonte de informação que dá visibilidade ao conhecimento latino-americano da área de saúde. A LILACS indexava, em junho de 2013, 860 revistas de 27 países latino-americanos, desde o ano 1982¹⁵. Dos 23 títulos, 10 foram publicados em revistas brasileiras, oito em revistas do Chile, dois em revistas peruanas, dois em revistas colombianas e um em revista mexicana. Essa supremacia brasileira nos resultados do nosso estudo pode ser justificada pelo número de registros localizados na LILACS e nas revistas publicadas no SciELO, como observamos na Tabela 23.

Tabela 23 - Países latino-americanos e número de registros LILACS e em títulos de revistas em SciELO

País	LILACS	SciELO
Brasil	331.379	316
Chile	52.120	102
Colômbia	36.985	152
México	31.357	110

Embora o uso de tabelas nesta parte da tese não seja recomendado, optamos por ilustrar os dados dessa maneira para que ficasse mais fácil a visualização da diferença entre os seis países que incluíram artigos no estudo. Brasil e Chile foram os países que mais incorporaram trabalhos à LILACS, desde o seu início. Foram, também, os dois primeiros países a criarem seus portais de revistas em acesso aberto. SciELO Brasil e SciELO Chile datam do ano de 1997. A Colômbia vem se destacando, com o seu SciELO publicando 152 revistas. Entretanto, o portal da Colômbia é recente e, como vimos, grande parte dos nossos estudos foi publicada no ano de 2008 e 2009. O portal do México foi um dos últimos a ser desenvolvido, sendo colocado no ar no ano de 2006, quase dez anos depois do Brasil e do Chile. O Peru ainda não obteve a certificação de seu Portal e é o país com menos revistas publicadas em acesso aberto no Portal Regional, que reúne todos os portais nacionais.

¹⁵ <http://lilacs.bvsalud.org>

Como nos interessava analisar artigos publicados em revistas latino-americanas, encontrar artigos nas fontes de informação da região foi um ponto forte para o nosso trabalho, que incluiu 34 artigos na revisão sistemática. Após avaliação dos artigos, nossa hipótese número dois (H2) foi corroborada, pois constatamos que, estatisticamente, os artigos publicados em revistas não latino-americanas e latino-americanas tiveram escores equivalentes na Escala de Loney et al. (1998).

6.4 Publicação em inglês e impacto dos artigos latino-americanos

Identificamos 11 estudos latino-americanos publicados em inglês, em revistas estrangeiras. Oito desses artigos apresentaram mais de 10 citações no Google Acadêmico, ou seja, têm Índice !10, como são classificados os trabalhos com mais de 10 citações nos perfis de autores do Google Acadêmico. Seis artigos tiveram mais de 10 citações no Scopus e quatro na WOK. O artigo mais citado no Google Acadêmico recebeu 85 citações e é um estudo chileno do ano de 1985.

Dos 23 artigos publicados na língua portuguesa e espanhola em revistas dos próprios países, 11 tiveram Índice !10 no Google Acadêmico. Os dois mais citados alcançaram a marca de 108 e 88 citações na data pesquisada. Oito revistas que publicaram os artigos são indexadas na Scopus e duas na WOK. Devido a essa disparidade em relação à indexação, não temos parâmetros que permitam comparações para discutirmos os dados de citações. Entretanto, os dados nos provocam a ampliar a discussão sobre autores brasileiros publicarem em revistas estrangeiras e revistas brasileiras publicarem seus artigos em inglês.

Um artigo publicado na Revista Latinoamericana de Psicopatologia Fundamental reflete sobre a questão da internacionalização da Psicologia e argumenta que a disciplina já nasceu internacionalizada. “Há nesse sentido inclusive a intenção, que é uma boa intenção, de que se publiquem revistas brasileiras em inglês, língua que é tida por muitos como a que traz em si a linguagem da ciência.” (Lo Bianco, 2009, p.452). A autora explica também que há diversas subáreas da Psicologia que não sofrem qualquer influência anglo-saxônica e que há subáreas inteiras cujas fontes não são norte-americanas. A conclusão é que “Vários Programas poderiam exigir de seus docentes que se esforcem para publicar nos veículos acadêmicos de língua inglesa, no entanto, estariam retirando tempo e

investimento de ações que realmente atingem seus objetivos.” (Lo Bianco, 2009, p.452).

A polêmica sobre publicar, ou não, em inglês segue com bastante força. De um lado, há os que defendem a publicação em inglês e explicam que sendo a ciência universal, os temas mais relevantes das pesquisas são aqueles que interessam a todos, independente da nacionalidade, e que não há uma ciência regional. Portanto, os cientistas devem publicar em inglês, para dialogarem com a comunidade científica internacional, e manterem os olhos para o que é publicado no mundo desenvolvido. Do outro lado, estão os contrários a essa forma de capitalizar o saber e que priorizam a produção do conhecimento voltada para quem efetivamente sustenta o desenvolvimento das pesquisas nas universidades públicas: a população que paga os impostos diretos e que, quase nunca, sabe ler o inglês. Toda essa discussão pode impactar a ciência dos países em desenvolvimento de duas maneiras: 1) positivamente: o conhecimento científico dos países falantes de outras línguas transpõe as fronteiras e contribui com o avanço da ciência mundialmente; 2) negativamente: a internacionalização das publicações em inglês pode levar ao declínio da ciência nacional e culminar com o fim das revistas dos países em desenvolvimento (Rodrigues, 2008; Ryder, 2001).

Outro ponto que nos instiga à discussão é o uso das bases de dados WOK e Scopus na avaliação de pesquisadores e cientistas no Brasil. Comentários ao modelo de avaliação focado no FI constam da Introdução desta tese.

Em entrevista à revista *Ciência e Cultura*, Eugene Garfield, criador do FI do *International Science Institute* (ISI, hoje Thomson Reuters), no ano de 1955, esclareceu que as bases de dados terão que arrumar um jeito de indexar as informações de livros e outros documentos publicados em formato impresso e em acesso aberto. Todas as vezes que surgirem novas formas de publicações, as bases de dados terão que se adaptar a elas ou não serão aceitas pela comunidade científica (Barata, 2010).

Discutindo o FI, Pinto e Andrade (1999) apontaram quatro motivos para que tomemos cuidado com esse índice de citação: 1) número de revistas muito diferentes entre áreas do conhecimento. Áreas com maior número de publicações têm mais chance de terem seus artigos citados; 2) o número de referências por artigo varia de área para área, o que pode interferir no número de citações; 3) algumas pesquisas são de caráter mais regional que outras, mas nem por isso

menos importantes para a sociedade na qual está inserida; 4) artigos com erros conceituais ou com erros na interpretação dos resultados experimentais são muito citados por serem contestados cientificamente. Os autores concluem que a ciência pode se beneficiar dos parâmetros da cientometria, mas seu crescimento e qualidade exigem que áreas importantes não se atrofiem por não estarem na moda e que congregam o maior número de cientistas que só publicam em revistas com F1 alto e que se transformam em verdadeiras estrelas, a exemplo “dos atores de Hollywood.” Completam explicando que “não se pode esquecer todo o jogo comercial que está envolvido na venda destes bancos de dados.” Algumas áreas já apresentam retração de lideranças científicas que estão deixando de prestigiar as revistas com fator de impacto baixo. Pinto e Andrade (1999, p. 453) diziam que havia o perigo de as “agências de financiamento à pesquisa no Brasil começarem a discriminar as revistas científicas brasileiras sob o pretexto de que tudo que é publicado nessas revistas é ruim, ou de qualidade inferior ao que é publicado no exterior nos países do primeiro mundo”. Passados 13 anos da publicação do artigo, Pinto e Andrade (1999) devem saber que seus temores não eram infundados e que, atualmente, as regras do jogo da ciência no Brasil seguiram os rumos apontados por eles. Podemos exemplificar com o que acontece atualmente com a avaliação das revistas na área de Psicologia, cujos extratos se baseiam em indexações em fontes de informação de difícil penetração para nossas revistas. Outro exemplo é o de agências de fomento que solicitam a inclusão do fator de impacto nos artigos incluídos no Currículo Lattes.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) solicita a inclusão do Índice H e o H10 dos pesquisadores quando ocorre a solicitação de recursos para projetos de pesquisas, como pode ser observado nos registros dos projetos na Biblioteca Virtual da Fundação (www.bv.fapesp.br). Essa postura demonstra uma mudança de comportamento em relação aos consagrados índices de impacto WOK e Scopus e nos anima a acreditar que os esforços para disponibilizar o conhecimento da região em acesso aberto finalmente estão surtindo efeito. A própria FAPESP suporta financeiramente o SciELO, um dos modelos mundiais para acesso aberto ao conhecimento científico. Muito natural que solicite o impacto do Google Acadêmico para avaliar as publicações dos pesquisadores.

Mesmo no Google Acadêmico, observamos dois artigos que não receberam nenhuma citação, publicados na revista *Ciência, Cuidado e Saúde* e na *Revista de*

Psiquiatria Clínica. A primeira não está no SciELO, o que pode ser um motivo para que o trabalho esteja menos visível. A segunda, mesmo estando em acesso aberto no SciELO, não apresentou nenhuma citação. O Google Acadêmico permite que possamos acompanhar as citações aos artigos de autores brasileiros e de outros países da América Latina mesmo que as revistas não estejam indexadas nas famosas WOK e Scopus. Pode ser que, mediante essa situação, aconteça a remissão das revistas científicas dos países não desenvolvidos que, pela publicação em acesso aberto, poderão maximizar a acessibilidade, o uso, a visibilidade e o impacto do conhecimento, retroalimentando com suas contribuições à atividade de pesquisa científica e de apoio aos processos de ensino e aprendizagem. Sendo assim, Victora e Moreira (2006), que denunciaram o racismo editorial do Norte contra os autores do Sul, devem estar satisfeitos em ver que, finalmente, o item que se refere à visibilidade das revistas latino-americanas estaria resolvido e essa parte do preconceito contra os artigos do Sul foi sanado. Entretanto, se lessem nosso trabalho os autores ficariam decepcionados ao verificar que as medidas para melhorar a qualidade dos artigos latino-americanos recomendadas por eles não estão sendo aplicadas e continuam sendo necessárias.

6.5 Avaliação de conteúdo e busca por evidência científica

A busca constante por evidência científica de qualidade nos leva, invariavelmente, à avaliação do conteúdo dos trabalhos científicos. Um estudo com desenho metodológico adequado, que responda corretamente à pergunta de pesquisa e que finalize com a entrega de resultados que possam servir de apoio à tomada de decisão para a prática clínica, ou gestão pública, são características de uma boa evidência. Para tanto, essas evidências devem estar acessíveis para que tenham sua qualidade avaliada e sejam oferecidos produtos em forma de revisões recomendáveis para os profissionais e gestores públicos.

Os oitos itens da Escala de Loney et al. (1998) utilizados para avaliar a qualidade metodológica e da apresentação dos resultados dos estudos sobre prevalência de DPP permitiram comparar estudos latino-americanos e não-latino-americanos. Observamos que, estatisticamente, não houve diferenças entre as duas categorias, como demonstrado na figura 6. Os dados nos autorizam a dizer que tanto os estudos latino-americanos, quanto não latino-americanos tiveram o mesmo

desempenho, corroborando nossa hipótese número um (H1). No entanto, as duas categorias estão entregando evidências de baixa qualidade metodológica e resultados inadequados para a tomada de decisão em gestão pública. Problemas com o método de amostragem, base amostral, tamanho da amostra, critérios objetivos e adequados para medir o desfecho e o tipo de instrumento para medir a DPP inviabilizaram o agrupamento dos estudos em evidências homogêneas, que facilitem o processo decisório. Mas, ao selecionar os estudos, analisar e reunir seus dados, certamente, superamos alguns desafios e apresentamos algumas orientações que compartilhamos, juntamente com Silva (2013), visando contribuir com futuras revisões de estudos observacionais de prevalência com maior chance de apoio aos gestores públicos e aos políticos no processo decisório.

Explorar a heterogeneidade metodológica dos artigos, a partir da análise de cada item da Escala de Loney et al. (1998), referentes ao delineamento do estudo e método de amostragem, base amostral, tamanho da amostra, critérios para medir o desfecho de saúde, taxa de resposta, apresentação das estimativas de prevalência e incidência com intervalos de confiança e detalhadas por subgrupo, e apresentação das características de perfil dos participantes para permitir possíveis generalizações dos resultados, foi um ponto forte deste estudo. No entanto, a metodologia do presente estudo está limitada pelo uso da Escala de Loney et al. (1998), que é composta por apenas oito itens de pontuação, com pesos iguais. Os resultados poderiam ser diferentes em estudo que utilizasse a nova escala desenvolvida por Giannakopoulos et al. (2012), que pontua os estudos em quatro subgrupos, atribuindo de 0 a 19 pontos.

Buscando estudos primários sobre prevalência de DPP, vimos que esses são de fácil acesso e recuperação, pois a revisão sistemática incluiu 337 estudos que foram aprovados nos critérios de inclusão. Entretanto, a qualidade fraca dos estudos comprometeu a recomendação da aplicabilidade dos resultados para a tomada de decisão em gestão ou políticas públicas. A revisão sistemática de Dennis e Hodnett (2009), publicada na Cochrane Library, concorda com nosso estudo. Os autores concluem que embora a qualidade metodológica dos estudos não tenha sido forte, os resultados da metanálise sugerem que intervenções psicossociais e psicológicas são uma opção eficaz de tratamento para as mulheres que sofrem de depressão pós-parto. Concluimos que, mesmo não apresentando qualidade metodológica

adequada diante da avaliação dos autores, as evidências trazem contribuições que podem ser aproveitadas por profissionais da área de Psicologia.

Comparando objetivos e resultados, Vieira (2006), avaliou a adequação metodológica de estudos clínicos sobre voz e laringe publicados em revistas brasileiras e internacionais e, também, concluiu que os artigos apresentaram metodologia fraca, com os desenhos e objetivos não compatíveis. A autora explica que o panorama dificulta o uso de evidências científicas nos tratamentos clínicos, assim como a revisão sistemática da qual participamos, que não permite recomendar o uso dos dados em gestão pública. Nossa hipótese de número três (H3) buscava verificar se artigos que declararam ter como objetivo avaliar a prevalência de DPP teriam o mesmo desempenho em relação aos que não declararam ter esse objetivo. A hipótese foi negada, pois o objetivo declarado não influenciou na qualidade metodológica e nem na apresentação dos resultados dos nossos estudos. Não verificamos se os objetivos e os desenhos das pesquisas eram compatíveis, como Vieira (2006), o que pode ser apontado como um ponto fraco do nosso estudo.

Em estudo que analisou os resultados de 1016 revisões sistemáticas disponíveis na Cochrane Library, issue 4, 2004, foi concluído que a maior parte das revisões não apresentou evidências suficientes para sugerir a tomada de decisão clínica com base nos artigos, por causa da baixa qualidade dos estudos primários (ensaios clínicos) (El Dib, Atallah, & Andriolo, 2007). Nosso estudo concorda com os resultados encontrados a pesquisa anterior e confirma que os trabalhos que revisamos também são metodologicamente inadequados e não fornecem respostas confiáveis para a tomada de decisão. Concluímos então que ser incluído em revisão sistemática não atribui qualidade ao estudo científico, mas simplesmente diz que foi aprovado nos critérios de inclusão. Na verdade, o julgamento da qualidade do estudo científico começa com a inclusão do trabalho, quando inicia a avaliação da qualidade metodológica e dos resultados.

Em uma revisão sistemática sobre prevalência de DPP que incluiu 26 estudos, Vigod et al. (2010) informaram que desistiram da metanálise e fizeram uma análise qualitativa dos estudos em virtude da heterogeneidade dos trabalhos revisados. Os autores concluíram que a falta de padronização das metodologias dos estudos impossibilitou que os resultados fossem usados na formulação de políticas públicas. Na revisão da qual participamos, foi possível utilizar a metanálise para apresentar a

prevalência média pontual e os intervalos de confiança por país dos 337 estudos incluídos (Silva, 2013), o que atribuiu força ao trabalho.

A Pirâmide de Evidências utilizada na área de ciências da saúde hierarquiza o estudo científico com base na metodologia utilizada, como vimos na Introdução desta tese. No topo da Pirâmide estão as revisões sistemáticas de ensaios clínicos aleatorizados, consideradas as evidências com maior força na escala de evidência. Demonstramos no nosso estudo e vimos nos trabalhos de El Dib, Atallah e Andriolo (2007); Dennis e Hodnett (2009) que fazer parte de uma revisão sistemática não atribui qualidade metodológica aos textos científicos. Assim, voltamos ao já citado no início desta tese e reiteramos a importância dos esforços relatados por Witter (2006), Volpato (2007, 2009, 2010a, 2010b, 2011, 2013), Sabadini et al. (2009a) e Trzesniak et al. (2012) para minimizar as dificuldades dos autores na redação de artigos com mais qualidade metodológica e rigor na apresentação de resultados.

Avaliamos todos os artigos latino-americanos incluídos na revisão e pareamos a amostra de comparação. A amostra aleatória é a que permite maiores generalizações, seguida das amostras estratificadas, sistemática e de grupo (Lay & Reis, 2005). Na população total da revisão sistemática (337), o desempenho dos estudos foi semelhante ao do nosso universo, ou seja, metodologicamente fraco (Silva, 2013). Podemos assinalar que a aleatorização de parte de nossa amostra foi um ponto forte do estudo, pois manteve os dados de desempenho dos estudos da subamostra semelhantes aos da amostra total. Seria interessante que outros estudos dessa natureza, com populações diferentes de artigos e escalas apropriadas, fossem desenvolvidos, objetivando comparar os dados aqui levantados.

Outro ponto forte do estudo foi a não limitação do período de busca nas bases de dados, que possibilitou a inclusão de trabalhos desde o ano de 1980. A ampliação dos termos de buscas em bases de dados que não apresentavam palavras controladas por vocabulários também foi um diferencial deste estudo, embora tenha demandado muito trabalho na seleção dos estudos.

Percorrermos todas as etapas de uma revisão que nos possibilitaram vivenciar os caminhos que levam à geração de uma revisão sistemática de estudos exploratórios e, com base nas discussões, esperamos poder ampliar as orientações aos autores na criação de evidências de qualidade. Para tanto, foi necessário acompanhar uma revisão sistemática que seguisse instruções confiáveis como as de Higgins e Green (2006), que exige rigor no processo de busca, seleção e análise

dos trabalhos. Acreditamos que o conhecimento adquirido no processo da revisão sistemática poderá colaborar para oferecermos orientações mais dirigidas aos editores e autores da região nos cursos de organização de uma revista científica e preparo de artigo científico. Com base no aprendizado poderemos, também, revisar diversas orientações do livro *Publicar em Psicologia: um Enfoque para a Revista Científica* (Sabadini et al., 2009b).

Realizamos todo o percurso da revisão sistemática e, ao avaliarmos os artigos desenvolvidos em nossa região, observamos que nossos autores estão contribuindo aquém do desejado em relação ao oferecimento de evidências científicas de qualidade. No entanto, o mais desalentador é que, em relação à prevalência de DPP, tanto nossa região como outras estão gerando artigos fracos, oriundos de estudos metodologicamente inadequados, o que compromete a sistematização desses resultados em evidências recomendáveis. A construção de guias para aumentar a qualidade dos estudos primários, que os tornem capazes de serem incluídos em revisões sistemáticas, com resultados possíveis de reduzir as incertezas geradas da consulta a estudos fracos, é uma das orientações que tiramos de nossa experiência. Na verdade, estamos falando da etapa que antecede a redação do artigo científico, uma vez que foi a estruturação da metodologia de pesquisa que prejudicou o bom desempenho dos nossos artigos.

Acreditamos que nossa contribuição amplia a discussão sobre a Prática Baseada em Evidências, ou Psicologia Baseada em Evidências. Considerando nossos resultados, as competências dos autores da região, em relação à produção de estudos com metodologias adequadas que possibilitem resultados fortes o suficiente para serem utilizados na tomada de decisão, ainda precisam ser aprimoradas. Claro que estamos falando de um pequeno recorte de artigos sobre prevalência de DPP. O panorama de qualidade dos estudos pode ser outro, bem diferente em outras áreas de estudos que não de prevalência. Nesse caso, vale visitar o estoque de conhecimento gerado na área em busca de avaliação das evidências científicas em outros ramos da ciência psicológica.

A busca por uma ciência sustentável, que revise e revise sistematicamente o conhecimento já publicado será um grande apoio à tomada de decisão na prática profissional e na gestão pública. Antes, porém, precisamos aprimorar a capacidade dos autores para a produção de artigos rigorosamente elaborados, com metodologias consistentes e resultados confiáveis.

6.6 Qualidade de forma: títulos, resumos palavras-chave

Os títulos, resumos e palavras-chave foram elemento de grande importância no nosso estudo. Desde a etapa das buscas nas bases de dados, até a seleção dos trabalhos para inclusão na revisão sistemática, essas partes do texto científico foram decisivas. Podemos enumerar três fatores que tornam esses elementos chave no processo de revisões: 1) títulos, resumos e palavras-chave são indexadores nas bases de dados; 2) são elementos utilizados na inclusão/exclusão em revisões pelos pesquisadores; 3) podem ser decisivos na publicação de manuscritos em revistas nacionais e estrangeiras.

Os serviços de indexação para geração das bases de dados dependem muito da acurácia dos títulos dos artigos, pois, como observamos na análise dos artigos diversas revistas não publicam mais palavras-chave, o que torna esse elemento ainda mais fundamental. Além do próprio artigo, as fontes de informação rastreiam artigos nas referências também, o que eleva o valor do título para a visibilidade do trabalho. O próprio Centro Cochrane Iberoamericano (2011) recomenda que os revisores consultem as referências dos artigos para chegarem a outros trabalhos de interesse. Nesse caso, o fator de decisão será, obrigatoriamente, o título.

Um artigo que não tivesse as palavras do título escolhidas adequadamente, no caso do nosso estudo, pode ter sido excluído, sem chance de ser submetido às demais etapas da avaliação.

Dos 68 artigos avaliados, analisamos os 30 que declararam ter como objetivo analisar a prevalência de DPP apenas 19 foram considerados com títulos adequados de acordo com as normas da APA. A não adequação dos 11 reprovados se deu em relação ao número de palavras. A APA considera que um título não deve ter mais que 12 palavras. Vale lembrar que as normas da APA foram desenvolvidas para orientar publicações em língua inglesa e as línguas latinas são mais extensas que o inglês. Assim, encaixar os títulos dos artigos latino-americanos no padrão 12 palavras pode ser complicado.

Em relação aos resumos, dos 68 resumos 49 atenderam as recomendações da APA, inclusive em relação ao tamanho, máximo de 250 palavras. Podemos dizer que os autores estão desenvolvendo resumos com bastante qualidade, o que pode ter beneficiado a inclusão dos estudos na revisão. Entretanto, no item palavras-

chave apenas um dos nossos artigos seguiu a orientação de uso dos vocabulários controlados.

À medida que mais informações migram para a Internet as chaves para recuperação dos dados ganham relevância. Com as mudanças no sistema de organização da informação e a nova cultura de operação no mundo digital, muito do trabalho que era feito por pessoas, agora é realizado por robôs, que organizam e recuperam a informação. Para que esse modelo possa efetivamente funcionar precisamos estruturar a informação de forma sistematizada e com regras muito claras. Assim, novas linguagens vão substituindo as antigas formas de organização do conhecimento. Atualmente, no ambiente de organização da informação temos lido e ouvido falar um termo antes utilizado na classificação das espécies em botânica ou zoologia: taxonomia. Na organização da informação na web semântica a taxonomia tem sido entendida como classificação sistemática dos termos para busca nas mais diversas fontes de informação.

Apesar de considerarmos que as técnicas de navegação por taxonomias não são superiores às outras formas de busca, como por exemplo, as buscas por palavras-chave sendo uma complementar à outra, com o advento dos sítios e portais, a utilidade das taxonomias tem-se tornado mais importante e difundida, pois a navegação é uma interface bem projetada sobre a informação desses sítios ou portais. Então, faz-se necessário que o classificacionista se pautem em princípios classificatórios explícitos e consistentes. (Campos & Gomes, 2008)

Quanto mais informatizado for o processo, maior será a necessidade de imprimir qualidade às partes pré-textuais dos artigos. Caso contrário, corremos o risco de, após um esforço enorme para conseguir publicar um artigo, este fique invisível na Internet. Caso seja recuperado, poderá ser excluídos das revisões por não ter sido devidamente intitulado, resumido, ou indexado com palavras-chave adequadas.

7 CONCLUSÃO

Para conclusão deste trabalho voltamos à primeira frase da Introdução: “A qualidade de um artigo científico é diretamente proporcional à qualidade metodológica e à apresentação dos resultados da pesquisa que o originou”. Outro fator de qualidade indispensável é a inserção epistemológica da questão investigada (Trzesniak, Plata-Caviedes & Córdoba-Salgado, 2012), aspecto o qual, em nossa pesquisa, não se constituiu em uma variável, já que nos fixamos na prevalência da

DPP. Percorremos um longo percurso para chegar aos resultados finais da nossa investigação e concluirmos que os artigos deste estudo tiveram um desempenho fraco em relação à qualidade desejada, ou seja, às pontuações nos oito itens da Escala de Loney et al. (1998). Poderíamos concluir que a Escala foi muito rígida, daí o fraco desempenho dos nossos artigos. No entanto, com base em estudos citados na Discussão, os quais se depararam com resultados semelhantes, parece-nos que o problema não reside na Escala, embora seja aconselhável uma revisão de seus oito itens. O uso de outras escalas com maior amplitude na pontuação também é aconselhável em estudos futuros. Podemos levantar outras questões em relação aos resultados e sugerir algumas possibilidades para incentivar estudos complementares ao nosso.

Alcançar pontos na Escala não influenciou na inclusão do artigo na revisão sistemática sobre prevalência de DPP. Entretanto, a heterogeneidade dos estudos foi grande e percebida até mesmo dentro de uma mesma cidade. As prevalências identificadas foram tão diferentes que a revisão sistemática concluída não fornece dados suficientes para a tomada de decisão na gestão pública como, por exemplo, propor uma política para prevenção ou tratamento da DPP. Mesmo reunindo em um único trabalho 337 estudos levantados em diversas partes do mundo e em diferentes culturas, os resultados não ofereceram um retrato confiável do panorama da prevalência de DPP, capaz de apoiar o tomador de decisão na definição de políticas dirigidas às mães que sofrem desse mal, que tanto prejudica as famílias e, principalmente, as crianças. Se os estudos incluídos tivessem seguido um protocolo metodológico orientado para a obtenção de resultados mais homogêneos, certamente teríamos um retrato mais confiável e o recomendaríamos como apoio na tomada de decisão dos gestores públicos e agentes governamentais.

Avaliar os estudos já publicados para inclusão em revisões sistemáticas que apoiem a tomada de decisão do gestor público e do profissional é importante e precisa receber incentivo para sua expansão enquanto linhas de pesquisas em universidades e centros de pesquisas. Entretanto, seguir roteiros que avaliem os estudos antes da publicação nos parece prioritário, pois, depois de publicado, a única coisa que se pode dizer é que não está adequado, mas o prejuízo não pode ser sanado. Os roteiros para avaliação de manuscritos por pareceristas devem se especializar e avançar rumo à adequação metodológica dos estudos para que apresentem resultados passíveis de reunião e análise em estudos de revisões e,

preferencialmente, que possibilitem aplicar estatísticas de metanálise. Uma tarefa para pesquisadores pode ser a de desenvolver guias que possibilitem o desenvolvimento de pesquisas com critérios de qualidade que, uma vez transformadas em artigos, permitam oferecer resultados confiáveis e adequados, possibilitando que a ciência possa ser mais aproveitada em prol da melhoria da vida das pessoas.

A Psicologia intensificou sua atuação na América Latina, enquanto ciência e profissão, a partir dos anos de 1960, sendo incluída na prática médica, nas orientações vocacionais, nas questões da aprendizagem e, principalmente, inserindo seus profissionais nos hospitais e ambulatórios para o tratamento da saúde mental.

O trabalho do psicólogo vem adquirindo, nos últimos anos, reconhecido valor na promoção de saúde e na melhoria da qualidade de vida do ser humano, mediante ações educativas, preventivas e da própria intervenção psicológica. Como resultado dessa presença no cenário da saúde, a Psicologia latino-americana construiu um importante corpo de conhecimento que está sendo organizado pela BVS-Psi ULAPSI, SciELO, LILACS e outras importantes fontes de informação. Entretanto, o excesso de contribuições impede aos psicólogos, pesquisadores e gestores utilizarem esse conhecimento de forma rápida e precisa para o tratamento dos pacientes, avanço da ciência psicológica e na tomada de decisão em gestão pública. Revisitar e sistematizar esse valioso conhecimento em revisões sistemáticas, narrativas e exploratórias é um caminho para o registro e análise constante dos procedimentos utilizados e intervenções que promoverão melhores resultados para os pacientes e facilitarão as atividades dos gestores na área. A Psicologia latino-americana necessita se apoderar das técnicas de revisão sistemática para aplicação da Psicologia Baseada em Evidências como ferramenta de apoio ao avanço das pesquisas e na solução de problemas comuns da população da região. Outros trabalhos como o nosso podem revisitar esse estoque de conhecimento e trazer à tona lacunas importantes para estudos e pesquisas na área.

A nossa pergunta de pesquisa indagou sobre a equivalência da qualidade de artigos latino-americanos em comparação com os de outras regiões. Concluímos que o desempenho dos artigos foi equivalente, de acordo com Escala de Loney e colaboradores. Lamentavelmente, tanto os artigos latino-americanos, quando os de outras regiões, apresentaram desempenho fraco e são de baixa qualidade.

7.1 Direções Futuras

Como todo estudo, este responde algumas poucas questões e suscita muitas outras. A questão que queremos destacar é a equivalência da qualidade dos estudos latino-americanos e não latino-americanos. Uma pergunta que continua nos intrigando é: como produzir artigos com qualidade suficientemente adequada para que sejam boas evidências científicas e que possam ser sintetizados? Estudos futuros podem propor a criação de protocolos de verificação de qualidade metodológica e da apresentação de resultados para utilização por pesquisadores ainda na fase de definição da metodologia de pesquisa. Dessa maneira, teremos pesquisas com metodologia adequada gerando resultados confiáveis para publicação em revistas científicas. Tal providência, certamente, assegurará maior êxito na construção de revisões que, de forma efetiva, servirão de apoio à tomada de decisão dos profissionais e gestores públicos.

Outro aspecto inexplorado neste trabalho e que pode gerar estudos interessantes é a aplicação de técnicas de revisões integrativas na área da Psicologia. Dentre os métodos de revisão, a integrativa parece ser bem adequada à Psicologia, pois permite a inclusão de estudos quantitativos e qualitativos e com delineamentos experimental e quase-experimental, proporcionando uma compreensão mais ampla do tema de interesse. O método de revisão integrativa permite a combinação de dados da literatura empírica e teórica que podem ser direcionados para a definição de conceitos. A revisão integrativa tem diferentes finalidades: definição de conceitos, identificação de lacunas para áreas de estudos, revisão de teorias, análise metodológica dos estudos sobre um determinado tópico. O conhecimento revisado com rigor reduz incertezas sobre recomendações práticas e permite generalizações mais precisas sobre um fenômeno em relação às intervenções, resultando em cuidado mais efetivo e de custo benefício mais acessível. A Psicologia construiu um corpo de conhecimento que precisa ser revisitado com métodos rigorosos para seu reaproveitamento. A aplicação das melhores evidências na solução dos problemas da população dos países latino-americanos pode representar um caminho para se evitar a duplicação de esforços e repetição de experiências exitosas e sem êxito. O aproveitamento das práticas efetivas e de sucesso significa economia de tempo, que poderá resultar em ganhos de saúde para o ser humano.

Sugerimos, também, que parcerias entre pesquisadores de Psicologia e profissionais da informação sejam ampliadas visando à construção de estudos acadêmicos nos quais os conhecimentos se somem em busca de novos projetos para as duas áreas do conhecimento: Psicologia e Ciência da Informação. Psicólogos, gestores públicos e políticos precisam entender que as revisões sistemáticas são um primeiro lugar para se buscar evidências de qualidade na tomada de decisão. Para avaliar a aplicabilidade dos resultados de uma revisão sistemática, é necessário ter habilidades para avaliar a qualidade dos estudos que a compõem. A soma da experiência das duas categorias profissionais pode facilitar o processo de aprendizagem nessas questões.

Finalmente, sugerimos que a comunidade científica, as agências de fomento e os programas de pós-graduação revejam alguns critérios de avaliação da produção científica dos cientistas, pesquisadores e docentes, considerando a qualidade dos estudos e desestimulando o produtivismo exagerado que valoriza os aspectos quantitativos em detrimento dos qualitativos.

REFERÊNCIAS

- Alhabib, S., Nur, U., & Jones, R. (2010). Domestic violence against women: Systematic review of prevalence studies. *Journal of Family Violence*, 25, 369-382. doi: 10.1007/s10896-009-9298-4
- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663-694.
- American Psychiatric Association. (1994). *DSM-IV. Manual de diagnóstico e estatística dos distúrbios mentais* (4a ed.). Lisboa: Climepsi.
- American Psychological Association, Council of Editors. (1952). Publication Manual of the American Psychological Association [Supplement, Pt 2]. *Psychological Bulletin*, 49(4), 389-449.
- American Psychological Association. (1957). *Publication manual of the American Psychological Association* (Rev. ed.). Washington, DC: Author.
- American Psychological Association. (1967). *Publication manual of the American Psychological Association* (Rev. ed.). Washington, DC: Author
- American Psychological Association. (1974). *Publication manual of the American Psychological Association* (2nd ed.). Washington, DC: Author
- American Psychological Association. (1983). *Publication manual of the American Psychological Association* (3rd ed.). Washington, DC: Author
- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- American Psychological Association. (2012). *Manual de publicação da American Psychological Association* (6a ed., D. Bueno, trad.). Porto Alegre: Penso. (Tradução da 6a ed.: *Publication manual of the American Psychological Association*, 2010)
- American Psychological Association Task Force on Issues of Sexual Bias in Graduate Education, Washington DC US. (1975). Guidelines for nonsexist use of language. *American Psychologist*, 30(6), 682-684. doi:10.1037/h0076869
- Angelini, A. L. (1954). O emprêgo do ácido glutâmico na terapia da debilidade mental. *Ciência e Cultura*, 6(3), 123-131.

- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Report standards for research in Psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839-851.
- Barata, G. (2010). América Latina e o impacto de suas publicações científicas. *Ciência e Cultura*, 62(2), 2010. Recuperado de <http://cienciaecultura.bvs.br/pdf/cic/v62n2/a07v62n2.pdf>
- Barone, L. M. C. (2011). A escala de Procusto ou à procura de um leito para a psicanálise. *Jornal de Psicanálise*, 80(44), 87-93. Recuperado de <http://pepsic.bvsalud.org/pdf/jp/v44n80/v44n80a08.pdf>
- Barreto, M. L. (2004). O conhecimento científico e tecnológico como evidência para políticas e atividades regulatórias em saúde. *Ciência & Saúde Coletiva*, 9(2), 329-338. doi: <http://dx.doi.org/10.1590/S1413-81232004000200010>
- Bégault, B. (2009). O periódico científico, um papel para a mediação de informação entre pesquisadores: qual seu futuro no ambiente digital? *RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 3(3), 91-96. doi: 10.3395/reciis.v3i3.284pt
- Benjamin Jr., L. T., & VandenBos, G. R. (2006). The window on psychology's literature: A history of psychological abstracts. *American Psychologist*, 61(9), 941-954.
- Bennet, H., Einarson, A., Taddio, A., Koren, G., & Einarson, T. (2004). Prevalence of depression during pregnancy: Systematic review. *American College of Obstetricians and Gynecologists*, 103, 698-709.
- Biruel, E., & Pinto, R. (2011). Bibliotecário um profissional a serviço da pesquisa. In *Anais, XXIV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação*. Maceió.
- Botella, J., & Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical and Health Psychology*, 6(2), 425-440. Recuperado de http://www.aepc.es/ijchp/articulos_pdf/ijchp-185.pdf
- Boutron, I., Moher, D. Douglas, D. G., Schult, K. F., & Ravaud, P. (2008). Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals Internal of Medicine*, 148: 295-309.
- Boyle, H. M. (1998). Guidelines for evaluating prevalence studies. *Evidence Based Mental Health*, 1, 37-39. doi: 10.1136/ebmh.1.2.37
- Buela-Casal, G. (2003). Evaluación de la calidad de los artículos y de las revistas científicas: propuesta del factor de impacto ponderado y de um índice de calidad. *Psicothema*, 15(1), 23-35. Recuperado de <http://www.psicothema.com/pdf/400.pdf>

- Campos, M. L. A., & Gomes, H. E. (2008). Taxonomia e classificação: o princípio de categorização. *DataGramaZero - Revista de Ciência da Informação*, 9(4). Recuperado de http://www.dgz.org.br/ago08/Art_01.htm
- Candeiro, G. T. de M., França, R. C., Cremonini, C. C., Miranda, L. F. de, Provenzano, M. de M., Hotta, P. T. H., ... Lima, L. A. P. A. de. (2011). Fontes de financiamento em ensaios clínicos: análise crítica. *Journal of Helth Sciences Institute*, 29(4), 248-249. Recuperado de http://www.unip.br/comunicacao/publicacoes/ics/edicoes/2011/04_out-dez/V29_n4_2011_p248-249.pdf
- Cantilino, A., Zambaldi, C., Albuquerque, T., Paes, J., Montenegro, A., & Sougey, E. (2010). Postpartum depression in Recife - Brazil: Prevalence and association with bio-socio-demographic factors. *Jornal Brasileiro de Psiquiatria*, 59, 1-9. doi: <http://dx.doi.org/10.1590/S0047-20852010000100001>
- Cantisano, T. G., Morales, D. M., & Dominguez, M. (2007). Mobbing: A meta-analysis and integrative model of its antecedents and consequences. *Psicothema*, 19(1), 88-94.
- Castro, A. A., Saconato, H., Guidugli, F., & Clark, O. A. C. (2002). *Curso de revisão sistemática e metanálise* [Online]. São Paulo: LED-DIS/UNIFESP. Curso disponível em <http://www.virtual.epm.br/cursos/metanalise>
- Castro, R. (2011). Indexação de revistas científicas em bases de dados. In G. P. Witter, D. A. Población, L. M. S. V. Ramos, & V. M. B. Funaro, *Revistas científicas - dos processos tradicionais às perspectivas alternativas de comunicação* (pp. 109-126). São Paulo: Ateliê.
- Centro Cochrane Iberoamericano. (2011). *Manual Cochrane de revisiones sistemáticas de intervenciones* (Versión 5.1.0) [Atualizada em marzo de 2011]. Barcelona: Centro Cochrane Iberoamericano. Recuperado de http://www.cochrane.es/files/handbookcast/Manual_Cochrane_510.pdf
- Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (2011). *Evidências na gestão municipal da saúde* (Pt 1-3) [Video]. Recuperado de http://www.youtube.com/watch?v=WXdkKzg_Mhl
- Chow, S. L. (2002). Methods in psychological research. In *Encyclopedia of life support systems (EOLSS)*. Oxford: Eolss Publishers. Recuperado de <http://cogprints.org/2643/1/EOLSSrm.pdf>
- Chubin, D., & Hackett, E. (1990). *Peerless science: Peer review and US science policy*. Albany: State University of New York Press.
- Classificação Internacional de Doenças: CID 10. *MedicinaNet*. Recuperado de <http://www.medicinanet.com.br/cid10/a.htm>

- Cordeiro, A. M., Oliveira, G. M., Rentería, J. M., & Guimarães, C. A. (2007). Revisão sistemática: uma revisão narrativa. *Revista do Colégio Brasileiro de Cirurgiões*, 34(6), 428-431. doi: <http://dx.doi.org/10.1590/S0100-6991200700060001>
- Costa, A. L. F., & Yamamoto, O. H. (2008). Publicação e avaliação de periódicos científicos: paradoxos da avaliação *Qualis* de Psicologia. *Psicologia em Estudo*, 13(1), 13-24. doi: [doi.org/10.1590/S1413-73722008000100003](http://dx.doi.org/10.1590/S1413-73722008000100003)
- Coutinho, D. S., Baptista, M. N., & Moraes, P. R. (2002). Depressão pós-parto: prevalência e correlação com o suporte social. *Infanto - Revista de Neuropsiquiatria da Infância e Adolescência*, 10(2), 63-71.
- 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>
- Cruz, E. B. S., Simões, G. L., & Faisal-Cury, A. (2005). Rastreamento da depressão pós-parto em mulheres atendidas pelo Programa de Saúde da Família. *Revista Brasileira de Ginecologia e Obstetrícia*, 27(4), 181-188. doi: <http://dx.doi.org/10.1590/S0100-72032005000400004>
- Cunha, M. B., & Cavalcanti, C. R. O. (2008). *Dicionário de biblioteconomia e arquivologia*. Brasília, DF: Briquet de Lemos/Livros.
- Davyt, A., & Velho, L. (2000). A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro?. *História, Ciências, Saúde-Manguinhos*, 7(1), 93-116. doi: <http://dx.doi.org/10.1590/S0104-59702000000200005>
- Delgado López-Cozar, E., & Repiso, R. (2013). El impacto de las revistas de comunicación: comparando Google Scholar Metrics, Web of Science y Scopus. *Comunicar. Revista Científica de Comunicación y Educación*, 21(41). doi: 10.3916/C41-2013-04
- Dennis, C.-L., & Hodnett, E. (2009) Psychosocial and psychological interventions for treating postpartum depression [First published: October 17, 2007]. *Cochrane Database of Systematic Reviews*, (7). (The Cochrane Library. Cochrane Reviews Art. No: CD006116). doi: 10.1002/14651858.CD006116.pub2
- Dias, R. C., & Dias, J. M. D. (2006). Prática baseada em evidências: uma metodologia para a boa prática em fisioterapêutica. *Fisioterapia em Movimento*, 19(1), 11-16.
- Dvorak, E. A. (1956). General guide to a study of research reports. *Peabody Journal of Education*, 34(3), 141-144.
- El Dib, R. P. (2007). Como praticar a medicina baseada em evidências [Editorial]. *Jornal Vascular Brasileiro*, 6(1), 1-4. doi: <http://dx.doi.org/10.1590/S1677-54492007000100001>

- El Dib, R. P., Atallah, A. N., & Andriolo, R. B. (2007). Mapping the Cochrane evidence for decision making in health care. *Journal of Evaluation in Clinical Practice*, 13(4), 689-692. doi: 10.1111/j.1365-2753.2007.00886.x/pdf
- Eugenio, B. (2000). On the usage of Kappa to evaluate agreement on coding tasks. In *Proceedings of the Second International Conference on Language Resources and Evaluation*. Athens, Greece: Athena Research Center.
- Evidências científicas: informação para a gestão. (2009). *Revista de Saúde Pública*, 43(2), 1084-1085. Recuperado de <http://www.scielo.br/pdf/rsp/v43n6/23.pdf>
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42), 17028-17033.
- Ferreira, A. S., & Abreu, M. L. T. (2007). Desconstruindo um artigo científico. *Revista Brasileira de Zootecnia*, 36(Suplemento Especial), 377-385. doi: <http://dx.doi.org/10.1590/S1516-35982007001000034>
- Fleck, M. P., Berlim, M. T., Lafer, B., Sougey, E. B., Del Porto, J. A., Brasil, M. A., ... Hetem, L. A. (2009). Revisão das diretrizes da Associação Médica Brasileira para o tratamento da depressão (Versão integral). *Revista Brasileira de Psiquiatria*, 31(Suplemento 1). doi: <http://dx.doi.org/10.1590/S1516-44462009000500003>
- Foo, J. Y. A. (2011). A retrospective analysis of the trend of retracted publications in the field of biomedical and life sciences. *Science and Engineering Ethics*, 17(3), 459-468.
- Froemming, L. M. S., Luce, F. B., Perin, M. G., Sampaio, C. H., Beber, S. J. N., & Trez, G. (2000). Análise da qualidade dos artigos científicos da área de marketing no Brasil: as pesquisas survey na década de 90. *Revista de Administração Contemporânea*, 4(3), 201-219. doi: <http://dx.doi.org/10.1590/S1415-65522000000300011>
- Gabarito, R. M., Gómez, S. T., López-González, M., Macías, L. M., D'Agostino, M., & Cabo, J. V. (2009). Revisiones sistemáticas exploratorias. *Medicina y Seguridad del Trabajo*, 55(216), 12-19. doi: <http://dx.doi.org/10.4321/S0465-546X2009000300002>
- Galvão, C. M., Sawada, N. O., & Rossi, L. A. (2002). A prática baseada em evidências: considerações teóricas para sua implementação na enfermagem perioperatória. *Revista Latino-Americana de Enfermagem*, 10(5), 690-695. doi: <http://dx.doi.org/10.1590/S0104-11692002000500010>
- Giannakopoulos, N., Rammelsberg, P., Eberhard, L., & Schmitter, M. (2012). A new instrument for assessing the quality of studies on prevalence. *Clinical Oral Investigations*, 16, 781-788.
- Gibbs, W. W. (1995) Trends in scientific communication. Lost science in the third world. *Scientific American*, 273(2), 76-83.

- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research*, 5(10), 3-8.
- Goldenberg, S. (2001). *Publicação do trabalho científico: compromisso ético*. São Paulo. Recuperado de http://metodologia.org/wp-content/uploads/2010/08/saul_etica.PDF
- Gonçalves, D. F. R. (2010). *Depressão pós-parto* (Dissertação de Mestrado). Faculdade de Medicina da Universidade do Porto, Porto. Recuperado de <http://repositorio-aberto.up.pt/bitstream/10216/52156/2/Depresso%20Psparto.pdf>
- Goodheart, C. D., Kazdin, A. E., & Sternberg, R. J. (Eds.). (2006). *Evidence-based Psychotherapy: Where practice and research meet*. Washington, DC: American Psychological Association.
- Granja, E. C. (1995). *Análise da produção científica do curso de pós-graduação do Instituto de Psicologia da USP, no período de 1980 a 1989* (Tese de Doutorado). Instituto de Psicologia da Universidade de São Paulo, São Paulo. Recuperado de <http://www.teses.usp.br/teses/disponiveis/47/47131/tde-21012009-153405/pt-br.php>
- Greene, L. J. (1998). O dilema do editor de uma revista biomédica: aceitar ou não aceitar. *Ciência da Informação*, 27(2), 230-232. Recuperado de <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/372/333>
- Greenhalgh, T. (2008). *Como ler artigos científicos: fundamentos da medicina baseada em evidências* (3a ed., A. P. Fajardo, trad.). Porto Alegre: Artmed.
- Hall, B. W., Ward, A. W., & Comer, C. B. (1988). Published educational research: An empirical study of its quality. *Journal of Educational Research*, 81(3), 182-189.
- Hasbún H., J., Risco N., L., Jadresic M., E., Galleguillo U., T, González A., M, & Garay S., J. (1999). Depresión postparto: prevalencia y factores de riesgo. *Revista Chilena de Obstetricia y Ginecología*, 64(6), 466-470.
- Haslam, N., & Laham, S. M. (2010). Quality, quantity, and impact in academic publication. *European Journal of Social Psychology*, 40, 216-220. doi: 10.1002/ejsp.727
- Higgins, J. P. T., & Green, S. (Eds.). (2006). *Cochrane handbook for systematic reviews of interventions 4.2.6* [updated September 2006] (*The Cochrane Library*, Issue 4). Chichester, UK: John Wiley & Sons. Recuperado de <http://www.cochrane.org/sites/default/files/uploads/Handbook4.2.6Sep2006.pdf>
- Hoppen, N. (1998). Sistemas de informação no Brasil: uma análise dos artigos científicos dos anos 90. *Revista de Administração Contemporânea*, 2(3), 151-177. doi: <http://dx.doi.org/10.1590/S1415-65551998000300009>

- Iribarry, I. N. (2003). O que é pesquisa psicanalítica? *Ágora: Estudos em Teoria Psicanalítica*, 6(1), 115-138. doi: <http://dx.doi.org/10.1590/S1516-14982003000100007>
- Johnson, G. B. (1957). A method for evaluating research articles in education. *Journal of Educational Research*, 51, 353-354.
- Klein, V. C., & Linhares, M. B. M. (2006). Prematuridade e interação mãe-criança: revisão sistemática da literatura. *Psicologia em Estudo*, 11(2), 277-284. doi: <http://dx.doi.org/10.1590/S1413-73722006000200006>
- Kronick, D. A. (1990). Peer review 18th century scientific journalism. *JAMA: The Journal of the American Medical Association*, 263(10), 1321-1323.
- Lay, M. C. D., & Reis, A. T. L. (2005). Análise quantitativa na area de estudos ambiente-comportamento. *Ambiente Construido*, 5(2), 21-36. Recuperado de <http://seer.ufrgs.br/ambienmteconstruido/article/view/3616/1998>
- Levant, R. F. (2005). *Report of the 2005 task force on evidence-based practice*. Recuperado de <http://www.apa.org/practice/resources/evidence/evidence-based-report.pdf>
- Lewis, R. W. (2011). Epidemiology of sexual dysfunction in Asia compared to the rest of the world. *Asian Journal of Andrology*, 13(1), 152-158. Recuperado de <http://www.nature.com/aja/journal/v13/n1/pdf/aja2010108a.pdf>
- Lobato, G., Moraes, C., & Reichenheim, M. (2012). Magnitude da depressão pós-parto no Brasil: uma revisão sistemática. *Revista Brasileira de Saúde Materno Infantil*, 11(4), 369-179. doi: dx.doi.org/10.1590/S1519-38292011000400003
- Loney, P. L., Chambers, L. W., Bennett, K. J., Roberts, J. G., & Stratford, P. W. (1998). Critical appraisal of the health research literature: Prevalence or incidence of a health problem. *Chronic Diseases in Canada*, 19(4), 170-176.
- Loney, P. L., & Stratford, P. W. (1999). The prevalence of low back pain in adults: A methodological review of the literature. *Physical Therapy*, 79(4), 384-396.
- Lopes, M. A. (2002). *Revisão sistemática dos estudos de prevalência de demência: período de 1994 à 2000* (Dissertação de Mestrado). Faculdade de Medicina da Universidade de São Paulo, São Paulo.
- Lopes, M. M., & Piscitelli, A. (2004). Revistas científicas e a constituição do campo de estudos de gênero: um olhar desde as “margens”. *Estudos Feministas*, 12(Número Especial), 115-121. Recuperado de <http://www.redalyc.org/articulo.oa?id=38114353013>
- Luiz, B. A. J. (2002). Meta-Análise: definição, aplicações e sinergia com dados espaciais. *Cadernos de Ciência e Tecnologia*, 19(3), 407-428. Recuperado de http://webnotes.sct.embrapa.br/pdf/cct/v19/cc19n3_03.pdf

- Lunsford, T. R., & Lunsford, B. R. (1995). Research forum - The research sample. Part 1: Sampling. *Journal of Prosthetics and Orthotics*, 7(3), 105-112. Recuperado de http://www.oandp.org/jpo/library/1995_03_105.asp
- Marconi, M. A., & Lakatos, E. M. (2010). *Fundamentos de metodologia científica* (7a ed.). São Paulo: Atlas.
- Manes, A. G., & Flach, L. (2012). Revisão sistemática da produção científica sobre fraudes em periódicos internacionais de contabilidade indexados ao ISI e Scopus. *Revista Contabilidade Vista & Revista*, 23(2), 163-189. Recuperado de <http://www.redalyc.org/articulo.oa?id=197026277007>
- Marques, F. (2013). Os limites do índice-h: supervalorização do indicador que combina quantidade e qualidade da produção científica gera controvérsia. *Pesquisa FAPESP*, 207, 35-39. Recuperado de http://revistapesquisa.fapesp.br/wp-content/uploads/2013/05/035-039_IndiceH_207.pdf
- Matínez, G. S. (2012). La revisión por pares y la selección de artículos para publicación. *Revista Colombiana de Psicología*, 21(1), 27-35. Recuperado de <http://www.scielo.org.co/pdf/rcps/v21n1/v21n1a03.pdf>
- McGraw, K., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. Recuperado de <http://psycnet.apa.org/journals/met/1/1/30.pdf>
- Melnik, T. (2006). *A efetividade da psicoterapia no tratamento da disfunção erétil: revisão sistemática e metanálise* (Tese de Doutorado). Instituto de Psicologia da Universidade de São Paulo, São Paulo.
- Melnik, T., & Atallah, A. N. (2011). *Psicologia baseada em evidências: provas científicas da efetividade da psicoterapia*. São Paulo: Santos.
- Melo Junior, E. F. (2011). *Prevalência de depressão perinatal e fatores associados* (Tese de Doutorado). Universidade Estadual de Campinas, Campinas, SP.
- Meneghini, R., & Packer, A. L. (2007). Is there science beyond English? *EMBO Reports*, 8(2), 112-116. doi:10.1038/sj.embor.7400906
- Mendes, K. S., Silveira, R. C. C. P., & Galvão, C. M. (2008). Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. *Texto & Contexto - Enfermagem*, 17(4), 758-764. <http://dx.doi.org/10.1590/S0104-07072008000400018>
- Moraes, I. G. S., Pinheiro, R. T., Silva, R. A., Horta, B. L., Sousa, P. L. R., & Faria A. D. (2006). Prevalência da depressão pós-parto e fatores associados. *Revista de Saúde Pública*, 40(1), 65-70. doi: <http://dx.doi.org/10.1590/S0034-89102006000100011>

- Murray, C. J. L., & Lopez, A. D. (1997). Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*, 349(9064), 1498-1504.
- Oliveira, N. S., Oliveira, J. M., & Bergamaschi, D. P. (2006). Concordância entre avaliadores na seleção de artigos em revisões sistemáticas. *Revista Brasileira de Epidemiologia*, 9(3), 309-315. doi: <http://dx.doi.org/10.1590/S1415-790X2006000300005>
- Peccin, M. S. (2005). *Estudos clínicos randomizados em ortopedia e traumatologia publicados em revistas indexadas no Thomsom Institute for Scientific Information (ISI®) e MEDLINE: estudo comparativo* (Tese de Doutorado). Escola Paulista de Medicina da Universidade Federal de São Paulo, São Paulo.
- Pessanha, C. (1998). Critérios editoriais de avaliação científica: notas para discussão. *Ciência da Informação*, 27(2), 226-229. doi: <http://dx.doi.org/10.1590/S0100-19651998000200020>
- Pheula, G. F., & Isolan, L. R. (2007). Psicoterapia baseada em evidências em crianças e adolescentes. *Revista de Psiquiatria Clínica*, 34(2), 74-83. doi: <http://dx.doi.org/10.1590/S0101-60832007000200003>
- Pinto, A. C., & Andrade, J. B. (1999). Fator de impacto de revistas científicas: qual o significado deste parâmetro? *Química Nova*, 22(3), 448-453. doi: <http://dx.doi.org/10.1590/S0100-40421999000300026>
- Pinto, A. L., & Pontes Júnior, J. (2008). Índice científico da Ciência da Informação: comparativo dos autores com maior visibilidade no ISI e no Google. *Informação & Informação*. Recuperado de <http://www.uel.br/revistas/uel/index.php/informacao/article/view/1801/1537>
- Price, D. S. (1976). *Ciência desde a Babilônia* (L. Hegenberg & O. S. da Mota, trads.). Belo Horizonte: Itatiaia.
- Prins, J., Blanker, M., Bohnen, A., Thomas, S., & Bosch, J. (2002). Prevalence of erectile dysfunction: A systematic review of population-based studies. *International Journal of Impotence Research*, 14(6), 422-432.
- Reiners, A. A. O., Costa, A. L. R. C., Zborowski, I. P., Lourenço, R., Hortense, P., & Sant'Anna, R. P. M. (2002). Avaliação em ciência e tecnologia: o papel dos pares na elaboração de pareceres com fim de publicação em revistas científicas. In 8. *Proceedings, Brazilian Nursing Communication Symposium* (pp. 1-4). Recuperado de http://www.proceedings.scielo.br/scielo.php?script=sci_arttext&pid=MSC0000000052002000100002&lng=en&nrm=iso&tlng=pt
- Reis, F. B., Ciconelle, R. M., & Faloppa, F. (2002). Pesquisa científica: a importância da metodologia. *Revista Brasileira de Ortopedia*, 37(3), 51-55. Recuperado de <http://alessandro.ccbs.uepb.edu.br/pdf/texto3.pdf>

- Rodrigues, A. V. F., Crespo, I. M., & Miranda, C. L. (2006). Ética em pesquisa em publicações científicas. *Em Questão*, 21(1), 33-50. Recuperado de <http://200.144.189.42/ojs/index.php/revistaemquestao/article/viewFile/3708/3496>
- Rodrigues, L. O. C. (2008). Publicar mais, ou melhor? O Tamanduá olímpico. *Revista Brasileira de Ciências do Esporte*, 29(1), 35-47. Recuperado de <http://rbceonline.org.br/revista/index.php/RBCE/article/view/8/14>
- Rodrigues, R. S., & Oliveira, A. B. (2012). Periódicos científicos na América Latina: títulos em acesso aberto indexados no ISI e SCOPUS. *Perspectivas em Ciência da Informação*, 17(4), 77-99. Recuperado de <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1593/1100>
- Ryder, E. (2001). Predominio del idioma inglés y no de la lengua nativa, el español, en trabajos de impacto en salud pública en Venezuela. *INCI*, 26(12), 619-623. Recuperado de http://www.scielo.org.ve/scielo.php?pid=S0378-18442001001200008&script=sci_arttext
- Sabadini, A. A. Z. P., Nascimento, M. M., & Cadidé, I. (2007). Análise das normas de referências adotadas pelas revistas brasileiras de psicologia. In *Seminário Internacional de Bibliotecas Digitais*. Recuperado de <http://www.bibliotecadigital.unicamp.br/document/?code=32073>
- Sabadini, A. A. Z. P., Sampaio, M. I. C., & Koller, S. H. (2009a). Preparando um artigo científico. In A. A. Z. P. Sabadini, M. I. C. Sampaio, & S. H. Koller (Orgs.), *Publicar em psicologia: um enfoque para a revista científica* (pp. 117-162). São Paulo: Associação Brasileira de Editores Científicos de Psicologia; Instituto de Psicologia da Universidade de São Paulo. Recuperado de <http://www.ip.usp.br/portal/images/stories/biblioteca/Publicarempsicologiaversao2012.pdf>
- Sabadini, A. A. Z. P., Sampaio, M. I. C., & Koller, S. H. (Orgs.). (2009b). *Publicar em psicologia: um enfoque para a revista científica*. São Paulo: Associação Brasileira de Editores Científicos de Psicologia; Instituto de Psicologia da Universidade de São Paulo. Recuperado de <http://www.ip.usp.br/portal/images/stories/biblioteca/Publicarempsicologiaversao2012.pdf>
- Sampaio, A. A. S., Azevedo, F. H. B., Cardoso, L. R. B., Lima, C., Pereira, M. B. R., & Andery, M. A. P. A. (2008). Uma introdução aos delineamentos experimentais de sujeito único. *Interação em Psicologia*, 12(1), 151-164. Recuperado de <http://ojs.c3sl.ufpr.br/ojs2/index.php/psicologia/article/view/9537/9218>
- Sampaio, M. I. C., & Sabadini, A. A. Z. P. (2009). Indexação e fator de impacto. In A. A. Z. P. Sabadini, M. I. C. Sampaio, & S. H. Koller (Orgs.), *Publicar em psicologia: um enfoque para a revista científica* (pp. 103-116). São Paulo: Associação Brasileira de Editores Científicos de Psicologia; Instituto de Psicologia da Universidade de São Paulo. Recuperado de <http://www.ip.usp.br/portal/images/stories/biblioteca/Publicarempsicologiaversao2012.pdf>

- Sampaio, M. I. C., & Sabadini, A. A. Z. P. (2012). La merecida visibilidad de las revistas latinoamericanas de Psicología. *Revista Colombiana de Psicología*, 21(1), 111-123. Recuperado de <http://www.scielo.org.co/pdf/rcps/v21n1/v21n1a08.pdf>
- Sánchez-Meca, J., Boruch, R. F., Petrosino, A., & Rosa, A. I. (2002). La Colaboración Campbell y la práctica basada en la evidencia. *Papeles del Psicólogo*, 83, 44-48. Recuperado de <http://www.um.es/metaanalysis/pdf/7073.pdf>
- Sawyer, A., Ayers, S., & Smith, H. (2010). Pre-and postnatal psychological wellbeing in Africa: A systematic review. *Journal of Affective Disorders*, 123(1-3), 17-29.
- Shamliyan, T., Kane, R., & Dickinson, S. (2010). 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.
- Shamu, S., Abrahams, N., Temmerman, M., Musekiwa, A., & Zarowsky, C. (2011). A systematic review of African studies on intimate partner violence against pregnant women: Prevalence and risk factors. *Plos One*, 6(3), 1-9. Recuperado de <http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0017591&representation=PDF>
- Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Silva, G. A. (2013). *Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise* (Tese de Doutorado). Instituto de Psicologia da Universidade de São Paulo, São Paulo. Recuperado de <http://www.teses.usp.br/teses/disponiveis/47/47132/tde-24052013-143728/pt-br.php>
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760. Recuperado de <http://psycnet.apa.org/journals/amp/32/9/752.pdf>
- Soares, A. B., Victoria, M. S., Cavalieri, A. M. A. P., & Bottino, A. G. (2006). A Psicologia divulgada através de periódicos científicos indexados e através de revistas de grande circulação do Brasil: um estudo preliminar. *Psicologia para América Latina*, (7). Recuperado de <http://www.psicolatina.org/Siete/divulgada.html>
- Stumpf, I. R. C. (1996). Passado e futuro das revistas científicas. *Ciência da Informação*, 25(3), 383-386. Recuperado de <http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/463/422>
- Thesaurus of psychological index terms*. (2007). (11th ed.). Washington, DC: American Psychological Association.
- Trzesniak, P., Plata-Caviedes, T., & Córdoba-Salgado, O. (2012). A qualidade de conteúdo, o grande desafio para os editores científicos. *Revista Colombiana de*

- Psicología*, 21(1), 57-78. Recuperado de <http://www.scielo.org.co/pdf/rcps/v21n1/v21n1a05.pdf>
- Tyrer, P. (2005). Combating editorial racism in psychiatric publications [Editorial]. *British Journal of Psychiatry*, 186, 1-3. Recuperado de <http://bjp.rcpsych.org/content/186/1/1.full.pdf+html>
- Urdaneta M., J., Rivera S., A., García I., J., Guerra V., M., Baabel Z., N., & Contreras B., A. (2011). Factores de riesgo de *depresión* posparto en puérperas venezolanas valoradas por medio de la escala de Edimburgo. *Revista Chilena de Obstetricia y Ginecología*, 76(2), 102-112. doi: <http://dx.doi.org/10.4067/S0717-75262011000200007>
- Vázquez, C., & Moreno, M. N. (2003). Psicología (clínica) basada en la evidencia (PBE): una revisión conceptual y metodológica. In J. L. Romero (Ed.), *Psicópolis: paradigmas actuales y alternativos de la Psicología contemporánea*. Barcelona: Kairos. Recuperado de http://pendientedemigracion.ucm.es/info/psisalud/carmelo/PUBLICACIONES_pdf/2003-Psicologia%20basada%20en%20la%20evidencia.pdf
- Victora, C. G., & Moreira, C. B. (2006). Publicações científicas e as relações Norte-Sul: racismo editorial? *Revista de Saúde Pública*, 40(Número Especial), 36-42. doi: <http://dx.doi.org/10.1590/S0034-89102006000400006>
- Vieira, V. P. (2006). *Avaliação da adequação metodológica de estudos clínicos sobre voz e laringe publicados em revistas brasileiras e internacionais* (Dissertação de Mestrado). Escola Paulista de Medicina da Universidade Federal de São Paulo, São Paulo.
- Vigod, S., Villegas, L., Dennis, C.-L., & Ross, L. (2010). Prevalence and risk factors for postpartum depression among women with preterm and low birth-weight infants: A systematic review. *International Journal of Obstetrics and Gynaecology*, 117(5), 540-550.
- Volpato, G. (2007). *Bases teóricas para redação científica*. São Paulo: Cultura Acadêmica.
- Volpato, G. (2009). *Administração da vida científica*. São Paulo: Cultura Acadêmica.
- Volpato, G. (2010a). *Dicas para redação científica*. São Paulo: Cultura Acadêmica.
- Volpato, G. L. (2010b). *Pérolas da redação científica*. São Paulo: Cultura Acadêmica.
- Volpato, G. L. (2011). *Método lógico para redação científica*. Botucatu, SP: Best Writing.
- Volpato, G. (2013). *Ciência: da filosofia à publicação* (6a ed.). São Paulo: Cultura Acadêmica.

- Wandt, E., Adams, G. W., Collett, D. M., Michael, W. B., Ryans, D. G., & Shay, C. B. (1967). An evaluation of educational research published in journals. Report of the Committee on Evaluation of Research. *American Educational Research Association* [Unpublished]. Recuperado de <http://www.indiana.edu/~educy520/readings/wandt65.pdf>
- Ward, A. W., Hall, B. W., & Schramm, C. (1975). Evaluation of published educational research: A national survey. *American Educational Research Journal*, 12(2), 109-128.
- Werneck, G. L. (2006). Ética na publicação científica [Editorial]. *Cadernos Saúde Coletiva*, 14(1), 5-6. Recuperado de http://www.iesc.ufrj.br/cadernos/images/csc/2006_1/artigos/editorial_2006_1.pdf
- Witter, G. P. (1990). Pesquisa bibliográfica, pesquisa documental e busca da informação. *Estudos Psicologia*, 7(1), 5-30.
- Witter, G. P. (2006). Produção científica: escalas de avaliação. In D. A. Población, G. P. Witter, & J. F. M. Silva, *Comunicação e produção científica: contexto, indicadores e avaliação* (pp. 287-311). São Paulo: Angelara.
- Witter, G. P. (2007). Psicoterapia: pesquisa e prática [Resenha do livro *Evidence-based Psychoterapy. Where practice and research meet*, de C. D. Goodheart, A. E. Kazdin, & R. J. Sternberg (Eds.)]. *Psicologia: Reflexão e Crítica*, 20(2), 523-524. doi: <http://dx.doi.org/10.1590/S0102-79722007000300021>
- Yamamoto, O. H., Menandro, P. R. M., Koller, S. H., Lo Bianco, A. C., Hutz, C. S., Bueno, J. L. O., & Guedes, M. C. (2002). Avaliação de periódicos científicos brasileiros da área da psicologia. *Ciência da Informação*, 31(2), 163-177. doi: <http://dx.doi.org/10.1590/S0100-19652002000200017>
- Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalization, structure and functions of the referee system. *Minerva*, 9(1), 66-100. Recuperado de <http://link.springer.com/content/pdf/10.1007%2F01553188.pdf>

APÊNDICE A - Instruções para seleção e análise dos estudos

Parte 1 – Seleção dos artigos, de acordo com os critérios de inclusão da revisão sistemática.

Respostas:

1 = verdadeiro

2 = falso

3 = não está claro

4 = não se aplica

- Critério A: O estudo é primário e quantitativo, com delineamento transversal ou de coorte longitudinal.

Excluir:

- Revisões de literatura
- Estudos qualitativos

Atenção: se este item for marcado com "falso" ou "não está claro", todos os demais itens deverão ser marcados com "não se aplica".

- Critério B: Foi avaliada a prevalência de depressão pós-parto na amostra estudada.

Excluir:

- Estudos que objetivaram avaliar melancolia pós-parto, psicose pós-parto, ansiedade ou outro transtorno.

Observação: se um estudo contém a prevalência de depressão pós-parto, deverá ser incluído, ainda que mencione também a prevalência de outros transtornos.

- Critério C: Os participantes do estudo são mães humanas.

Excluir: estudos que tenham pesquisado, exclusivamente:

- Homens (depressão paterna)
- Animais não humanos

- Critério D: A amostra estudada não foi submetida a nenhuma intervenção com o objetivo de prevenir DPP.

Excluir:

- Estudos que tenham realizado alguma intervenção (psicoterapia, medicação ou outro tipo de intervenção) antes da avaliação para diagnóstico de DPP (na gestação ou após o parto), com o objetivo de prevenir a ocorrência de DPP.

- Critério E: O método de diagnóstico é mencionado claramente e baseia-se em instrumento objetivo de medida, ou em avaliação feita por profissional de saúde mental.

Excluir:

- Estudos que não mencionam claramente o método de diagnóstico.
- Estudos em que o diagnóstico foi feito a partir de prontuário médico ou psicológico, sem indicar a forma de diagnóstico que levou ao preenchimento do prontuário.
- Estudos em que o diagnóstico não foi feito por instrumento objetivo ou por profissional de saúde mental. Ex.: avaliação subjetiva de agentes comunitários de saúde.

- Critério F: O período do puerpério em que o diagnóstico foi feito está claramente mencionado. Pelo menos uma das avaliações de DPP ocorreu de 8 dias (inclusive) a 365 dias (inclusive) após o parto.

Excluir:

- Estudos em que o período em que o diagnóstico foi feito não está claro.
- Estudos em que o intervalo de tempo em que diagnóstico foi feito compreende período inferior a 8 dias ou superior a 365 dias após o parto.
- Critério G: O estudo foi realizado na população geral e não declara incluir população exclusivamente com características específicas, como: doenças crônicas (ex.: diabetes, depressão, hipertensão), outras doenças graves (ex.: câncer), complicações na gestação (ex.: eclampsia), bebês com condições de saúde alteradas (ex.: má formação fetal, baixo peso, prematuridade), homossexualidade, adoção, uso de drogas na gestação (álcool, drogas ilícitas).

Excluir:

- Estudos com amostras que apresentam problemas graves de saúde da mãe ou do bebê.
- Estudos sobre mães em relações homoafetivas.
- Estudos sobre mães adotivas.

Não excluir:

- Estudos sobre minorias étnicas.

- Critério H: Em pelo menos uma das avaliações para identificar DPP, o resultado de prevalência está relatado de forma clara, contendo todos os

seguintes dados: período após o parto, instrumento, tamanho de amostra e número absoluto ou porcentagem de participantes consideradas deprimidas.

- **Critério I:** O texto completo do estudo foi publicado em inglês, português ou espanhol.

Parte 2 – Instruções para o preenchimento do protocolo de avaliação dos estudos

1.Nome do revisor

Informar quem está coletando os dados.

2.Código do estudo

Informar, conforme a tabela, o código do estudo que está sendo revisado.

3.Ano em que o trabalho foi publicado

Informar com 4 dígitos.

4.Autores

Usar formato: Sobrenome, Iniciais. Ex.:

•Autor 1: Ades, C.

•Autor 2: Patto, M.H.S.

Informar na ordem de autoria que aparece no trabalho.

Caso existam mais de 10 autores, o campo 10 deverá ser usado para informar os autores excedentes. Nesse caso, separar os autores por ponto e vírgula.

Ex.: Autor 10: Ades, C.; Patto, M.H.S.

5.Título do trabalho em inglês

Informar o título em inglês, conforme aparecer na publicação. Caso não exista título em inglês, deixar esse campo em branco. Caso o título seja traduzido para outros idiomas, todos devem ser informados nos campos correspondentes.

6.Título do trabalho em português

Informar o título em português, conforme aparecer na publicação. Caso não exista título em português, deixar esse campo em branco. Caso o título seja traduzido para outros idiomas, todos devem ser informados nos campos correspondentes.

7.Título do trabalho em espanhol

Informar o título em espanhol, conforme aparecer na publicação. Caso não exista título em espanhol, deixar esse campo em branco. Caso o título seja traduzido para outros idiomas, todos devem ser informados nos campos correspondentes.

8. Tipo de publicação

A resposta a esta pergunta abrirá os campos necessários para montar as referências bibliográficas.

Considera-se que dissertação é referente ao nível de Mestrado; tese é referente ao nível de Doutorado, Pós-Doutoramento ou Livre-Docência; monografia se refere a outros tipos de trabalho (inclusive de graduação).

9. Palavras-chave

Usar as palavras-chave que constam no texto publicado (geralmente, folha de rosto do artigo). Não usar as definidas pelas bases de dados eletrônicas. Procurar copiar e colar, para manter exatamente a forma como o autor definiu as palavras-chave. Inclusive os eventuais erros de digitação devem ser mantidos. Inserir palavras-chave nos diversos idiomas, sempre que estiverem disponíveis no texto consultado.

10. Idiomas em que o resumo está disponível.

11. Idiomas em que o texto completo está disponível

12. País em que a coleta de dados foi realizada

A lista contém todos os países oficiais. Caso o nome de algum país não esteja disponível, usar opção "outro" e escrever o nome, de preferência em inglês. Caso o estudo não declare o local em que a coleta de dados foi realizada, usar opção "não declarou". Caso o estudo não seja primário (ex.: revisão de literatura), usar opção "outro" e indicar o delineamento do estudo.

13. Critérios de inclusão

Um estudo será excluído sempre que um ou mais critérios de inclusão for preenchido com "falso" ou "não está claro".

Critério A: incluir apenas estudos com delineamento transversal ou de coorte longitudinal. Excluir estudos com delineamento caso-controle ou ensaio clínico aleatorizado, por exemplo. Ver descrição dos delineamentos no item 14.

Excluir estudos retrospectivos, ou seja: aqueles em que a mãe é entrevistada e questionada sobre seu estado de humor em período anterior e descontinuado com o período da entrevista. Por exemplo, devem ser excluídos estudos com as seguintes características:

- A mãe foi entrevistada quando o bebê tinha 26 meses, com perguntas que avaliavam como ela se sentia quando o bebê tinha de 2 a 4 meses.
- A mãe foi entrevistada quando o bebê tinha 9 meses, com perguntas que

avaliavam como ela se sentia quando o bebê tinha 5 meses.

Se o Critério A for preenchido com "falso" ou "não está claro", preencher os demais critérios com "não se aplica".

Critério B: deve ser marcado como "falso" nos seguintes casos:

- Estudo avaliou apenas a prevalência de outros transtornos que não a DPP, como por exemplo: melancolia pós-parto, psicose pós-parto, transtorno de estresse pós-traumático etc.
- Estudo avaliou depressão apenas durante a gestação, ou após o bebê ter completado um ano de vida.
- Estudo não avaliou a prevalência de DPP, mas sim outros atributos, como por exemplo: associação com fatores de risco, validade de escala de rastreamento de DPP.
- Estudo avaliou a incidência de DPP e não é possível deduzir, pelos dados apresentados, o resultado referente à prevalência. Lembrar que incidência é o número de casos novos de uma determinada doença, enquanto prevalência é a proporção de casos em um determinado instante.

Critério C: excluir estudos cujos participantes são: pais (e que não tenham mães), animais não humanos (ratas, vacas etc.), crianças etc.

Critério D: excluir estudos que tenham feito alguma intervenção com o objetivo de prevenir DPP, como por exemplo: atendimentos individuais ou em grupos, visitas domiciliares etc.

Critério E: devem ser incluídos estudos com critérios de diagnóstico replicáveis.

Excluir estudos com as seguintes características:

- identificação de mães com DPP através de pesquisa em prontuários, quando não é possível saber o critério usado pelos profissionais de saúde para diagnosticar DPP.
- identificação de mães com DPP através do julgamento de um profissional, sem informações que permitam saber qual foi o critério usado.

Critério F: verificar se o período após o parto foi mencionado e atende ao critério.

Ter cuidado com as seguintes situações:

- Estudos que dizem que avaliaram DPP um ano após o parto, ou 12 meses

após o parto: se o bebê já havia completado 12 meses no momento da avaliação, o estudo deve ser excluído.

- Estudos que dizem ter avaliado DPP "até 1 mês após o parto", ou "*within one month postpartum*": verificar se parte das avaliações pode ter sido feita antes de o bebê completar uma semana de vida. Caso positivo, excluir.

Critério G: apesar de várias situações serem contempladas, observar que:

- Os estudos incluídos devem ser feitos na população geral.
- Os estudos cujos participantes são, exclusivamente, mulheres com problemas de saúde, ou que têm bebês com problemas de saúde serão excluídos.
- Não serão excluídos estudos feitos exclusivamente em populações com maior vulnerabilidade social, como: minorias étnicas, baixo nível socioeconômico, adolescentes etc. Essas situações serão apontadas nos campos do formulário referentes ao perfil das participantes.
- Serão excluídos casos de adoção e de mães lésbicas, por serem situações muito específicas e menos passíveis de generalização.

Critério H: este critério não existia nas avaliações de títulos e resumos, pois só pode ser verificado diante do texto completo. Quando o resultado de prevalência não estiver claro, contendo todas as informações necessárias para definir esse resultado, o estudo deverá ser excluído.

Importante

- Quando um estudo tiver a mesma amostra, total ou parcialmente, que outro estudo, apenas um deles deverá ser incluído. O(s) demais poderá(ão) ser usado(s) para complementar informações.
- Quando um estudo for excluído, indicar o motivo pelo qual não preencheu um, ou mais¹ critérios de inclusão, no campo "observações". Esse campo também pode ser usado para outros comentários a respeito da inclusão ou exclusão de um estudo.
- Os estudos cujos participantes são, exclusivamente, mulheres com problemas de saúde ou que têm bebês com problemas de saúde serão excluídos.
- Não serão excluídos estudos feitos exclusivamente em populações com

maior vulnerabilidade social, como: minorias étnicas, baixo nível socioeconômico, adolescentes etc. Essas situações serão apontadas nos campos do formulário referentes ao perfil das participantes.

- Serão excluídos casos de adoção e de mães lésbicas, por serem situações muito específicas e menos passíveis de generalização.

14. Delineamento

Considerar

- Transversal: as participantes são submetidas a apenas uma avaliação ao longo do tempo.
- Longitudinal: as participantes são seguidas ao longo do tempo e submetidas a mais de uma avaliação, em diferentes períodos.

Alguns estudos transversais avaliam algumas mães 2 meses após o parto, outras 5 meses, outras 10 meses. Lembrar que delineamento longitudinal significa que a mesma mãe foi avaliada mais de uma vez. Ou seja, para ser longitudinal, o estudo deve avaliar a mesma mãe aos 2 meses, 5 meses e 10 meses.

15. Amostra

- Censo: admitir como censo a pesquisa que afirmou ter procurado avaliar todas as integrantes de uma determinada região geográfica. Ex.: se o estudo buscou determinar a prevalência de DPP na cidade X, todas as mães que deram à luz na cidade X em um determinado período devem ter sido entrevistadas para que o estudo se caracterize como censo.
- Amostra aleatória: configura-se quando uma seleção aleatória de pessoas de um grupo alvo foi convidada a participar da pesquisa. Foi considerada amostra aleatória aquela em que as participantes foram sorteadas previamente, por um método que deu a cada mulher do grupo alvo chance igual de ser escolhida. Ex.: sorteio por computador, ou por bolas retiradas de uma urna. A estratificação ou amostragem por clusters não compromete a classificação da amostra como aleatória, desde que sejam cumpridos os demais requisitos.
- Não aleatória: qualquer outro tipo de amostra que não os dois acima.

Considerar que a amostra é aleatória apenas quando houver descrição da metodologia usada para a aleatorização. Apenas dizer que a amostra é aleatória, ou "*random sample*", não é suficiente para que a amostra do estudo seja classificada como aleatória. Quando isso acontecer, deve ser usada a opção "não está claro".

Caso seja marcada a opção "amostra aleatória", será aberto um item no formulário que solicitará a descrição do método de aleatorização. Transcrever fielmente a descrição presente no estudo (de preferência, copiar e colar).

Caso seja marcada a opção "amostra não aleatória", será aberto no formulário o item "Método de amostragem", com as opções:

- Consecutiva: são abordadas todas as mães que comparecem a um determinado serviço (ex.: maternidade, clínica de pré-natal) em um determinado período.
- Bola de neve: é solicitado a uma entrevistada que convide outras conhecidas, e a estas conhecidas que convide outras, e assim por diante.
- Sistemática: configura-se quando a amostra foi selecionada por um método sistemático, mas está sujeita a vieses. Ex.: entrevistar todas as participantes que tiverem bebês em determinados dias da semana ou do mês, entrevistar mulheres cujos prontuários tenham números pares ou ímpares, usar um intervalo numérico fixo para selecionar as participantes em uma lista (ex.: entrevistar a 5ª, a 10ª, a 15ª, etc.). A estratificação ou amostragem por clusters não compromete a classificação da amostra como sistemática, desde que sejam cumpridos os demais requisitos.
- Amostra de conveniência: qualquer outro método de seleção de participantes foi considerado como amostra de conveniência.
- Amostra sistemática: configura-se quando a amostra foi selecionada por um método sistemático, mas está sujeita a vieses. Ex.: entrevistar todas as participantes que tiverem bebês em determinados dias da semana ou do mês, entrevistar mulheres cujos prontuários tenham números pares ou ímpares, usar um intervalo numérico fixo para selecionar as participantes em uma lista (ex.: entrevistar a 5ª, a 10ª, a 15ª etc.). A estratificação ou amostragem por clusters não compromete a classificação da amostra como sistemática, desde que sejam cumpridos os demais requisitos.

- Amostra de conveniência: qualquer outro método de seleção de participantes foi considerado como amostra de conveniência.

Esse item também envolve um campo para que seja transcrita a descrição que o estudo fornece sobre o processo de amostragem. Procurar transcrever fielmente, de preferência, copiar e colar.

16.Critérios de inclusão / exclusão. Procurar transcrever fielmente, de preferência, copiar e colar.

17.Confiabilidade (em inglês, *reliability*): são medidas de quanto o instrumento de diagnóstico é replicável. Existem várias medidas de confiabilidade, sendo que as mais comuns são: concordância teste-reteste, *Split-half*, *alpha de cronbach* (ou somente alfa).

O item permite as seguintes opções de resposta:

- Não mencionou confiabilidade: não há nenhuma menção a confiabilidade no texto.
- Mencionou confiabilidade obtida em estudos prévios: estudo menciona valores (números) indicando a confiabilidade obtida em estudos anteriores. Ex.: em estudo anterior, o *alfa de cronbach* foi 0,86. Não considerar neste item estudos que indicam que a confiabilidade foi "alta", "média", "aceitável" etc., caso a descrição não seja acompanhada por números.
- Não explicitou a confiabilidade, mas citou outro(s) artigo(s) que avaliou(aram) confiabilidade: considerar neste item estudos que mencionaram a palavra "confiabilidade" (*reliability*), e/ou o nome de um teste de confiabilidade (ex.: concordância teste-reteste, *Split-half*, *alpha de cronbach*) e indicam uma referência bibliográfica em que o valor da confiabilidade estaria presente. Não será checada a referência a que o estudo se refere, por indisponibilidade de tempo. Mencionar uma referência sobre a escala ou outro instrumento de diagnóstico de DPP sem mencionar que a confiabilidade pode ser encontrada nessa referência não é suficiente para enquadrar o estudo nesta alternativa.
- Verificou confiabilidade no estudo em questão: os dados do estudo avaliado são usados para calcular a confiabilidade. Ex.: o *alfa de cronbach* neste

estudo é 0,86.

18.Validade: indica que o instrumento mede o que se propôs a medir. As principais medidas de validade são: sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo. Também é comum aparecer a medida de acurácia ou AUC (*area under curve*). As opções de resposta são semelhantes ao item anterior:

- Não mencionou validade: não há nenhuma menção a validade no texto.
- Mencionou validade obtida em estudos prévios: estudo menciona valores (números) indicando a validade obtida em estudos anteriores. Ex.: em estudo anterior, a sensibilidade foi 86%. Não considerar neste item estudos que indicam que a validade foi "alta", "média", "aceitável", etc., caso a descrição não seja acompanhada por números.
- Não explicitou a validade, mas citou outro(s) artigo(s) que avaliou(aram) validade: considerar neste item estudos que mencionaram a palavra "validade" (*validity*), e/ou o nome de um atributo de validade (ex.: sensibilidade, especificidade) e indicam uma referência bibliográfica em que o valor da validade estaria presente. Não será checada a referência a que o estudo se refere, por indisponibilidade de tempo. Mencionar uma referência sobre a escala ou outro instrumento de diagnóstico de DPP, sem mencionar que a validade pode ser encontrada nessa referência não é suficiente para enquadrar o estudo nesta alternativa.
- Verificou validade no estudo em questão: os dados do estudo avaliado são usados para calcular a validade. Ex.: a sensibilidade neste estudo é 86%.

19.Cálculo do tamanho da amostra: verificar se o estudo menciona ter usado algum critério para calcular o tamanho da amostra.

20.Informações sobre participantes perdidas: as perdas do estudo deverão ser calculadas. Não confiar na porcentagem fornecida no texto. O cálculo das perdas deve ser feito da seguinte maneira:

A) Determinar o número de participantes elegíveis, ou seja: todas as participantes que foram abordadas e que poderiam ser entrevistadas. Vamos chamar esse valor de E. Não computar em E o número de participantes excluídas por não se adequarem aos critérios de inclusão definidos para o estudo.

B) Determinar o número de participantes perdidas no recrutamento (que será chamado de R), por diferentes razões, como por exemplo:

- Recusa em participar do estudo
- Dados incompletos
- Não encontradas (quando o recrutamento é feito por telefone ou endereço)

C) Determinar o número de participantes perdidas no seguimento longitudinal, desde a primeira até a última avaliação. Esse valor será chamado de L.

D) Realizar o seguinte cálculo: $P = (R+L)/E \times 100$

Caso o estudo não tenha informações suficientes para realizar esse cálculo, marcar resposta "não" na pergunta "Existem informações sobre a proporção de participantes perdidas...".

Se essa pergunta for marcada com "sim", outros itens do formulário serão abertos. O primeiro deles será um campo para colocar o resultado do cálculo acima. Outra pergunta é se as características das participantes perdidas foram descritas. Considerar, nesse caso, se o estudo comparou respondentes e não respondentes em termos de perfil sociodemográfico. A seguir, será perguntado se há diferença entre respondentes e não respondentes. Considerar que qualquer diferença estatisticamente significativa entre perfis, seja em qualquer variável analisada (ex.: idade, estado civil, etc.) implica em responder "sim" a este item.

21.Avaliações de DPP: indicar se o diagnóstico de DPP foi feito por instrumento de autoavaliação, entrevista clínica ou ambos. Caso seja marcada a opção "entrevista clínica", serão abertos outros itens:

Proporção de participantes em que foi aplicada entrevista clínica; alguns estudos aplicam entrevista clínica em uma subamostra, que pode ser definida por sorteio, ou pelo resultado de um instrumento de autoavaliação.

Taxa de concordância entre entrevistadores: marcar "sim" caso mais de um entrevistado tenha avaliado cada entrevistada, gerando uma taxa de concordância (porcentagem de avaliações coincidentes). Marcar "não" caso mais de um entrevistado tenha participado das entrevistas, mas não foi avaliada a concordância entre eles. Marcar "não se aplica" caso apenas um entrevistador tenha feito as entrevistas.

Informações sobre cegamento: caso o estudo tenha usado mais de um instrumento para avaliar DPP, é importante que os entrevistadores estejam cegos para as avaliações anteriores, ou para as avaliações de seus pares ao verificar a taxa de

concordância. Essa informação precisa ser fornecida de forma explícita. Caso o estudo trabalhe apenas com uma avaliação clínica, este item deve ser marcado com "não se aplica".

22. Escala de Loney - delineamento

Considerar adequado estudo que usou:

- Delineamento transversal ou de coorte longitudinal
- Amostra aleatória ou censitária

Considerar que a amostra é aleatória apenas quando houver descrição da metodologia usada para a aleatorização. Apenas dizer que a amostra é aleatória, ou "*random sample*", não é suficiente para que a amostra do estudo seja classificada como aleatória. Quando isso acontecer, deve ser usada a opção "não está claro".

Considerar censitária a pesquisa que avaliou todas as mães de uma determinada localidade geográfica (município, vila, estado) que deram à luz a filhos vivos em um determinado período. Levar em consideração o quanto o universo em que foi feito censo é representativo da população da localidade geográfica. Por exemplo, uma pesquisa realizada em uma única unidade de saúde, de uma cidade em que existem outras unidades, não será considerada censitária, mesmo que todas as mães que deram à luz naquela unidade em determinado período tenham sido avaliadas quanto à presença de DPP.

23. Escala de Loney - base amostral

Considerar adequadas as seguintes bases amostrais:

- Dados do censo
- Dados de pesquisa porta-a-porta
- Outra base amostral com chance de viés bastante reduzida

Considerar inadequadas as seguintes bases amostrais:

- Lista telefônica (exclui mulheres sem telefone)
- Prontuários de maternidade (excluem partos caseiros e os realizados em outros hospitais)
- Participantes que decidiram se voluntariar por internet
- Outras bases amostrais com grande chance de viés

24. Escala de Loney - Tamanho da amostra

Considerar se o estudo tem pelo menos uma avaliação válida com 470 participantes ou mais. Considera-se avaliação válida aquela que foi feita de 8 a

365 dias após o parto.

25. Escala de Loney - Confiabilidade e validade do instrumento de pesquisa

Para que este item seja pontuado, é necessário que estejam presentes ambas as informações: confiabilidade E validade. Se apenas uma informação estiver presente, o item não deve ser pontuado.

Não é preciso mencionar, necessariamente, os valores de sensibilidade e confiabilidade. O artigo pode indicar que o instrumento é válido e confiável ("*reliable*") e indicar uma referência bibliográfica.

Se o instrumento de pesquisa for entrevista clínica, considerar que é válido e confiável se for mencionada concordância entre entrevistadores.

Se for usado mais de um instrumento, atribuir pontuação a este item se mencionou confiabilidade e validade em pelo menos um dos instrumentos de pesquisa.

26. Escala de Loney - Taxa de resposta: lembrar que, para que este item seja pontuado, é necessário que duas condições sejam satisfeitas simultaneamente:

A. A taxa de perda deve ser menor que 30%

B. Deve ser feita a comparação entre respondentes e não respondentes.

O item deve ser pontuado, mesmo que a comparação indique existir diferença significativa entre os dois grupos (respondentes e não respondentes).

27. Escala de Loney - Intervalo de confiança e detalhamento por subgrupo

Este item também é pontuado quando duas condições são satisfeitas simultaneamente:

A. Deve ser considerado se, ao fornecer o resultado de prevalência, o estudo fornecer o intervalo de confiança, e não apenas a estimativa pontual. Lembrar que o intervalo de confiança precisa se referir à prevalência, e não a outras medidas, como as *odds ratio* referentes a fatores de risco.

B. Deve ser considerado se os subgrupos importantes do estudo foram considerados, de forma que se forneçam resultados de prevalência específicos para esses subgrupos, com seus respectivos intervalos de confiança.

28. Escala de Loney - Perfil sociodemográfico

Considerar que está presente caso o estudo forneça informações como por exemplo: idade das participantes, estado civil, escolaridade, renda, paridade etc.

APÊNDICE B - Referências dos estudos avaliados

- Affonso, D. D., De, A. K., Horowitz, J. A., & Mayberry, L. J. (2000). Na international study exploring levels of postpartum depressive symptomatology. *Journal of Psychosomatic Research*, 49, 207-216.
- Alvarado, R., Perucca P., E., Rojas, M., Monardes, J., Olea, E., Neves G. , E., & Vera C., A. (1993). Aspectos gineco-obstetricos en mujeres que desarrollan una depresión postparto. *Revista Chilena de Obstetricia y Ginecología*, 58(3), 239-244.
- Alvarado-Esquivel, C., Sifuentes-Álvarez, A., Estrada-Martínez, S., Salas- Martínez, C., Hernández-Alvarado, A. B., Ortiz-Rocha, S. G., ... Sandoval-Herrera, F. (2010). Prevalencia de depresión posnatal en mujeres atendidas en hospitales públicos de Durango, México. *Gaceta Médica de México*, 146(1), 1-9.
- Alvarado-Esquivel, C., Sifuentes-Alvarez, A., Salas-Martinez, C., & Martínez-García, S. (2006). Validation of the Edinburgh postpartum depression scale in a population of puerperal women in Mexico. *Clinical Practice and Epidemiology in Mental Health*, 2(33).
- Anderson, L. N., Campbell, M. C., da Silva, O., Freeman, T., & Xie, B. (2008). Effect of maternal depression and anxiety on use of health services for infants. *Canadian Family Physician*, 54, 1718-1719.
- Aramburú, P., Arellano, R., Jáuregui, S., Pari, L., Salazar, P., & Sierra, O. (2008). Prevalencia y factores asociados a depresión posparto en mujeres atendidas en establecimientos de salud del primer nivel de a. *Revista Peruana de Epidemiología*, 12(3), 1-5.
- Barbosa, E. M. S., Petribú, K., Mariano, M. H. A., Ferreira, M. N. L., & Almeida, A. (2008). Quality of life in postpartum depressed adolescents. *Revista Brasileira de Psiquiatria*, 30(1), 86-90.
- Beeghly, M., Olson, K. L., Weinberg, K., Pierre, S. C., Downey, N., & Tronick, E. Z. (2003). Prevalence, stability, and socio-demographic correlates of depressive symptoms in Black mothers during the first 18 months postpartum. *Maternal and Child Health Journal*, 17(3), 157-168.
- Boyce, P., & Hickey, A. (2005). Psychosocial risk factors to major depression after childbirth. *Social Psychiatry and Psychiatric Epidemiology*, 40, 605-612.
- Bunevičius, A., Kusminskas, L., & Bunevičius, R. (2009). Validation of the Lithuanian version of the Edinburgh Postnatal Depression Scale. *Medicina (Kaunas)*, 45(7), 544-548.
- Canaval, G. E., González, M. C., Schallmoser, L. M., Tovar, M. C., & Valencia, C. (2000). Depresión postparto, apoyo social y calidad de vida en mujeres de Cali, Colombia. *Colombia Médica*, 31(1), 4-10.

- Cantilino, A., Carvalho, J. A., Maia, A., Albuquerque, C., Cantilino, G., & Sougey, E. B. (2007). Translation, validation and cultural aspects of Postpartum Depression Screening Scale in Brazilian Portuguese. *Transcultural Psychiatry*, 44(4), 672-684.
- Cantilino, A., Zambaldi, C. F., Albuquerque, T. L. C., Paes, J. A., Montenegro, A. C. P., & Sougey, E. B. (2010). Postpartum depression in Recife - Brazil: Prevalence and association with bio-socio-demographic factors. *Jornal Brasileiro de Psiquiatria*, 59(1), 1-9.
- Castañón S. , C., & Pinto L. , J. (2008). Use to the Edinburgh Postnatal Depression Scale to detect postpartum depression. *Revista Médica de Chile*, 136, 851-858.
- Chaudron, L. H., Szilagyi, P. G., Tang, W., Anson, E., Talbot, N. L., Wadkins, H. I. M., Wisner, K. L. (2010). Accuracy of depression screening tools for identifying postpartum depression among urban mothers. *Pediatrics*, 125(3), 609-617.
- Chee, C. Y. I., Lee, D. T. S., Chong, Y. S., Tan, L. K., Ng, T. P., & Fones, C. S. L. (2005). Confinement and other psychosocial factors in perinatal depression: A transcultural study in Singapore. *Journal of Affective Disorders*, 89, 157-166.
- Coutinho, D. S., Baptista, M. N., & Morais, P. R. (2002). Post-partum depression: Prevalence and social support correlation. *Infanto - Revista de Neuropsiquiatria da Infância e Adolescência*, 10(2), 63-71.
- Crotty, F., & Sheehan, J. (2004). Prevalence and detection of postnatal depression in an Irish community sample. *Irish Journal of Psychological Medicine*, 21(4), 117-121.
- Cruz, E. B. S., Simões, G. L., & Faisal-Cury, A. (2005). Rastreamento da depressão pós-parto em mulheres atendidas pelo Programa de Saúde da Família. *Revista Brasileira de Ginecologia e Obstetrícia*, 27(4), 181-188.
- Póo, F. A. M., Espejo, S. C., Godoy, P. C., Gualda, C. M., Hernández, O. T., & Pérez, H. C. (2008). Prevalence and risk factors associated with postpartum depression in puerperal women consulting in primary care. *Revista Médica de Chile*, 136, 44-52.
- Dunn, S., Davies, B., McCleary, L., Edwards, N., & Gaboury, I. (2006). The relationship between vulnerability factors and breastfeeding outcome. *Journal of Obstetric Gynecologic and Neonatal Nursing (JOGNN)*, 35(1), 87-97.
- Edge, D., Baker, D., & Rogers, A. (2004). Perinatal depression among black Caribbean women. *Health and Social Care in the Community*, 12(5), 430-438.
- Edhborg, M. (2008). Comparisons of different instruments to measure blues and to predict depressive symptoms 2 months postpartum: A study of new mothers and fathers. *Scandinavian Journal of Caring Sciences*, 22, 186-195.

- Edhborg, M., Lundh, W., Seimyr, L., & Widström, A. -M. (2001). The long-term impact of postnatal depressed mood on mother± child interaction: A preliminary study. *Journal of Reproductive and Infant Psychology*, 19(1), 61-71.
- Eilat-Tsanani, S., Merom, A., Romano, S., Reshef, A., Lavi, I., & Tabenkin, H. (2006). The effect of postpartum depression on Womens' consultations with physicians. *Israel Medical Association Journal*, 8, 406-410.
- Eisenach, J. C., Pan, P. H., Smiley, R., Lavand'homme, Landau, R., & Houle, T. T. (2008). Severity of acute pain after childbirth, but not type of delivery, predicts persistent pain and postpartum depression. *Pain*, 140, 87-94.
- Ekuklu, G., Tokuc, B., Eskiocak, M., Berberoglu, U., & Saltik, A. (2004). Prevalence of postpartum depression in Edirne, Turkey, and related factors. *Journal of Reproductive Medicine for the Obstetrician and Gyneco*, 49(11), 908-914.
- Escobar, J., Pacora, P., Custodio, N., & Villar, W. (2009). Escobar, J., Pacora, P., Custodio, N., & Villar, W. (2009). Depresión posparto: ¿se encuentra asociada a la violencia basada en género? *Anales de la Facultad de Medicina (Perú)*, 70(2), 115-118.
- Escribà-Agüir, V., & Artazcoz, L. (2011). Gender differences in postpartum depression: A longitudinal cohort study. *Journal of Epidemiology and Community Health*, 65, 320-326.
- Evans M., G., Vicuña M., M., & Marín, R. (2003). Depresion postparto realidade en el sistema publico de atencion de salud. *Revista Chilena de Obstetricia y Ginecología*, 68(6), 491-494.
- Evins, G. G., Theofrastous, J. P., & Galvin, S. L. (2000). Postpartum depression: A comparison of screening and routine clinical evaluation. *American Journal of Obstetrics and Gynecology*, 182(5), 1080-1082.
- Faisal-Cury, A., Tedesco, J. J. A., Kahhale, S., Menezes, P. R., & Zugaib, M. (2004). Postpartum depression: In relation to life events and patterns of coping. *Archives of Women's Mental Health*, 7, 123-131.
- Figueira, P., Corrêa, H., Malloy-Diniz, L., & Romano-Silva, M. (2009). Edinburgh Postnatal Depression Scale for screening in the public health system. *Revista de Saúde Pública*, 43, 79-84.
- Fonseca, M. O., Tavares, D. M. S., & Rodrigues, L. R. (2009). Investigação dos fatores indicativos de depressão pós-parto em dois grupos de puérperas. *Ciência, Cuidado e Saúde*, 8(3), 321-328.
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Research*, 5, 3-8.

- Grazioli, R., & Terry, D. J. (2000). The role of cognitive vulnerability and stress in the prediction of postpartum depressive symptomatology. *British Journal of Clinical Psychology*, 39, 329-347.
- Hasbún H., J., Risco N., L., Jadresic Marinovic, E., Galleguillo U., T, González A., M, & Garay S., J. (1999). Depresión postparto: prevalencia y factores de riesgo. *Revista Chilena de Obstetricia y Ginecología*, 64(6), 466-470.
- Hullfish, K. L., Fenner, D. E., Sorser, S. A., Visger, J., Clayton, A., & Steers, W. D. (2007). Postpartum depression, urge urinary incontinence, and overactive bladder syndrome: Is there an association? *International Urogynecology Journal*, 18, 1121-1126.
- Jadresic M. , E., & Araya B. , R. (1995). Prevalencia de depresión postparto y factores asociados en Santiago, Chile. *Revista Médica de Chile*, 123, 694-699.
- Jadresic, E., Araya, R., & Jara, C. (1995). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Chilean postpartum women. *Journal of Psychosomatic Obstetrics & Gynecology*, 16, 187-191.
- Kim, Y. -K., Hur, J. -W., Kim, K. -H., Oh, K. -S., & Shin, Y. -C. (2008). Prediction of postpartum depression by sociodemographic, obstetric and psychological factors: A prospective study. *Psychiatry and Clinical Neurosciences*, 62, 331-340.
- Kron, T., & Brosh, A. (2003). Can dreams during pregnancy predict postpartum depression? *Dreaming*, 13(2), 67-81.
- Latorre-Latorre, J. F., Contreras-Pezzotti, L. M., García-Rueda, S., & Arteaga-Medina. (2006). Postpartum depression in adolescent women from Bucaramanga, Colombia. *Revista Colombiana de Obstetricia y Ginecología*, 57(3), 156-162.
- Lee, D. T. S., Yip, A. S. K., Chiu, H. F. K., Leung, T. Y. S., & Chung, T. K. H. (2001). A psychiatric epidemiological study of postpartum Chinese women. *American Journal of Psychiatry*, 158, 220-226.
- Lobato, G., Moraes, C. L., Dias, A. S., & Reichenheim, M. E. (2011). Postpartum depression according to time frames and sub-groups: a survey in primary health care settings in Rio de Janeiro, Brazil. *Archives of Women's Mental Health*, 14, 187-193.
- Lo Bianco, A. C. (2009). Da globalização inevitável à internacionalização desejável. *Revista Latinoamericana de Psicopatología Fundamental*, 12(3), 445-453. Recuperado de <http://www.scielo.br/pdf/rlpf/v12n3/v12n3a02.pdf>
- Ludermir, A. B., Lewis, G., Valongueiro, S. A., Araújo, T. V. B., & Araya, R. (2010). Violence against women by their intimate partner during pregnancy and postnatal depression: A prospective cohort study. *Lancet*, 376, 903-910.
- Matijasevich, A., Golding, J., Smith, G. D., Santos, I. S., Barros, A. J. D., & Victora, C. G. (2009). Differentials and income-related inequalities in maternal depression

during the first two years after childbirth: Birth cohort studies from Brazil and the UK. *Clinical Practice and Epidemiology in Mental Health*, 5, 12.

- Matthey, S., Barnett, B. E. W., & Elliott, A. (1997). Vietnamese and Arabic women's responses to the Diagnostic interview Schedule (depression) and self-report questionnaires: Cause for concern. *Australian and New Zealand Journal of Psychiatry*, 31, 360-369.
- Mohammad, K. I., Gamble, J., & Creedy, D. K. (2011). Prevalence and factors associated with the development of antenatal and postnatal depression among Jordanian women. *Midwifery*, 27, 238-245.
- Monti, F., Agostini, F., Marano, G., & Lupi, F. (2008). The course of maternal depressive symptomatology during the first 18 months postpartum in an Italian sample. *Archives of Women's Mental Health*, 11, 231-238.
- Moraes, I. G. S., Pinheiro, R. T., Silva, R. A., Horta, B. L., Sousa, P. L., & Faria, A. D. (2006). Prevalence of postpartum depression and associated factors. *Revista de Saúde Pública*, 40(1), 65-70.
- Pinheiro, K. A. T., Pinheiro, R. T., Silva, R. A., Coelho, F. M. C., Quevedo, L. A., Godoy, R. V., Oses, J. P. (2011). Chronicity and severity of maternal postpartum depression and infant sleep disorders: A population-based cohort study in southern Brazil. *Infant Behavior and Development*, 34, 371-373.
- Pop, V. J. M., Essed, G. G. M., Degeus, C. A., Vanson, M. M., & Komproe, I. H. (1993). Prevalence of postpartum depression - Or is it post-puerperium depression? *Acta Obstetricia et Gynecologica Scandinavica*, 72, 354-358.
- Reading, R., & Reynolds, S. (2001). Debt, social disadvantage and maternal depression. *Social Science & Medicine*, 53, 441-453.
- Ruschi, G. E. C., Chambô Filho, A., Lima, V. J., Yazaki-Sun, S., Zandonade, E., & Mattar, R. (2009). Thyroid dysfunction: A risk factor associated with postpartum depression? *Revista Brasileira de Saúde Materno Infantil*, 9(2), 207-213.
- Santos, M. F. S., Martins, F. C., & Pasquali, L. (1999). Post-natal Depression Self-Rating Scales: Brazilian study. *Revista de Psiquiatria Clínica*, 26, 90-95.
- Saraiva, E. R. A., & Coutinho, M. P. L. (2007). A estrutura das representações sociais de mães puérperas acerca da depressão pós-parto. *Psico-USF*, 12(2), 319-326.
- Sebastián Romero, E., Mas Lodo, N., Martin Blazquez, M., Raja Casillas, M. I., Izquierdo Zamarriego, M. J., Valles Fernandez, N., & Metola Gomez, M. (1999). Postpartum depression in the health area of Toledo. *Atencion primaria - Sociedad Española de Medicina de Familia y Comunitaria*, 24(4), 215-219.

- Seimyr, L., Edhborg, M., Lundh, W., & Sjögren, B. (2004). In the shadow of maternal depressed mood: Experiences of parenthood during the first year after childbirth. *Journal of Psychosomatic Obstetrics & Gynecology*, 25, 23-34.
- Skari, H., Skreden, M., Malt, U. F., Dalholt, M., Ostensen, A. B., Egeland, T., & Emblem, R. (2002). Comparative levels of psychological distress, stress symptoms, depression and anxiety after childbirth - A prospective population-based study of mothers and fathers. *British Journal of Obstetrics and Gynaecology (BJOG)*, 109, 1154-1163.
- Söderquist, J., Wijma, B., Thorbert, G., & Wijma, K. (2009). Risk factors in pregnancy for post-traumatic stress and depression after childbirth. *British Journal of Obstetrics and Gynaecology (BJOG)*, 116, 672-680.
- Stiles, W. B. (2007). What qualifies as research on which to judge effective practice. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 56-130). Washington, DC: American Psychological Association. doi:10.1037/11265-002
- Sword, W., Landy, C. K., Thabane, L., Watt, S., Krueger, P., Farine, D., & Foster, G. (2011). Is mode of delivery associated with postpartum depression at 6 weeks: A prospective cohort study. *British Journal of Obstetrics and Gynaecology (BJOG)*, 118, 966-977.
- Tannous, L., Gigante, L. P., Fuchs, S. C., & Busnello, E. D. A. (2008). Postnatal depression in Southern Brazil: Prevalence and its demographic and socioeconomic determinants. *BMC Psychiatry*, 8, 1.
- Wang, S. -Y., & Chen, C. -H. (2006). Psychosocial health of Taiwanese postnatal husbands and wives. *Journal of Psychosomatic Research*, 60, 303-307.
- Wissart, J., Parshad, O., & Kulkarni, S. (2005). Prevalence of pre- and postpartum depression in Jamaican women. *BMC Pregnancy and Childbirth*, 5, 15.
- Yamashita, H., Yoshida, K., Nakano, H., & Tashiro, N. (2000). Postnatal depression in Japanese women: Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. *Journal of Affective Disorders*, 58, 145-154.
- Zaconeta, A. M., da Motta, L. D. C., & França, P. S. (2004). Depresión postparto: Prevalencia de test de rastreo positivo en púerperas del hospital universitario de Brasilia, Brasil. *Revista Chilena de Obstetricia y Ginecología*, 69(3), 209-213.
- Zubaran, C., Foresti, K., Schumacher, M. V., Amoretti, A. L., Müller, L. C., Thorell, M. R., Madi, J. M. (2009). Validation of a screening instrument for postpartum depression in Southern Brazil. *Journal of Psychosomatic Obstetrics & Gynecology*, 30(4), 244-254.
- Urdaneta M., J., Rivera S., A., García I., J., Guerra V., M., Baabel Z., N., & Contreras B., A. (2010). Prevalencia de depresión posparto en primigestas y multíparas

valoradas por la escala de edimburgo. *Revista Chilena de Obstetricia y Ginecología*, 75(5), 312-320.