UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE PSICOLOGIA

GABRIELA ANDRADE DA SILVA

Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise

São Paulo

2013

GABRIELA ANDRADE DA SILVA

Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise

(Versão corrigida)

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

Área de concentração:

Psicologia Experimental

Orientadora: Profa. Titular Emma Otta

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 divulgada a fonte.

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

Silva, Gabriela Andrade da.
Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise / Gabriela Andrade da Silva; orientadora Emma Otta. -- São Paulo, 2013.
240 f.

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

1. Depressão pós-parto 2. Prevalência 3. Metanalise 4. Psicologia baseada em evidências 5. Países em desenvolvimento I. Título.

RC537

Nome: Silva, Gabriela Andrade da

Título: Prevalência de depressão pós-parto em países desenvolvidos e em desenvolvimento: contribuições metodológicas de uma metanálise

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

Aprovado em:

Banca examinadora:

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Agradecimentos

Esta Tese contém registros de lições e experiências diárias vivenciadas junto a colegas, amigos e familiares. Às tantas pessoas que compartilharam dessas experiências, expresso aqui minha gratidão.

Agradeço à Professora Titular Emma Otta, com quem tenho o prazer de pesquisar e aprender há dez anos, e que muito contribuiu para minha formação acadêmica.

À Maria Imaculada Cardoso Sampaio, segunda revisora, grande companheira de assuntos acadêmicos, cafés e conversas desde a elaboração da metodologia até a finalização da Tese, e que muito engrandeceu este trabalho.

À Ágatha S. Rodrigues, pela dedicação e criatividade na análise estatística, cerne do presente trabalho.

Aos Professores Doutores Tamara Melnik e José de Oliveira Siqueira, pelas valiosas sugestões na Qualificação do presente trabalho.

A meu esposo, Daniel Barbosa do Nascimento, pela elaboração dos *softwares* que facilitaram o trabalho e, principalmente, pela presença e apoio em todos os momentos, com companheirismo, atenção, compreensão e paciência.

A meu irmão, Samuel Andrade da Silva, pela elaboração dos mapas que ilustraram as análises descritivas e pelo fraterno apoio e incentivo, desde que nasceu.

A minha mãe, Rita de Cassia Padula, e a meu pai, José Rocha Andrade da Silva, pela educação que recebi, pelo apoio em todos os momentos e por serem meus principais exemplos de pessoas engajadas em transformar nossa realidade através de educação, pesquisa e muito trabalho.

A meu primo, Artur Andrade da Silva Machado, pela revisão dos resumos em língua estrangeira.

A Aluysio Marcos Robalinho de Azevedo, pela revisão do texto final e pela amizade.

A Luana Fernandes Galindo de Oliveira, pelo auxílio na organização do material para coleta de dados.

Aos amigos e colegas do IPUSP, pelas conversas, cafés, risadas nos corredores e trabalhos realizados em conjunto.

Aos amigos e colegas da Pró-Reitoria de Assuntos Comunitários e Políticas Afirmativas da Universidade Federal do ABC, pelo incentivo, compreensão e apoio que me permitiram concluir o Doutorado simultaneamente com as atividades de trabalho.

Aos funcionários da Biblioteca Dante Moreira Leite, que me auxiliaram na busca de estudos revisados para o presente trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa de estudos recebida nos dois primeiros anos do Doutorado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por disponibilizar recursos através do Projeto Temático “Depressão pós-parto como um fator de risco para o desenvolvimento do bebê: Estudo interdisciplinar dos fatores envolvidos na gênese do quadro e em suas consequências”.

Ao Departamento de Psicologia Experimental da USP, que destinou recursos para obtenção de artigos por comutação bibliográfica, aumentando o número de trabalhos disponíveis para a revisão sistemática.

RESUMO

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, Universidade de São Paulo, São Paulo.

Introdução: A depressão pós-parto (DPP) é um transtorno de humor que pode ocorrer em mulheres no primeiro ano após o parto. Estudos epidemiológicos anteriores apresentaram variações em relação à metodologia adotada e resultados conflitantes quanto à prevalência de DPP. Não há consenso sobre a possibilidade de que a prevalência de DPP seja maior nos países em desenvolvimento do que nos desenvolvidos. **Objetivos:** Usando revisão sistemática da literatura seguida de metanálise, objetivou-se identificar variáveis regionais e metodológicas dos estudos primários que contribuíram para as diferenças nas prevalências de DPP; e verificar se haveria diferença entre a prevalência de DPP em países desenvolvidos e em desenvolvimento, controlando-se variáveis metodológicas. **Método:** Trata-se de pesquisa documental, para a qual foi realizada busca em 14 bases de dados eletrônicas, usando estratégias com os termos "depressão pós-parto" e "prevalência", sem restrição de tempo. Critérios de inclusão foram aplicados por duas revisoras independentes aos registros obtidos, em três etapas: títulos, resumos e texto completo. Dados sobre a prevalência de DPP e a metodologia adotada foram coletados dos estudos incluídos. A Escala de Loney foi usada para avaliar a qualidade metodológica. Técnicas estatísticas foram aplicadas para analisar a heterogeneidade entre os trabalhos incluídos e sintetizar os resultados, gerando médias ponderadas das prevalências para subgrupos de análise. Um modelo de metarregressão foi construído, tendo as variáveis metodológicas dos estudos como preditores e a prevalência de DPP como variável dependente. **Resultados:** Foram obtidos 1.881 registros de estudos originais, publicados entre 1980 e 2012, dos quais 778 tiveram o texto completo avaliado e 337 foram incluídos. A qualidade metodológica dos trabalhos foi considerada fraca. A prevalência de DPP variou entre 0,5% em Singapura e 62,8% nos Estados Unidos da América, observando-se alta heterogeneidade. O modelo de metarregressão indicou que as variáveis instrumento para identificar DPP, índice de desenvolvimento humano (IDH) do país, delineamento de pesquisa e tamanho amostral foram preditores significativos da prevalência de DPP. A cidade onde o estudo foi realizado, inserida como cluster para controlar a dependência entre estudos, foi significativa. Maior prevalência média foi encontrada em trabalhos que usaram instrumentos de autoavaliação para identificar DPP, comparados aos que usaram entrevista clínica; em pesquisas transversais, comparados aos longitudinais; em países com menores IDHs; e em estudos com menor tamanho amostral. **Discussão:** Foi construído um modelo explicativo da prevalência de DPP a partir de características metodológicas e regionais dos estudos primários, indicando que os métodos empregados nesses estudos geraram vieses nos resultados. Mesmo controlando as variáveis metodológicas, a prevalência de DPP foi mais elevada em países em desenvolvimento que nos desenvolvidos, portanto, aspectos sociais, econômicos e culturais podem ter influência sobre a ocorrência desse transtorno. **Conclusões:** A DPP foi mais prevalente em países em desenvolvimento do que nos países

desenvolvidos, mesmo levando em conta características metodológicas dos estudos incluídos. No entanto, essa conclusão não pode ser considerada definitiva, devido à qualidade fraca dos estudos primários.

Palavras-chave: Depressão pós-parto. Prevalência. Metanálise. Psicologia baseada em evidências. Epidemiologia. Países em desenvolvimento.

ABSTRACT

Silva, G. A. (2013). *Prevalence of postpartum depression in developed and developing countries: methodological contributions of a meta-analysis.* PHD Thesis, Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Introduction: Postpartum depression (PPD) is a mood disorder that might occur in women in the first year after delivery. Previous epidemiological studies varied with regards to methodology and showed conflicting results regarding the prevalence of PPD. There is no consensus on whether the prevalence of PPD is greater in developing countries than in developed ones, or not. **Objective:** Using systematic literature review followed by meta-analysis, we aimed to identify regional and methodological variables of primary studies that contributed to the differences in the prevalence of PPD; and check if there was a difference between the prevalence of PPD in developed countries and developing ones, keeping control of methodological variables. **Methods:** This thesis made a documental research, which searched 14 electronic databases for specialized studies, using strategies with the terms "postpartum depression" and "prevalence", without time restrictions. Inclusion criteria were applied for two reviewers to obtained records in three stages: titles, abstracts and full text. Data on the prevalence of PPD as well as methodology from the included studies were collected. The Loney Scale was used to assess the methodological quality. Statistical techniques were applied to analyze the heterogeneity among the included studies and synthesize the results, generating weighted average prevalence subgroups for analysis. A meta-regression model was constructed, using methodological variables of the studies as predictors and the prevalence of PPD as a dependent variable. **Results:** Records from 1881 original studies published between 1980 and 2012 were obtained, of which 778 were evaluated for full text and 337 were included. The methodological quality of the studies was considered weak. The prevalence of PPD ranged from 0.5% in Singapore to 62.8% in United States of America, observing high heterogeneity. The meta-regression model indicated that the variables tool to identify DPP, human development index (HDI) of the country, research design and sample size were significant predictors of the prevalence of PPD. The city where the study was conducted, inserted as a cluster to control the dependency between studies, was significant. Higher average prevalence was found in studies that used self-assessment tools to identify PPD, compared to those which used clinical interview; in cross-sectional surveys, compared to longitudinal; in countries with lower HDIs, and in studies with smaller sample sizes. **Discussion:** We constructed an explanatory model for the prevalence of PPD from regional and methodological characteristics of primary studies, indicating that method led to biases in the results of those studies. Even controlling for methodological variables, the prevalence of PPD was higher in developing countries than in developed ones. Therefore, social, economic and cultural rights may influence the occurrence of this disorder. **Conclusions:** PPD was more prevalent in developing countries than in developed ones, even taking into account methodological issues of the included studies. However, this conclusion cannot be considered definitive, due to the poor quality of the primary studies.

Keywords: Depression, postpartum. Prevalence. Meta-analysis. Evidence-based psychology. Epidemiology. Developing countries.

RESUMEN

Silva, G. A. (2013). *Prevalencia de la depresión post-parto en los países desarrollados y en desarrollo: contribuciones metodológicas de un meta-análisis*. Tese de Doctorado, Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Introducción: La depresión posparto (DPP) es un trastorno del estado de ánimo que puede se producir en mujeres en el primer año después del parto. Estudios epidemiológicos anteriores han mostrado variaciones en la metodología y resultados contradictorios con respecto a la prevalencia de DPP. No hay consenso sobre la posibilidad de que la prevalencia de DPP sea mayor en los países en desarrollo que en los desarrollados. **Objetivo:** Mediante revisión sistemática seguida por meta-análisis, se objetivó identificar las variables regionales y metodológicas de los estudios primarios que contribuyeron a las diferencias en la prevalencia de DPP; y comprobar si había una diferencia entre la prevalencia de DPP en los países desarrollados y en desarrollo, manteniendo el control de variables metodológicas. **Método:** Esta Tesis consiste en una investigación documental, en que se ha llevado a cabo una búsqueda en 14 bases de datos electrónicas, utilizando estrategias con los términos "depresión posparto" y "prevalencia", sin restricciones de tempo. Los criterios de inclusión han sido aplicados por dos revisoras a los registros obtenidos, en tres pasos: títulos, resúmenes y texto completo. Los datos sobre la prevalencia de DPP y la metodología de los estudios incluidos han sido recogidos. La Escala de Loney se ha utilizado para evaluar la calidad metodológica. Las técnicas estadísticas se han aplicado para analizar la heterogeneidad entre los estudios incluidos y sintetizar los resultados, lo que ha generado promedios ponderados de prevalencia para los subgrupos de análisis. Se construyó un modelo de meta-regresión, usando las variables metodológicas de los estudios como predictores y la prevalencia de DPP como una variable dependiente. **Resultados:** Se han obtenido registros de 1881 estudios originales, publicados entre 1980 y 2012, de los cuales 778 han tenido el texto completo evaluado y 337 han sido incluidos. La calidad metodológica de los estudios ha sido considerada débil. La prevalencia de DPP ha variado de 0,5% en Singapur a 62,8% en Estados Unidos de la América, observándose una alta heterogeneidad. El modelo de meta-regresión ha indicado que las variables instrumento para identificar DPP, índice de desarrollo humano (IDH) en el país, diseño de la investigación y el tamaño de la muestra han sido predictores significativos de la prevalencia de DPP. La ciudad donde se ha realizado el estudio, como grupo introducido para controlar la dependencia entre los estudios, ha sido significativa. Mayor promedio de prevalencia se ha encontrado en los estudios que han utilizado instrumentos de autoevaluación para identificar DPP, en comparación con la entrevista clínica; en estudios transversales, en comparación con los longitudinales; en los países con menor IDH, y en estudios con muestras más pequeñas. **Discusión:** Se ha construido un modelo para explicar la prevalencia de DPP a partir de las características regionales y metodológicas de los estudios primarios, lo que permite indicar que los métodos utilizados en estos estudios han generado sesgos en los resultados. Incluso después de controladas las variables metodológicas, la prevalencia de DPP ha sido mayor en los países en desarrollo que en los desarrollados, por lo tanto, los aspectos sociales, económicos y culturales pueden influir en la aparición de ese trastorno. **Conclusiones:** La DPP fue más frecuente en los países en desarrollo que en los países desarrollados, incluso

teniendo en cuenta las características metodológicas de los estudios incluidos. Sin embargo, esta conclusión no puede considerarse definitiva, debido a la mala calidad de los estudios primarios.

Palabras clave: Depresión posparto. Prevalencia. Metanálisis. Psicología basada en la evidencia. Epidemiología. Países en desarrollo.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema representando a Psicologia Baseada em Evidências como a integração das evidências científicas, experiência clínica e contexto e preferências do paciente. 4

Figura 2 – Níveis de evidência de estudos sobre eficácia de intervenções (adaptado de Melnik & Atallah, 2011).....14

Figura 3 – Fluxograma de seleção de estudos para a presente revisão sistemática. 75

Figura 4 – Quadro apresentando os itens das diretrizes para avaliação crítica de estudos de prevalência ou incidência de um problema de saúde (Loney et al., 1998) 82

Figura 5 – Fluxograma de seleção de estudos para a revisão sistemática. 101

Figura 6 – Histograma representando as frequências de ano de publicação dos estudos incluídos.....105

Figura 7 – Histograma representando a distribuição das pontuações dos estudos na Escala de Loney et al. (1998).....116

Figura 8 – Mapa indicando as médias das estimativas pontuais de prevalência de DPP por país. 122

Figura 9 – Mapa indicando as médias das estimativas pontuais de prevalência de DPP por cidade. 123

Figura 10 – Gráfico de floresta indicando a prevalência média pontual e os intervalos de confiança por país. 124

Figura 11 – Gráficos de funil para avaliar heterogeneidade dos estudos incluídos. Gráfico A: prevalências dos estudos incluídos em relação ao erro padrão; Gráfico B: resíduos em relação ao erro padrão, usando país como moderador; Gráfico C: resíduos em relação ao erro padrão, usando cidade como moderador.....126

Figura 12 – Prevalência média de DPP por tipo de instrumento usado para identificar o transtorno. 127

Figura 13 – Prevalência média de DPP por grau de desenvolvimento do país onde o estudo primário foi realizado. 128

Figura 14 – Prevalência média de DPP por delineamento de pesquisa.....131

Figura 15 – Gráfico mostrando o valor predito da prevalência de DPP em relação ao valor observado. 138

LISTA DE TABELAS

Tabela 1 – Correspondência entre campos ampliados de pesquisa em saúde e delineamentos metodológicos mais adequados (adaptada de Greenhalgh, 2008). 10

Tabela 2 – Características da revisão narrativa e da revisão sistemática de literatura (adaptado de Cook, Mulrow, & Haynes, 1997).....12

Tabela 3 – Níveis de evidência para responder à pergunta "Quão comum é o problema?", baseado na proposta do OCEBM Levels of Evidence Working Group (2011).....23

Tabela 4 – Síntese de metanálises sobre prevalência de DPP (continua).....48

Tabela 5 – Estratégias de busca para as bases de dados consultadas na presente revisão.....73

Tabela 6 – Exemplos de registros eliminados pelo título. 78

Tabela 7 – Datas das buscas eletrônicas e número de registros recuperados. 99

Tabela 8 – Concordância entre avaliadoras quanto à inclusão de estudos pelos títulos, resumos e textos completos. 103

Tabela 9 – Número de estudos incluídos por país (continua) 107

Tabela 10 – Características metodológicas dos estudos incluídos. 109

Tabela 11 – Perfil das participantes dos estudos incluídos: renda e região urbana ou rural. 110

Tabela 12 – Perfil das participantes dos estudos incluídos.....111

Tabela 13 – Motivos que levaram à exclusão de estudos.....113

Tabela 14 – Concordância entre avaliadoras para os itens da Escala de Loney et al. (1998).....115

Tabela 15 – Número de estudos que obtiveram pontuação em cada item da Escala de Loney et al. (1998). 117

Tabela 16 – Prevalência média de DPP de acordo com país e tipo de instrumento usado para identificar DPP.....129

Tabela 17 – Prevalência média de DPP de acordo com país e técnica de amostragem. 132

Tabela 18 – Prevalência média de DPP de acordo com o contexto de pesquisa. ..134

Tabela 19 – Prevalência média de DPP de acordo com o ponto de corte da EDPE. 135

Tabela 20 – Ajuste final do Modelo Linear Misto Generalizado tendo DPP como variável resposta e características metodológicas dos estudos primários como preditores.....137

Tabela 21 – Resultados das análises de sensibilidade. 140

LISTA DE ABREVIATURAS E SIGLAS

AHRQ – *Agency for Healthcare Research and Quality*

AIC – Akaike (medida da qualidade de ajuste relativa de um modelo estatístico

APA – *American Psychological Association*

BDI – *Beck Depression Inventory*

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 (base de dados)

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CES-D – Center for Epidemiologic Studies Depression Scale

CID-10 – Classificação Internacional de Doenças

CINAHL – Cumulative Index to Nursing & Allied Health Literature (base de dados)

CIS – Clinical Interview Schedule

Comut – Programa de Comutação Bibliográfica

DACL - *Lubin Depression Adjective Check-List*

DP – Desvio padrão

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

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

EDPE – Escala de Depressão Pós-parto de Edimburgo (em inglês, EPDS – Edinburgh Postpartum Depression Scale)

EMBASE – *Excerpta Medica Database* (base de dados)

EPDS – Edinburgh Postpartum Depression Scale (em português, EDPE – Escala de Depressão Pós-parto de Edimburgo)

ERIC –*Education Resources Information Center* (base de dados)

ESTs – Empirically Supported Treatments

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

GHQ – *General Health Questionnaire*

GLM – Modelo Linear Geral (em inglês, *General Linear Model*)

IC – Intervalo de Confiança

IDH – Índice de Desenvolvimento Humano

IPUSP – Instituto de Psicologia da Universidade de São Paulo

ISI Web of Knowledge – Base de dados de trabalhos publicados em periódicos internacionais

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) (base de dados)

MLMG – Modelo Linear Misto Generalizado

NCBI – *National Center for Biotechnology Information*

NLM – *National Library of Medicine's*

OCEBM – Oxford Centre for Evidence-Based Medicine

PDSS – *Postpartum Depression Screening Scale*

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

PBE – Prática da Psicologia Baseada em Evidências

PPD – *Postpartum depression*

PRAMS – *Pregnancy Risk Assessment Monitoring System*

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

Psicodoc – Base de dados do Colégio de Psicólogos de Madrid

PsycINFO – Base de dados na área da Psicologia da *American Psychological Association* (Associação Americana de Psicologia).

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

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

RDC – Research Diagnostic Criteria

Redalyc – Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (base de dados)

SCAD – Serviço de Acesso ao Documento

SCID – *Structured Clinical Interview for DSM Disorders* (entrevista clínica estruturada com base nos critérios do DSM-IV)

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

SCOPUS – Base de dados multidisciplinar, produzida pela editora Elsevier

Sociofile – Sociological Abstracts (base de dados)

TDM - Transtorno Depressivo Maior

UNDP – United Nations Development Programme

SUMÁRIO

1 PREÂMBULO 1

2 INTRODUÇÃO 3

2.1 Psicologia Baseada em Evidências: decisões que combinam conhecimento cientifico e experiência profissional.....4

2.2 Níveis de evidência em pesquisa e Revisão Sistemática como "padrão-ouro" do conhecimento científico 8

2.2.1 Classificações dos estudos segundo delineamento e natureza.....8

2.2.2 Revisão sistemática de literatura 11

2.2.3 Metanálise: Aplicação de técnicas estatísticas à Revisão Sistemática.....13

2.2.4 Pirâmides de evidências 14

2.3 Procedimentos para a tomada da decisão conforme a Prática da Psicologia Baseada em Evidências (PBE) 16

2.3.1 Formulação da pergunta de pesquisa.....17

2.3.2 Busca da melhor evidência científica disponível.....18

2.3.3 Avaliação da qualidade da evidência disponível.....20

2.3.4 Integração dos resultados de diferentes pesquisas.....20

2.3.5 Tomada de decisão 21

2.4 Procedimentos para revisão sistemática com metanálise de estudos de prevalência.....21

2.4.1 Formulação da pergunta de pesquisa e avaliação da necessidade da revisão sistemática.....22

2.4.2 Níveis de evidência em pesquisas de prevalência 23

2.4.3 Definição dos critérios de inclusão e estratégia de busca de estudos.....24

2.4.4 Seleção de estudos e coleta de dados 24

2.4.5 Análise dos dados 26

2.4.6 Apresentação dos resultados 35

2.4.7 Atualização da revisão sistemática.....35

2.4.8 Aplicação da PBE ao presente trabalho 35

2.5 Depressão Pós-parto 36
2.5.1 Depressão pós-parto: definição e características 37
2.5.2 Métodos para identificar DPP 39
2.5.3 Epidemiologia 42
2.5.4 Fatores de risco 54
2.5.5 Acurácia de instrumentos de rastreamento 56
2.5.6 Consequências da DPP para o desenvolvimento infantil.....58
2.5.7 Eficácia e efetividade de intervenções para o tratamento da DPP 59
2.5.8 Eficácia e efetividade de intervenções para prevenção de DPP.....61

3 OBJETIVOS E HIPÓTESES 62

3.1 Objetivos 62

3.2 Hipóteses 62

4 JUSTIFICATIVA.....65

5 MÉTODO.....69

5.1 Critérios de inclusão 69
5.1.1 Tipo de estudo 70
5.1.2 Tipo de participantes.....71
5.1.3 Desfecho.....72

5.2 Estratégia de busca.....72

5.3 Seleção de estudos 74

5.4 Extração e gerenciamento de dados 76
5.4.1 Protocolo para avaliação do texto completo 78
5.4.2 Identificação de estudos com amostras duplicadas.....80
5.4.3 Método para estabelecer a concordância entre revisoras na avaliação do texto completo 81

5.5 Avaliação da qualidade metodológica dos estudos incluídos.....81

5.5.1 Outras variáveis controladas para avaliar a qualidade metodológica dos estudos incluídos 88

5.6 Análise estatística 88

5.6.1 Definição dos dados a serem usados na análise estatística 88

5.6.2 Determinação da concordância entre revisoras 90

5.6.3 Avaliação da heterogeneidade 90

5.6.4 Avaliação do viés de publicação 91

5.6.5 Síntese dos dados 91

6 RESULTADOS 98

6.1 Resultados da busca bibliográfica 98

6.2 Seleção de estudos 99

6.3 Análises descritivas 104

6.3.1 Características dos estudos incluídos 105

6.3.2 Características dos estudos excluídos 112

6.3.3 Qualidade metodológica dos estudos incluídos 114

6.3.4 Prevalência de DPP 120

6.4 Análises de heterogeneidade 125

6.5 Análises de subgrupos 127

6.6 Impacto da metodologia sobre os resultados de prevalência: metarregressão 135

6.7 Análises de sensibilidade 139

7 DISCUSSÃO 141

7.1 Forças e limitações da estratégia de busca 141

7.2 Viés de publicação contra o Sul 144

7.3 Concordância entre revisoras 146

7.4 Análise crítica da qualidade metodológica dos estudos incluídos 148

7.5 Prevalência de DPP no mundo.....150

7.6 Comparação da prevalência de DPP em países desenvolvidos e em desenvolvimento 152

7.7 Influência de fatores metodológicos sobre resultados de estudos de prevalência de DPP 155

7.7.1 Discussão das hipóteses de pesquisa 155

7.7.2 Discussão dos resultados da metarregressão 160

7.8 Implicações para a prática profissional 162

7.9 Implicações para pesquisas.....166

7.10 Recomendações para revisões sistemáticas de estudos de prevalência 172

8 Conclusões 175

REFERÊNCIAS.....176

APÊNDICE 1 – PROTOCOLO DE AVALIAÇÃO DE ESTUDOS POR TÍTULO E RESUMO.....189

APÊNDICE 2 – REFERÊNCIAS DOS ESTUDOS INCLUÍDOS.....192

APÊNDICE 3 – PREVALÊNCIA MÉDIA DE DPP POR CIDADE 219

1 Preâmbulo

O presente trabalho foi realizado no Programa de Pós-Graduação em Psicologia Experimental do Instituto de Psicologia da Universidade de São Paulo (IPUSP), no âmbito do Projeto Temático multicêntrico "Depressão pós-parto como um fator de risco para o desenvolvimento do bebê: estudo interdisciplinar dos fatores envolvidos na gênese do quadro e em suas consequências", desenvolvido com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Esse projeto acompanhou díades mãe-bebê, desde a gestação até três anos após o parto, buscando avaliar fatores de risco para a depressão pós-parto (DPP) e as consequências desse transtorno para a relação entre a mãe e a criança, levando em consideração os fatores hormonais, as condições de suporte social, o estilo de apego da mãe e da criança, entre outros. Atualmente, o projeto se encontra em fase de análise dos dados.

No âmbito desse Projeto Temático, do qual a presente autora participa desde 2006, foi desenvolvida a Dissertação de Mestrado (Silva, 2008), que objetivou identificar a prevalência e os fatores de risco para DPP em uma amostra de mães atendidas em três Unidades Básicas de Saúde da zona Oeste de São Paulo. Ao examinar a literatura sobre a depressão pós-parto, observou-se ampla variação entre os valores de prevalência relatados nos diferentes estudos, que poderiam estar relacionados com o nível de desenvolvimento socioeconômico dos países. No entanto, a comparação dos dados de acordo com variáveis geográficas foi prejudicada pela confusão com outras variáveis metodológicas: os trabalhos usaram diferentes instrumentos para identificação de DPP; o diagnóstico de DPP foi feito em diferentes momentos, variando desde os primeiros dias até um ano ou mais após o parto; a composição das amostras variou em relação ao perfil das participantes, como por exemplo, renda, escolaridade e idade; e a maior parte das amostras foi selecionada por conveniência. Assim, não foi possível avaliar se os resultados realmente variaram de acordo com o nível socioeconômico dos países, ou se as variações nas características metodológicas dos estudos produziram essa impressão.

Sabe-se que os dados epidemiológicos devem nortear os investimentos públicos na área da saúde. O amplo conjunto de fatores que poderiam influenciar os

resultados de estudos de prevalência de DPP fez com que a comparabilidade entre eles não fosse intuitiva, de forma que o leitor poderia superestimar ou subestimar a quantidade de mães com esse transtorno. Assim, surgiu o seguinte questionamento: seria possível comparar os dados dos estudos sobre prevalência de DPP, levando em consideração as diferenças metodológicas?

Considerou-se que a melhor ferramenta para responder a essa pergunta seria a revisão sistemática de literatura, se possível com metanálise, metodologia comumente adotada por trabalhos que têm como pano de fundo a Psicologia Baseada em Evidências. Essa abordagem, relativamente nova, foi reconhecida oficialmente pela *American Psychological Association (APA)* em 2005 (APA, 2005), e tem potencial para integrar a pesquisa e a prática profissional em psicologia, ao promover a aplicação dos conhecimentos científicos em prevenção, tratamento e formulação de políticas públicas de saúde mental. As revisões sistemáticas têm se tornado mais frequentes nas pesquisas em psicologia, geralmente, com foco em avaliar efetividade e/ou eficácia de intervenções na área de saúde mental.

As fases de elaboração da estratégia de busca, classificação dos estudos de acordo com os critérios de inclusão e coleta de dados da presente revisão sistemática foram feitas em colaboração com a Doutoranda do IPUSP Maria Imaculada Cardoso Sampaio, Bacharel em Biblioteconomia. Sua pesquisa de Doutorado, sob orientação da Professora Doutora Sonia Meyer, pretende avaliar a qualidade dos estudos de psicologia publicados em revistas latino-americanas. Após a finalização da coleta de dados, os artigos aprovados nos critérios de inclusão constituíram universo de trabalhos explorados no presente estudo, enquanto o extrato de artigos publicados em periódicos científicos latino-americanos constituiu o universo de pesquisa de Maria Imaculada Cardoso Sampaio.

Espera-se que, além da contribuição para os estudiosos e os profissionais de saúde que lidam com a DPP, o presente estudo colabore para a reflexão sobre a Prática da Psicologia Baseada em Evidências e sua consolidação como método de trabalho do psicólogo.

2 Introdução

O volume de publicações científicas tem crescido constantemente nas últimas décadas e o acesso a esse conhecimento tem sido facilitado pelo aprimoramento e ampliação da abrangência e do acesso aberto às bases de dados informatizadas. De acordo com Breakwell, Fife-Schaw, Hammond e Smith (2010): "*O número de periódicos que estão sendo publicados cresce de ano para ano e, no entanto, parece haver mais artigos do que espaço disponível nos periódicos*" (p. 456). O crescente número de informações disponíveis significa que pesquisadores e técnicos têm maior facilidade de acesso às descobertas nas áreas de ciência e tecnologia, o que pode contribuir para sua prática profissional. No entanto, devido ao grande número de trabalhos publicados sobre cada tema de interesse e às diferenças metodológicas entre eles, extrapolar os resultados das pesquisas e determinar sua aplicabilidade não é uma tarefa trivial. Parte da dificuldade se dá porque a maior disponibilidade de estudos não está, necessariamente, relacionada com a qualidade dos mesmos.

Na maior parte das vezes, as informações usadas nos processos de tomada de decisão que ocorrem no cotidiano de trabalho são oriundas de livros didáticos, conselhos de colegas mais experientes e a própria experiência do profissional. Todas essas fontes são possivelmente relevantes, mas não estão, necessariamente, baseadas diretamente no conhecimento científico e podem rapidamente se tornar desatualizadas. Portanto, há uma lacuna entre a ciência e sua aplicação prática.

Observa-se, na área de saúde, um aumento da consciência de que decisões referentes às políticas e ações governamentais devem ser cada vez mais fundamentadas em evidências científicas. No Brasil, foi lançado em 29 de maio de 2012 o Portal Saúde Baseada em Evidências (http://periodicos.saude.gov.br/, acessado em 06/12/2012), iniciativa do Ministério da Saúde em parceria com a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes/MEC). O objetivo do portal é contribuir para o aprimoramento dos profissionais da área de saúde, fornecendo acesso rápido ao conhecimento científico por meio de publicações atuais e sistematicamente revisadas, esperando-se que possa apoiar a prática clínica e a tomada de decisão para a gestão em saúde. Para ter acesso ao

portal, o único requisito é que o usuário esteja cadastrado em um dos Conselhos Profissionais da área de saúde.

Especificamente na área de Psicologia, vem ganhando força a Psicologia Baseada em Evidências (PBE), uma ferramenta metodológica mediante a qual se pretende modificar o modo de trabalho do profissional de Psicologia, de forma que leve em consideração, em suas decisões cotidianas, as melhores evidências científicas sobre um determinado assunto (Sanchez-Meca & Botella, 2010).

2.1 Psicologia Baseada em Evidências: decisões que combinam conhecimento cientifico e experiência profissional

De acordo com a *APA Presidential Task Force on Evidence-Based Practice*, a Prática da Psicologia Baseada em Evidências é a integração da melhor evidência de pesquisa disponível com a experiência clínica no contexto das características do paciente, de sua cultura e suas preferências. Esse conceito pode ser visualizado, de forma esquemática, na Figura 1. A PBE é uma ferramenta que pode ser usada em diversos espaços de trabalho do psicólogo, como hospitais, clínicas, escolas, instalações militares, instituições de saúde pública, institutos de reabilitação e unidades básicas de saúde (American Psychological Association, 2005).

Figura 1 – Esquema representando a Psicologia Baseada em Evidências como a integração das evidências científicas, experiência clínica e contexto e preferências do paciente.

Embora o termo “Prática da Psicologia Baseada em Evidências” tenha sido adotado oficialmente pela *American Psychological Association* (APA) apenas em 2005, desde a década de 1970, observa-se um interesse em buscar evidências científicas da eficácia e efetividade[1] dos tratamentos psicoterápicos a partir da síntese estatística de resultados de estudos anteriores. Smith e Glass (1977) fizeram uma revisão sistemática de literatura buscando estudos controlados que compararam, ao menos, um grupo tratado com psicoterapia com um grupo sem tratamento ou que recebeu outro tipo de tratamento. Foram encontrados 375 trabalhos considerados adequados, e realizou-se uma síntese estatística dos resultados, observando-se que os indivíduos tratados com psicoterapia tiveram, em média, desfechos melhores que 75% dos indivíduos não tratados com psicoterapia. Esse resultado se tornou uma das primeiras evidências científicas da eficácia do tratamento psicoterápico.

O artigo de Smith e Glass (1977) foi publicado em um volume especial do periódico científico “*American Psychologist*”, que discutia a possibilidade de inclusão da psicoterapia no Seguro Nacional de Saúde dos EUA (Cummings, 1977). Para que a Psicologia pudesse fazer parte dos sistemas de saúde, era preciso mostrar empiricamente a sua eficácia e efetividade, bem como suas vantagens econômicas. Apesar de discutido há mais de 30 anos, o tema da avaliação científica da efetividade da prática psicológica para permitir sua inclusão nos convênios e investimentos públicos mantém sua importância, visto que os tratamentos psicológicos ainda estão pouco representados nos serviços públicos e privados de saúde.

O reconhecimento da importância das diretrizes de conduta (*guidelines*) pela APA pode ser considerado como um dos primeiros passos para o surgimento da PBE. Desde o princípio, foi reconhecido o risco de que essas diretrizes fossem usadas de forma inapropriada por organizações com fins comerciais, para impor formatos específicos de tratamento, limitando a possibilidade de escolha dos profissionais e dos clientes e restringindo o acesso destes últimos aos cuidados em

---

[1] A distância entre a pesquisa e a aplicação prática dos resultados levou Flay (1986) a definir eficácia e efetividade como termos distintos: ensaios sobre eficácia testam se um programa traz mais benefícios que prejuízos, quando executado em condições ótimas; ensaios sobre efetividade, por sua vez, testam se um programa traz mais benefícios que prejuízos quando executado sob condições reais.

saúde. Assim, formaram-se forças-tarefa para estudo do tema e foi elaborado o documento "*Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of pshysical disorders*" (American Psychological Association Task Force on Psychological Intervention Guidelines, 1995), que descreveu a variedade de evidências que deveriam ser consideradas no desenvolvimento das diretrizes de conduta, alertando que deveriam sempre ser baseadas na ponderação dos resultados de pesquisa e da experiência clínica (American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006).

Paralelamente, a Divisão 12 da APA (*Promotion and Dissemination of Psychological Procedures*) passou a elencar, com base em critérios previamente estabelecidos, os chamados "*Empirically Supported Treatments (ESTs)*", ou seja: tratamentos cuja eficácia pudesse ser sustentada por evidências científicas (American Psychological Association Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Essas evidências eram encontradas, sobretudo, nos resultados de ensaios clínicos controlados, envolvendo a aplicação de intervenções descritas em manuais de padronização para tratar uma condição específica em uma população com características conhecidas (APA Presidential Task Force on Evidence-Based Practice, 2006).

Desde a publicação dos primeiros resultados da Divisão 12 da APA, em 1995, ocorreram controvérsias. Parte da comunidade científica mostrou entusiasmo pelo reconhecimento da efetividade de tratamentos psicológicos a partir de dados empíricos. Outros pesquisadores criticaram o foco exclusivo em tratamentos breves e baseados em manuais, a ênfase nos efeitos específicos do tratamento em contraposição a efeitos de fatores comuns que são responsáveis por grande parte da variância dos desfechos, e o baixo potencial de aplicação dos resultados a grupos diversos de clientes, variando em comorbidade, personalidade, raça, etnia e cultura (Chambless & Ollendick, 2001). Assim, diversas forças-tarefa foram criadas em resposta à anterior, como a da Divisão 29 da APA (Psicoterapia), que passou a estudar o efeito da relação terapêutica nos desfechos dos tratamentos (APA Presidential Task Force on Evidence-Based Practice, 2006).

Ao mesmo tempo em que psicólogos discutiam a melhor forma de usar evidências científicas em decisões relacionadas ao cotidiano da profissão, a Prática Baseada em Evidências foi ganhando força em diversas áreas da saúde, tornando-

se uma ferramenta chave para as políticas públicas e os sistemas de saúde, em geral. Nesse contexto, foi criada, em 2005, uma nova força-tarefa da APA, chamada "*APA Presidential Task Force on Evidence-Based Practice*", que pretende estimular a busca de fundamentos empíricos para as decisões a respeito de práticas clínicas. Ao mesmo tempo, essa força-tarefa busca o cuidado para que as evidências científicas não sejam erroneamente tomadas como justificativa para a restrição ao acesso dos clientes aos cuidados em saúde, bem como à sua liberdade de escolha (APA Presidential Task Force on Evidence-Based Practice, 2006).

A PBE, portanto, não deixa em segundo plano a experiência do profissional ou a individualidade do cliente, mas tem como meta fazer com que os conhecimentos científicos atualizados sejam incorporados como mais um fator a ser considerado na atuação profissional. Para isso, é preciso que os psicólogos tenham bom conhecimento dos métodos de busca de estudos científicos e da forma de avaliar a sua qualidade e potencial para responder às perguntas de pesquisa propostas.

2.2 Níveis de evidência em pesquisa e Revisão Sistemática como "padrão-ouro" do conhecimento científico

A definição da PBE menciona que a experiência clínica e as características do cliente devem ser integradas com *a melhor evidência de pesquisa disponível* (APA Presidential Task Force on Evidence-Based Practice, 2006), o que implica na necessidade de se definir o que se considera por melhor evidência. Atualmente, considera-se que os trabalhos científicos podem ser classificados em diferentes graus ou níveis, de acordo com o quanto se acredita que a metodologia usada está adequada para produzir resultados confiáveis, isto é, com menores chances de apresentar viés. Assim, para eleger a melhor evidência científica disponível, é preciso conhecer metodologia de pesquisa, incluindo os conceitos de natureza, delineamento, campo ampliado de pesquisa e a relação entre eles.

2.2.1 Classificações dos estudos segundo delineamento e natureza

Os trabalhos científicos podem ser classificados, de acordo com sua natureza, em primários ou secundários (Glass, 1976; Cunha & Cavalcanti, 2008):

- Estudos primários: o pesquisador entra em contato diretamente com o objeto de estudo, como em experimentos (procedimentos em ambientes artificiais, controlados), ensaios clínicos (intervenção e acompanhamento dos participantes) e levantamentos (mensuração de algo em uma amostra de participantes). Portanto, são as fontes originais da informação.
- Estudos secundários: são documentos gerados a partir da análise dos documentos primários. Integram resultados de outros estudos, podendo gerar uma nova conclusão. Dentre os estudos secundários, encontram-se as diretrizes, análises econômicas e revisões sistemáticas de literatura.

Outra forma de classificar pesquisas científicas agrupá-las em estudos experimentais e observacionais. Nos estudos experimentais, há uma intervenção aplicada pelo pesquisador aos participantes da investigação, enquanto em estudos observacionais, o pesquisador apenas observa e registra as informações, sem aplicar intervenções, ao menos de forma intencional (Bastos & Duquia, 2007).

De forma geral, as perguntas de pesquisa em saúde podem ser classificadas nos seguintes campos ampliados[2]: diagnóstico, rastreamento, psicometria, causalidade, prognóstico e tratamento. A cada um desses campos, corresponde um delineamento metodológico considerado mais adequado. Os principais delineamentos de pesquisa quantitativa são os seguintes (Greenhalgh, 2008):

- Transversal: é uma pesquisa observacional em que uma amostra representativa de participantes é investigada quanto à presença de uma determinada condição, através de levantamentos (questionários), exames, testes, e outros procedimentos. A coleta de dados é feita em um único momento, podendo se referir a dados atuais ou do passado (por exemplo, levantamento em prontuários). Um estudo aprofundado sobre pesquisas transversais pode ser encontrado no artigo de Bastos e Duquia (2007).
- Coorte longitudinal: é um estudo observacional em que dois ou mais grupos de participantes são definidos de acordo com suas diferenças em relação a uma determinada condição (ex.: exposição a uma vacina, presença de uma doença) e acompanhados ao longo do tempo, realizando-se sucessivas avaliações de um determinado desfecho definido em protocolo.
- Caso-controle: é um estudo observacional em que os participantes com uma determinada condição (grupo caso) são emparelhados com outros participantes que não têm essa condição (grupo controle), mas são semelhantes aos primeiros em determinadas variáveis (ex.: sexo, idade, condição socioeconômica). Dados do passado desses participantes são investigados de forma retrospectiva, para levantar se foram expostos a um determinado evento que, hipoteticamente, é o agente causal da condição estudada.
- Ensaio clínico aleatorizado: é um estudo experimental em que os participantes são aleatoriamente alocados em dois ou mais grupos, que recebem intervenções distintas. Os grupos são acompanhados por um período de tempo e analisados em termos de desfechos previamente estabelecidos em protocolo. Como os critérios de inclusão asseguram que os

[2] No presente trabalho, são apresentados conceitos referentes a métodos quantitativos de pesquisa. A pesquisa qualitativa é válida para responder a outros tipos de perguntas e deve ter sua qualidade avaliada por outros critérios.

grupos comparados são semelhantes, exceto pela intervenção, espera-se que qualquer diferença entre eles seja atribuível à intervenção.

A Tabela 1 mostra a correspondência entre campos ampliados de pesquisa e delineamentos metodológicos (Greenhalgh, 2008). Estudos com esses delineamentos são considerados primários e quantitativos.

Tabela 1 – Correspondência entre campos ampliados de pesquisa em saúde e delineamentos metodológicos mais adequados (adaptada de Greenhalgh, 2008).

Campo ampliado de pesquisa	**Descrição do campo**	**Delineamento**
Diagnóstico	Acurácia[3] de um instrumento de diagnóstico.	Estudo transversal
Rastreamento	Teste da acurácia de instrumentos que podem ser aplicados a grandes populações, para identificar pessoas com maiores chances de apresentarem determinada doença ou condição (e que, portanto, devem ser encaminhadas para realização de teste diagnóstico).	Estudo transversal
Psicometria	Mensuração de atitudes, crenças, preferências e características de personalidade.	Estudo transversal
Causalidade	Investigação de relação de causa e efeito entre dois eventos (ex.: um suposto agente prejudicial e uma determinada doença)	Estudo de coorte longitudinal ou estudo caso-controle
Prognóstico	Previsão do que pode acontecer a um indivíduo a partir da detecção de uma doença.	Estudo de coorte longitudinal
Tratamento	Teste da eficácia de tratamentos medicamentosos, psicoterápicos ou outro tipo de intervenção.	Ensaio clínico aleatorizado

[3] O conceito de acurácia se refere ao quanto o resultado de uma medida se aproxima do real valor medido. Pesquisas sobre acurácia de instrumentos de diagnóstico buscam avaliar o quanto esses instrumentos são eficazes ou eficientes para distinguir pacientes com determinada condição daqueles que não apresentam essa condição (Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi, 2010).

2.2.2 Revisão sistemática de literatura

Revisões sistemáticas de literatura são estudos secundários (isto é, cujas unidades de análise são outros estudos científicos), que usam métodos sistemáticos de busca de estudos científicos e critérios de seleção definidos *a priori*, com o objetivo de responder a uma pergunta específica. Portanto, o conceito de revisão sistemática pode ser definido como:

> Uma revisão de uma pergunta claramente formulada que usa métodos sistemáticos e explícitos para identificar, selecionar e avaliar criticamente pesquisas relevantes; e coletar e analisar dados dos estudos incluídos na revisão. Métodos estatísticos (metanálise) podem ou não ser usados para analisar e sumarizar os resultados dos estudos incluídos.[4] (The Cochrane Collaboration, 2005,45).

Apesar de serem chamadas de estudos secundários, as revisões sistemáticas de literatura são reconhecidas, para grande parte das perguntas de pesquisa, como "padrão-ouro" do conhecimento científico, isto é, são consideradas as evidências científicas que devem ser consideradas mais relevantes ao se tomar uma decisão. Isso se deve à sua capacidade de reunir resultados de outros estudos científicos usando metodologia que diminui a possibilidade de viés, ao contrário das chamadas revisões narrativas.

As revisões sistemáticas e narrativas se diferenciam pela metodologia adotada (Cook, Mulrow, & Haynes, 1997). A revisão narrativa, mais frequente na literatura científica, é aquela em que o pesquisador faz uma busca nas fontes de informação sem criar uma estratégia prévia e definir a priori os critérios de seleção dos estudos encontrados. Esse tipo de revisão é adequado para apresentar conceitos gerais, mas está sujeito a vieses, porque o revisor pode incluir ou dar mais destaque para os estudos que confirmarem sua hipótese de pesquisa e deixar de mencionar estudos que refutem essa hipótese. A revisão sistemática, por sua vez, estabelece, a priori, métodos sistemáticos de busca e seleção de estudos científicos, sendo mais adequada para responder a uma pergunta específica e subsidiar uma

---

[4] No original: A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies.

decisão. A Tabela 2 é um quadro comparativo entre revisões sistemáticas e narrativas.

Tabela 2 – Características da revisão narrativa e da revisão sistemática de literatura (adaptado de Cook, Mulrow, & Haynes, 1997)

	Revisão narrativa	**Revisão sistemática**
Objetivos	Apresenta conceitos gerais	Busca responder a uma pergunta específica
Fontes e busca	Não explicita métodos sistemáticos de busca, seleção e análise	Explicita métodos sistemáticos de busca, seleção e análise
Seleção de estudos	Estudos incluídos na revisão são selecionados por conveniência	Estudos incluídos na revisão são selecionados por critérios estabelecidos *a priori*
Peso dos estudos	Todos os estudos têm o mesmo peso	Os resultados dos estudos podem ser ponderados, atribuindo-se maior peso àqueles com qualidade superior ou tamanho de amostra maior
Síntese	Geralmente qualitativa	Qualitativa ou quantitativa (metanálise)
Inferências	Método subjetivo para chegar à conclusão	Método objetivo e replicável para chegar à conclusão

Revisões sistemáticas de literatura têm sido cada vez mais comuns, pelo seu alto potencial para resultar em conclusões aplicáveis ao cotidiano e que possam nortear decisões e investimentos. Esforços têm sido feitos para ampliar o número de revisões sistemáticas e elevar a sua qualidade metodológica, contando com o trabalho, principalmente, de organizações voltadas para a saúde baseada em evidências. Na área de medicina, uma organização internacional, chamada Colaboração Cochrane, objetiva preparar, manter e assegurar o acesso a revisões sistemáticas sobre efeitos de intervenções em saúde. Com características

semelhantes, a Colaboração Campbell, fundada em 2000, na Filadélfia (Pensilvânia), objetiva oferecer apoio a políticos, gestores e demais profissionais na tomada de decisões em relação a programas. Suas atividades são preparar, manter e promover acesso a revisões sistemáticas nas áreas de educação, justiça criminal, política, serviço social, psicologia e economia (Sánchez-Meca, Boruch, Petrosino, & Rosa, 2002).

Uma das ferramentas que facilitam a tomada de decisões a partir das revisões sistemáticas é a aplicação de técnicas estatísticas que permitem sintetizar os resultados dos diferentes estudos. Assim, a pergunta de pesquisa que levou à revisão pode ser respondida a partir de critérios estatísticos, como o tamanho de efeito. A isso se dá o nome de metanálise.

2.2.3 Metanálise: Aplicação de técnicas estatísticas à Revisão Sistemática

O termo metanálise surgiu na década de 1970, com a publicação do artigo "Prymary, secondary and meta-analysis of research" (Glass, 1976), que buscava incentivar a aplicação de metanálise na área da educação. A metanálise pode ser definida como "Conjunto de procedimentos estatísticos que permite combinar os resultados de vários estudos fornecendo uma avaliação geral das relações entre as variáveis" (Cozby, 2009, 425). O Glossário da Colaboração Cochrane define metanálise como "O uso de técnicas estatísticas em uma revisão sistemática para integrar os resultados de estudos incluídos"[5] (The Cochrane Collaboration, 2005, 25).

Ou seja, a metanálise consiste em sumarizar resultados de dois ou mais estudos conduzidos de forma independente, aplicando-se técnicas estatísticas e permitindo que seja extraída uma conclusão, que pode, inclusive, ser diferente das apontadas pelos estudos primários incluídos. Essa técnica pode ser usada para sintetizar resultados de estudos com quaisquer tipos de delineamentos e em qualquer área da ciência, realizando-se adaptações metodológicas de acordo com a pergunta de pesquisa e as características dos trabalhos que se pretende sintetizar.

[5] No original: "The use of statistical techniques in a systematic review to integrate the results of included studies.".

2.2.4 Pirâmides de evidências

Pirâmides de evidências são representações gráficas de ordenações das pesquisas em níveis hierárquicos, conforme o campo ampliado correspondente à pergunta de pesquisa, a natureza e o delineamento do estudo. Nessas figuras, o topo da pirâmide é ocupado por estudos que oferecem maior confiabilidade e a base ocupada por estudos menos confiáveis em termos metodológicos. Na área da Psicologia, ainda não foram estabelecidas as pirâmides indicativas dos níveis de evidência para avaliar quais são os estudos com características metodológicas mais adequadas para responder às perguntas de pesquisa propostas. A pirâmide de evidências mais conhecida (Figura 2) é proveniente da Medicina e apresenta os níveis hierárquicos para avaliação da qualidade do delineamento de estudos de eficácia de intervenções.

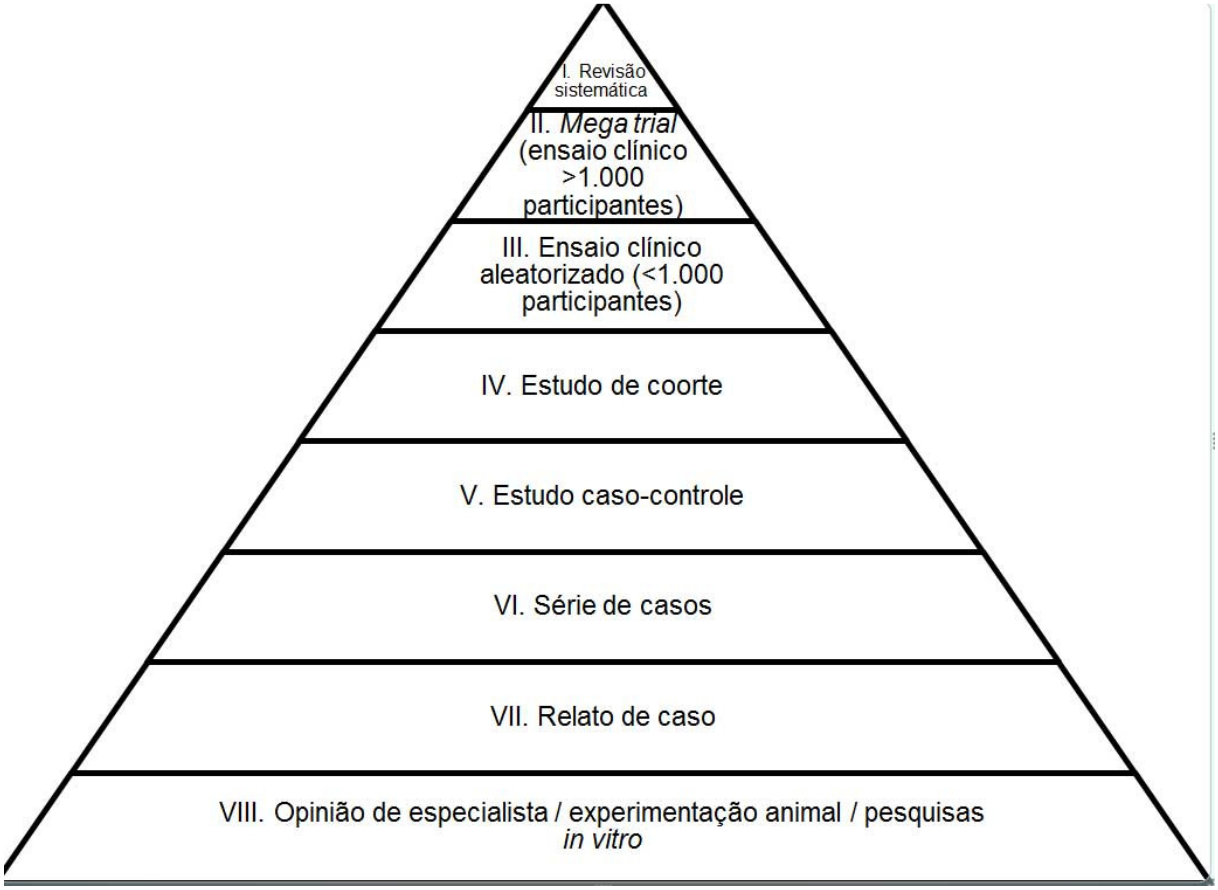

Figura 2 – Níveis de evidência de estudos sobre eficácia de intervenções (adaptado de Melnik & Atallah, 2011).

Embora tenham vantagens do ponto de vista didático, as hierarquias estabelecidas pelas pirâmides de evidências devem ser relativizadas. Em razão da complexidade da área de saúde, o estabelecimento de uma hierarquia rígida de níveis de evidência baseada no tipo de pesquisa pode ser considerado uma abordagem demasiadamente simplista (Concato, 2004). Ensaios clínicos aleatorizados, apesar de reduzirem a chance de viés na obtenção de resposta para a pergunta de pesquisa, podem restringir a amostra de pacientes selecionados, o tipo de intervenção e os desfechos avaliados, reduzindo a possibilidade de generalização dos resultados. Uma revisão sistemática de literatura, buscando metanálises baseadas em ensaios clínicos aleatorizados e em estudos observacionais, concluiu que:

> [...] contrariando crenças anteriores, os resultados médios de estudos observacionais (de coorte ou caso-controle) bem delineados não superestimaram sistematicamente a magnitude das associações entre exposição e desfecho relatadas em ensaios clínicos aleatorizados. Em vez disso, os resultados sintéticos de ensaios clínicos aleatorizados e estudos observacionais foram muito semelhantes para cada questão clínica estudada (Concato, 2004, p. 343).[6]

Na mesma linha de argumentação, considerou-se que as pirâmides de evidências foram planejadas para questões de eficácia terapêutica, não se aplicando ou tendo aplicação reduzida a outras questões de pesquisa na área de saúde, como por exemplo: efetividade de intervenções, ou seja, como elas se comportam em situações reais, e/ou naquelas em que ensaios clínicos aleatorizados não são possíveis. Nesse caso, em vez de usar classificações rígidas como as pirâmides de evidência, sugere-se considerar o quanto os delineamentos de estudos disponíveis na literatura seriam apropriados para responder à pergunta de pesquisa (Ho, Peterson, & Masoudi, 2008).

Pode-se acrescentar que as pirâmides de evidências mais usadas na área de saúde não são aplicáveis a outros campos ampliados de pesquisa, como por

---

[6] No original: […] contrary to prevailing beliefs, average results from well-designed observational (cohort and case-control) studies did not systematically overestimate the magnitude of exposure-outcome associations reported in randomized, controlled trials. Rather, the summary results from randomized, controlled trials and observational studies were remarkably similar for each clinical question addressed.

exemplo: diagnóstico, rastreamento, causalidade, prognóstico e estudos de prevalência e incidência. Classificações mais abrangentes e flexíveis foram propostas pelo Oxford Centre for Evidence-Based Medicine (OCEBM Levels of Evidence Working Group, 2011), indicando níveis de evidência para sete perguntas: estudos epidemiológicos (quão comum é o problema?), diagnóstico (o teste de diagnóstico ou monitoramento é acurado?), prognóstico (o que acontecerá se não for adicionado um tratamento?), benefícios de um tratamento (a intervenção ajuda?), prejuízos de um tratamento (duas perguntas: quão comuns são os prejuízos? Quais são os prejuízos raros?) e rastreamento (o teste [de detecção precoce] vale à pena?). O documento advertiu que os níveis de evidência ocupados por cada delineamento podem variar. Por exemplo, se forem encontrados estudos com qualidade baixa, imprecisão ou inconsistência entre eles, ou com tamanho de efeito muito pequeno, estes devem ser graduados em níveis de evidência mais baixos. Da mesma maneira, os estudos podem ser graduados em níveis de evidência mais elevados, se for encontrado tamanho de efeito grande. Em todos os campos ampliados de pesquisa, uma revisão sistemática geralmente ocupa nível mais alto de evidência que um estudo individual.

2.3 Procedimentos para a tomada da decisão conforme a Prática da Psicologia Baseada em Evidências (PBE)

Dada a importância das evidências científicas para a tomada de decisões na prática profissional, considera-se fundamental que profissionais de saúde, sobretudo gestores, conheçam as etapas envolvidas no processo de decisão baseado em evidências científicas (American Psychological Association, 2005):

1. Formulação da pergunta de pesquisa
2. Busca da melhor evidência científica disponível
3. Avaliação da qualidade da evidência disponível
4. Integração de resultados de diferentes pesquisas
5. Tomada de decisão.

2.3.1 Formulação da pergunta de pesquisa

O primeiro passo para os profissionais ou o gestores que desejam buscar evidências científicas para subsidiar uma decisão é formular a pergunta de pesquisa, usando os elementos importantes para o delineamento do estudo. Por exemplo, em estudos sobre efetividade e/ou eficácia de intervenções, tradicionalmente, se usa o modelo conhecido como "PICO": Participantes, Intervenção, Controle e *Outcomes* ou desfecho (Castro, Saconato, Guidugli, & Clark; Higgins & Green, 2011; Greenhalgh, 2008). Esses quatro elementos constitutivos de qualquer pergunta sobre intervenção clínica devem ser pensados da seguinte maneira:

1. Participantes: nesse campo, o pesquisador descreve quais são as características dos participantes que ele deseja contemplar em sua investigação (ex.: adultos de sexo masculino, gestantes, idosos), bem como a condição ou problema investigado (ex.: hipertensão arterial, depressão).

2. Intervenção: é o tratamento ou outro tipo de ação aplicado aos participantes (ex.: uso de determinado medicamento, psicoterapia, exercícios físicos).

3. Controle: é o tratamento alternativo usado para comparação com o tratamento investigado (ex.: medicação vs. placebo, medicação vs. psicoterapia).

4. Desfecho (*outcomes*): é o resultado da intervenção, ou, mais precisamente, o indicador clínico da modificação do paciente após intervenção. Ao buscar estudos sobre efetividade ou eficácia de intervenções, deve-se definir previamente o tipo de desfecho de interesse. No entanto, o desfecho não deve ser incluído na pergunta de pesquisa, pois ele é o resultado que deverá responder a essa pergunta.

Nem todas as pesquisas seguem o modelo PICO, porque na área de Psicologia, assim como em outras ciências, existem estudos cujo foco não é a intervenção. Por exemplo, no caso de levantamentos de prevalência ou incidência, a intervenção e o controle não fazem parte do delineamento. O mesmo se dá com outros campos ampliados de pesquisa: diagnóstico, rastreamento, psicometria, causalidade e prognóstico. Nesses casos, pode ser feita uma adaptação do modelo PICO, que foi chamada de PVO (Biruel & Pinto, 2011). As iniciais P (participantes) e O (outcomes) se mantém as mesmas e foi inserida a inicial V, de Variáveis, "que podem ou não ser propostas pelo pesquisador e utilizadas como limitador na

construção da estratégia de pesquisa, ou identificadas como categorias em uma estratégia de pesquisa mais ampla". (p.3).

2.3.2 Busca da melhor evidência científica disponível

A partir da pergunta de pesquisa, as evidências científicas precisam ser localizadas nas bases de dados. A busca deve ser planejada para ser, ao mesmo tempo, abrangente, evitando que se percam informações importantes, e específica, para que o profissional ou gestor não seja sobrecarregado com um excesso de informações irrelevantes. Deve ser incluído o maior número possível de bases de dados que indexem trabalhos científicos, visto que, apesar de haver sobreposição entre os resultados retornados, certamente haverá estudos que não estarão referenciados em todas as bases. No caso da Psicologia, as principais bases de dados são:

- PsycINFO – Considerada a mais importante base de dados na área da Psicologia. Desenvolvida e mantida pela *American Psychological Association*, reúne, organiza e divulga a literatura relevante publicada internacionalmente em Psicologia e disciplinas correlatas.
- Psicodoc – Desenvolvida pelo Colégio de Psicólogos de Madrid, inclui periódicos, trabalhos publicados em congressos e livros na área, editados na Espanha e América Latina, desde 1975 até a atualidade.
- BVS-PSI ULAPSI – Reúne informações sobre periódicos, teses, livros, filmes e outros materiais publicados nos países da América Latina na área de Psicologia.

Além dessas bases, específicas para o campo da Psicologia, estudos relevantes podem ser encontrados em bases de dados eletrônicas da Medicina, da Educação ou de ciência em geral, como por exemplo:

- ISI Web of Knowledge – Uma das mais importantes bases de dados, com reconhecimento mundial, permite a recuperação de trabalhos publicados nos mais relevantes periódicos internacionais, em todas as áreas do conhecimento.

- SCOPUS – Base de dados multidisciplinar, produzida pela editora Elsevier desde 2004, com cobertura desde 1960. Indexa artigos de periódicos de diversas áreas e, a exemplo da Web of Science, faz análise das citações.
- Lilacs (Literatura Latino-Americana e do Caribe em Ciências da Saúde) – Base de dados cooperativa do Sistema BIREME, compreende teses, capítulos de teses, livros, capítulos de livros, anais de congressos ou conferências, relatórios técnico-científicos e publicações governamentais em Ciências da Saúde, publicada nos países da Região, a partir de 1982.
- PubMed – é uma das mais importantes bases de dados desenvolvidas no mundo. Publicada pelo *National Center for Biotechnology Information* (NCBI) na *National Library of Medicine* (NLM) disponível na *web*, indexa periódicos e outros materiais na área de ciências da saúde.
- 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 pesquisa. É produzida pelo *Education Resources Information Center*.

Em uma revisão sistemática de literatura, espera-se que todo o conhecimento científico disponível seja mapeado. Portanto, além do uso das bases de dados eletrônicas e nas listas de referências bibliográficas dos estudos encontrados, recomenda-se a busca na chamada "literatura cinzenta"[7], ou seja: teses e dissertações (podendo-se usar acervos eletrônicos como a Biblioteca Digital de Teses e Dissertações ou a *Dissertation Abstracts*), anais de eventos e contato com os principais especialistas da área, para verificar se conhecem outros estudos em andamento ou ainda não publicados (Higgins & Green, 2011). A literatura cinzenta deve ser analisada com cuidado, visto que não é submetida à revisão rigorosa que é realizada por pareceristas no caso dos artigos em periódicos. Para elaborar uma estratégia de busca adequada nas bases de dados, é fundamental consultar um bibliotecário, preferencialmente, com experiência em revisões sistemáticas.

[7] Em inglês, "*gray literature*".

2.3.3 Avaliação da qualidade da evidência disponível

A seleção da melhor evidência científica disponível inicia-se pela definição de critérios de inclusão de estudos, elaborados de acordo com a pergunta de pesquisa, a partir da qual se determina o delineamento metodológico mais adequado para respondê-la (Greenhalgh, 2008). Esses critérios, estabelecidos anteriormente à busca de trabalhos, devem estar relacionados com o delineamento da pesquisa, procurando evitar vieses metodológicos.

Após a seleção dos trabalhos, também é preciso avaliar a qualidade metodológica dos mesmos. Essa avaliação pode ser feita de forma qualitativa pelo revisor, ou de forma quantitativa, a partir de escalas padronizadas.

Assim, na etapa de avaliação da evidência científica disponível, o profissional ou gestor deve procurar responder a três questões: 1) Que campo ampliado de pesquisa o estudo abrange? 2) O delineamento do estudo foi adequado para responder à questão proposta? e 3) A qualidade metodológica do estudo é adequada? É fundamental que o revisor tenha bom conhecimento de metodologia de pesquisa, para fundamentar sua escolha de critérios de avaliação da qualidade metodológica e para que lance olhar crítico sobre os estudos encontrados, procurando identificar possíveis fontes de vieses e o quanto estes podem comprometer a aplicabilidade dos resultados à prática profissional.

2.3.4 Integração dos resultados de diferentes pesquisas

A quarta etapa da Prática Baseada em Evidências é a integração de resultados, que impõe ao profissional ou gestor o desafio de lidar com dados de diferentes pesquisas de forma sistemática e reduzindo, na medida do possível, o risco de viés. Essa tarefa pode ser facilitada pelos estudos secundários, ou seja, aqueles que se baseiam nos resultados de estudos primários, integrando-os para que se possa chegar a uma nova conclusão. A revisão sistemática é um exemplo de pesquisa secundária.

2.3.5 Tomada de decisão

A quinta e última etapa é tomar uma decisão com base em evidências científicas. As evidências científicas encontradas nos passos anteriores devem ser integradas às considerações em relação à experiência profissional, ao contexto cultural e às características e preferências do indivíduo ou grupo de indivíduos que será alvo de intervenções, norteando a decisão final (American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006).

2.4 Procedimentos para revisão sistemática com metanálise de estudos de prevalência

O presente trabalhou investigou a prevalência de DPP e o impacto de fatores metodológicos dos estudos nos resultados de prevalência, através de metanálise. Buscas na literatura sobre metodologia de pesquisa não localizaram diretrizes para revisões sistemáticas de estudos observacionais. As diretrizes mais consistentes na área de saúde direcionaram-se a estudos de eficácia e efetividade de intervenções, que são o foco da Colaboração Cochrane. Nesse caso, as unidades de análise são ensaios clínicos aleatorizados, pois pesquisas com esse delineamento têm maior chance de proporcionar informações confiáveis sobre as diferenças de efeito entre as formas de tratamento (Higgins & Green, 2011). Estudos de prevalência, por sua vez, são observacionais, apresentando delineamento transversal ou longitudinal e, portanto, têm particularidades que precisam ser levadas em conta ao se planejar uma metanálise.

A Colaboração Cochrane considerou que, nos casos em que ensaios clínicos aleatorizados não são possíveis, ou para perguntas de pesquisa que não estão relacionadas a eficácia ou eficiência de intervenções (ex.: pesquisa sobre efeitos adversos), decisões na área de saúde podem ser feitas com base em revisões sistemáticas de estudos observacionais, ou mesmo outros tipos de delineamento. A orientação, para esses casos, é que “os princípios básicos de revisão são os mesmos, qualquer que seja o tipo de evidência que está sendo revisto” (Higgins & Green, 2011, p. 16).

Portanto, na área de saúde, revisões sistemáticas de estudos observacionais têm sido feitas, principalmente, em situações em que a realização de ensaios

clínicos aleatorizados é difícil ou inadequada por razões éticas. Outras áreas que necessitam de revisões dessa natureza, como associação entre fatores de risco e ocorrência de uma doença, avaliação de testes de rastreamento e variáveis relacionadas ao prognóstico implicam em desafios metodológicos particulares. Por isso, a área demanda aprimoramento nas questões metodológicas sobre inclusão de estudos com diferentes delineamentos, populações e contextos, bem como nos métodos estatísticos para sintetizar os dados (Gülmezoglu, Say, Betrán, Villar, & Piaggio, 2004).

Na ausência de padrões internacionais que definam a metodologia mais adequada para estudos de prevalência, a metodologia do *Cochrane Handbook for Systematic Reviews of Intervention* (Higgins & Green, 2011) e de outras revisões de literatura sobre prevalência de transtornos mentais (ex.: Bennet, Einarson, Taddio, Koren, & Einarson, 2004; Gülmezoglu et al., 2004) foram adaptadas para o presente trabalho.

2.4.1 Formulação da pergunta de pesquisa e avaliação da necessidade da revisão sistemática

Como em todas as pesquisas, o projeto de revisão sistemática de literatura inicia-se pela formulação da pergunta, a partir da qual serão elaborados os critérios de busca e de seleção de artigos e tomadas decisões a respeito da extração e análise dos dados. Em seguida, é preciso avaliar a necessidade da revisão sistemática (Higgins & Green, 2011). Para isso, o primeiro passo é verificar a existência de outras revisões sistemáticas sobre o tema, consultando, principalmente, as bases de dados especializadas em estudos secundários (ex.: *The Cochrane Library* e a biblioteca da Colaboração Campbell). Caso elas existam, devem ser avaliadas quanto à sua qualidade e data da última atualização. Se for constatado que as revisões pré-existentes são de baixa qualidade, pode ser necessário realizar uma nova revisão. Se forem encontradas revisões de boa qualidade, mas desatualizadas, deve ser feito um processo de atualização.

2.4.2 Níveis de evidência em pesquisas de prevalência

A qualidade das evidências disponíveis tem relação, dentre outros fatores, com o delineamento dos estudos e os níveis que eles ocupam na hierarquia. A proposta de pirâmides de evidências para a área de saúde do Oxford Centre for Evidence-Based Medicine (OCEBM Levels of Evidence Working Group, 2011) inclui os estudos de prevalência, recomendando que para responder à pergunta: "Quão comum é o problema?", os delineamentos dos estudos correspondem aos níveis de evidência apresentados na Tabela 3. O nível mais alto da hierarquia (Nível 1) é ocupado por estudo de levantamento local e atual, visto que a epidemiologia das condições estudadas costuma variar de acordo com a localidade e ao longo do tempo. Esses estudos devem ser censitários ou ter amostra aleatória, pois as amostras de conveniência8 apresentam alto risco de viés.

Tabela 3 – Níveis de evidência para responder à pergunta "Quão comum é o problema?", baseado na proposta do OCEBM Levels of Evidence Working Group (2011)

Nível de evidência	**Tipo de estudo**
Nível 1	Estudo de levantamento local e atual, censitário ou com amostra aleatorizada
Nível 2	Revisão sistemática de levantamentos que permitem pareamento com as circunstâncias locais
Nível 3	Estudo local com amostra não aleatorizada
Nível 4	Série de casos

[8] Considera-se amostra de conveniência aquela em que participantes são selecionados de acordo com sua disponibilidade ao pesquisador, ou seja, porque são mais fáceis de serem abordados para participar da pesquisa (Lunsford & Lunsford, 1995).

2.4.3 Definição dos critérios de inclusão e estratégia de busca de estudos

Para definir critérios de inclusão de estudos na revisão sistemática, é preciso estabelecer: qual é ou quais são os tipos de delineamento mais adequados para responder à pergunta de pesquisa; qual é o tipo de participante que o estudo deve incluir; e qual é o desfecho estudado. Os critérios de inclusão devem ser definidos *a priori*, isto é, antes do início da coleta de dados, para evitar que a escolha dos estudos seja feita de forma a enviesar os resultados (Higgins & Green, 2011).

Objetivando encontrar o máximo possível de estudos sobre o tema pesquisado, é fundamental que a estratégia de busca inclua diversas bases de dados eletrônicas. Para tanto, o conhecimento das bases de dados da área em que está sendo feita a revisão é de suma importância. Preferencialmente, devem ser identificados e aceitos trabalhos em todos os idiomas, pois a limitação idiomática pode enviesar os resultados (Higgins & Green, 2011). Esforços devem ser empreendidos na identificação de estudos de países em desenvolvimento, pois têm maior chance de ser publicados em periódicos não indexados e em idiomas diferentes do inglês (Gülmezoglu et al., 2004).

Além das buscas em bases de dados eletrônicas, outras estratégias devem ser empregadas para identificar trabalhos que não estão indexados nas bases de dados: verificação das listas de referências dos artigos; e busca por estudos não publicados ou em andamento, que deve ser feita através da consulta aos anais dos principais eventos científicos da área e aos principais autores que estudam esse tema. Estudos em andamento, embora não tenham seus resultados incluídos na revisão sistemática de literatura, devem ser referenciados para contribuir com futuras atualizações (Higgins & Green, 2011).

2.4.4 Seleção de estudos e coleta de dados

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

Geralmente, o uso de múltiplas bases para a busca sistemática de trabalhos implica na localização de um grande número de registros, o que se deve, em parte, a uma particularidade das revisões sistemáticas de estudos observacionais: não é possível pesquisar de acordo com o tipo de estudo, como é o caso de revisões sobre eficácia e efetividade de intervenções, que buscam apenas ensaios clínicos controlados (Gülmezoglu et al., 2004). Sugere-se, portanto, que a seleção de estudos seja iniciada por um rastreamento por títulos e resumos, para selecionar os registros que têm chance de atender aos critérios de inclusão estabelecidos no protocolo.

A inclusão de mais de uma base de dados costuma resultar em sobreposições (o mesmo trabalho é identificado mais de uma vez). Estudos duplicados devem ser eliminados, tarefa que pode ser simplificada pelo uso de um *software* de gerenciamento, como por exemplo, o Reference Manager (usado por Gülmezoglu et al., 2004) ou o EndNote® Web (*software* gratuito, https://www.myendnoteweb.com/EndNoteWeb.html, acessado em 04/09/2012). O EndNote® Web apresenta a ferramenta "*Find duplicates*" (encontrar duplicatas), que compara os registros através dos seguintes campos: autor, ano, título e tipo de referência (ex.: artigo, livro, capítulo etc.). No entanto, a identificação automática de duplicatas pode ser prejudicada pela diferença de formato nos registros de diferentes bases de dados. Por exemplo, pode ocorrer dificuldade com os artigos escritos em espanhol, pois os autores costumam ter dois sobrenomes, sendo que nem sempre os registros diferenciam corretamente o sobrenome que deve ser usado nas citações. Dificuldades dessa natureza levaram Gülmezoglu et al. (2004) a afirmar:

> Essa ausência de uniformidade na formatação entre as bases de dados geraram diversos registros diferentes do mesmo artigo, se indexado em mais de uma base de dados. Nós vivenciamos um processo exaustivo de busca e detecção manual de registros duplicados para resolver esse problema. (p.6).[9]

Ao menos dois revisores devem participar do processo de seleção de estudos para a revisão sistemática, realizando todas as etapas de forma independente. Em

[9] No original: This lack of uniformity in formatting of citations across databases causes several different recordings of the same article if indexed in more than one database. We experienced an exhaustive process of manual searching and deleting of duplicate records to address this issue.

caso de divergências, pode-se buscar consenso, ou consultar um terceiro revisor (Higgins & Green, 2011). A tarefa de seleção resulta na classificação de estudos de acordo com os seguintes critérios:

1. **Estudos identificados** – são todos os localizados pela busca nas bases de dados. Deve-se registrar o número desses estudos para cada fonte;

2. **Estudos não selecionados** – dentre os estudos identificados, são aqueles que claramente não preenchem os critérios de inclusão. Essa constatação pode ser feita a partir da leitura do título e resumo dos artigos, não sendo necessário avaliar o texto completo. Deve-se citar o número de estudos nessas condições.

3. **Estudos selecionados** – dentre os estudos identificados, são os que, aparentemente, preenchem os critérios de inclusão. Nessa etapa, a seleção é feita a partir dos resumos, não havendo necessidade de ler o texto completo dos artigos.

4. **Estudos excluídos** – são os estudos selecionados pelo resumo, mas que não preenchem os critérios de inclusão, conforme avaliação do texto completo. Deve-se citar o número de estudos nessas condições.

5. **Estudos incluídos** – são os estudos selecionados em todas as etapas, ou seja, que atendem a todos os critérios de inclusão, após a avaliação do texto completo. Deve ser registrado o número de estudos nessas condições e as referências bibliográficas completas precisam constar do trabalho final.

O grau de concordância entre os revisores quanto à inclusão e exclusão de estudos por títulos, resumos e texto completo precisa ser avaliado através de testes estatísticos e apresentado ao relatar os resultados da revisão sistemática (Higgins & Green, 2011).

2.4.5 Análise dos dados

A fase de análise dos dados divide-se em duas partes: análise da qualidade dos estudos incluídos, de caráter descritivo, e análise estatística, de caráter quantitativo.

2.4.5.1 Análise de qualidade dos estudos incluídos

A análise da qualidade dos estudos incluídos abrange duas tarefas: avaliar o grau em que as metodologias dos estudos estão apropriadas para evitar vieses nos resultados; e avaliar, por meio de técnicas estatísticas, qual é o nível de heterogeneidade dos resultados dos estudos incluídos. A primeira tarefa pode ser facilitada pelo uso de uma escala de avaliação da qualidade metodológica dos estudos incluídos, atribuindo-se pontuações para aspectos da metodologia de pesquisa considerados relevantes para assegurar que os resultados sejam replicáveis e apresentem o menor nível possível de viés. A opção pelo uso de uma escala deve ser feita antes do início da coleta de dados e apresentada no protocolo de pesquisa. As pontuações dos estudos nas escalas de qualidade podem ser usadas como fator de ponderação durante a análise estatística dos resultados, atribuindo-se maior peso aos trabalhos de melhor qualidade.

Até o presente momento, a maior parte dos esforços para criar procedimentos e guias para elaboração de revisões sistemáticas na área de saúde tem sido aplicada aos estudos sobre eficácia e efetividade de intervenções. Assim, não existe uma padronização das recomendações sobre a metodologia mais adequada para estudos epidemiológicos, o que põe em dúvida a confiabilidade de seus resultados (Giannakopoulos, Rammelsberg, Eberhard, & Schmitter, 2012).

As primeiras escalas para avaliar estudos epidemiológicos de prevalência e incidência foram publicadas no ano de 1998 (Loney, Chambers, Bennett, Roberts, & Stratford, 1998; Boyle, 1998). Ambos os instrumentos apontam para elementos semelhantes a serem avaliados. Loney et al. (1998) criou um roteiro estruturado para a avaliação crítica da qualidade de estudos de prevalência ou incidência, dividido em três partes: A. A metodologia do estudo é válida? (seis questões), B. Qual é a interpretação dos resultados? (uma questão), e C. Qual é a aplicabilidade dos resultados? (uma questão). O uso desse instrumento foi exemplificado a partir da investigação de qualidade de artigos sobre prevalência de demência no Canadá. O roteiro de avaliação de estudos de prevalência de Boyle (1998) também se divide em três partes: 1. Amostragem (três questões), 2. Medição (três questões) e 3. Análise (duas questões).

A *Agency for Healthcare Research and Quality* (AHRQ) publicou, em 2010, uma revisão sistemática de instrumentos de avaliação de qualidade de estudos

observacionais de prevalência e incidência, fatores de risco, tratamentos ou intervenções em saúde (Shamliyan, Kane, & Dickinson, 2010). Foram encontradas 96 escalas, das quais apenas cinco objetivavam avaliar estudos de prevalência e incidência. Nenhum dos artigos referentes a essas escalas relatou como o instrumento foi desenvolvido e apenas a escala de Loney et al. (1998) se propôs a avaliar a qualidade de estudos de prevalência ou incidência de qualquer transtorno, e não de uma condição específica.

Revisões sistemáticas de literatura sobre prevalência de diversos transtornos e fenômenos variaram em relação à metodologia adotada para avaliação da qualidade dos estudos incluídos. Por exemplo, estudos sobre prevalência de violência doméstica contra a mulher (Alhabib, Nur, & Jones, 2010) e prevalência de violência doméstica contra mulheres grávidas (Shamu, Abrahams, Temmerman, Musekiwa, & Zarowsky, 2011) usaram a escala de Loney et al. (1998), enquanto outros trabalhos usaram instrumentos construídos pelo próprio autor (Beck, 2001; Prins, Blanker, Bohnen, Thomas, & Bosch, 2002; Bennett et al., 2004).

Em uma revisão sistemática sobre prevalência de depressão pré-natal (Bennett et al. 2004), foi desenvolvido um *checklist* de 12 itens, adaptando-se variáveis do *NHS Centre for Reviews and Dissemination Guidance for Undertaking Systematic Reviews* (NHS Centre for Reviews and Dissemination, 2001 *apud* Bennett et al., 2004). Considerou-se que o maior potencial para viés em estudos observacionais seria a seleção de participantes e a taxa de perda, que foram avaliadas a partir das seguintes variáveis: estratégia de recrutamento (considerada adequada se foi fornecida evidência de que foi estudada uma amostra aleatória da população, ou de que a amostra é representativa da população geral), critérios de inclusão e exclusão (que deveriam constar do texto de forma explícita), avaliação do desfecho, informações sobre não respondentes e participantes perdidos ao longo do estudo (deveriam constar razões para não participação, número de perdas ao longo do estudo e dados perdidos) e informações demográficas sobre os participantes. Para cada item do *checklist*, foi atribuído um ponto, de forma que poderia ser atingida a pontuação máxima de 12 pontos na avaliação de qualidade.

Beck (2001) criou uma escala para avaliar a qualidade de estudos incluídos em uma revisão sistemática de literatura sobre fatores de risco para DPP. Sua escala de qualidade, portanto, objetivou avaliar a confiabilidade não apenas do resultado referente ao desfecho (DPP), mas também da medição dos possíveis

fatores preditivos. Foram considerados 11 critérios, que poderiam totalizar 32 pontos. Um ponto forte da escala foi ter considerado o tipo de avaliação de DPP, atribuindo maior qualidade a estudos que usaram entrevista clínica. No entanto, o instrumento é passível de críticas, visto que foram usados itens como “nível de especialização do primeiro autor” e “suporte financeiro”, que não estão, necessariamente, relacionados com a qualidade metodológica. Não se pode determinar, por exemplo, em que grau existe (ou mesmo, se existe) associação entre a qualidade da pesquisa e o nível de especialização do primeiro autor ou o suporte financeiro recebido. Também se pode questionar o uso de números arbitrários para definir as faixas em que o tamanho de amostra seria classificado e pontuado. A pontuação máxima foi atribuída a estudos com mais de 100 participantes, o que seria uma pontuação pouco discriminativa atualmente, visto que a maior parte das pesquisas epidemiológicas sobre DPP tem mais de 100 participantes.

Recentemente, foi publicado um novo instrumento de avaliação de qualidade de estudos de prevalência (Giannakopoulos et al., 2012), com itens avaliados em três blocos principais: amostragem, medição e análise. No primeiro bloco, considerou-se a representatividade da amostra, avaliando se o estudo apresentou definição clara da população alvo (características como: idade; sexo; ocupação; classe social, educacional ou financeira; etnia; região de residência urbana, suburbana ou rural; dados relevantes sobre a saúde dos participantes; critérios de inclusão e exclusão bem definidos), o método de amostragem (uso de amostra probabilística) e se os participantes correspondiam à população alvo (taxa de resposta e descrição dos participantes perdidos). O segundo bloco considerou o quanto o instrumento usado para o estudo poderia resultar em medidas válidas e confiáveis do transtorno. O terceiro bloco, por fim, avaliou os procedimentos estatísticos, considerando que deveria ser relatado o intervalo de confiança do resultado de prevalência, o procedimento de recrutamento (deve ser feito na população geral, de forma que a amostra seja representativa) e o poder estatístico da amostra. Também foi incluído um item verificando se o estudo foi aprovado por comissão de ética em pesquisa, argumentando-se que essa medida poderá incentivar os pesquisadores a protegerem os indivíduos em estudos futuros. A pontuação dos itens foi somada para formar uma “Pontuação Total de Qualidade” (PTQ), que varia de zero (muito ruim) a 19 (excelente). A escala foi testada através

da obtenção da concordância entre dois avaliadores, ambos dentistas, em uma amostra de 400 artigos sobre transtorno temporomandibular, selecionados aleatoriamente. Os resultados indicaram que a concordância entre avaliadores foi satisfatória.

2.4.5.2 Avaliação do grau de heterogeneidade dos estudos incluídos

Faz parte da avaliação de qualidade dos estudos incluídos determinar o nível de homogeneidade entre eles, pois caso tenham metodologias demasiadamente heterogêneas, não há sentido em sintetizar estatisticamente seus resultados. Considera-se heterogeneidade qualquer tipo de variabilidade entre os estudos incluídos em uma revisão sistemática. Heterogeneidade estatística é o termo específico que se refere a uma consequência da variabilidade clínica (variabilidade entre participantes) e/ou metodológica (diferenças quanto ao delineamento e qualidade dos estudos incluídos) (Higgins & Green, 2011).

Críticas à técnica de metanálise frequentemente levantaram o problema conhecido como “misturar maçãs e laranjas”, ou seja, de combinar estudos sobre desfechos distintos, com metodologias distintas e cuja síntese em valor único, portanto, teria significado duvidoso do ponto de vista prático:

> Uma crítica comum às metanálises é que elas ‘combinam maçãs com laranjas’. Se estudos são clinicamente diversos, então uma metanálise pode não ter sentido, e diferenças genuínas entre os efeitos podem ser obscurecidas. Um tipo particularmente importante de diversidade são as comparações feitas pelos estudos primários. Frequentemente, é despropositado combinar todos os estudos incluídos em uma única metanálise. Às vezes, há uma mistura de comparações de diferentes tratamentos com diferentes comparadores, e pode ser preciso considerar cada combinação dos mesmos separadamente. Além disso, é importante não combinar desfechos que são muito diversos.[10] (Higgins & Green, 2011, p. 99).

[10] No original: A common criticism of meta-analysis is that they ‘combine apples with oranges’. If studies are clinically diverse then a meta-analysis may be meaningless, and genuine differences in effects may be obscured. A particularly important type of diversity is in the comparisons being made by the primary studies. Often is nonsensical to combine all the included studies in a single meta-analysis. Sometimes there is a mix of comparisons of different treatments with different comparators, each combination of which may need to be considered separately. Further, it is important not to combine outcomes that are too diverse.

A heterogeneidade em estudos de prevalência costuma ser grande, devido às diferenças metodológicas entre os estudos primários. Referindo-se a estudos de prevalência, Giannakopoulos et al. (2012) afirmaram que:

> As potenciais conclusões a partir dos resultados são limitadas por deficiências metodológicas. A grande heterogeneidade desses estudos é a razão pela qual a combinação estatística dos dados que deles derivam parece impossível e não deve ser um componente proeminente de revisões sistemáticas de estudos observacionais. Assim, o esforço científico parece inútil, visto que nenhuma combinação ou comparação estatística pode ser feita e, portanto, nenhuma conclusão a partir do quadro geral pode ser desenhada.[11] (p. 781).

A opinião de Giannakopoulos et al. (2012) parece pessimista, visto que metanálises sobre prevalência têm sido publicadas na literatura, inclusive levando em conta a heterogeneidade no modelo (ex.: Bennett et al., 2004; Waraich, Goldner, Somers, & Hsu, 2004; Gavin et al., 2005; Alhabib et al., 2010; Villegas, McKay, Dennis, & Ross, 2011). No entanto, as diferenças metodológicas entre estudos primários, de fato, dificultaram a síntese estatística das prevalências, levando alguns revisores a desistir de realizar metanálise, devido à alta heterogeneidade (ex.: Sawyer, Ayers, & Smith, 2010; Schram et al., 2010; Shamu et al., 2011; Lobato, Moraes, & Reichenheim, 2012). Essas diferenças também podem ter comprometido os resultados de diversas metanálises que não relataram ter realizado análise de heterogeneidade (ex.: O'Hara & Swain, 1996; Prins et al., 2002; Lewis, 2011; Melo Junior, 2011).

Assim, nas metanálises de estudos de prevalência, é fundamental que a heterogeneidade seja avaliada qualitativamente e quantitativamente, usando-se ferramentas estatísticas. Uma das formas de se quantificar a heterogeneidade é através de testes de qui-quadrado, procedimento que testa a hipótese de que os desfechos são iguais em todos os estudos. Outro parâmetro que pode ser usado é o I^2 ($I^2 = [(Q - gl)/Q] \times 100\%$, onde Q = valor do qui-quadrado, gl = graus de liberdade do teste), que descreve a porcentagem de variabilidade nas estimativas de efeito que se devem à heterogeneidade, e não ao erro de amostragem (acaso). Considera-

[11] No original: The potential conclusions from the results are limited by methodological shortcomings. The great heterogeneity of such studies is the reason that the statistical combination of data derived from them seems impossible and should not be a prominent component of systematic reviews of observational studies. Thus, great scientific effort remains useless, as no statistical combination or comparison can be conducted, and hence, no conclusions about the general picture can be drawn.

se que valores acima de 50% indicam heterogeneidade substancial (Higgins & Thompson, 2002; Higgins & Green, 2011).

Visualmente, a heterogeneidade pode ser observada através dos gráficos de floresta (Lewis & Clarke, 2001; Anzures-Cabrera & Higgins, 2010), que são construídos a partir da estimativa pontual de cada estudo e do intervalo de confiança (95%). Exemplos de gráficos de floresta podem ser encontrados na Figura 10 (p. 124) do presente trabalho. As estimativas pontuais de prevalência foram representadas pelos quadrados, enquanto as linhas horizontais representaram as estimativas intervalares (intervalo de confiança a 95%). O valor de prevalência média, que sintetizou os resultados dos estudos primários, foi representado pelo losango na parte inferior do gráfico.

Outliers podem ser identificados visualmente nos gráficos de floresta, mas recomenda-se que não sejam excluídos em busca de maior homogeneidade, pois o procedimento de excluir estudos com base em seus resultados pode gerar vieses (Higgins & Green, 2011). Recomenda-se, nesse caso, que seja feita análise de sensibilidade, isto é, a análise dos resultados que seriam obtidos com a exclusão dos *outliers*, e que deve ser comparado ao resultado obtido com a inclusão dos mesmos.

Um dos fatores que podem enviesar os resultados de uma metanálise é o viés de publicação, conceito definido como um desvio nos resultados que ocorre quando os estudos publicados não representam adequadamente o universo de trabalhos realizados sobre um determinado assunto (Delgado-Rodríguez, 2001). No caso de estudos sobre eficácia e efetividade de intervenções, espera-se viés de publicação, porque os trabalhos que concluem que o efeito da intervenção foi não significativo teriam menor chance de ser publicados, seja porque o pesquisador não submeteu o trabalho a periódicos científicos ou porque foi recusado pelos pareceristas (Rosenthal, 1979; Csada, James, & Espie, 1996). Provavelmente, esse viés é menos comum em revisões sistemáticas de estudos prevalência e incidência, visto que não envolvem intervenção. No entanto, revisões sistemáticas de estudos observacionais podem estar sujeitas a outro tipo de viés de publicação, relacionado ao país de origem.

A expressão "viés de publicação contra o Sul" (Victora & Moreira, 2006) foi usada para referir-se à desproporção entre a produção científica de países de renda alta (referidos como “Norte”) e os de renda média ou baixa (referidos como “Sul”).

Foram apontados quatro motivos para essa diferença: 1. a produção científica em países “do Sul” seria menor que em países “do Norte”, devido a menores investimentos financeiros e consequente “fuga de cérebros”; 2. haveria menor pressão em países “do Sul” que em países “do Norte” pela publicação de artigos; 3. os delineamentos escolhidos por autores de países “do Sul” seriam, em geral, os menos valorizados na literatura internacional, como estudos qualitativos e observacionais em detrimento dos experimentais; 4. ocorrência do chamado “racismo editorial”, ou a menor chance de publicação de artigos de países “do Sul” nos periódicos “do Norte”, que são os de maior visibilidade no meio científico, devido a preconceito dos editores e pareceristas.

Dados estatísticos corroboraram a afirmação de que países em desenvolvimento apresentam menor produção científica: uma revisão internacional de artigos da área de saúde publicados entre 1992 e 2001 verificou que apenas 7,9% dos artigos científicos eram de países de renda média e 2,7% de países de renda baixa. Mais de 90% dos artigos foram produzidos em 20 países, e mais de um terço da produção mundial era proveniente dos Estados Unidos da América (Paraje, Sadana, & Karam, 2005).

O viés de publicação pode ser investigado através de um gráfico de funil (Egger, Smith, Schneider, & Minder, 1997; Higgins & Green, 2011; Anzures-Cabrera & Higgins, 2010), como os que podem ser encontrados na Figura 11 (p. 124) do presente trabalho. Trata-se de um gráfico de distribuição que relaciona os resultados dos estudos com o tamanho da amostra ou o erro padrão. Parte-se do pressuposto de que estudos com amostras maiores terão maior precisão, enquanto estudos com pequenas amostras terão uma dispersão aleatória simétrica em relação ao resultado médio encontrado na metanálise. Nesse caso, os pontos do gráfico assumirão o formato de um funil. Formatos assimétricos podem indicar vieses de publicação, mas também podem ser produzidos por outras razões, como por exemplo: baixa qualidade metodológica dos estudos com amostras pequenas, ou variabilidade ao acaso (Higgins & Green, 2011). Ou seja, o gráfico de funil é uma forma de verificar a heterogeneidade dos resultados dos estudos incluídos na metanálise, que pode ou não ser causada por viés de publicação.

Embora a síntese dos resultados dos estudos incluídos em uma revisão sistemática possa ser prejudicada pela heterogeneidade, é possível usar técnicas estatísticas que a levem em consideração, permitindo que seja feita metanálise. O

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

2.4.5.3 Análise estatística

Após verificar a qualidade dos estudos incluídos, o nível de heterogeneidade e o viés de publicação, o revisor pode considerar adequado sintetizar os resultados dos diferentes estudos em uma metanálise. Nesse caso, será necessário usar informações coletadas em cada estudo incluído, como o tamanho do efeito da variável independente sobre as dependentes, o erro padrão e o tamanho da amostra. As análises variam de acordo com o tipo de variável em questão (nominal, ordinal ou contínua) e com a pergunta de pesquisa. A metanálise deve obter a síntese geral dos resultados dos estudos incluídos e os resultados por subgrupos de análise, que foram definidos no momento da elaboração do protocolo de pesquisa, isto é, antes da realização da coleta de dados.

Uma técnica multivariada, chamada metarregressão, pode ser usada para verificar, simultaneamente, os efeitos de múltiplos fatores sobre os resultados da metanálise. O conceito de metarregressão foi definido como "[…] um conjunto de procedimentos estatísticos (regressão linear logística ponderada/ não ponderada) para avaliar a heterogeneidade, na qual o tamanho de efeito de um estudo i é regredido a uma ou diversas covariáveis, com um valor definido para cada estudo I."[12] (Delgado-Rodríguez, 2001, 535). Essa técnica pode ser considerada uma extensão da análise de subgrupos, que avalia o efeito de variáveis contínuas ou categóricas sobre os resultados (Thompson & Sharp, 1999; Thompson & Higgins, 2002). Em sua essência, é similar às regressões simples, nas quais um modelo é construído para prever uma variável desfecho de acordo com uma ou mais variáveis explicativas (Higgins & Green, 2011).

[12] No original, "[…] a collection of statistical procedures (weighted/unweighted linear, logistic regression) to assess heterogeneity, in which the effect size of study i is regressed on one or several covariates, with a value defined for each study I."

2.4.6 Apresentação dos resultados

Os resultados de uma metanálise são apresentados em três partes: descrição dos estudos, qualidade dos estudos e resultado das análises estatísticas. Os dados são interpretados levando em consideração a qualidade dos estudos e o poder estatístico observado. Além do resultado geral, indicando uma estimativa pontual e um intervalo de confiança para o desfecho, deverão ser apresentadas as análises por subgrupos, quando conveniente, e as análises de sensibilidade.

Análises de sensibilidade são ferramentas para investigar o quanto os resultados encontrados na metanálise são sensíveis à metodologia que foi usada na revisão sistemática de literatura. Por exemplo, mudanças nos critérios de inclusão ou na forma de coletar os dados podem afetar os resultados? Essa investigação é fundamental para que se tenha segurança na aplicação prática dos resultados da metanálise (Higgins & Green, 2011). Assim, devem ser feitas análises levando em conta variáveis metodológicas, qualidade dos estudos e exclusão de *outliers*, discutindo-se a diferença entre os resultados encontrados nas análises de sensibilidade e o resultado geral apresentado na metanálise.

2.4.7 Atualização da revisão sistemática

A revisão sistemática, após concluída, não deve ser abandonada: precisa ser continuamente atualizada, acrescentando-se a ela os novos estudos publicados. Deve-se buscar incluir a revisão em bases de dados específicas, como por exemplo, as bases de dados da Colaboração Cochrane e da Colaboração Campbell, para evitar a duplicação de esforços por outros autores. A revisão sistemática de literatura, portanto, é considerada uma linha de pesquisa à qual o autor deverá dar continuidade.

2.4.8 Aplicação da PBE ao presente trabalho

No presente trabalho, a prevalência de depressão pós-parto (DPP) foi investigada em âmbito mundial, usando revisão sistemática de literatura. Para compreender de que forma esse procedimento foi realizado, é preciso conhecer o objeto de estudo e a literatura recente referente a ele. O conceito de DPP e as

características desse transtorno serão apresentados na seção seguinte (2.5 Depressão Pós-parto) do presente trabalho, usando-se, sempre que possível, revisões sistemáticas para fundamentar a discussão.

2.5 Depressão Pós-parto

A experiência de maternidade é culturalmente considerada como positiva para a mulher. No entanto, em pesquisa qualitativa, verificou-se que a vivência da chegada de um novo bebê representou para as famílias a realização do sonho da maternidade e paternidade, mas com rompimento das rotinas anteriores, desencadeando sentimentos ambíguos de alívio, alegria, angústia e ansiedade (Martins et al., 2008). O puerpério se configurou como um período de transição ao papel materno, podendo ser considerado uma crise de desenvolvimento que exige adaptação ao novo ser e conciliação do bebê fantasiado durante o período de gestação com o bebê real (Zagoneli, Martins, Pereira, & Athayde, 2003). Assim, sentimentos ambivalentes marcaram o período que se seguiu ao nascimento do bebê, mesmo em famílias consideradas saudáveis. Sabe-se que também é um período de vulnerabilidade, em que a mulher pode apresentar um espectro de transtornos psicóticos e não psicóticos (Austin, 2004), dentre os quais está incluído o Transtorno Depressivo Maior (TDM), que nesse período, é conhecido como depressão pós-parto.

A depressão tem sido considerada 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. A Organização Mundial de Saúde (OMS) estimou que o TDM foi a quarta maior causa de mortalidade no ano de 1990 (Murray & Lopez, 1996), e projeções para 2020 estimam que poderá ser a segunda maior causa de mortalidade no mundo e a primeira em países em desenvolvimento (Murray & Lopez, 1997). Mulheres têm risco cerca de duas a três vezes maior que o dos homens de apresentar TDM (Bromet et al., 2011). O TDM é causador de prejuízos à qualidade de vida e dificuldades de desempenhar normalmente as tarefas do dia-a-dia. Os gastos com TDM são expressivos, devido à perda da capacidade produtiva e ao investimento necessário nos serviços de saúde,

incluindo tratamentos ambulatoriais, hospitalizações e medicamentos (Greenberg, Stiglin, Finkelstein, & Berndt, 1993; Sobocki, Jönsson, Angst, & Rehnberg, 2006). No caso da DPP, além dos danos citados, também podem ocorrer consequências negativas para a criança e para a vida familiar e conjugal (Beck, 1998; Beck, 1999).

As pesquisas sobre DPP são fundamentais para o desenvolvimento e aprimoramento de intervenções para prevenção e tratamento, que podem se consolidar em políticas públicas, a partir de compreensão mais aprofundada desse fenômeno. Nas seções 2.5.1 a 2.5.8 do presente trabalho, serão apresentados o conceito de DPP e as mais relevantes fontes de evidências científicas sobre esse transtorno, sobretudo as revisões sistemáticas de literatura.

2.5.1 Depressão pós-parto: definição e características

O TDM é um transtorno de humor, de grau moderado a severo, cujos principais sintomas são humor deprimido e perda de interesse e prazer em quase todas as atividades (American Psychiatric Association, 1994). Outros sintomas são alterações do sono, apetite e peso, sensação de diminuição da energia, agitação ou retardo psicomotor, sentimentos de culpa, comportamento suicida (podendo incluir pensamentos, planos e tentativas) e dificuldade para pensar, concentrar-se ou tomar decisões. O diagnóstico de transtorno depressivo é feito quando são identificados, pelo menos, cinco desses sintomas, sendo que um deles deve ser, obrigatoriamente, tristeza ou diminuição de interesse e prazer em quase todas as atividades. Também é necessário que os sintomas permaneçam por, no mínimo, duas semanas e que causem sofrimento ou prejuízo significativo no funcionamento social ou ocupacional do indivíduo.

De acordo com critérios da *American Psychiatry Association* (APA, 1994), a DPP consiste em episódio depressivo ocorrido no período de até quatro semanas após o nascimento do bebê, quando se pode atribuir o especificador “com início no pós-parto” ao diagnóstico de TDM. Atualmente, a literatura tem usado uma definição mais ampla, considerando que a DPP faz parte de um espectro de transtornos psicóticos e não psicóticos, que podem incluir depressão maior ou menor, transtorno de ajustamento, transtorno do pânico, transtorno obsessivo-compulsivo e agorafobia. Além disso, observou-se que os sintomas de depressão após o parto encontram-se fortemente associados a sintomas depressivos na gravidez, sugerindo-se o uso da

nomenclatura “depressão perinatal” para abordar o período da gestação até um ano após o parto (Austin, 2004).

Considera-se que os sintomas de DPP não são diferentes dos sintomas de depressão ocorrida em outros momentos da vida da mulher (Austin, 2004; Schmidt, Piccoloto, & Müller, 2005). No entanto, na identificação do transtorno, pode ocorrer confusão de algumas características comuns do puerpério, como alteração de sono, alteração de apetite e fadiga, com sintomas de depressão (Pearlstein, Howard, Salisbury, & Zlotnick, 2009; Mathey, 2010).

O diagnóstico diferencial deve considerar as características de dois outros transtornos puerperais: melancolia pós-parto e psicose pós-parto. A melancolia pós-parto caracteriza-se por disforia, labilidade de humor, humor deprimido, choro, ansiedade, insônia, redução do apetite, irritabilidade e confusão (O’Hara, Zekoski, Philipps, & Wright, 1990; Murray & Cooper, 1997), ocorrendo nos primeiros dez dias após o nascimento da criança (Henshaw, 2003). O choro excessivo, que pode ocorrer de 50% a 80% das mães, não está necessariamente associado a sintomas de depressão, e algumas mulheres podem apresentar alegria e disforia ao mesmo tempo, com diversas mudanças no estado de humor ao longo do dia (Henshaw, 2003). A prevalência de melancolia pós-parto é relativamente elevada, e as estimativas variam entre 15% e 85% das mães (Pearlstein et al., 2009). Diferentemente da DPP, costuma haver remissão espontânea dos sintomas de melancolia em poucos dias (Schimidt et al., 2005), trazendo aparentemente pouca ou nenhuma consequência desfavorável para a mãe e para o bebê em longo prazo (Murray & Cooper, 1997), mas seu reconhecimento é importante por se tratar de fator de risco para DPP (Pearlstein et al., 2009).

A psicose pós-parto, por sua vez, caracteriza-se por sintomas como confusão mental, labilidade de humor, alucinações, delírios, paranoia e comportamento desorganizado, resultando em extremo prejuízo de funcionamento (Murray & Cooper, 1997; Pearlstein et al., 2009) e havendo o risco de a mãe machucar o bebê ou mesmo cometer infanticídio, em casos extremos (Schimidt et al., 2005; Pearlstein et al., 2009). O surgimento do transtorno costuma ser rápido, entre duas e quatro semanas após o parto. A prevalência tem sido estimada em cerca de 1 a cada 500 mães (Pearlstein et al., 2009), e o tratamento geralmente requer internação hospitalar (Murray & Cooper, 1997; Schimidt et al., 2005).

2.5.2 Métodos para identificar DPP

Os dois principais métodos usados em pesquisas para identificar DPP têm sido o diagnóstico por meio de entrevista clínica (Ex.: Moraes et al, 2006; Ballestrem, Strauß, & Kächele, 2005; Bantia et al., 2011) e a aplicação de escalas de autoavaliação (Ex.: Cruz, Simões, & Faisal-Cury, 2005; Santos, Martins, & Pasquali, 1999; Georgiopoulos et al., 1999; Josefsson, Berg, Nordin, & Sydsjö, 2001; Faisal-Cury, Tedesco, Kahhale, Menezes, & Zugaib, 2004; Zaconeta, Motta, & França, 2004; Ruschi et al., 2007; Silva, 2008). Entrevistas clínicas implicam em dificuldades de ordem prática: é necessária a participação de um profissional de saúde mental devidamente treinado e um tempo maior para a aplicação dos instrumentos, o que torna o trabalho de pesquisa mais custoso. Assim, escalas de autoavaliação são usadas com maior frequência em pesquisas.

Essas escalas fornecem pontuações que indicam a intensidade dos sintomas depressivos. Pontos de corte são definidos em estudos de validação, em que as pontuações obtidas por uma amostra de participantes são comparadas com um padrão-ouro, ou seja, uma avaliação considerada mais fidedigna (ex.: diagnóstico de um psiquiatra cego para as pontuações da escala). A validação busca determinar o ponto de corte ideal para que a escala tenha alta sensibilidade[13], especificidade[14], valor preditivo positivo[15] e valor preditivo negativo[16], ou seja: o mínimo possível de falsos positivos e falsos negativos. No entanto, considera-se que a tradução das escalas para outros idiomas, somada às variáveis culturais, podem afetar a pontuação obtida pelos respondentes. Por isso, costumam ser feitas validações da

---

[13] Sensibilidade é a proporção de casos positivos verdadeiros, que indica qual é a capacidade do teste para detectar pessoas com o problema estudado (Greenhalgh, 2008). No caso de escalas de avaliação de DPP, é a proporção de parturientes consideradas deprimidas segundo o padrão-ouro que também foram identificadas como deprimidas pelo resultado da escala.

[14] Especificidade é proporção de casos negativos verdadeiros, que indica qual é a capacidade do teste para excluir corretamente as pessoas sem o problema estudado (Greenhalgh, 2008). No caso de escalas de avaliação de DPP, é a proporção de parturientes consideradas não deprimidas segundo o padrão-ouro que também foram identificadas como não deprimidas pelo resultado da escala.

[15] Valor preditivo positivo é a probabilidade pós-teste de um teste positivo, ou seja, a probabilidade de que a pessoa tenha o problema caso tenha tido um resultado positivo no teste (Greenhalgh, 2008).

[16] Valor preditivo negativo é a probabilidade pós-teste de um teste negativo, ou seja, a probabilidade de que a pessoa não tenha o problema caso tenha tido um resultado negativo no teste (Greenhalgh, 2008).

escala em diferentes países, que procuram determinar o ponto de corte mais adequado para o seu contexto.

Do ponto de vista clínico, nenhum inventário ou escala substitui a avaliação realizada por profissional de saúde mental, mas instrumentos de autoavaliação podem ser usados por psicólogos, médicos, enfermeiros, assistentes sociais, agentes comunitários de saúde e outros profissionais para rastrear as mães que têm maior chance de apresentar sintomas depressivos, encaminhando-as para avaliação e, se necessário, tratamento. Por serem instrumentos de rastreamento, o ponto de corte costuma ser calculado de forma a identificar o máximo possível de mães com DPP, ou seja: os falsos positivos são mais aceitáveis que os falsos negativos. Assim, existe uma tendência de superestimar a prevalência nas pesquisas que usam apenas instrumentos de rastreamento como critério para identificar DPP.

A Escala de Depressão Pós-Parto de Edimburgo (EDPE – Cox, Holden, & Sagovsky, 1987) é a escala mais usada nos estudos sobre prevalência de DPP. Sua popularização se deu pela facilidade e agilidade na aplicação e disponibilidade de estudos de validação em diversos países. Trata-se de um inventário de dez itens, cada um apresentando quatro possibilidades de resposta, que indicam a intensidade dos sintomas (pontuadas de zero a três). A respondente deve escolher a resposta que mais se aproxima da forma como se sentiu ao longo da última semana. Ao final da aplicação, somam-se as pontuações dos itens da escala, obtendo-se um valor de zero a 30 pontos, que é usado para classificar as participantes em potencialmente deprimidas ou não deprimidas de acordo com o ponto de corte definido no estudo de validação.

A versão original da EDPE (Cox, Holden, & Sagovsky, 1987) foi validada em amostra de parturientes de Livingston e Edimburgo, tomando como padrão-ouro os resultados de entrevista estruturada (*Goldberg's Standardised Psychiatric Interview*). Concluiu-se que o ponto de corte mais adequado foi ≥13, para o qual a escala apresentou sensibilidade de 86%, especificidade de 78%, valor preditivo positivo de 73% e valor preditivo negativo de 88%[17]. Estudos sobre a prevalência de DPP usaram a EDPE com ponto de corte entre 9 e 15 (Melo Junior, 2011), variando de

[17] O valor preditivo negativo não foi apresentado por Cox, Holden e Sagovsky (1987), mas foi calculado pela autora do presente estudo com base em outros dados disponíveis no artigo.

acordo com os objetivos do trabalho e com resultados da validação da escala realizada na região.

A EDPE foi traduzida para o português e teve sua validade estudada no Brasil pela primeira vez por Santos, Martins e Pasquali (1999) em uma amostra de puérperas selecionadas por conveniência na cidade de Brasília, usando o mesmo padrão-ouro definido por Cox, Holden e Sagovsky (1987). A maior parte das participantes era de classe média ou alta. Concluiu-se que o ponto de corte ≥12 seria o mais adequado para a amostra brasileira, encontrando-se sensibilidade de 72%, especificidade de 88%, valor preditivo positivo de 78% e valor preditivo negativo de 85%.

Posteriormente, Santos et al. (2007) fizeram nova tradução e validação da EDPE, em uma amostra aleatória de puérperas de Pelotas (RS). O padrão-ouro foi entrevista psiquiátrica semiestruturada, baseada na Classificação de Transtornos Mentais e de Comportamento da Classificação Internacional de Doenças (CID-10). Concluiu-se que o melhor ponto de corte para rastreamento foi ≥ 10, apresentando sensibilidade de 82,6% e especificidade de 65,4%.

Um terceiro estudo (Figueira, Corrêa, Malloy-Diniz, & Romano-Silva, 2009) de validade da EDPE em amostra de mães brasileiras foi realizado em Belo Horizonte (MG), usando a versão traduzida para o português por Santos, Martins e Pasquali (1999) e tendo como padrão-ouro entrevista psiquiátrica usando o Mini-Plus (instrumento baseado no DSM-IV). Concluiu-se que o melhor ponto de corte para rastreamento foi ≥11, apresentando sensibilidade de 72,7%, especificidade de 95,5%, valor preditivo positivo de 85% e valor preditivo negativo de 90%.

As diferenças entre os resultados dos três estudos de validação da EDPE no Brasil (Santos et al., 1999; Santos et al., 2007; Figueira et al., 2009) podem ter ocorrido devido a variações metodológicas, como método de amostragem, características das participantes (renda, escolaridade, estado civil, idade) e variações na tradução da escala, pois a versão em português usada por Santos et al. (1999) e por Figueira et al. (2009) é ligeiramente diferente da versão usada por Santos et al. (2007). O padrão-ouro usado para comparação com a EDPE também não foi o mesmo nos três estudos: Santos et al. (1999) aplicaram a *Clinical Interview Schedule* (CIS), Santos et al. (2007) usaram critérios do CID-10 e Figueira et al. (2009) usaram a Mini-Plus 5.0. Além das prováveis influências das características

metodológicas dos trabalhos de validação, podem existir diferenças culturais entre participantes das diferentes cidades (Brasília, Belo Horizonte e Pelotas).

Os métodos usados para diagnosticar DPP têm sido criticados, sugerindo-se que a prevalência desse transtorno, provavelmente, tem sido superestimada (Mathey, 2010). Os principais fatores que levariam a esse viés seriam: 1. identificação de todas as mulheres com pontuações positivas nas escalas de rastreamento como deprimidas, sendo que cerca de 50% das mulheres que obtém pontuação maior ou igual a 13 na EDPE não apresentam DPP; 2. ausência de distinção entre sentimentos passageiros de ansiedade ou depressão e sintomas permanentes; 3. uso de pontos de corte incorretos para os instrumentos de autoavaliação; 4. os critérios do DSM-IV para episódio depressivo contêm sintomas que podem ser considerados estados normais em mães e pais, como por exemplo: perda de peso em puérperas, alterações no sono devido aos cuidados com o bebê, fadiga, dificuldade de concentração (decorrente de fadiga e alteração no sono); 5. o rastreamento de mulheres com risco de DPP pode exagerar a validade preditiva dos fatores de risco, identificando que quase todas as mulheres apresentam riscos. A possibilidade de que a prevalência de DPP possa ter sido superestimada não é pequena, visto que dados epidemiológicos indicam que esse transtorno ocorreu com frequência relativamente elevada, conforme será apresentado na próxima seção do presente trabalho.

2.5.3 Epidemiologia

Na área de epidemiologia, o conhecimento dos conceitos de prevalência e incidência é fundamental. Prevalência é definida como “A proporção de uma população que apresenta uma condição ou característica particular: por ex., a porcentagem de pessoas em uma cidade com uma doença específica, ou que fumam”[18] (The Cochrane Collaboration, 2005, 34). Incidência, por sua vez, é definida como “O número de novas ocorrências de algo na população durante um determinado período de tempo, ex.: o número de casos de uma doença em um país

[18] No original: The proportion of a population having a particular condition or characteristic: e.g. the percentage of people in a city with a particular disease, or who smoke.

durante um ano”[19] (The Cochrane Collaboration, 2005, 21). De forma sucinta, a prevalência de uma determinada condição é a proporção de pessoas que apresentam essa condição em um determinado instante, enquanto a incidência se refere ao número de novos casos de uma condição na população em um determinado intervalo de tempo. Outra maneira de compreender esses conceitos é considerar que a prevalência mede quantas pessoas estão doentes, enquanto a incidência mede quantas pessoas se tornaram doentes (Pereira, Paes, & Okano, 2000). Por fim, pode-se fazer uma analogia em que a prevalência seria a fotografia da população, enquanto a incidência seria o filme.

As prevalências de DPP em diferentes estudos foram sintetizadas, de forma pioneira, pela metanálise de O'Hara e Swain (1996). O trabalho reuniu dados de 59 pesquisas, totalizando 12.810 participantes, que usaram amostras selecionadas de forma aleatória ou quasi-aleatória. Todos os estudos incluídos foram realizados em países desenvolvidos. Foi encontrada prevalência média de 12,8% (IC 95% 12,3%-13,4%), variando de acordo com o instrumento de avaliação usado para identificar DPP, sendo que a menor prevalência média foi obtida com avaliações baseadas no DSM: 7,2% (IC 95% 3,7%-10,7%) e a maior, com avaliações baseadas no critério de Goldberg: 17,8% (IC 95% 15,8%-19,8%). A EDPE foi usada em 12 estudos, resultando em prevalência média de 12,0% (IC 95% 10,9%-13,1%). No entanto, essa revisão foi publicada há mais de 15 anos e os dados não foram atualizados. Também foram observadas deficiências no relato, que indicaram possíveis limitações metodológicas: não foram mencionadas a estratégia de busca, as bases de dados consultadas e a qualidade dos estudos incluídos.

O'Hara e Swain (1996) investigaram se havia diferenças na prevalência relacionadas a fatores metodológicos: país no qual o estudo foi realizado, método de definição da DPP (por meio de autorrelato ou por entrevista clínica), tamanho do intervalo de tempo no qual a depressão foi avaliada (ex.: se um estudo aplicou uma determinada escala em um período de 2 a 6 semanas após o parto, então o intervalo seria de 4 semanas ou 28 dias) e tempo após o parto no qual o instrumento foi aplicado (ex.: quatro semanas após o parto). Os resultados mostraram que trabalhos que usaram escalas baseadas em autorrelato resultaram em prevalência maior que

[19] No original: The number of new occurrences of something in a population over a particular period of time, e.g. the number of cases of a disease in a country over one year.

aqueles que diagnosticaram depressão através de entrevista clínica. Uma análise de regressão mostrou que a variável que mais contribuiu para a variância nas estimativas de prevalência foi o tamanho do intervalo de tempo no qual a depressão foi avaliada (aproximadamente 17% da variância). Como era esperado, quanto maior o período em avaliação, maior a prevalência de DPP. A forma de diagnóstico (autorrelato ou entrevista clínica) também entrou no modelo final, contribuindo para explicar mais 9% da variância. O país e o número de dias após o parto em que a depressão foi avaliada não foram estatisticamente significativos para predizer a variância.

Gavin et al. (2005) fizeram outra metanálise, incluindo apenas estudos realizados em países desenvolvidos, publicados em inglês e que usaram entrevista clínica como critério de diagnóstico (excluindo, portanto, estudos que usaram apenas instrumentos de autoavaliação). Chegaram à prevalência de 1,0% a 5,9% para depressão maior e 6,5% a 12,9% para depressão maior e menor.

No ano seguinte, Halbreich e Karkun (2006) publicaram outra revisão sistemática sobre prevalência de DPP, com abordagem transcultural. Os estudos incluídos foram realizados em 40 países desenvolvidos e em desenvolvimento. A DPP foi menos prevalente em países como Singapura, Malta, Dinamarca e Malásia (0,5% a 9%) e mais prevalente em países como Guyana, Costa Rica, Itália, Chile, Coreia e Taiwan (34% a 57%). Apenas um estudo foi realizado no Brasil (Da-Silva, Moraes-Santos, Carvalho, Martins, & Teixeira, 1998), encontrando-se prevalência de 42,8%. Com base nos resultados, foi contestada a afirmação, frequentemente encontrada em introduções de artigos sobre DPP, de que a prevalência desse transtorno seria de 10% a 15%. Dada a ampla faixa de resultados encontrada pelas diferentes pesquisas, essa estimativa não seria representativa da prevalência global. Nessa revisão sistemática, não foi usada metanálise para sintetizar os resultados de prevalência, devido à heterogeneidade entre os estudos. As diferenças entre os trabalhos foram classificadas em três níveis: conceitual (definição de DPP e diagnóstico diferencial em relação a outros transtornos comuns no puerpério), transcultural (as diferenças de interpretação de sentimentos e sintomas relativos ao nascimento do bebê entre as culturas não estariam, necessariamente, contempladas pelos instrumentos de avaliação de DPP) e metodológico (constatou-se necessidade de se desenvolver e usar instrumentos adaptados às diferentes culturas; realizar

estudos prospectivos iniciando durante a gestação e acompanhando as participantes durante o puerpério; e estabelecer distinção entre prevalência e incidência de DPP).

A revisão sistemática de Sawyer et al. (2010) teve como foco estudos sobre bem-estar psicológico no período perinatal, no continente africano. Foram encontrados 21 estudos, realizados em seis países: África do Sul, Etiópia, Gambia, Marrocos, Nigéria e Uganda. As prevalências pontuais de DPP variaram entre 3,2% em mães de bebês de 6 a 18 meses na Gambia e 48% em mães de bebês de uma semana, nascidos pré-termo na África do Sul. A prevalência média foi estimada em 18,3% (IC 95% 17,6%-19,1%), valor superior ao encontrado em metanálise de estudos de prevalência de DPP realizados em países desenvolvidos (Gavin et al., 2005).

Vigod, Villegas, Dennis e Ross (2010) investigaram a prevalência de DPP em mães de recém-nascidos pré-termo através de revisão sistemática, incluindo 26 estudos. Os autores pretendiam usar metanálise, mas não foi possível devido à heterogeneidade, de forma que os dados foram apresentados de forma qualitativa. As estimativas pontuais de prevalência variaram de zero a 44%. Essa variação foi tão ampla, que os resultados dificilmente poderiam ser usados na formulação de políticas de saúde. Assim, a falta de padronização de metodologias dos estudos primários levou a resultados com baixo potencial de contribuição para o avanço da ciência e da prática profissional.

Villegas et al. (2011) investigaram, através de metanálise, a prevalência de DPP em comunidades rurais de países desenvolvidos e em desenvolvimento, selecionando estudos em inglês ou francês, que incluíssem uma medida padronizada de DPP (por autoavaliação ou entrevista clínica) no primeiro ano após o parto. Foram incluídos 17 trabalhos, sendo sete de países em desenvolvimento. A metanálise indicou prevalência média global de 27% (IC 95% 18,8%-37,2%), sendo 21,5% em países desenvolvidos (IC 95% 10,9%-38,0%) e 31,3% em países em desenvolvimento (IC 95% 21,3%-43,5%). Apesar da diferença de cerca de dez pontos percentuais na estimativa pontual, a sobreposição de intervalos de confiança não permitiu afirmar que a prevalência de DPP diferiu estatisticamente entre os dois grupos. Os dados dos poucos estudos que compararam a prevalência de DPP entre mulheres de zona rural e urbana sugeriram que moradoras da zona rural podem ter risco aumentado de desenvolver DPP.

Trabalhos de autores brasileiros também têm usado revisão sistemática para localizar e sintetizar estudos sobre DPP. A tese de Melo Junior (2011) realizou metanálise de estudos nacionais e internacionais sobre prevalência de DPP. Foram incluídos 102 trabalhos, totalizando 120.936 mulheres, e os resultados indicaram prevalência média de 12,61% (IC 99% 12,36-12,86). Não foram fornecidas informações detalhadas sobre o método de avaliação de qualidade dos estudos incluídos. Os dados indicaram não haver diferenças estatisticamente significativas entre as prevalências avaliadas com instrumentos de autoavaliação ou de entrevista clínica, mas observou-se que a prevalência de DPP foi maior em trabalhos com amostras menores e com pontos de corte mais baixos nas escalas de rastreamento. Os resultados também mostraram que houve diferença entre a prevalência média em países desenvolvidos (11,22%, IC 99% 10,97%-11,47%) e em desenvolvimento (20,09%, IC 99% 19,28%-20,90%).

Lobato et al. (2012) realizaram revisão sistemática de estudos de prevalência e/ou incidência de DPP em mães brasileiras, publicados em revistas nacionais ou internacionais. A busca foi feita nas bases de dados Lilacs e SciELO, incluindo estudos publicados até 16 de agosto de 2010, bem como busca secundária nas listas de referências dos trabalhos encontrados. Não foram avaliados artigos que estavam indisponíveis em formato digital nas bibliotecas nacionais. Foram incluídos 14 estudos, sendo 13 de prevalência e um de incidência de DPP. As estimativas de prevalência de DPP variaram entre 7,2% e 39,4%, com a maior parte dos trabalhos indicando valores entre 15% e 28%. Lobato et al. (2012) chamaram a atenção para a heterogeneidade de metodologias, que dificultou a síntese dos resultados em uma única estimativa de DPP através de metanálise. Além disso, foi constatado que poucos trabalhos usaram técnicas adequadas para estimar a prevalência, como amostra aleatória de base populacional (três estudos) e aplicação de entrevista clínica para identificação de DPP (três estudos). Ainda assim, foi possível verificar que os níveis de prevalência brasileiros foram similares aos obtidos em países com nível de desenvolvimento socioeconômico semelhante, como Itália, Chile, África do Sul, Índia e Turquia, e maiores que os encontrados em países ricos. Algumas limitações dessa revisão são: ausência de segundo revisor (o procedimento para aumentar a confiabilidade foi replicar a seleção de estudos, de forma cega, pelo mesmo revisor) e ausência de avaliação da qualidade metodológica dos trabalhos incluídos.

Uma síntese das revisões sistemáticas sobre prevalência de DPP é apresentada na Tabela 4, a seguir. De forma geral, as metanálises sobre prevalência de transtorno depressivo maior ou menor, para mães com bebês nascidos a termo, concluíram que a prevalência média é de cerca de 12% (O'Hara & Swain, 1996; Gavin et al., 2005; Melo Junior, 2011). Com exceção de Melo Junior (2011) e Lobato et al. (2012), que são brasileiros, as revisões sistemáticas de literatura não usaram bases de dados latino-americanas em sua estratégia de busca. Além disso, a maior parte dos trabalhos usou poucas bases eletrônicas, não esgotando as fontes de informação. O impacto do método de avaliação de sintomas depressivos (autoavaliação ou avaliação pelo observador) sobre os resultados se mostrou controverso, pois um estudo (O'Hara & Swain, 1996) apontou diferenças na prevalência média de acordo com o instrumento usado, enquanto outro (Melo Junior, 2011) não encontrou essa diferença.

Nas buscas de revisões sistemáticas sobre prevalência e/ou incidência de DPP, foi encontrada uma *overview*, isto é, um estudo que realiza busca sistemática de revisões sistemáticas sobre um determinado assunto, analisando-as criticamente. Esse trabalho (Mann, Gilbody, & Adamson, 2010) incluiu cinco revisões e identificou deficiências na qualidade metodológica das mesmas, bem como quantidade pequena de estudos primários incluídos, características que impõem limitações às evidências encontradas sobre a epidemiologia da DPP. Foram apresentadas críticas quanto ao viés de seleção, visto que poucos estudos primários de países em desenvolvimento foram incluídos, impedindo a generalização dos resultados para esses contextos. Concluiu-se que as revisões sistemáticas existentes são de baixa qualidade e, portanto, o conhecimento sobre incidência e prevalência de DPP, atualmente, é pobre e insuficiente para informar autoridades em políticas de saúde pública, clínicos e outros profissionais.

Tabela 4 – Síntese de metanálises sobre prevalência de DPP (continua)

Autores, ano	Critérios de inclusão	Bases de dados incluídas na estratégia de busca	Total de estudos incluídos	Prevalência média
O'Hara e Swain, 1996	• Declarou relação estatística entre a variável de interesse (fatores de risco) e a depressão pós-parto • Variável de interesse avaliada durante a gestação ou no parto • Participantes recrutadas por técnicas de amostragem aleatória ou quasi-aleatória • Depressão avaliada pelo menos duas semanas após o parto • DPP avaliada por um método validado ou padronizado	Não declarou	59	12,8% (IC 95% 12,3%-13,4%)
Gavin et al., 2005	• Relatou dados originais • Publicado em inglês • Publicado entre janeiro de 1980 e março de 2004 • Realizado em país desenvolvido • Avaliou depressão em mulheres durante na gestação ou até um ano após o parto • Confirmou diagnóstico através de avaliação clínica ou entrevista estruturada • Exclusão de estudos em que não fosse possível distinguir resultados dos desfechos de interesse (depressão maior ou menor) dos relativos a outros transtornos.	• MEDLINE • Cumulative Index to Nursing & Allied Health Literature (CINAHL) • PsycINFO • Sociofile • Buscas secundárias nas referências bibliográficas dos estudos incluídos • Consultas a especialistas	28	1,0% a 5,9% para depressão maior; 6,5% a 12,9% para depressão maior e menor

Tabela 4 – Síntese de metanálises sobre prevalência de DPP (continuação)

Autores, ano	**Critérios de inclusão**	**Bases de dados incluídas na estratégia de busca**	**Total de estudos incluídos**	**Prevalência média**
Halbreich e Karkun, 2006	• Publicado de 1980 a 2005 • Avaliou depressão até um ano após o parto • Excluídos estudos de mulheres que já estavam referenciadas para clínicas psiquiátricas	• MEDLINE • PsycINFO • CINAHL • EMBASE • Buscas secundárias nas referências bibliográficas dos estudos incluídos	Cerca de 140	Não declarou
Sawyer et al., 2010	• Coleta de dados na África • Relatou estimativas de prevalência pontual de transtornos psicológicos durante a gestação ou após o parto • Publicado em inglês • Excluídos estudos de intervenção que só incluíram medidas de DPP pós-intervenção	• MEDLINE • PsycINFO • PsycArticles APA • PubMed • Web of Science • Scopus • African Index Medicus • Buscas secundárias nas referências bibliográficas dos estudos incluídos e de revisões de literatura	35 (21 sobre DPP, os demais sobre outros transtornos mentais no puerpério)	18,3% (IC 95% 17,6%-19,1%)

Tabela 4 – Síntese de metanálises sobre prevalência de DPP (continuação)

Autores, ano	**Critérios de inclusão**	**Bases de dados incluídas na estratégia de busca**	**Total de estudos incluídos**	**Prevalência média**
Vigod et al., 2010	• Artigo publicado e com revisão de pares • Publicado em inglês ou francês • Definição de nascimento pré-termo como menos de 38 semanas de idade gestacional ou definição de baixo peso ao nascer como peso menor que 2500 gramas no nascimento • Uso de medida padronizada de depressão (instrumento de autoavaliação ou preenchido por observador) • Relatou prevalência de DPP ou pontuação média em escala padronizada e com ponto de corte validado • Excluídos estudos que limitaram a amostra a mães com gestações múltiplas	• MEDLINE • CINAHL • EMBASE • PsycINFO • The Cochrane Library • Buscas secundárias nas referências bibliográficas dos estudos incluídos e de revisões de literatura	26	Não declarou
Villegas et al., 2011	• Artigo publicado e com revisão de pares • Publicado em inglês ou francês • Uso de método padronizado para avaliar DPP • Avaliação de DPP realizada no primeiro ano após o parto	• PsycINFO • MEDLINE • CINHAL • EMBASE • The Cochrane Library	17	27% (IC 95% 18,8%-37,2%)

Tabela 4 – Síntese de metanálises sobre prevalência de DPP (conclusão)

Autores, ano	Critérios de inclusão	Bases de dados incluídas na estratégia de busca	Total de estudos incluídos	Prevalência média
Melo Junior, 2011	• Publicado de janeiro de 2000 a junho de 2010 • Delineamento transversal, de coorte, caso-controle, ensaio clínico aleatorizado ou análise de banco de dados, com informação sobre instrumento usado para a coleta dos dados • Avaliou prevalência de depressão na gestação ou DPP • Excluídos estudos com foco apenas em mães adolescentes, HIV positivas ou pertencentes a minorias étnicas • Excluídos estudos em que faltavam informações sobre delineamento, com amostras menores que 150, ou que incluíram participantes com diagnóstico prévio de depressão.	• MEDLINE • EMBASE • ISI Web of Knowledge • CINAHL • SciELO • Buscas secundárias nas referências bibliográficas dos estudos incluídos	102	12,61% (IC 99% 12,36-12,86)
Lobato et al., 2012	• Artigo cujo título e/ou resumo sugerisse a presença, no texto principal, de resultados originais sobre a prevalência ou incidência de DPP na população de estudo. • Excluídos artigos que estavam indisponíveis em formato digital nas bibliotecas nacionais.	• LILACS • SciELO • Buscas secundárias nas referências bibliográficas dos estudos incluídos	14	Não declarou

Foi encontrada apenas uma metanálise sobre incidência (isto é, número de casos novos) de DPP, a qual sintetizou estudos realizados em países desenvolvidos que aplicaram entrevista clínica como método de diagnóstico. Nos três primeiros meses após o parto, 14,5% das mães apresentaram novos casos de episódio depressivo maior ou menor nos três primeiros meses após o parto (Gavin et al., 2005). Considerando apenas episódio depressivo maior, esse valor foi de 6,5%. Os resultados podem não ser generalizáveis para populações de países em desenvolvimento, que não foram incluídas nessa revisão. Uma revisão sistemática brasileira (Lobato et al., 2012) localizou apenas um artigo sobre esse assunto, cuja estimativa foi de 42,8%, resultado que pode não ser representativo da verdadeira incidência do transtorno, pois o tamanho da amostra foi restrito (n=21) e ocorreram muitas perdas de seguimento (27,6%). Por existirem poucas revisões sistemáticas sobre esse assunto, o estudo de coorte longitudinal de Bantia et al. (2011) pode ser considerado uma boa evidência, visto que avaliou incidência de DPP em amostra de grande porte, formada por 1.066 mulheres recrutadas na região da Toscana (Itália) e acompanhadas do terceiro mês de gestação até o 12º mês após o parto. A incidência cumulativa[20] de DPP em mães que não estavam deprimidas durante a gestação foi de 6,8% (IC 95% 4,6% a 9,2%) novos casos em um ano.

2.5.3.1 Prevalência de DPP: possíveis diferenças entre países desenvolvidos e em desenvolvimento

A ideia de que a DPP seria mais prevalente entre países em desenvolvimento tem sido difundida em estudos internacionais (ex.: Walker et al., 2007; Sawyer et al., 2010; Villegas et al., 2011; Melo Junior, 2011; Lobato et al., 2012). No entanto, alguns pesquisadores têm observado grande divergência entre resultados, afirmando, por exemplo, que com base nos resultados de estudos anteriores, "*as taxas de prevalência em países em desenvolvimento variam de valores iguais a quase o dobro das verificadas em países desenvolvidos*"[21] (Husain et al., 2006, p. 197). Em verdade, a observação dos resultados dos trabalhos existentes é

[20] Foi calculada a taxa cumulativa de ocorrência de novos episódios depressivos entre o parto e o 12º mês após o parto, dividida pelo número total de participantes que completaram a avaliação realizada no oitavo mês após o parto e não apresentaram depressão durante a gestação.

[21] No original: "*prevalence rates in the developing world range from being equal to almost double that of developed countries*".

insuficiente para que se verifique, em uma revisão narrativa, se as diferenças nas prevalências de DPP se devem às características culturais dos países, ao perfil das participantes ou a fatores metodológicos.

A importância de se levar em consideração a metodologia se torna evidente nas conclusões de um estudo realizado em Recife (Cantilino et al., 2010), que encontrou prevalência de DPP de 7,2% (IC 95% 4.7%-9.8%). Comparando esse valor com o comumente relatado em estudos internacionais, os autores consideraram que a ideia de que a prevalência de DPP é maior no Brasil do que em outros países seria um mito, e que a prevalência de DPP no Brasil seria equivalente à de países com prevalências consideradas moderadas, caso o instrumento usado para identificar episódio depressivo fosse o SCID-I (entrevista clínica com base nos critérios do DSM-IV).

Uma metanálise brasileira (Melo Junior, 2011) verificou que a prevalência média de DPP em países em desenvolvimento foi de 20,09% (IC 99% 19,28-20,90), quase o dobro da encontrada em países desenvolvidos: 11,22% (IC 99% 10,97-11,47). Outra metanálise da prevalência de DPP encontrou prevalência de 31,3% (IC 95% 21,3-43,5) em zonas rurais de países em desenvolvimento e 21,5% (IC 95% 10,9-38,0) em zonas rurais de países desenvolvidos (Villegas et al., 2011), mostrando uma importante diferença entre as estimativas pontuais, mas com sobreposição entre os intervalos de confiança. Ambas as metanálises não levaram em consideração a possibilidade de que aspectos metodológicos dos estudos incluídos atuariam como variáveis de confusão.

Uma pesquisa transcultural sobre prevalência de DPP (Affonso, De, Horowitz, & Mayberry, 2000), que aplicou o mesmo protocolo de coleta de dados em nove países, verificou que a prevalência de DPP foi menor na Europa Ocidental (Suécia, Finlândia e Itália) e na Austrália, intermediária nos Estados Unidos da América e maior na América do Sul (Guiana) e países da Ásia (Índia, Coreia do Sul e Taiwan). Os autores concluíram que as menores prevalências de DPP na Austrália e países da Europa ocidental podem ser resultado de programas de rastreamento e intervenção em saúde mental, e que as prevalências mais altas na Ásia e América do Sul podem ser causadas pelo menor reconhecimento da DPP como um transtorno de saúde importante. No entanto, as amostras foram selecionadas por conveniência e não foi discutido o possível efeito de diferenças de perfil das participantes em cada país (ex.: a amostra da Guiana tinha a menor escolaridade,

enquanto a da Austrália tinha a maior), nem as implicações de diferentes formas de coleta de dados adotadas (entrevista ou envio do questionário pelo correio), de forma que não fica claro se pode ter ocorrido viés na seleção de participantes.

2.5.4 Fatores de risco

Fatores de risco para DPP foram investigados de forma pioneira pela metanálise de O'Hara e Swain (1996), a mesma que compilou as prevalências de DPP em diferentes estudos. Foram significativos fatores como: renda familiar baixa, ocupação da mãe, complicações no parto, eventos de vida negativos, dificuldades no ajustamento conjugal, falta de apoio social, falta de apoio do pai do bebê e histórico de transtorno psiquiátrico. Outros fatores investigados não se mostraram significativamente associados com DPP, como: idade da mãe, nível de escolaridade, estado civil, duração do relacionamento, trabalhar na gestação, número de filhos, paridade e ter familiar com histórico de transtorno psiquiátrico.

Publicada no mesmo ano, outra metanálise (Beck, 1996) incluiu 44 estudos, procurando avaliar a magnitude do efeito das seguintes variáveis sobre a DPP: depressão durante a gestação, histórico de depressão, apoio social, estresse em eventos de vida, estresse nos cuidados com o bebê, melancolia pós-parto, satisfação conjugal e ansiedade durante a gestação. A atualização desse trabalho (Beck, 2001) incluiu 84 estudos e identificou quatro fatores de risco adicionais: autoestima, estado civil, nível socioeconômico e gestação não planejada ou não desejada.

A relação entre variáveis medidas na gestação e a ocorrência de DPP foi investigada por Robertson, Grace, Wallington e Stewart (2004), que concluíram que depressão, ansiedade, eventos de vida negativos, apoio social insuficiente e histórico de depressão (anterior à gestação) apresentaram associação forte a moderada com DPP; neuroticismo e relacionamento conjugal insatisfatório apresentaram associação moderada; e baixo nível socioeconômico e complicações obstétricas apresentaram associação leve com esse transtorno. O artigo deixou de apresentar informações importantes sobre a metodologia, como estratégia de busca, quais bases de dados foram consultadas, quantos estudos foram incluídos e qual era o nível de qualidade metodológica dos mesmos.

Para investigar se haveria associação entre cesariana e DPP, Carter, Frampton, e Mulder (2006) fizeram revisão sistemática de publicações em inglês. Sintetizando os resultados dos 24 estudos incluídos, não foi possível encontrar evidências de que parto cesáreo e DPP estivessem associados.

Revisão sistemática também foi usada para investigar o papel de elementos culturais sobre a DPP. Wong e Fisher (2009) avaliaram o efeito das práticas de confinamento, que são rituais comuns entre mulheres chinesas no período do puerpério, sobre a prevalência de DPP. Essas práticas incluem mudanças de comportamento, como não tomar banho ou lavar o cabelo, não sair de casa, evitar comidas cruas ou frias, comer comidas quentes e abster-se de relações sexuais. As mulheres que aderem a esse costume, em geral, permanecem reclusas por um mês após o parto e recebem auxílio de parentes de sexo feminino para as tarefas de casa, preparação de alimentação especial e cuidados com o bebê. Essa revisão sistemática não usou metanálise para sintetizar estatisticamente os resultados, mas apresentou análise descritiva. Não foi possível determinar se as práticas de confinamento apresentaram, de fato, o papel de fator de proteção para DPP.

Recentemente, foram publicadas duas revisões sistemáticas que investigaram a associação entre violência e DPP. Uma delas incluiu sete estudos, concluindo que mulheres que foram agredidas pelo parceiro íntimo tiveram risco de 1,5 a duas vezes maior de desenvolver DPP, quando comparadas às que não estiveram expostas a essa forma de violência (Beydoun, Beydoun, Kaufman, Lo, & Zonderman, 2012). A outra incluiu seis artigos que compararam mulheres expostas à violência com as não expostas e encontraram razão de chance de 3,47 (IC 95% 2,13-5,64), indicando que a violência foi um fator de risco para DPP (Wu, Chen, & Xu, 2012).

Países latino-americanos estiveram pouco representados nas revisões sistemáticas internacionais voltadas à identificação de fatores de risco para DPP, apesar de existirem diversos estudos que poderiam ser incluídos (ex.: Hasbún et al., 1999; Coutinho, Baptista, & Morais, 2002; Cruz et al., 2005; Moraes et al., 2006; Urdaneta et al., 2011). Algumas explicações possíveis seriam a data de publicação das revisões, pois a mais recente é de 2004 e a maior parte dos estudos latino-americanos sobre DPP foi 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, Psicodoc e Redalyc.

Recentemente, foi publicada uma revisão sistemática sobre fatores de risco para DPP, realizada por autora brasileira (Aliane, Mamede, & Furtado, 2011), em que a busca foi feita nas bases de dados Pubmed, SciELO e Lilacs e foram admitidos estudos publicados em inglês, espanhol ou português. Foram incluídos 35 artigos (30 deles em língua inglesa), os quais identificaram 70 fatores de risco, que foram agrupados nas seguintes categorias: genéticos, hormonais, obstétricos, psicológicos/psiquiátricos, saúde do bebê, saúde materna, socioeconômico/cultural e suporte social/relações interpessoais. Os fatores mais citados foram: histórico de depressão, depressão durante a gestação, baixo suporte social, insatisfação com o marido ou com o relacionamento conjugal e ansiedade durante a gestação. Por não ter sido feita metanálise, não foi possível avaliar o tamanho de efeito da associação desses fatores com a DPP.

Em síntese, revisões sistemáticas e metanálises têm sido realizadas de forma exploratória para identificar quais são os fatores de risco para DPP, de acordo com a ampla literatura existente sobre esse tema (O'Hara & Swain, 1996; Beck, 1996; Beck, 2001; Aliane et al., 2011); ou, de forma mais específica, testar hipóteses de que determinadas situações, como por exemplo, parto cesáreo (Carter et al., 2006), práticas chinesas de confinamento (Wong & Fisher, 2009) ou violência (Beydoun et al., 2012; Wu et al., 2012) são fatores de risco ou de proteção para DPP. Enquanto a revisão sistemática permite compilar os fatores de risco para DPP já identificados na literatura, a aplicação de metanálise e obtenção de valores de tamanho de efeito trazem a compreensão adicional da magnitude da associação entre os fatores pesquisados e a DPP, bem como a direção dessa associação, ou seja: se o fator é de risco ou de proteção.

2.5.5 Acurácia de instrumentos de rastreamento

Conforme já foi explicado na seção “2.3.3 Avaliação da qualidade da evidência disponível” do presente trabalho, a validade de uma escala de autoavaliação como instrumento de rastreamento e de pesquisa é verificada por estudos quantitativos, em que as pontuações da escala obtidas por uma amostra de participantes são comparadas com seus diagnósticos clínicos. Dessa maneira, se estabelece um ponto de corte ótimo, procurando-se minimizar o número de falsos positivos (resultado positivo na escala e negativo no diagnóstico clínico) e falsos

negativos (resultado negativo na escala e positivo no diagnóstico clínico). Os parâmetros de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo podem ser sintetizados em revisões sistemáticas de literatura, com ou sem metanálise. Normas para revisões de acurácia de testes diagnósticos já estão sendo desenvolvidas pela Colaboração Cochrane (Diagnostic Test Accuracy Working Group, 2011). O "*Handbook for DTA Reviews*" está sendo escrito e os capítulos são publicados no site da Colaboração Cochrane conforme ficam prontos (ver: http://srdta.cochrane.org/, acessado em 29/11/2012).

A revisão sistemática de Gaynes et al. (2005) objetivou avaliar a acurácia de diferentes instrumentos para avaliação de depressão na gestação e no puerpério. Para isso, buscaram estudos prospectivos em que foram usadas entrevistas clínicas estruturadas para avaliar as propriedades de um instrumento de avaliação de DPP. Foram coletados os dados de sensibilidade e especificidade presentes nos dez estudos identificados pela busca sistemática. A qualidade metodológica desses trabalhos foi considerada de moderada a boa, mas a heterogeneidade dos estudos impossibilitou realização de metanálise. Para todos os instrumentos de rastreamento avaliados (Beck Depression Inventory - BDI, Postpartum Depression Screening Scale - PDSS, Edinburgh Postpartum Depression Scale – EPDS), a especificidade para depressão maior foi relativamente alta, indicando baixo risco de falsos positivos. A sensibilidade, por sua vez, foi variável, e os intervalos de confiança dos três instrumentos se sobrepuseram, de forma que não foi possível identificar qual foi mais sensível. Algumas limitações dos trabalhos incluídos foram levantadas, como o pequeno número de participantes em cada trabalho, que pode ter gerado imprecisões na determinação da sensibilidade; e a baixa diversidade étnica, pois como as populações estudadas foram quase inteiramente brancas, não há evidências de que esses resultados possam ser generalizados para populações com outras características.

A escala mais usada nos estudos de DPP, atualmente, é a EDPE. Uma revisão sistemática de literatura buscou avaliar a acurácia desse instrumento de autoavaliação (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009). Foram incluídos 37 estudos com variedade de contextos e idiomas de administração, sendo três realizados com gestantes e 34 com puérperas. Buscou-se compilar os parâmetros de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo da EDPE usando os pontos de corte mais comuns na

literatura: ≥10 e ≥13. A maior parte dos trabalhos incluídos teve a qualidade considerada razoável. Ocorreu ampla variação nos parâmetros avaliados, indicando heterogeneidade na performance da EDPE. Por exemplo, usando o ponto de corte ≥10 para avaliação de puérperas quanto à DPP, obteve-se sensibilidade entre 59% e 100% e especificidade entre 44% e 97%. Constatou-se que, quando o ponto de corte foi ≥10, o idioma em que a EDPE foi aplicada não resultou em diferenças substanciais nos parâmetros; mas usando o ponto de corte ≥13, estudos que aplicaram a EDPE em inglês apresentaram sensibilidade variando de 76% a 100%, enquanto os que a aplicaram em outros idiomas apresentaram variação entre 49% e 100%. O mesmo se deu para a especificidade, que variou de 70% a 99% para a escala aplicada em inglês e de 34% a 100% para a escala aplicada em outros idiomas. Não foi possível testar o efeito dos diferentes países sobre esses resultados, devido à insuficiência de dados.

A extensão da heterogeneidade entre os estudos de validade da EDPE levaram Gibson et al. (2009) a sugerir que talvez não existam evidências suficientes para assegurar a validade da EDPE entre diferentes grupos culturais. A variabilidade entre os resultados pode ter ocorrido devido à diversidade de populações e metodologias, com uso de diferentes critérios de diagnóstico e tempos após o parto. Também podem ter ocorrido diferenças de compreensão semântica da escala, visto que os termos usados para definir depressão nos países ocidentais podem não detectar, de forma acurada, os mesmos sintomas em outras sociedades. Essa heterogeneidade também pode ser observada no Brasil, onde três estudos avaliaram a validade da EDPE, recomendando o uso de diferentes pontos de corte: ≥10 (Santos et al., 2007), ≥11 (Figueira et al., 2009) e ≥12 (Santos et al., 1999).

2.5.6 Consequências da DPP para o desenvolvimento infantil

Revisões sistemáticas encontraram evidências de que a DPP pode ter efeito sobre o desenvolvimento infantil, nas áreas de cognição (Beck, 1998; Grace, Evindar, & Stewart, 2003), desenvolvimento emocional (Beck, 1998), estilo de apego (Martins & Gaffan, 2000) e problemas comportamentais (Grace et al., 2003), inclusive em longo prazo, conforme constatado por revisão sistemática que investigou os efeitos da DPP sobre o comportamento dos filhos em período de um a 18 anos após o parto (Beck, 1999). Os tamanhos de efeito dessas metanálises têm

sido considerados fracos ou moderados, e mostraram-se menores em estudos com amostras maiores e melhor qualidade metodológica. A heterogeneidade de metodologias prejudicou a confiabilidade desses resultados, que devem ser interpretados com cautela. As revisões sistemáticas sobre esses assuntos necessitam de atualização, de forma a contemplar dados mais recentes. Não foram encontradas revisões sistemáticas de efeitos da DPP sobre o desenvolvimento infantil que incluíssem estudos latino-americanos ou de outros países em desenvolvimento, portanto, o efeito das diferenças culturais, caso existente, permanece ignorado.

2.5.7 Eficácia e efetividade de intervenções para o tratamento da DPP

Revisões sistemáticas têm evidenciado que diversas formas de intervenção foram eficazes em reduzir a DPP, como: tratamento psicoterápico (Dennis & Hodnett, 2007; Lumley, Austin, & Mitchell, 2004; Bledsoe & Grote, 2006; Cuijpers, Brännmark, & van Straten, 2008; Sockol, Epperson, & Barber, 2011), farmacológico (Hofbrand, Howard, & Crawley, 2001; Bledsoe & Grote, 2006; Cuijpers et al., 2008; Sockol et al., 2011) e exercícios físicos (Daley, Jolly, & MacArthur, 2009). Dentre as abordagens psicoterápicas, destacaram-se os efeitos significativos das terapias cognitivo-comportamental e interpessoal na redução da DPP (Bledsoe & Grote, 2006; Dennis & Hodnett, 2007; Sockol et al., 2011). Do ponto de vista da prática profissional, esses resultados indicam que podem ser criados programas de tratamento para mães com DPP, baseados em uma das técnicas que se mostraram significativas, ou em uma combinação delas.

Considerando-se os possíveis efeitos negativos da DPP sobre as crianças, uma metanálise avaliou a eficácia de intervenções para aumentar a sensibilidade das mães aos seus filhos (Kersten-Alvarez, Hosman, Riksen-Walraven, van Doesum, & Hoefnagels, 2011), definida como apoio emocional à criança, na forma de responsividade apropriada aos sinais e necessidades da mesma, ou expressão de afeto positivo ou gratidão. Os resultados indicaram que as intervenções foram eficazes, com tamanho de efeito moderado quando foi incluída massagem feita pela mãe no bebê e fraco quando foi incluída psicoterapia individual para as mães. Essas evidências não foram consideradas consistentes, devido à variabilidade entre resultados.

Faltam informações sobre técnicas alternativas para tratamento de DPP. Uma revisão sobre intervenções não farmacológicas, psicoterápicas ou psicossociais para tratar DPP teve resultados inconclusivos, pois incluiu apenas um estudo, que evidenciou que técnicas de massagem e acupuntura não tiveram resultados significativos na redução de sintomas depressivos (Dennis & Allen, 2008). Além disso, as revisões sistemáticas sobre intervenções para prevenir ou tratar DPP que foram publicadas até o momento (Dennis & Creedy, 2004; Dennis & Hodnett, 2007; Dennis & Allen, 2008; Cuijpers et al., 2008; Daley et al., 2009; Sockol et al., 2011) não incluíram estudos realizados em países latino-americanos, de forma que não é indicada a generalização dos resultados para esse contexto.

De fato, ensaios clínicos aleatorizados sobre intervenções para prevenir ou tratar DPP são raros em países em desenvolvimento, mas uma busca na BVS-Psi pelos termos “depressão pós-parto” e “tratamento” localizou um registro de ensaio clínico controlado realizado no Chile, que avaliou a efetividade de um programa de tratamento da DPP na atenção primária (Guajardo et al., 2007). Esse estudo comparou os desfechos de 230 mulheres, distribuídas aleatoriamente entre duas modalidades de intervenção: o tratamento multicomponente (grupo experimental), que incluía administração de fluoxetina ou sertralina para casos mais graves, intervenções psicoeducativas em grupo e monitoramento telefônico; e o tratamento habitual (grupo controle), que recebeu os cuidados que normalmente são dispensados às mulheres com depressão pós-parto nos centros de atenção primária, podendo incluir controles médicos, administração de antidepressivos, psicoterapia individual ou em grupo e encaminhamento a serviços secundários (ambulatoriais ou hospitalares) de saúde. A alocação das participantes nos grupos foi feita de forma aleatória, usando sistema informatizado e envelopes lacrados, e a taxa de perda foi inferior a 10%. Concluiu-se que o grupo experimental apresentou menor intensidade de sintomatologia depressiva nas avaliações pós-tratamento que o grupo controle, sendo que a diferença entre ambos os grupos foi maior aos três meses do que aos seis meses após o tratamento, embora ambas tenham sido significativas. A existência desse ensaio clínico aleatorizado sobre tratamento DPP (Guajardo et al., 2007), publicado em espanhol, em revista latino-americana e não incluído em revisões sistemáticas sobre esse tema, mostra a importância de se realizar buscas mais abrangentes ao revisar a literatura, incluindo as bases de dados latino-americanas, como a BVS.

Uma discussão pormenorizada a respeito de revisões sistemáticas da Cochrane sobre intervenções para prevenir ou tratar depressão no período gestacional ou no puerpério pode ser encontrada em Silva, Sampaio e Otta (2011).

2.5.8 Eficácia e efetividade de intervenções para prevenção de DPP

Dennis e Creedy (2004) sintetizaram ensaios clínicos aleatorizados que usaram intervenções psicológicas ou psicossociais durante a gestação ou nas primeiras semanas após o parto para prevenir DPP. Verificou-se que o formato mais efetivo de intervenção com essa finalidade foi atendimento em grupo, abordando mulheres em situação de risco de DPP (qualquer que fosse o critério adotado para definir situação de risco) e iniciado após o parto.

Howard, Hoffbrand, Henshaw, Boath e Bradley (2009) procuraram avaliar a eficácia de antidepressivos administrados na gestação para prevenir DPP, em mulheres com histórico de transtorno depressivo, e incluíram dois ensaios clínicos aleatorizados. Embora o número de trabalhos e de participantes seja pequeno para embasar orientações para a prática clínica, observou-se que a Sertralina foi eficaz em reduzir a recorrência de depressão, mas a Nortriptilina não teve esse efeito.

Portanto, foram encontradas evidências de que algumas intervenções podem ser eficazes para prevenir DPP. A combinação desses resultados com as evidências referentes a fatores de risco para DPP podem ser usadas para a elaboração de programas de prevenção. No entanto, ambas as revisões incluíram poucos estudos e são necessários novos ensaios clínicos aleatorizados, sobretudo nos países em desenvolvimento, que não foram contemplados até o presente momento.

3 Objetivos e hipóteses

3.1 Objetivos

O objetivo geral do presente trabalho foi 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.

Os objetivos específicos foram:

1. determinar a prevalência média geral de DPP;
2. comparar a prevalência média de DPP em países desenvolvidos e em desenvolvimento;
3. comparar a prevalência média de DPP de acordo com as variáveis metodológicas dos estudos incluídos;
4. avaliar a qualidade metodológica de estudos de prevalência de DPP e verificar se ocorreu associação entre a qualidade dos estudos e seus resultados.
5. construir um modelo explicativo dos resultados de prevalência de DPP com base em fatores metodológicos e geográficos.

3.2 Hipóteses

As hipóteses do presente estudo foram:

1. Quanto à prevalência média geral: a prevalência média de DPP, ponderada pelo tamanho amostral, seria de 12% a 14% (tomando-se por base os estudos de O'Hara & Swain, 1996 e Melo Junior, 2011).

2. Quanto ao grau de desenvolvimento dos países:

2A. o valor da prevalência média em estudos realizados em países desenvolvidos seria menor que o valor da prevalência média em estudos realizados em países em desenvolvimento (conforme encontrado por Melo Junior, 2011);

2B. controlando-se variáveis metodológicas, a diferença entre países desenvolvidos e em desenvolvimento desapareceria.

3. Quanto às variáveis metodológicas:

3A. período após o parto: considerando que a DPP é um transtorno que pode durar cerca de seis meses a um ano, sua prevalência média seria maior no terceiro quadrimestre após o parto (241 a 365 dias após o parto), intermediária no segundo quadrimestre (121 a 240 dias após o parto) e menor no primeiro quadrimestre (nove a 120 dias após o parto);

3B. instrumento de avaliação de DPP: estudos que usaram entrevista clínica como método para avaliar DPP teriam prevalência média menor que estudos que usaram escalas de autoavaliação;

3C. delineamento: seria encontrada maior prevalência média de DPP em estudos transversais, quando comparados a estudos de coorte longitudinal. Essa hipótese se baseia no fenômeno conhecido como viés de prevalência, que prevê que os indivíduos que manifestam o desfecho por um maior período de tempo têm maior chance de serem incluídos nos estudos transversais, de forma que não se pode diferenciar entre o surgimento de novos casos e a duração dos mesmos (Bastos & Duquia, 2007). Considera-se que a DPP é um transtorno de duração prolongada, que pode persistir desde as primeiras semanas de puerpério até um ano ou mais após o parto, estando sujeito, portanto, ao fenômeno do viés de prevalência;

3D. técnica de amostragem: a técnica de amostragem teria influência sobre os resultados de prevalência de DPP. A direção do efeito não foi prevista.

3E. tamanho da amostra: o tamanho da amostra não teria associação com a prevalência de DPP, quando controladas outras variáveis metodológicas (período de diagnóstico, tipo de diagnóstico, tipo de amostra, perfil das participantes incluídas). Essa hipótese questiona a afirmação de Melo Junior (2011) de que estudos com maiores amostras estão associados com menores prevalências de DPP;

3F. planejamento do tamanho da amostra: a prevalência média de DPP em estudos que planejaram o tamanho da amostra com base em análise de poder seria semelhante à dos que não planejaram o tamanho da amostra.

3G. base amostral: o uso de base amostral adequada ou inadequada, conforme avaliado no item 2 da Escala de Loney et al. (1998), teria influência sobre os resultados de prevalência de DPP. A direção do efeito não foi prevista;

3H. contexto de pesquisa: o contexto de pesquisa se refere ao local onde os dados foram coletados ou à abrangência do estudo, que poderia ser nacional (realizado em várias províncias, estados ou regiões de um país), regional (realizado em instalações médicas e/ou cidades localizadas em diversas províncias, estados ou regiões de um país), em uma cidade, em uma instalação médica, ou com outra abrangência (ex.: dados coletados em mais de um país; dados coletados por internet). Foi previsto que o contexto não teria associação com a prevalência média de DPP;

3I. ponto de corte da Escala de Depressão Pós-parto de Edimburgo (EDPE): dentre os estudos que usaram a EDPE, quanto menor o ponto de corte estipulado, maior seria a prevalência de DPP;

4. Quanto à qualidade dos estudos:

4A. a qualidade dos estudos realizados em países desenvolvidos, medida pela escala de Loney et al. (1998), seria equivalente à qualidade dos estudos realizados em países em desenvolvimento;

4B. estudos mais recentes teriam maior qualidade que estudos mais antigos, conforme avaliação pela escala de Loney et al. (1998);

4C. a qualidade do estudo, medida pela Escala de Loney et al. (1998), teria associação com a prevalência de DPP. Essa hipótese se baseou na inferência de que estudos com maiores pontuações teriam, com mais frequência, características que evitariam vieses, como: amostras aleatórias ou censitárias; tamanho amostral adequado; e avaliação de DPP feita por meio de entrevista clínica, declarando-se a concordância entre entrevistadores e o cegamento dos mesmos. A direção da associação não foi prevista.

5. Quanto à possibilidade de construir modelo explicativo dos resultados de prevalência de DPP: seria possível construir um modelo explicativo dos resultados de prevalência de DPP com base em fatores metodológicos e geográficos. O tipo de instrumento usado para avaliar DPP e a técnica de amostragem seriam fatores que contribuem mais para explicar os resultados de prevalência que o nível de desenvolvimento socioeconômico do país.

4 Justificativa

Os sintomas de DPP, além de desagradáveis para a mulher que os vivencia, podem causar prejuízos à sua vida familiar e social, bem como consequências indesejáveis para o desenvolvimento do bebê (Murray & Cooper, 1997; Righetti-Veltema, Bousquet, Conne-Perréard, & Manzano, 2001; Righetti-Veltema, Bousquet, & Manzano, 2003; Schwengber & Piccinini, 2003). Assim, pesquisadores têm buscado alternativas para prevenir e tratar a DPP (Silva et al., 2011), algumas delas consideradas bem sucedidas.

Conhecer a prevalência de um transtorno é fundamental para decidir sobre a necessidade ou a viabilidade de implantação de um programa para prevenção ou tratamento em uma determinada população, bem como para calcular os custos de implantação e a infraestrutura necessária. Em 1996, Murray e Lopez afirmaram que "A formulação de políticas públicas de saúde precisa desesperadamente de informações independentes e objetivas sobre a magnitude dos problemas de saúde e suas prováveis tendências, baseadas em unidades de medida padrão e métodos comparáveis."[22] (p.740). Dez anos depois, Halbreich & Karkun (2006) afirmaram que:

> O mapeamento diversificado e global das desordens pós-parto e suas naturezas é de extrema importância para a elucidação dos mecanismos subjacentes e distinções entre processos universais e culturalmente específicos. Além disso, as determinações especificamente culturais da magnitude dos problemas proporcionaria ferramentas para influenciar as decisões políticas e o desenvolvimento de serviços de saúde para mulheres que necessitem.[23] (p. 109-110).

Revisões sistemáticas sobre prevalência de DPP, com e sem metanálise, já foram publicadas (O'Hara & Swain, 1996; Gavin et al., 2005; Halbreich & Karkun, 2006; Sawyer et al., 2010; Vigod et al., 2010; Villegas et al., 2011; Melo Junior, 2011;

---

[22] No original: "Public health policy formulation desperately needs independent, objective information on the magnitude of health problems and their likely trends, based on standard units of measurement and comparable methods.

[23] No original: The diversified global mapping of postpartum disorders and their natures is of utmost importance for further elucidation of underlying mechanisms and distinctions between universal and culturally-specific processes. Indeed, the culturally specific determinations of the magnitude of the problems would provide tools for influencing policy decisions and development of necessary health delivery systems for women in need.

Lobato et al., 2012), mas apresentaram limitações do ponto de vista de descrição de metodologia e das técnicas adotadas, que implicaram em risco de viés (Mann et al., 2010). As principais limitações constatadas foram estratégias de busca usando poucas bases de dados e incluindo apenas estudos em inglês; ausência de avaliação da heterogeneidade dos estudos incluídos; deficiências no relato da metodologia e dos resultados; número insuficiente de estudos realizados em países em desenvolvimento, prejudicando a possibilidade de generalização dos resultados para esse contexto; e ausência de um segundo avaliador para possibilitar a verificação da confiabilidade da seleção dos estudos primários para inclusão.

A interpretação dos resultados dos estudos sobre prevalência de DPP tem sido dificultada pela falta de consenso conceitual e metodológico, que tem sido alvo de debates científicos. Em carta enviada ao editor do Jornal Brasileiro de Psiquiatria, Cantilino (2010) argumentou:

> Parece relevante que uma das condições clínicas mais estudadas da psiquiatria, como é a depressão pós-parto, ainda padeça de considerável carência de consenso em importantes questões que guiariam a metodologia das diversas pesquisas. Qual o período do pós-parto ideal para que se realize triagem? Deve-se diagnosticar como depressão pós-parto apenas aqueles quadros incidentes no puerpério? Até que mês após o parto se pode considerar que uma condição clínica tem relação com ele?
>
> O problema instalado é que cada pesquisador acaba utilizando sua própria perspectiva para projetar seu estudo. Isso gera grande dificuldade de comparação entre os resultados. [...] Assim, além de podermos promover dados atualizados no futuro, apontamos para a necessidade de criarmos “consensos metodológicos” no sentido de termos dados comparáveis entre centros e percebermos melhor quais variáveis realmente fazem a diferença. (p.257)

Da variabilidade entre metodologias usadas em estudos epidemiológicos decorrem dúvidas sobre a possibilidade de que países desenvolvidos e em desenvolvimento apresentem diferenças significativas quanto à prevalência de DPP. Algumas metanálises (Melo Junior, 2011; Lobato et al., 2012) verificaram que a prevalência de DPP foi maior em países em desenvolvimento do que em países desenvolvidos. Porém, se os resultados referentes à prevalência de DPP avaliada por instrumento de autoavaliação forem, de fato, mais elevados do que os resultados obtidos por entrevista clínica (O'Hara & Swain, 1996; Mathey, 2010), esta seria uma variável de confusão que deveria ser levada em consideração. Estudos que usaram

autoavaliação podem ser mais comuns em países em desenvolvimento, devido ao baixo custo dessa metodologia, e tenderiam a superestimar a prevalência de DPP.

Até o momento, não se tem conhecimento de outra metanálise que tenha levado em consideração os fatores metodológicos ao comparar países desenvolvidos e em desenvolvimento quanto à prevalência de DPP. A realização do presente estudo contribuiu para identificar as variáveis que, como sugeriu Cantilino (2010), “realmente fazem a diferença” (p.259), auxiliando na construção de um possível consenso metodológico. Ao mesmo tempo, dados sobre a prevalência do transtorno em diversos locais puderam ser atualizados, fornecendo subsídios para o planejamento de programas de saúde mental.

Portanto, mesmo havendo metanálises anteriores sobre a prevalência de DPP, o presente estudo justificou-se pela atualização das informações e pelo aprimoramento do método, que buscou seguir, sempre que pertinente, as recomendações da Colaboração Cochrane sobre metodologia de revisões sistemáticas de literatura (Higgins & Green, 2011). O presente trabalho trouxe contribuições originais para a compreensão do fenômeno da prevalência de DPP, a saber:

1. **Método de busca**: ampliação, em relação às revisões sistemáticas anteriores, das bases de dados eletrônicas consultadas durante a fase de busca de trabalhos. Foram realizadas buscas em 14 bases de dados: Biblioteca Digital de Teses e Dissertações, BVS-PSI ULAPSI, CINAHL, EMBASE, ERIC, ISI Web of Knowledge, LILACS, ProQuest, Psicodoc, PsycInfo, PubMed, Redalyc, SciELO e Scopus. Não foi estabelecido limite para o período de abrangência das publicações.

2. **Avaliação de qualidade metodológica dos trabalhos incluídos**: critérios claros e objetivos de avaliação de qualidade metodológica dos trabalhos sobre prevalência de DPP incluídos.

3. **Verificação da confiabilidade entre revisoras**: participação de duas revisoras independentes na seleção de estudos para inclusão e na posterior análise de qualidade metodológica, usando-se técnicas estatísticas para determinar o grau de concordância entre ambas. Esse procedimento verifica a confiabilidade do protocolo de coleta de dados da revisão sistemática.

4. **Análise estatística**: uso de técnicas estatísticas para investigar evidências de heterogeneidade; análises de subgrupo e de sensibilidade.

5. **Metarregressão**: uso de análise multivariada, explorando a heterogeneidade entre os trabalhos de forma a avaliar a contribuição de variáveis metodológicas e geográficas sobre os resultados de prevalência.

6. **Detalhamento dos resultados geográficos**: apresentação de estatísticas descritivas para os diferentes países e cidades; apresentação de resultados para países desenvolvidos e em desenvolvimento, de acordo com o Índice de Desenvolvimento Humano (IDH).

5 Método

5.1 Critérios de inclusão

Foram definidos nove critérios de inclusão de estudos primários na presente revisão:

- **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.

A delimitação do período de oito a 365 dias após o parto foi definida para atender a um entendimento, comum na área de pesquisa sobre DPP, de que o transtorno pode ter início até um ano após o nascimento do bebê (Austin, 2004). Ao mesmo tempo, excluiu-se o período imediatamente posterior ao parto (até oito dias), para evitar confusão com a melancolia pós-parto, que é um transtorno com prevalência sabidamente maior que a DPP, sendo estimada entre 15% a 85% das mães (Pearlstein et al., 2009).

Para manter a independência dos dados, foi definido que cada pesquisa seria incluída apenas uma vez, mesmo que tivesse gerado mais de uma publicação. Duplicações ou sobreposições de amostra foram localizadas ao se comparar, sistematicamente, os dados referentes ao local, data da coleta de dados, tamanho da amostra, perfil das participantes e resultados encontrados. Nos casos em que se constatou sobreposição parcial das participantes, foi incluída apenas a publicação com a maior amostra. Informações que constavam de artigos excluídos por razão de duplicação ou sobreposição de amostra foram usadas para complementar o formulário de coleta de dados, quando necessário.

Com esses critérios, pretendeu-se determinar que os estudos incluídos teriam as características relatadas a seguir, nas seções 5.1.1 a 5.1.3 do presente trabalho.

5.1.1 Tipo de estudo

Foram incluídos estudos primários e quantitativos com delineamento transversal ou de coorte longitudinal (critério A), que investigaram a prevalência de DPP (critério B) entre oito e 365 dias após o parto (critério F), através de avaliação clínica por profissional habilitado ou escalas de autoavaliação (critério E).

Foram excluídos, de acordo com esses critérios:

- Revisões de literatura, estudos qualitativos e outros que não usaram delineamento transversal ou de coorte longitudinal (ex.: caso-controle, ensaio clínico aleatorizado), de acordo com o critério A.
- Estudos que não investigaram a prevalência de DPP, ainda que tenham investigado outros transtornos puerperais, como psicose, melancolia ou

transtorno de estresse pós-traumático no puerpério, de acordo com o critério B.

- Estudos que não mencionam claramente o período após o parto em que a investigação foi realizada, ou que avaliaram a presença de DPP unicamente em período inferior a oito e/ou superior a 365 dias após o parto, conforme critério F.
- Estudos que avaliaram DPP por método não replicável, como a percepção de leigos a respeito do estado das participantes; dados de prontuário em que não foi especificado o método de diagnóstico; ou instrumentos considerados subjetivos e não validados, de acordo com o critério E.

5.1.2 Tipo de participantes

Os participantes dos estudos incluídos na presente revisão sistemática foram mulheres (critério C) não submetidas a intervenção prévia com objetivo de prevenir ou tratar DPP (critério D) e que não pertenciam a grupos com características específicas que se diferenciam da população geral (critério G).

De acordo com esses critérios, foram excluídos:

- Estudos que avaliaram a presença de DPP em homens (depressão paterna) ou em animais não humanos (ex.: ratos, vacas), conforme o critério C.
- Estudos que avaliaram amostra submetida a alguma intervenção (ex.: psicoterapia ou uso de medicamentos psicoativos) com o objetivo de prevenir DPP, conforme critério D. Foram admitidos estudos em que as participantes receberam o tratamento usual dispensado às gestantes e puérperas, como consultas ginecológicas e acompanhamento pré e pós-natal da mãe e do bebê pela equipe de saúde local.
- Estudos que declararam ter estudado participantes com características específicas, que se diferenciavam da população geral, conforme o critério G. Trabalhos que não declararam ter avaliado uma amostra com características específicas foram incluídos na revisão sistemática, admitindo-se que tenham estudado amostra com características próximas às da população geral.

5.1.3 Desfecho

O desfecho estudado foi a prevalência de DPP na amostra. Compreendeu-se prevalência como a proporção de mulheres identificadas com DPP em amostra tomada da população de estudo (ou na população toda, em estudos censitários), usando critério para identificação de DPP estabelecido pelo trabalho avaliado.

5.2 Estratégia de busca

Foram consultadas 14 bases de dados eletrônicas e as estratégias de busca específicas para cada base de dados foram definidas junto a uma bibliotecária, usando os termos “depressão pós-parto” (em inglês: “postpartum depression”; em espanhol: “depresión postparto”) e “prevalência” (em inglês: “prevalence”; em espanhol: “prevalencia”). As estratégias foram elaboradas para atender às particularidades de cada base de dados, como idioma de busca, presença ou ausência de um *thesaurus*, diferenças nos comandos de truncagem e configuração do mecanismo de busca para permitir o uso de lógica booleana. As estratégias de busca para cada base de dados estão descritas na Tabela 5.

Tabela 5 – Estratégias de busca para as bases de dados consultadas na presente revisão

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[24]
ERIC	(postpartum depression or postnatal depression or puerperal depression) AND preval*
ISI Web of Science	(postpartum depression OR postnatal depression OR puerperal depression) AND preval*
LILACS	(postpartum depression OR depresion postparto OR depressao pos-parto OR depressao puerperal OR puerperal depression OR depressao pos-natal or postnatal depression OR depresion postnatal) AND preval$
ProQuest	(postpartum depression OR postnatal depression OR puerperal depression) AND (prevalence)
Psicodoc	Depresión postparto
PsycInfo	((postpartum depression OR postnatal depression OR puerperal depression) AND preval*):Any Field
PubMed	"postpartum depression" in Mesh AND prevalence in Mesh[25]
Redalyc	Depression postparto
Scielo	prevalencia depressao pos-parto
Scopus	("postpartum depression" OR "postnatal depression" OR "puerperal depression") AND preval*

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

[25] MESH significa Medical Subjects Hearing, 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.

5.3 Seleção de estudos

Inicialmente, todos os trabalhos identificados na busca sistemática foram reunidos, eliminando-se as duplicidades. Prosseguiu-se com o trabalho de seleção dos estudos encontrados, conforme esquema apresentado na Figura 3, envolvendo as seguintes etapas:

1. **Seleção pelo título:** os títulos dos trabalhos identificados foram avaliados independentemente por duas revisoras, excluindo-se aqueles que claramente não preenchiam os critérios de inclusão. Em caso de divergência, os registros foram avaliados na próxima etapa (seleção pelos resumos).
2. **Seleção pelos resumos:** todos os resumos dos trabalhos que foram selecionados na etapa 1 foram lidos pelas duas revisoras, que aplicaram novamente os critérios de inclusão para eliminar os estudos que claramente deveriam ser excluídos. Em caso de divergência, os registros foram avaliados na próxima etapa (avaliação do texto completo).
3. **Avaliação do texto completo:** realizada usando-se formulário eletrônico[26]. O primeiro passo, nessa fase, foi uma nova avaliação do preenchimento dos critérios de inclusão, agora partindo das informações do texto completo e buscando esclarecer as eventuais dúvidas levantadas durante a avaliação do resumo. A coleta de dados foi encerrada nos casos em que foi constatado que o estudo não atendia a, pelo menos, um critério de inclusão. Os trabalhos que atendiam a todos os critérios de inclusão, conforme avaliação do texto completo, foram incluídos na revisão sistemática, prosseguindo-se com a coleta de dados. Essa etapa foi realizada, em sua totalidade, pela primeira revisora. Objetivando verificar a taxa de concordância entre avaliadoras, a segunda revisora avaliou amostra pareada desse universo de estudos, formada pela totalidade dos artigos com população latino-americana e número equivalente do restante dos estudos eleitos para a revisão. Os casos de divergência foram decididos por consenso.

[26] Acesso ao formulário eletrônico de avaliação do texto completo está disponível no link: http://revsist.limequery.com/62197/lang-pt-BR (acessado em 17/01/2013).

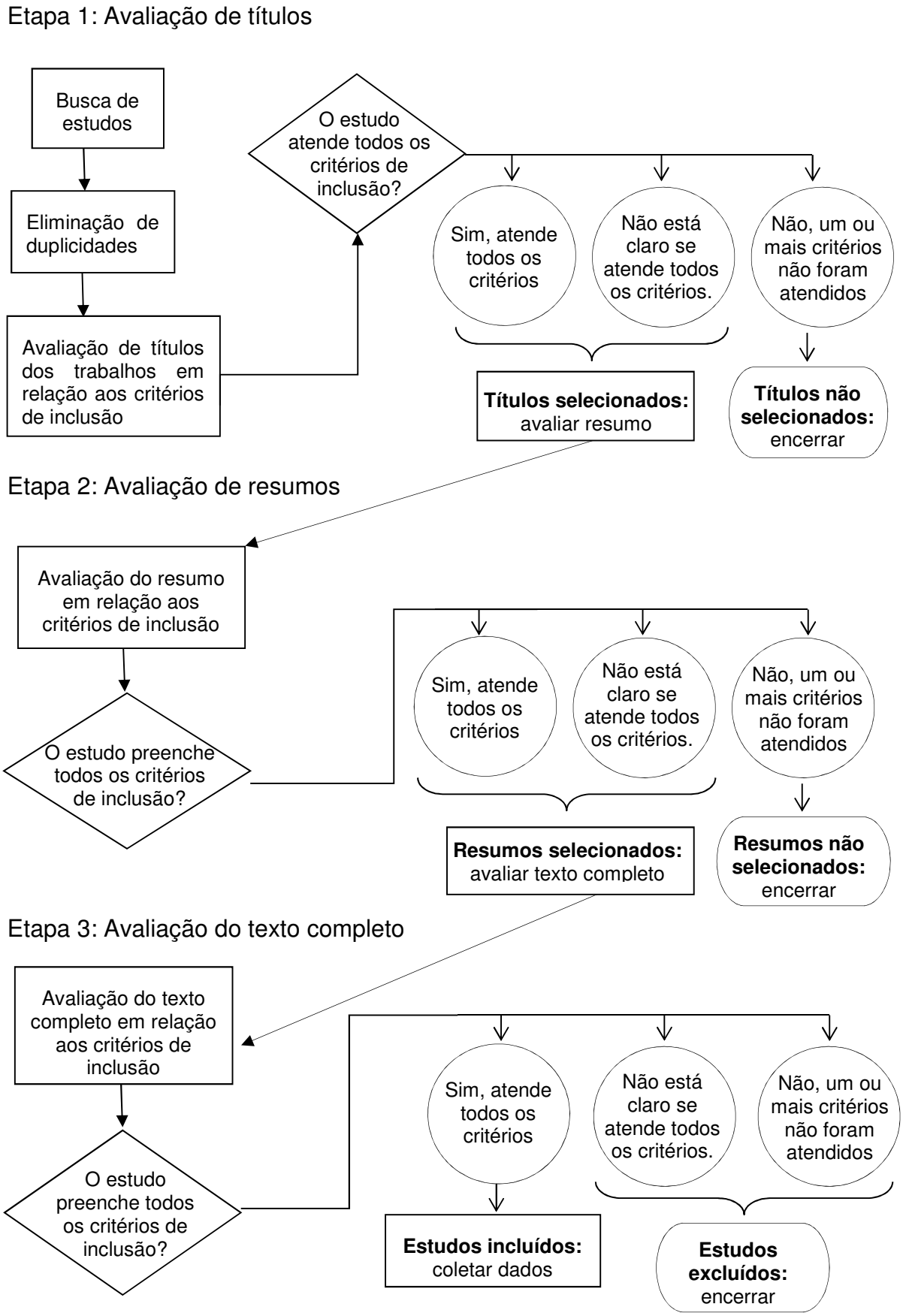

Figura 3 – Fluxograma de seleção de estudos para a presente revisão sistemática.

5.4 Extração e gerenciamento de dados

Todos os registros foram salvos no Endnote®Web (www.myendnoteweb.com, acesso em 05/01/2011), *software* disponível gratuitamente, que permite coletar e organizar estudos identificados em bases de dados virtuais, bem como fazer citações e referências bibliográficas. Diversas bases de dados (ex.: ISI Web of Knowledge, PsycInfo) permitem exportar os registros diretamente para o Endnote®Web. Nas plataformas em que essa opção não estava disponível (ex.: BVS-PSI ULAPSI), os registros foram inseridos manualmente.

Para identificar os registros duplicados, foi usada a ferramenta "*Find duplicates*" do Endnote®Web, que detecta os registros idênticos, comparando os seguintes campos: autor, ano, título e tipo da referência. Por esse método, foram descartadas cerca de 50% das duplicatas. As demais não foram identificadas automaticamente, pois há divergências na forma como alguns campos são apresentados em diferentes bases de dados, e precisaram ser excluídas manualmente.

Os registros não duplicados do Endnote®Web foram exportados em estilo RIS, gerando um arquivo em formato de texto (extensão .txt). Um *software* desenvolvido especialmente para o presente estudo, em linguagem Ruby, leu o arquivo em formato .txt, gerando um arquivo com extensão .xls, que foi aberto usando o Microsoft Excel 2010. Essa planilha, que continha os campos: autores, ano, título e nome do periódico, serviu como referência para as próximas etapas, que envolveram a seleção dos estudos por título, resumo e texto completo. Automaticamente, o *software* em Ruby atribuiu um código numérico sequencial a cada registro, que foi usado como chave para a identificação dos estudos em todos os formulários e planilhas usados no presente estudo.

Para localizar os textos completos dos trabalhos selecionados pelo título e resumo, foram usadas diversas ferramentas. A maior parte dos textos foi localizada via internet, em acesso aberto (ex.: SciELO) ou restrito (mas disponível à Universidade de São Paulo) pelo Portal de Periódicos da Capes. Esse portal, inteiramente financiado pelo governo brasileiro, é uma biblioteca virtual que reúne e disponibiliza produção científica internacional a instituições de ensino e pesquisa no Brasil (www.periodicos.capes.gov.br, acessado em 23 de outubro de 2012). Os

trabalhos que não puderam ser obtidos por esses meios foram solicitados através do Programa de Comutação Bibliográfica (Comut) e do Serviço de Acesso ao Documento (SCAD), que viabilizam a obtenção de cópias de materiais científicos disponíveis nos acervos das principais bibliotecas brasileiras e em serviços internacionais de informação, incluindo periódicos técnico-científicos, teses e dissertações, anais de conferência, relatórios técnicos e capítulos de livros (http://www.ibict.br, acessado em 23/12/2012).

A avaliação dos estudos por título foi realizada com auxílio de planilha do Microsoft Excel. Cada registro foi julgado quanto aos sete primeiros critérios de inclusão (A, B, C, D, E, F, G), conforme instruções do "Protocolo de avaliação de registros por título e resumo", que se encontra no Apêndice 1. Sempre que as revisoras verificaram, através do título do trabalho, que ele violou um ou mais dos sete critérios estabelecidos, foi indicado o código "F" (falso) na célula correspondente ao critério não atendido. Poderia ser indicada a violação de mais de um critério, se necessário. Uma coluna preenchida com fórmula indicou VERDADEIRO para todo estudo que foi selecionado pelo título (que deveria, portanto, ser avaliado através do resumo) e FALSO para o estudo não selecionado. Alguns exemplos de registros eliminados pelo título estão apresentados na Tabela 6.

Na segunda etapa, foram avaliados os resumos de todos os estudos selecionados pelo título. Para isso, foi usada a mesma planilha da etapa 1. As células correspondentes a cada um dos sete primeiros critérios de inclusão (A, B, C, D, E, F, G) foram preenchidas com um dos seguintes códigos: 1 = verdadeiro, 2 = falso, 3 = não está claro, 4 = não se aplica. A alternativa "não se aplica" esteve disponível somente para os critérios B a G e deveria ser assinalada sempre que o resumo indicasse que o estudo era uma revisão de literatura. O protocolo de coleta de dados usado nessa etapa consta do Apêndice 1.

.

Tabela 6 – Exemplos de registros eliminados pelo título.

Título do estudo	Motivo da exclusão
Postpartum depression among rural women from developed and developing countries: A systematic review (Villegas et al., 2011)	Violou critério A: trata-se de revisão de literatura, e não de estudo primário e quantitativo.
Antenatal depression in socially high-risk women in Canada (Bowen, Stewart, Baetz, & Muhajarine, 2009)	Violou critério B: avaliou depressão gestacional, e não DPP.
Abnormal, abusive, and stress-related behaviors in baboon mothers (Brent, Koban, & Ramirez, 2002)	Violou critério C: sujeitos são animais não-humanos.
Post-traumatic stress symptoms in mothers of very low birth weight infants 2-3 years post-partum (Ahlund, Clarke, Hill, & Thalange, 2009)	Violou critério B: avaliou transtorno de estresse pós-traumático, e não DPP. Violou critério F: o período após o parto em que a avaliação foi realizada é superior a um ano. Violou critério G: a população estudada é de alto risco (recém-nascidos com baixo peso).
Retention rates and potential predictors in a longitudinal randomized control trial to prevent postpartum depression (Asunción Lara, Navarro, Navarrete, & Le, 2010)	Violou critério A: é ensaio clínico aleatorizado, e não estudo transversal ou de coorte longitudinal. Violou critério D: foi usada intervenção para prevenir DPP.

Após o preenchimento de cada uma das duas primeiras etapas, as respostas independentes de ambas as revisoras foram comparadas, obtendo-se o nível de concordância. Os casos em que foram identificadas divergências foram decididos por consenso.

5.4.1 Protocolo para avaliação do texto completo

Para a terceira etapa, em que se fez a coleta de dados dos textos completos, foi usado formulário eletrônico, hospedado no servidor Lime Service (www.limeservice.com, acessado em 14/01/2012). As respostas inseridas nos

formulários foram armazenadas em planilha eletrônica, que foi exportada em formato compatível com o *Microsoft Excel* e com *softwares* de análise de dados, como o *IBM SPSS*. O Formulário de avaliação do texto completo foi aplicado a todos os trabalhos selecionados na etapa de avaliação de resumos, ou seja, aos estudos em que as informações do resumo não indicaram violação de critérios de inclusão.

Um Formulário Piloto foi desenvolvido e testado a partir de sua aplicação a 22 estudos aleatórios (aleatorização feita pelo *software IBM SPSS Statistics 20*). A partir dessa experiência, foram identificadas necessidades de alterações para otimizar o tempo de coleta de dados, deixar os campos para preenchimento mais claros e objetivos e incluir perguntas ou respostas pré-codificadas para contemplar possibilidades que não haviam sido previstas. Após as adequações, foi criado o Formulário Final, aplicado aos 778 estudos (inclusive os 22 que já haviam sido avaliados pelo Formulário Piloto), e que continha campos para armazenar os seguintes dados:

1. **Identificação:** código do estudo (o mesmo usado nas fases de avaliação de títulos e resumos) e identificação do revisor (Gabriela, Imaculada).

2. **Avaliação do cumprimento dos critérios de inclusão:** os critérios de inclusão A, B, C, D, E, F e G foram reapresentados. Na avaliação do texto completo, também foi apresentado o critério H, que diz respeito à clareza dos resultados e completude das informações necessárias para a revisão sistemática. A coleta de dados foi encerrada nos casos em que um ou mais critérios foram marcados com “falso” ou “não está claro”. Caso o estudo estivesse adequado em relação a todos os critérios de inclusão, prosseguiu-se com a coleta de dados. Um campo para observações foi usado para justificar os motivos que levaram à exclusão dos estudos.

3. **Avaliação de qualidade metodológica:** foram apresentados os oito itens da escala de Loney et al. (1998), além de questões complementares para avaliação da adequação da metodologia dos estudos para a investigação da prevalência de DPP.

4. **Características do estudo:** objetivos (transcritos exatamente como foram apresentados no texto completo do trabalho), total de participantes, período em que foi realizada a coleta de dados, número de participantes elegíveis, contexto de pesquisa, observações.

5. **Perfil das participantes:** informações sobre raça/etnia, imigrantes, idades (presença de adolescentes e/ou mulheres com mais de 40 anos), estado civil,

paridade, histórico de depressão gestacional e depressão anterior à gestação, renda e região em que viviam (rural ou urbana). A renda foi classificada como “baixa”, “média” ou “alta”, segundo as avaliações constantes dos próprios trabalhos. Expressões como “em situação de pobreza”, ou classes socioeconômicas de acordo com critérios de poder aquisitivo (geralmente, variando de classe A a classe E), também foram usadas para estabelecer a classificação. Os trabalhos que mencionaram os valores de renda em unidade monetária local, em dólares ou salários mínimos, mas não indicaram uma classificação que permitisse inferir o nível socioeconômico dentro das categorias “baixo”, “médio” e “alto”, não foram incluídos na contagem, porque a variação entre as condições sociais e econômicas dos países, bem como as variações monetárias e de contexto ao longo do tempo, não permitiriam a comparação entre os estudos.

6. **Resultados de prevalência de DPP:** período após o parto em que a avaliação foi feita; método usado para identificar DPP (autoavaliação, entrevista clínica, autoavaliação seguida por entrevista clínica); nome do instrumento usado; ponto de corte do instrumento, para as escalas de autoavaliação; número de participantes avaliadas; número de participantes consideradas deprimidas; informações geográficas (prevalência de DPP por cidade ou país em que os dados foram coletados). Foram previstos campos para armazenamento de dados de até 12 resultados de prevalência por estudo.

Informações sobre IDH, grau de desenvolvimento do país (desenvolvido, em desenvolvimento) foram acrescentadas ao banco de dados após a finalização da coleta de dados.

5.4.2 Identificação de estudos com amostras duplicadas

Após a coleta de dados, a planilha final foi checada de forma sistemática, buscando identificar casos diferentes publicações usaram a mesma amostra. Para isso, foram comparadas as seguintes informações: nomes dos autores, data e local da coleta de dados, tamanho amostral, características das participantes e prevalência encontrada, entre outras. Nos casos em que foi identificada duplicação de amostra, apenas um dos estudos foi incluído na revisão sistemática.

5.4.3 Método para estabelecer a concordância entre revisoras na avaliação do texto completo

A concordância entre as revisoras na avaliação do texto completo foi estabelecida a partir das avaliações independentes de critérios de inclusão e Escala de Loney et al. (1998) em uma amostra do total de estudos que foram selecionados para essa etapa. Foi selecionada amostra pareada, seguindo os seguintes passos: 1. Foram selecionados todos os estudos em que a coleta de dados foi realizada, total ou parcialmente, na América Latina; 2. Foram selecionados, aleatoriamente, trabalhos em que a coleta de dados foi realizada em países não latino-americanos, em número idêntico ao de estudos latino-americanos. Para a seleção de estudos aleatórios, foi usado o *software IBM SPSS Statistics 20*.

5.5 Avaliação da qualidade metodológica dos estudos incluídos

Todos os trabalhos incluídos na presente revisão sistemática passaram por avaliação da qualidade metodológica, usando-se uma adaptação da escala de Loney et al., (1998), que pode ser visualizada na Figura 4. A cada questão com resposta afirmativa, foi atribuído um ponto, totalizando oito pontos. A opção por esse instrumento se deveu à facilidade de aplicação, visto que são apenas oito itens, e à possibilidade de se obter uma pontuação passível de ser usada em análises quantitativas.

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, se adequado?

C. Qual é a aplicabilidade dos resultados?

8. Os participantes e o contexto são descritos em detalhe e similares aos que interessam a você?

Figura 4 – Quadro apresentando os itens das diretrizes para avaliação crítica de estudos de prevalência ou incidência de um problema de saúde (Loney et al., 1998)

A Escala de Loney et al. (1998) foi desenvolvida para avaliação de qualquer estudo de prevalência, exigindo adaptações relativas ao tipo de estudo e de desfecho que está sendo avaliado. A seguir, será descrita a forma como cada item foi avaliado na presente revisão sistemática.

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 do presente estudo, 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, primeiramente, a amostra foi classificada nas seguintes categorias:

- 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. Levou-se em consideração o quanto o universo em que foi feito censo era representativo da população da 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 foi declarado na publicação. Nos casos em que o trabalho declarou que a amostra foi 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 seguintes categorias:

- 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, 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.).
- Outra: qualquer tipo de amostra que não se enquadrou 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 devido à frequente má compreensão do conceito de amostra aleatória por pesquisadores, constatada a partir de estudos que claramente o usaram de forma errônea. A citação abaixo é um exemplo de compreensão inadequada do conceito de aleatorização, pois o fato de as entrevistas terem ocorrido em qualquer dia da semana não seria suficiente para definir a amostra como aleatória:

> As entrevistas usando o SCID [Structured Clinical Interview for DSM Disorders] tiveram lugar em 25 dos 45 dias do estudo principal. Elas foram conduzidas ao longo do período do estudo principal e tiveram lugar em qualquer dia da semana, indicando que as participantes convidadas eram uma subamostra aleatória da amostra total do estudo.[27] (Stewart et al., 2009,126).

[27] No original: SCID interviews took place on 25 of the 45 days of the main study. They were conducted throughout the period of the main study and took place on any day of the week, indicating that the invited subjects were a random subsample of the total study sample.

Item 2: A base amostral foi adequada?

Considerou-se 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. Considerou-se inadequada 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 por 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 se dar início à coleta de dados de um estudo de prevalência, é preciso calcular o tamanho de amostra adequado usando-se análise de poder, levando em consideração o grau de precisão desejado, a taxa de erro aceitável e uma estimativa de prevalência da condição na população. Para exemplificar a avaliação de qualidade de estudos sobre prevalência de demência, os autores calcularam que com uma prevalência estimada de 8% (baseada em estudo anterior), taxa de erro < 3% e 95% de confiança, a amostra mínima necessária seria de 314 participantes.

Para o presente estudo, o tamanho adequado da amostra para investigar a prevalência de DPP foi calculado usando o instrumento oferecido pelo *website* LEE (Laboratório de Epidemiologia e Estatística, disponível em: http://www.lee.dante.br/pesquisa/amostragem/di_1_pro_est.html#pro-pop, acesso em 05/07/2011), que calcula o tamanho da amostra para estudos epidemiológicos a partir da prevalência estimada (a partir de estudos anteriores), precisão da estimativa (indica o quanto a estimativa pode se distanciar da verdadeira proporção) e nível de significância (porcentagem dos casos na população que estarão fora do intervalo considerado para a proporção), a partir das fórmulas indicadas por Lwanga e Lemeshow (1991). Considerando-se prevalência estimada de DPP de 12,61% (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, no entanto, é 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). Considerou-se 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). Considerou-se que escalas de autoavaliação 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 recusaram 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 participar do estudo para o qual foram convidados; 2. deixaram de participar de uma ou mais etapas de um estudo com seguimento longitudinal. Loney et al. (1998) sugeriram que uma taxa de resposta maior ou igual a 70% seria aceitável. No presente trabalho, o item 6 foi avaliado positivamente quando ambas as condições abaixo 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.

Item 8: Os participantes e o contexto foram descritos em detalhe e similares aos que interessam a você?

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.

5.5.1 Outras variáveis controladas para avaliar a qualidade metodológica dos estudos incluídos

Constatando-se as vantagens de outros métodos que têm sido usados na avaliação da qualidade metodológica de estudos epidemiológicos (Boyle, 1998; Prins et al., 2002; Bennett et al., 2004; Alhabib et al., 2010; Lewis, 2011; Shamu et al., 2011; Shamliyan et al., 2010), foram coletadas informações adicionais sobre a qualidade dos artigos sobre prevalência de DPP:

1. Tipo de instrumento para identificar DPP: considerou-se que estudos que usaram entrevista clínica têm qualidade superior aos que usaram instrumento de autoavaliação.
2. Planejamento do tamanho amostral: foi registrado se os autores do trabalho indicaram ter calculado o tamanho de amostra mínima para o estudo de prevalência, com base em análise de poder, isto é, usando parâmetros de prevalência estimada, erro máximo aceito e grau de confiança.
3. Critérios de inclusão e exclusão: foi registrado se o estudo relatou os critérios de inclusão ou exclusão, que são importantes na definição da população estudada, constituindo uma variável que pode ter implicações sobre o resultado de prevalência. Esse critério foi usado em escalas de avaliação de qualidade de estudos de prevalência criadas por outros autores (Prins et al, 2002; Bennett et al, 2004; Lewis, 2011).

5.6 Análise estatística

A metanálise é um conjunto de procedimentos estatísticos aplicados à revisão sistemática, em que a unidade de observação é o estudo científico. A metodologia das análises estatísticas aplicadas na presente Tese será descrita a seguir.

5.6.1 Definição dos dados a serem usados na análise estatística

Parte dos estudos realizou mais de uma avaliação de DPP, seja usando instrumentos diferentes ou períodos após o parto diferentes. Por exemplo, o mesmo estudo poderia ter avaliado DPP usando a Escala de Depressão Pós-Parto de

Edimburgo (EDPE) e o Inventário de Depressão de Beck (BDI). Nos estudos longitudinais, as mesmas mulheres poderiam ter sido avaliadas em diferentes momentos após o parto, como por exemplo: 30 dias, 60 dias e 180 dias. O formulário para coleta de dados do texto completo previa campos para armazenar os dados de até 12 avaliações de DPP em um mesmo estudo. No entanto, para assegurar a independência das medidas, cada estudo poderia ser considerado apenas uma vez na metarregressão. Por isso, nos estudos que relataram dados de mais de uma avaliação de DPP, foram usados os seguintes critérios para a escolha dos dados que seriam considerados para efeito de análise:

1. Entrevista clínica teve prioridade sobre autoavaliação. Ex.: se o estudo usou a EDPE e a entrevista estruturada seguindo os critérios do DSM IV, foram usados os dados da entrevista estruturada.
2. O período mais próximo do parto teve prioridade sobre o período mais distante do parto. Ex.: se o estudo avaliou DPP aos 30, 60 e 180 dias, foi escolhida a avaliação aos 30 dias. A escolha pelo período mais próximo ao parto foi feita porque, em estudos longitudinais, as mães identificadas como deprimidas são encaminhadas para tratamento medicamentoso e/ou psicoterápico, por razões éticas, o que pode contribuir para diminuir a prevalência de DPP em avaliações subsequentes.
3. Nos casos em que foi usada apenas autoavaliação, os instrumentos mais frequentes tiveram precedência sobre os menos frequentes. A frequência de uso dos instrumentos foi determinada de acordo com análise descritiva dos dados do presente trabalho.
4. Nos casos em que foram usados diferentes pontos de corte dos instrumentos, foi selecionado:
    a. O ponto de corte indicado pelos autores como critério para definir depressão menor;
    b. Se não foi indicado um ponto de corte como critério, foi selecionado o ponto de corte mais próximo da validação original do instrumento.
5. Se, após análise de todos os critérios anteriores, ocorresse empate, seria feito sorteio. No entanto, essa etapa não foi necessária.

5.6.2 Determinação da concordância entre revisoras

A inclusão ou exclusão dos estudos de acordo com os critérios estabelecidos e a avaliação de qualidade usando a escala de Loney et al. (1998) foi feita por duas revisoras, de forma independente, para que pudesse ser avaliada a confiabilidade das medidas. Em caso de divergência entre as avaliações, foi buscado consenso.

A mensuração de confiabilidade pelo índice de kappa foi descartada, pois em situações em que há uma desproporção entre a ocorrência das categorias do objeto avaliado (por exemplo, na avaliação de um critério de inclusão, haveria maior número de estudos classificados como “verdadeiro” do que como “falso”), 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). Assim, esse índice seria de difícil interpretação no presente estudo.

Optou-se por avaliar a concordância entre avaliadoras em relação aos critérios de inclusão e a cada item da escala de avaliação de qualidade, usando os dados referentes à porcentagem de estudos em que as avaliadoras apresentaram avaliações concordantes.

A concordância da pontuação total (isto é, a soma das pontuações nos oito itens) da Escala de Loney et al. (1998), por ser uma variável de contagem, foi mensurada obtendo-se o 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. Considerou-se que um valor acima de 0,7 seria aceitável.

5.6.3 Avaliação da heterogeneidade

Análise de heterogeneidade foi feita através do parâmetro do I^2, que descreve a porcentagem de variabilidade nas estimativas de efeito que se devem à heterogeneidade, e não ao erro de amostragem (acaso). Estabeleceu-se o critério de que valores acima de 50% seriam considerados como heterogeneidade substancial (Higgins & Thompson, 2002; Higgins & Green, 2011).

Visualmente, foram usados gráficos de funil (Lewis & Clarke, 2001; Anzures-Cabrera & Higgins, 2010) para avaliar a heterogeneidade. A heterogeneidade entre

os estudos foi posteriormente explorada através de análises de subgrupos, metarregressão e análises de sensibilidade.

5.6.4 Avaliação do viés de publicação

Considerou-se que a ocorrência de viés de publicação em razão dos resultados dos estudos seria improvável, pois as unidades de análise da presente revisão sistemática são estudos observacionais, em que não há resultados considerados menos favoráveis para publicação, como no caso dos delineamentos de pesquisa que envolvem intervenção. A possibilidade de viés de publicação regional foi considerada na Discussão da presente Tese (ver seção 7.2 Viés de publicação contra o Sul).

5.6.5 Síntese dos dados

Para sintetizar os resultados dos trabalhos sobre prevalência de DPP, foi feita metanálise, usando modelo de efeitos aleatórios, por entender que os estudos não foram idênticos em seus métodos e/ou suas características de inclusão de indivíduos.

A prevalência pontual de DPP em cada estudo foi calculada pela seguinte fórmula:

$$\widehat{p_{\mathrm{i}}} = \frac{d_{\mathrm{i}}}{m_{\mathrm{i}}} \qquad \text{... (1)}$$

Em que $\hat{p}_i$ é a prevalência estimada da população do i-ésimo estudo, d_i é o número de mulheres consideradas deprimidas observadas no i-ésimo estudo, m_i é o tamanho da amostra do i-ésimo estudo, em que *i* variou de 1 a *n* e *n* corresponde ao total de estudos incluídos.

Supondo que seja de interesse a prevalência geral de um grupo específico, a prevalência estimada ponderada pelo tamanho da amostra é dada por:

$$\widehat{p_p} = \frac{\sum_{\mathrm{j}}^{\mathrm{k}} \widehat{p_{\mathrm{j}}}.m_{\mathrm{j}}}{\sum_{\mathrm{j}}^{\mathrm{k}} \mathrm{m_j}}, \qquad \text{... (2)}$$

em que j=1,...k, e k é número de estudos considerados no subgrupo.

Substituindo 1 em 2, a fórmula simplificada da prevalência média ponderada foi obtida dividindo-se o número de mulheres deprimidas pelo número total de participantes em todos os estudos, ou seja:

$$\widehat{p_p} = \frac{\sum_{\mathrm{j}}^{\mathrm{k}} \mathrm{d}_j}{\sum_{\mathrm{j}}^{k} m_j} \qquad \text{... (3)}$$

Seja D_i uma variável aleatória que representa o número de mulheres deprimidas no i-ésimo estudo, em que i=1,..., n, e n é o número total de estudos. A variável D_i segue distribuição binomial (m_i, p_i), em que m_i é o número de mulheres amostradas no i-ésimo estudo e p_i a probabilidade de uma mulher do i-ésimo estudo ser considerada deprimida.

Ainda, seja $\widehat{p_i} = \frac{D_i}{m_i}$ um estimador de p_i, em que:

$$\widehat{p_i} \sim N\left(p_i\,, \frac{p_i(1-p_i)}{m_i}\right) \qquad \text{... (4)}$$

A estimativa intervalar para p_i, com 95% de confiança, é dada por:

$$IC\,(p_i\,, 95\%) = \left[\widehat{p_i} \pm 1{,}96\sqrt{\frac{\widehat{p_i}\,(1-\widehat{p_i})}{m_i}}\right] \qquad \text{... (5)}$$

Mas supondo que existe interesse em um subgrupo de estudos, e por isso se deseja obter uma prevalência ponderada pelo tamanho da amostra dos estudos, o estimador da prevalência ponderada é dado por:

$$\widehat{p_p} = \frac{\sum_{j=1}^{k} m_j\,\widehat{p_j}}{\sum_{j=1}^{k} m_j} = \frac{\sum_{j=1}^{k} m_j \frac{D_j}{m_j}}{\sum_{j=1}^{k} m_j} = \frac{\sum_{j=1}^{k} D_j}{m} \qquad \text{... (6)}$$

em que j=1,..., k, k é o número de estudos do subgrupo e m é o tamanho da amostra do subgrupo, em que $\sum_{j=1}^{k} m_j = m$.

Ainda,

$$E\left[\widehat{p_p}\right] = E\left\{\frac{\sum_{j=1}^{k} D_j}{m}\right\} = \frac{\sum_{j=1}^{k} E\{D_j\}}{m} = \frac{1}{m}\sum_{j=1}^{k} m_j p_j$$

$$Var\left[\widehat{p_p}\right] = Var\left\{\frac{\sum_{j=1}^{k} D_j}{m}\right\} = \frac{1}{m^2}\sum_{j=1}^{k} Var\{D_j\} = \frac{1}{m^2}\sum_{j=1}^{k} m_j\, p_j(1-p_j)$$

Assim,

$$\widehat{p_p} \sim N\left(\frac{1}{m}\sum_{j=1}^{k} m_j p_j\ ,\frac{1}{m^2}\sum_{j=1}^{k} m_j p_j(1-p_j)\right)$$

$$\mu_p \qquad\qquad \sigma_p^2$$

Ainda, seja os estimadores de μ_p e σ_p^2dados por:

$$\hat{\mu}_p = \frac{1}{m}\sum_{j=1}^{k} m_j \frac{d_j}{m_j} = \frac{\sum_{j=1}^{k} d_j}{m}$$

$$\hat{\sigma}_p^2 = \frac{1}{m^2}\sum_{j=1}^{k} m_j \frac{d_j}{m}\left(1 - \frac{d_j}{m_j}\right),$$

em que d_j é o número de mulheres deprimidas observadas no j-ésimo estudo.

Assim, um intervalo de confiança para p_p é dado por:

$$IC\left(p_p\ ,95\%\right) = \left[\hat{\mu}_p \pm 1{,}96\sqrt{\hat{\sigma}_p^2}\right]$$

As análises estatísticas foram no software *IBM SPSS Statistics 20*, com exceção dos gráficos de floresta e de funil, que foram feitos no *software* livre R 2.15.2 (http://www.r-project.org/, acessado em 14/11/2012), usando o pacote *Metafor*. Dois mapas para representar as prevalências de DPP em países e cidades foram feitos usando os *softwares* Arcmap 10 e ArcCatalog 10.

5.6.5.1 Análises de subgrupos

Foram estabelecidos os seguintes subgrupos para comparação:

a) tipo de instrumento para identificar DPP: instrumento de autoavaliação ou entrevista clínica;
b) grau de desenvolvimento do país onde os dados foram coletados: países desenvolvidos, países em desenvolvimento;
c) período após o parto: primeiro quadrimestre (de nove a 120 dias após o parto), segundo quadrimestre (de 121 a 240 dias após o parto), terceiro quadrimestre (de 241 a 365 dias após o parto);
d) delineamento: transversal ou de coorte longitudinal;
e) amostragem: aleatória, censitária, não aleatória ou não estava clara; outra análise foi realizada com a variável binária: adequada (amostra aleatória ou censitária) ou inadequada (outros tipos de amostra, ou estudos em que a amostragem não estava clara);
f) tamanho da amostra: menor que 470 participantes ou maior ou igual a 470 participantes;
g) planejamento do tamanho da amostra: planejou a amostra com base em análise de poder ou não planejou a amostra;
h) base amostral: usou base amostral adequada ou inadequada;
i) contexto de pesquisa: nacional, província / estado / região, cidade, instalação médica, outro, não está claro;
j) ponto de corte da EDPE: comparação dos diferentes pontos de corte da EDPE nos estudos que usaram esse instrumento;

Os países foram classificados quanto ao nível de desenvolvimento socioeconômico de acordo com a tabela de Índice de Desenvolvimento Humano (IDH) publicada pela *United Nations Development Programme* – UNDP (United Nations Development Programme, 2011). De acordo com o glossário dessa organização, o IDH é:

> Uma medida composta de realizações em três dimensões básicas do desenvolvimento humano: uma vida longa e saudável, acesso à educação e um padrão de vida decente. Para facilidade de comparação, o valor médio de realizações nessas três dimensões é colocado em uma escala de 0 a 1, em

que valores maiores são melhores, e esses indicadores são agregados usando médias geométricas.[28] (fonte: http://hdr.undp.org/en/reports/glossary/, acesso em 30/11/2012).

Foram considerados desenvolvidos os países classificados pela UNDP como "nível de desenvolvimento humano muito alto", compreendendo IDHs de 0,793 (Barbados) a 0,943 (Noruega). Nesse grupo, encontram-se 47 países, sendo que Noruega, Austrália, Holanda, Estados Unidos da América e Nova Zelândia são os cinco primeiros da lista, e apenas dois países da América Latina estão incluídos: Argentina e Chile. Foram considerados "em desenvolvimento" os países que foram classificados com índice de desenvolvimento humano "alto" (IDH de 0,698 a 0,783), "médio" (IDH de 0,522 a 0,698); "baixo" (IDH de 0,286 a 0,510); e "outros", grupo que contém países em que não foi possível calcular o IDH, devido à indisponibilidade de dados estatísticos (ex.: Coreia do Norte e Somália). O Brasil e grande parte dos demais países da América Latina, como México e Venezuela, estão incluídos no grupo com IDH considerado alto.

5.6.5.2 Metarregressão

Usando a técnica de metarregressão, buscou-se um modelo que pudesse predizer os resultados de prevalência a partir das variáveis relacionadas à metodologia do estudo. Para isso, foi usado Modelo Linear Misto Generalizado (MLMG) no *software IBM SPSS Statistics 20*. O Modelo Linear Misto Generalizado é um modelo de regressão linear que prevê a existência de dados hierárquicos, isto é, com mais de um nível de dependência entre observações.

A variável resposta do modelo foi a prevalência de DPP nos estudos incluídos. Assumiu-se distribuição normal logarítmica, visto que os dados de prevalência, por serem uma proporção, são sempre positivos ou nulos, variando entre zero e um.

Foram assumidos dois níveis hierárquicos de dependência, referentes ao país e cidade onde os estudos foram realizados. Conforme verificado nas análises de

[28] No original: A composite measure of achievements in three basic dimensions of human development — a long and healthy life, access to education and a decent standard of living. For ease of comparability, the average value of achievements in these three dimensions is put on a scale of 0 to 1, where greater is better, and these indicators are aggregated using geometric means.

heterogeneidade, essas duas variáveis atuaram como moderadoras da prevalência de DPP.

Os preditores foram as variáveis metodológicas dos estudos incluídos. Foram testados diversos modelos, diversificando as variáveis incluídas e usando como critério heurístico o menor valor de Akaike (AIC) [29]. A regressão foi feita pelo método *backward*, isto é: todos os preditores foram inseridos no modelo inicial de uma só vez, e depois foram retirados, um a um, os que apresentaram p-valor mais alto, até que restaram apenas preditores significativos (p-valor<0,05) no modelo.

O uso da cidade como um nível de dependência no modelo hierárquico implicou na perda de 103 estudos incluídos, devido a valores faltantes (*missing values*) nessa variável. Isso ocorreu porque diversos estudos não indicaram em que cidade a coleta de dados foi realizada, ou coletaram dados em mais de uma cidade. Considerou-se, no entanto, que essa variável foi relevante e deveria ser mantida no modelo, devido a três argumentos: 1. os resultados das análises descritivas (ver Figura 8 e Figura 9) e de heterogeneidade (ver Figura 11) indicaram que a cidade foi um moderador importante da prevalência de DPP; 2. os valores de Akaike obtidos nas análises preliminares foram menores usando a cidade como um dos níveis de dependência do que usando apenas o país, ou nenhum moderador; 3. os gráficos bidimensionais representando o valor predito da prevalência de DPP em relação ao valor observado ajustaram-se melhor ao modelo linear quando a cidade foi inserida como um dos níveis de dependência. Assim, a metarregressão foi realizada com uma amostra de 234 estudos.

[29] O Akaike (AIC) é uma medida da qualidade de ajuste (*goodness of fit*) relativa de um modelo estatístico. Quanto menor o seu valor, considera-se que o modelo está mais ajustado.

5.6.5.3 Análises de sensibilidade

Foram realizadas as seguintes análises de sensibilidade:

1. Cálculo da prevalência média omitindo teses e dissertações.
2. Cálculo da prevalência média omitindo estudos publicados até 31 de dezembro de 1989. Essa análise foi realizada para checar se o período de publicação interferiria nos resultados da metanálise, visto que até a década de 1990, havia poucos estudos de países em desenvolvimento publicados sobre prevalência de DPP e as escalas de autoavaliação ainda não haviam sido amplamente estudadas (Melo Junior, 2011).
3. Cálculo da prevalência média omitindo estudos que não declararam ter como objetivo avaliar prevalência de DPP.
4. Cálculo da prevalência média de DPP omitindo estudos que tiveram pontuação na Escala de Loney et al. (1998) abaixo da mediana.

6 Resultados

O presente capítulo apresentará, inicialmente, os resultados da busca bibliográfica e seleção de estudos. Em seguida, serão apresentadas as análises descritivas das características dos estudos incluídos e excluídos, os resultados da avaliação da qualidade metodológica dos trabalhos incluídos e as análises descritivas da prevalência de DPP. Após as análises descritivas, serão apresentadas as avaliações do nível de heterogeneidade entre os estudos e as análises de subgrupos, que testaram as hipóteses da presente Tese; e o resultado da metarregressão, que consiste em análise multivariada da influência de fatores metodológicos e regionais sobre os resultados de prevalência. Por fim, serão apresentadas as análises de sensibilidade.

6.1 Resultados da busca bibliográfica

As buscas bibliográficas resultaram em 3.909 registros. A Tabela 7 mostra as datas das consultas e o número de registros recuperados em cada base de dados. Após eliminar duplicatas decorrentes de sobreposições entre bases de dados, restaram 1.881 registros de estudos originais.

Tabela 7 – Datas das buscas eletrônicas e número de registros recuperados.

Base	Data	Número de registros
BDTD	09/10/2011	17
CINAHL	17/10/2011	178
Portal Nacional BVS Brasil em Saúde	19/10/2011	405
EMBASE	30/09/2011	582
ERIC	03/10/2011	5
ISI Web of Knowledge	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

6.2 Seleção de estudos

A seleção dos estudos para inclusão na revisão sistemática obedeceu a três etapas: avaliação de títulos, avaliação de resumos e avaliação dos textos completos. O fluxograma apresentado na Figura 5 mostra o número de trabalhos selecionados e eliminados em cada etapa.

Na primeira etapa de seleção dos estudos, os títulos foram julgados de forma independente por duas revisoras, de acordo com os critérios de inclusão estabelecidos, excluindo os registros em que foi possível identificar claramente que não se adequaram a um ou mais critérios. Obteve-se 78,6% de concordância entre

os títulos selecionados. Na próxima etapa (avaliação dos resumos), foram avaliados os trabalhos cujos títulos foram considerados pertinentes por ambas as revisoras, adicionando-se os 21,4% de trabalhos em que ocorreu discordância entre as avaliadoras.

Após a eliminação de estudos pelos títulos, foram selecionados 1.314 registros, cujos resumos foram avaliados independentemente por ambas as revisoras, de acordo com os mesmos critérios de inclusão. Houve 80,8% de concordância entre os resumos avaliados. Os trabalhos cujos resumos foram considerados adequados por ambas as revisoras, mais os 19,2% de estudos em que ocorreram divergências, foram selecionados para a próxima etapa (avaliação do texto completo). No total, foram selecionados os resumos de 893 trabalhos.

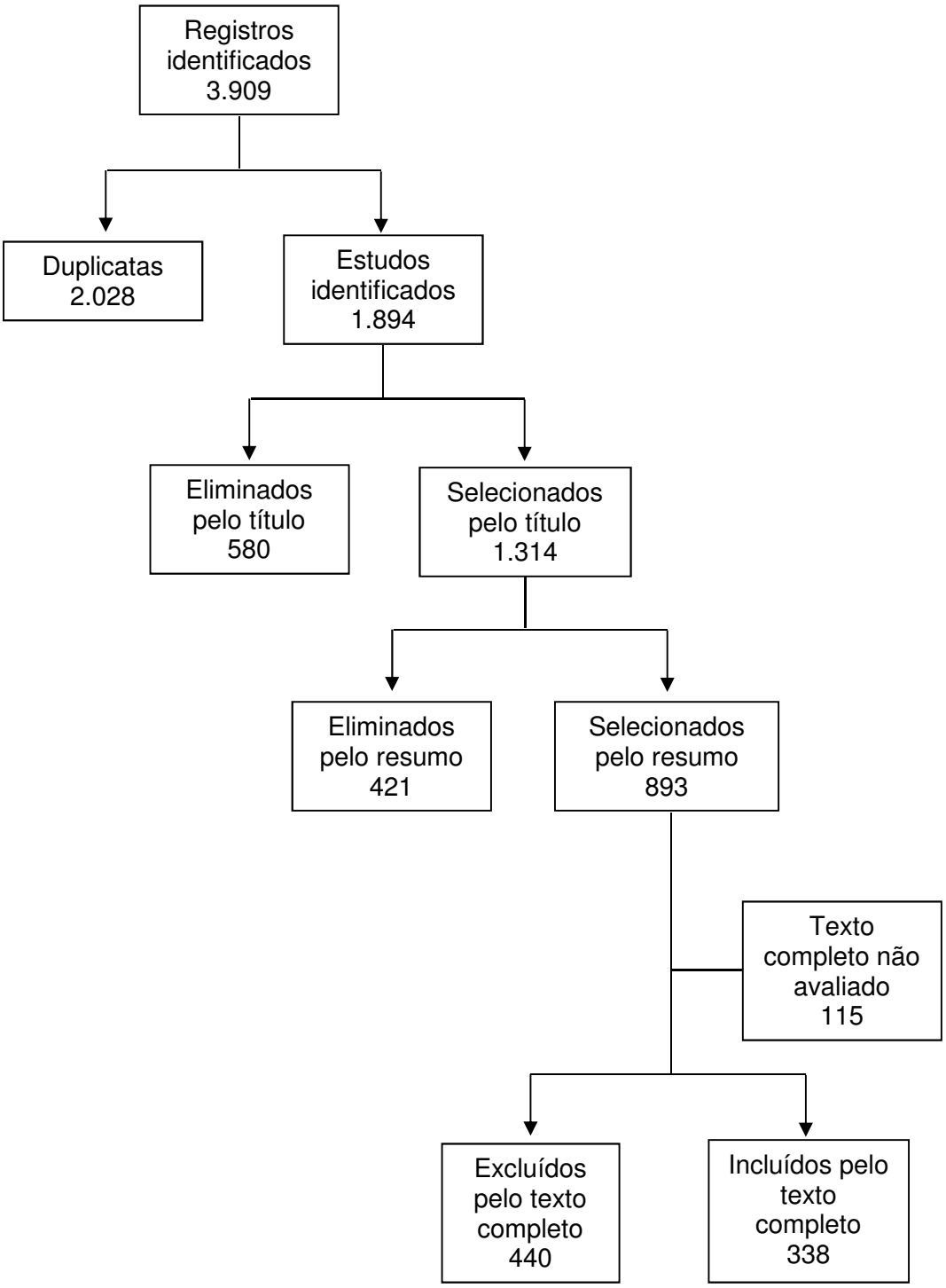

Figura 5 – Fluxograma de seleção de estudos para a revisão sistemática.

A Tabela 8 mostra a taxa de concordância entre avaliadoras, calculada a partir de avaliações cegas de 100% dos títulos e resumos por ambas as revisoras. A decisão final sobre selecionar ou não os estudos para a etapa seguinte teve concordância maior na avaliação dos resumos que na avaliação dos títulos. Porém, a concordância em cada item foi menor para os resumos do que para os títulos dos trabalhos, o que provavelmente decorreu da forma como foram feitas as avaliações. Ao avaliar os títulos, as revisoras observaram se o trabalho violou um ou mais dos sete critérios estabelecidos, indicando o código "F" (falso) na célula correspondente ao critério não atendido. A avaliação dos resumos foi mais detalhada, preenchendo-se cada um dos sete primeiros critérios de inclusão com as seguintes alternativas: verdadeiro, falso, não está claro, não se aplica. O maior número de opções gerou maior divergência na avaliação de cada critério.

Na próxima etapa, iniciou-se a avaliação dos 893 estudos selecionados na etapa de avaliação de resumos. A maioria deles foi localizada em formato eletrônico, por acesso aberto ou pelo Portal de Periódicos da Capes, e 122 trabalhos (13,7%) foram obtidos via Sistema de Comutação Bibliográfica (Comut) e através do Serviço de Acesso ao Documento (SCAD). Foram avaliados 778 textos, que corresponderam a 87,1% dos 893 trabalhos selecionados pelo resumo.

A avaliação de todos os textos completos foi realizada pela primeira revisora. A segunda revisora avaliou uma amostra dos 893 estudos selecionados para verificação da taxa de concordância, calculada de forma a contemplar todos os 69 estudos cuja coleta de dados ocorreu parcialmente ou totalmente em países da América Latina; e 69 selecionados aleatoriamente dentre os que coletaram dados em outros países[30]. A taxa de concordância entre as revisoras na avaliação dos textos completos foi apresentada na Tabela 8. A avaliação dos critérios de inclusão pela leitura do texto completo apresentou concordância entre revisoras maior que 70% em todos os itens, para as amostras latino-americana e de outros países, sendo considerada aceitável.

[30] Em ambos os grupos, foram inseridos apenas estudos que não apresentaram sobreposição total ou parcial de amostra.

Tabela 8 – Concordância entre avaliadoras quanto à inclusão de estudos pelos títulos, resumos e textos completos.

Critério de inclusão	**Títulos** (N=1.894) %	**Resumos** (N=1.314) %	**Texto completo: América Latina** (N=69) %	**Texto completo: outros países** (N=69) %
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 D: amostra não foi submetida a 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

Foram excluídos 39 estudos pelo Critério I (idioma do texto completo), que corresponderam a 4,4% dos trabalhos. Outros 29 trabalhos não foram localizados em forma eletrônica e não foram solicitados via Comut por três motivos: 1) eram textos provenientes de fontes que, em geral, não publicam pesquisas empíricas, como: revistas de divulgação científica, livros e capítulos de livro; 2) eram resumos em anais de conferência, provavelmente com informações insuficientes sobre metodologia e resultados; 3) foi possível identificar que não atendiam ao critério A de inclusão (por exemplo, revisões de literatura e pesquisa qualitativa), de acordo com informações consultadas nas bases de dados PubMed e PsycInfo. Somando-se os estudos avaliados (87,1%) e os não avaliados porque não atenderiam aos critérios de inclusão (7,6%), foram revisados 94,7% dos estudos. Os demais não foram avaliados devido a questões de ordem prática[31].

Foram observadas diferenças significativas ($F_{(2)}$ = 16,150, p<0,001) no Modelo Linear Geral entre as médias dos anos de publicação de três categorias de estudos: os que não foram encontrados; os que foram encontrados e avaliados; e os que foram excluídos devido ao idioma de publicação. As diferenças entre a categoria dos trabalhos não encontrados e as demais foram estatisticamente significativas no post-hoc de Bonferroni (p<0,001). Estudos que não foram encontrados foram publicados, em média, em 2000,8 (DP = 7,2), sendo significativamente mais antigos que estudos encontrados (média = 2005,2, DP = 5,2) e que os excluídos devido ao idioma (média = 2003,4, DP = 5,8).

6.3 Análises descritivas

As análises descritivas contemplaram as principais características dos estudos incluídos e excluídos, a avaliação da qualidade metodológica e as prevalências de DPP encontradas nos trabalhos incluídos. A apresentação dos dados de prevalência nessa etapa proporciona informações gerais sobre os

[31] Os 47 trabalhos restantes foram solicitados via Comut, mas não estavam disponíveis por essa via ou não puderam ser obtidos até o dia 9 de novembro de 2012, data em que a coleta de dados foi encerrada.

resultados dos estudos, mas não deve ser tomada como resultado final, visto que não considerou a heterogeneidade metodológica, que foi contemplada na etapa de metarregressão.

6.3.1 Características dos estudos incluídos

Dos 778 estudos avaliados quanto ao texto completo, foram incluídos 337 na presente revisão sistemática, cujas referências bibliográficas foram apresentadas no Apêndice 2. As amostras dos estudos incluídos totalizaram 296.139 participantes. Esses trabalhos foram publicados entre 1980 e 2012 (ver Figura 6) e a mediana do ano de publicação foi 2007, ou seja, metade foi publicada nos últimos cinco anos.

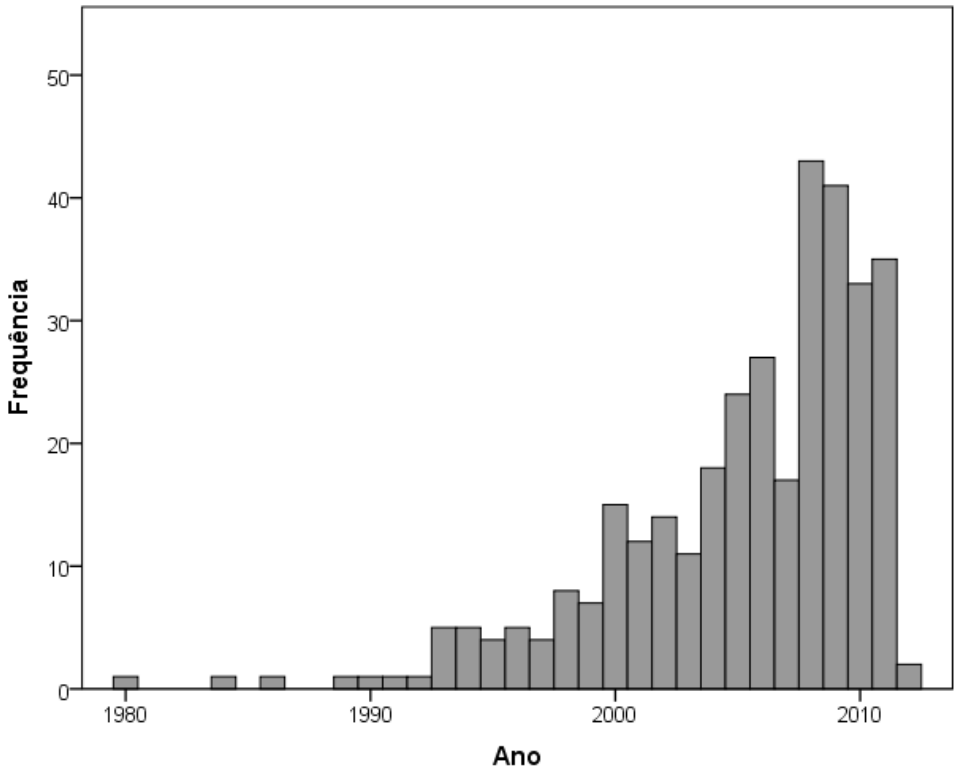

Figura 6 – Histograma representando as frequências de ano de publicação dos estudos incluídos.

Dentre os trabalhos incluídos, encontram-se 307 artigos em periódicos científicos, 13 dissertações de mestrado e 17 teses de doutorado. Os 307 artigos constavam de 147 periódicos científicos. Dentre eles, dois se destacaram por serem os mais frequentes meios de publicação dos estudos incluídos: *Journal of Affective Disorders* (23 artigos) e *Archives of Women's Mental Health* (20 artigos).

O idioma mais frequente para o texto completo dos estudos incluídos foi inglês (303 estudos, 90,0%), 18 (5,3%) estavam em português e 16 (4,7%) em espanhol. A maior parte dos trabalhos incluídos apresentou resumo em inglês (329, 97,6%), 16 (4,8%) apresentaram resumo em português, 14 (4,2%) em espanhol, cinco (1,5%) em francês, um (0,3%) em alemão, um (0,3%) em árabe, um (0,3%) em japonês e um em italiano (0,3%). Três estudos incluídos não apresentaram resumo.

Os estudos incluídos foram realizados em 61 países. A maior parte coletou dados nos Estados Unidos da América, seguido por Austrália e Brasil. Cinco realizaram coleta de dados em mais de um país, através de colaborações internacionais. A Tabela 9 mostra o número de estudos incluídos por país.

Classificando os países de acordo com critérios da UNDP (UNDP, 2011), 238 (70,6%) trabalhos coletaram dados em países desenvolvidos, 94 (27,9%) em países em desenvolvimento e três (0,9%) apresentaram dados coletados em países desenvolvidos e em desenvolvimento. Dentre as 296.139 participantes, 263.951 eram de países desenvolvidos, 31.068 eram de países em desenvolvimento e o restante das participantes compuseram as amostras dos três estudos mistos. Os anos de publicação de estudos publicados em países em desenvolvimento (média = 2006,9, DP = 3,6) foram, em média, mais recentes (T_{330} = -4,507, $p<0,001$) que os publicados em países desenvolvidos (média = 2004,6, DP = 5,7).

Os objetivos mais frequentes declarados pelos estudos incluídos foram: investigar fatores de risco para DPP (52,2%), investigar prevalência de DPP (48,1%), validar um instrumento de avaliação de DPP (8,6%) e investigar consequências da DPP (4,2%). Um mesmo estudo poderia apresentar mais de um objetivo.

Tabela 9 – Número de estudos incluídos por país (continua)

País	**N**	**%**
Estados Unidos da América	81	24,0
Austrália	26	7,7
Brasil	26	7,7
Reino Unido	25	7,4
Canadá	17	5,0
China	12	3,6
Turquia	11	3,3
Itália	10	3,0
Suécia	10	3,0
Chile	8	2,4
Espanha	8	2,4
Japão	8	2,4
Israel	7	2,1
Portugal	7	2,1
Noruega	6	1,8
Hong Kong	5	1,5
Irã	5	1,5
Irlanda	5	1,5
Alemanha	4	1,2
Grécia	4	1,2
Holanda	4	1,2
Paquistão	4	1,2
Suíça	4	1,2
África do Sul	3	0,9
Emirados Árabes Unidos	3	0,9
Índia	3	0,9
México	3	0,9
Nigéria	3	0,9
Nova Zelândia	3	0,9
Bangladesh	2	0,6
Colômbia	2	0,6
Coreia do Sul	2	0,6
Finlândia	2	0,6
França	2	0,6
Hungria	2	0,6
Malaísia	2	0,6
Nepal	2	0,6
Peru	2	0,6
Polônia	2	0,6
Singapura	2	0,6
Uganda	2	0,6

Tabela 9 – Número de estudos incluídos por país (conclusão)

País	**N**	**%**
Vietnam	2	0,6
Arábia Saudita	1	0,3
Armênia	1	0,3
Áustria	1	0,3
Bélgica	1	0,3
Dinamarca	1	0,3
Egito	1	0,3
Guiana	1	0,3
Islândia	1	0,3
Jamaica	1	0,3
Jordânia	1	0,3
Líbano	1	0,3
Lituânia	1	0,3
Malta	1	0,3
Marrocos	1	0,3
Mongólia	1	0,3
Tailândia	1	0,3
Venezuela	1	0,3
Zâmbia	1	0,3
Zimbábue	1	0,3
Base (total de estudos incluídos)[32]	337	100,0

As características metodológicas dos trabalhos incluídos foram apresentadas na Tabela 10. O delineamento da maior parte dos estudos foi de coorte longitudinal, e a maioria usou amostragem não probabilística. Mais de um terço dos estudos não apresentou informações claras sobre o método de amostragem. Quase três quartos usaram apenas instrumentos de autoavaliação para identificar as participantes com DPP. Os contextos de pesquisa mais comuns foram instalações médicas ou cidades.

[32] Alguns estudos realizaram a coleta de dados em mais de um país, de forma que a soma das frequências é maior que o total de 337 estudos.

Tabela 10 – Características metodológicas dos estudos incluídos.

Características	**N**	**%**
Delineamento		
• Transversal	156	46,3
• Coorte longitudinal	181	53,7
Amostragem		
• Censo	14	4,2
• Amostra aleatória	17	5,0
• Amostra sistemática	87	25,9
• Amostra de conveniência	83	24,6
• Outra[33]	11	3,3
• Descrição não estava clara	125	37,1
Método usado para identificar DPP		
• Instrumento de autoavaliação	248	73,5
• Entrevista clínica	19	5,6
• Misto (autoavaliação + entrevista clínica)	70	20,8
Contexto de pesquisa		
• Nacional	13	4,0
• Província / estado / região	51	15,7
• Cidade	108	33,2
• Instalação de saúde	109	33,5
• Outros[34]	15	4,6
• Não está claro	29	8,9

O tamanho de amostra variou entre 16 e 51.850 participantes, a média foi 874,2, o desvio padrão foi 3.124,2 e a mediana foi 261,0. Na presente revisão sistemática, considerou-se que o tamanho de amostra adequado para os estudos de prevalência de DPP seria acima de 470 participantes, e 116 (34,4%) atenderam a essa condição. Observou-se que a maior parte dos estudos usou amostras menores que 1.000 participantes, mas foram observados *outliers*. Um estudo apresentou tamanho de amostra de 51.850 participantes (Centers for Disease Control and Prevention, 2008), sendo considerado o mais discrepante em relação ao conjunto. Ele foi realizado em 17 estados dos EUA, usando dados do *Pregnancy Risk Assessment Monitoring*

[33] Outras técnicas de amostragem: amostra estratificada, cotas.

[34] Outros contextos: estudos realizados em mais de um país, dados coletados em duas ou mais instalações médicas em estados ou países diferentes, dados coletados por internet.

System (PRAMS), que é parte do *Centers for Disease Control and Prevention* (CDC), uma iniciativa com o objetivo de reduzir a mortalidade infantil e o baixo peso ao nascer (Gilbert, Shulman, Fischer, & Rogers, 1999).

Embora a maior parte dos estudos incluídos (94,1%) tenha relatado características de perfil das participantes, a forma de relatar o perfil e quais variáveis foram mensuradas variou entre os trabalhos. Foi registrado na presente revisão se o estudo incluiu informações sobre o nível de renda ou classe social das participantes; e a região em que residiam (urbana ou rural), que foram apresentados na Tabela 11. Grande parte dos trabalhos incluídos não continham essas informações.

Tabela 11 – Perfil das participantes dos estudos incluídos: renda e região urbana ou rural.[35]

Característica	**N**	**%**
Renda ou classe social baixa	64	19,0
Renda ou classe social média	42	12,5
Renda ou classe social alta	32	9,5
Residentes em região urbana	73	21,7
Residentes em região rural	55	16,3

Outras características referentes ao perfil das participantes dos estudos incluídos foram apresentadas na Tabela 12. Também foi observada ausência desses dados em grande parte dos trabalhos incluídos. Quase um terço dos trabalhos declarou que pelo menos parte da amostra foi composta por minorias étnicas. A maioria dos trabalhos (72,4%) não declarou se as participantes tinham histórico de depressão ou se apresentaram depressão na gestação, variáveis consideradas como importantes fatores de risco para DPP.

[35] Um mesmo estudo pode estar incluído em mais de uma categoria. Por ex.: estudos que declararam ter entrevistado puérperas de renda baixa, média e alta foram contabilizados nas três categorias.

Tabela 12 – Perfil das participantes dos estudos incluídos.

Características	**Não apresentou participantes com essa característica**		**Parte da amostra apresentou essa característica**		**Toda a amostra apresentou essa característica**		**Dados não foram apresentados ou não estão claros**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Pertencentes a comunidade indígena/aborígine	12	3,6	22	6,5	-	-	303	89,9
Pertencentes a outras minorias étnicas (em relação ao perfil da população local)	7	2,1	102	30,3	5	1,5	223	66,2
Imigrantes	7	2,1	47	13,9	2	0,6	281	83,4
Adolescentes (até 18 anos)	107	31,8	99	29,4	2	0,6	129	38,3
Mulheres acima de 40 anos	58	17,2	127	37,7	-	-	152	45,1
Mulheres casadas ou em relação estável	-	-	202	59,9	45	13,4	90	26,7
Primíparas	2	0,6	228	67,7	27	8,0	80	23,7
Com histórico de depressão	15	4,5	77	22,8	1	0,3	244	72,4
Que tiveram depressão durante a gestação	3	0,9	60	17,8	-	-	274	81,3

6.3.2 Características dos estudos excluídos

Após a avaliação por texto completo, foram excluídos 359 (46,1%) estudos, além de outros 80 (10,3%) que usaram amostras parcialmente ou totalmente idênticas às de outros estudos que foram incluídos na revisão. Assim, foi excluído após leitura do texto completo um total de 439 (56,4%) trabalhos[36].

A Tabela 13 mostra o número de estudos que foram excluídos em razão de cada um dos oito critérios de inclusão/exclusão. O critério A (delineamento) foi o principal motivo para a exclusão, seguido pelo critério B (estudos que não avaliaram a prevalência de DPP) e pelo critério F (período após o parto entre 8 e 365 dias). Foram excluídos 20 (4,4%) trabalhos pelo critério H, ou seja, porque o resultado de prevalência não foi apresentado de forma clara.

Dentre os estudos excluídos, alguns realizaram a coleta de dados em países que não foram contemplados dentre os estudos incluídos: Burkina Faso, Costa Rica, Etiópia, Gâmbia, Gana, Honduras, Indonésia e Malawi, todos considerados em desenvolvimento (UNDP, 2011). Não foram encontradas diferenças estatisticamente significativas entre os estudos incluídos e excluídos quanto ao ano de publicação, idioma do texto completo e nível de desenvolvimento socioeconômico do país em que a coleta de dados foi realizada.

[36] As referências bibliográficas completas e a razão pela qual cada estudo foi excluído estão disponíveis mediante contato com a presente autora, pelo e-mail: metanalise.dpp@gmail.com.

Tabela 13 – Motivos que levaram à exclusão de estudos.

Critério de inclusão	Avaliação "falso"		Avaliação "não está claro"	
	N	%	N	%
Critério A: estudo primário e quantitativo, delineamento transversal ou de coorte longitudinal	111	30,9	3	0,8
Critério B: avaliou prevalência de DPP	89	24,8	22	6,1
Critério C: participantes são mães humanas	2	0,6	1	0,3
Critério D: amostra não foi submetida a intervenção com objetivo de prevenir DPP	4	1,1	3	0,8
Critério E: método de diagnóstico mencionado claramente e baseia-se em instrumento objetivo de medida, ou em avaliação feita por profissional de saúde mental	16	4,5	8	2,2
Critério F: pelo menos uma das avaliações de DPP ocorreu de 8 dias (inclusive) a 365 dias (inclusive) após o parto	65	18,1	57	15,9
Critério G: O estudo foi realizado na população geral	17	4,7	1	0,3
Critério H: Em pelo menos uma das avaliações de prevalência de DPP, o resultado de prevalência está relatado de forma clara	20	4,4	-	-
Base[37]	459			

[37] Foram tomados como base os estudos excluídos e que não tinham sobreposição total ou parcial de amostra com estudos incluídos. Um estudo poderia ser classificado como "falso" ou "não está claro" em mais de um critério de inclusão.

6.3.3 Qualidade metodológica dos estudos incluídos

A qualidade metodológica dos estudos foi avaliada pela escala de Loney et al. (1998), cuja pontuação varia de zero a oito. A Tabela 14 mostra a concordância entre as avaliadoras para cada item da escala. A amostra para estabelecer o nível de concordância foi formada pelos estudos que atenderam a dois requisitos: 1. foram avaliados pelas duas revisoras; e 2. foram incluídos na presente revisão sistemática.

A maior parte dos itens alcançou concordância maior que 70%, considerada aceitável, mas o item 7 (apresentação dos resultados de prevalência com intervalo de confiança e detalhamento de subgrupos) apresentou taxas de concordância de 60,2% para amostra da América Latina e 50,0% para amostra de estudos de outros países. A pontuação final da escala, formada pela soma das pontuações nos oito itens, obteve coeficiente de correlação intraclasse de 0,508 (IC 95% 0,231-0,685). Esses resultados serão discutidos na seção “7.3 Concordância entre revisoras” do presente trabalho.

Tabela 14 – Concordância entre avaliadoras para os itens da Escala de Loney et al. (1998).

	Nível de concordância	
Item	**América latina N=43**	**Outros países N=36**
1. Delineamento e método de amostragem	83,7	97,2
2. Base amostral (*sampling frame*)	86,0	97,2
3. Tamanho da amostra	90,7	86,1
4. Apresentação de medidas de confiabilidade e validade do instrumento	72,1	77,8
5. Desfecho medido de forma a apresentar baixo risco de viés	100,0	100,0
6. Taxa de resposta adequada	79,1	83,3
7. Apresentação dos resultados de prevalência com intervalo de confiança e detalhamento de subgrupos	60,5	50,0
8. Apresentação do perfil sociodemográfico da amostra estudada	93,0	91,7

A média de qualidade, aferida pela escala de Loney et al. (1998), foi de 2,0, com desvio-padrão de 1,1. A distribuição das pontuações dos estudos na escala foi apresentada na Figura 7. Observa-se que a maior parte dos trabalhos avaliados (87,8%) se concentrou na faixa de pontuação de 1 a 3. Assim, pode-se considerar que a qualidade dos estudos incluídos foi fraca.

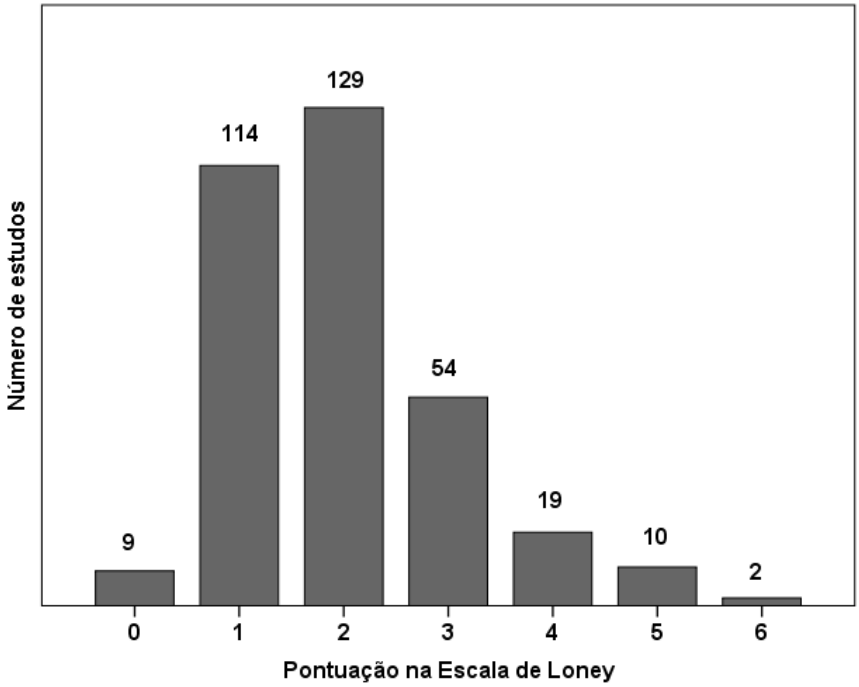

Figura 7 – Histograma representando a distribuição das pontuações dos estudos na Escala de Loney et al. (1998).

A Tabela 15 mostra o número de estudos que obtiveram pontuação em cada item da escala de Loney et al. (1998), separando os trabalhos de acordo com o grau de desenvolvimento do país onde os dados foram coletados. O item 8 (apresentação do perfil sociodemográfico da amostra estudada) foi aquele em que mais trabalhos obtiveram pontuação (94,1%), enquanto no item 5, o menor número de trabalhos obteve pontuação (1,8%).

Tabela 15 – Número de estudos que obtiveram pontuação em cada item da Escala de Loney et al. (1998).

Item	**Total**		**Países desenvolvidos**		**Países em desenvolvimento**	
	N	**%**	**N**	**%**	**N**	**%**
1. Delineamento e método de amostragem	29	8,6	19	7,9	10	10,8
2. Base amostral (*sampling frame*)*	23	6,8	12	5,0	11	11,8
3. Tamanho da amostra*	116	34,4	93	38,6	22	23,7
4. Apresentação de medidas de confiabilidade e validade do instrumento*	90	26,7	55	22,8	33	35,5
5. Desfecho medido de forma a apresentar baixo risco de viés	6	1,8	4	1,7	2	2,2
6. Taxa de resposta adequada*	51	15,1	42	17,4	8	8,6
7. Apresentação dos resultados de prevalência com intervalo de confiança e detalhamento de subgrupos*	40	11,9	22	9,1	18	19,4
8. Apresentação do perfil sociodemográfico da amostra estudada	317	94,1	223	92,5	91	97,8

* Item com diferença estatisticamente significativa entre países desenvolvidos e em desenvolvimento ($p<0,05$ no teste de qui-quadrado).

Não foi observada diferença estatisticamente significativa entre as médias das pontuações na Escala de Loney et al. (1998) de acordo com o nível de desenvolvimento socioeconômico do país (desenvolvido ou em desenvolvimento), resultado que está de acordo com a Hipótese 4A do presente trabalho, que previu que a qualidade dos estudos realizados em países desenvolvidos seria equivalente à qualidade dos estudos realizados em países em desenvolvimento. No entanto, comparando países desenvolvidos e em desenvolvimento em cada item da Escala de Loney et al. (1998) e usando testes de qui-quadrado, foram encontradas diferenças significativas nos itens 2 (base amostral), 3 (tamanho da amostra), 4 (confiabilidade e validade do instrumento), 6 (taxa de resposta) e 7 (intervalo de confiança e análise de subgrupos). Países em desenvolvimento usaram com maior frequência bases amostrais adequadas (Item 2, $\chi^2_{(1)}$=4,909, p=0,027); relataram com maior frequência a confiabilidade e validade do instrumento de medida (Item 4, $\chi^2_{(1)}$=5,544, p=0,019); e relataram com maior frequência o intervalo de confiança e os resultados de análises de subgrupos (Item 7, $\chi^2_{(1)}$=6,657, p=0,010). Entre os estudos de países desenvolvidos, por sua vez, foi mais comum o uso de amostras maiores ou iguais a 470 participantes (Item 3, $\chi^2_{(1)}$=6,629, p=0,010); e taxas de resposta consideradas adequadas (≥70%), com descrição das características das participantes perdidas (Item 6, $\chi^2_{(1)}$=4,106, p=0,043).

Não foi observada correlação entre a pontuação total na Escala de Loney et al. (1998) e o ano de publicação. Esse resultado contrariou a Hipótese 4B do presente trabalho, que previu que estudos mais recentes teriam maior qualidade que os mais antigos.

Além dos resultados da Escala de Loney et al. (1998), é possível ter um panorama da qualidade dos trabalhos considerando-se as análises descritivas de fatores metodológicos: técnica de amostragem, tipo de avaliação de DPP, planejamento do tamanho da amostra e apresentação dos critérios de inclusão e/ou exclusão de participantes. Essas análises serão apresentadas a seguir:

1. Técnica de amostragem

Conforme apresentado na Tabela 10, a maior parte dos trabalhos (53,8%) usou amostra não aleatória e em 37,1%, a técnica de amostragem não estava clara. Dentre eles, muitos informaram que usaram amostra aleatória, mas não esclareceram qual foi o método de aleatorização.

A tabela de níveis de evidência científica da Oxford Centre for Evidence-Based Medicine (OCEBM Levels of Evidence Working Group, 2011) indica que as pesquisas censitárias ou com amostragem aleatória, realizadas no local de interesse, ocupam o nível 1 de evidência em estudos de prevalência. De acordo com esse critério, apenas 9,2% dos estudos incluídos podem ser considerados evidências de nível 1. Os demais trabalhos devem ser considerados evidências de nível 3 (amostras locais e não aleatórias).

2. Tipo de instrumento para identificar DPP

A maior parte dos estudos (73,6%) usou instrumentos de autoavaliação para avaliar quais mulheres apresentavam DPP, técnica considerada menos adequada em termos metodológicos que a aplicação de entrevista clínica. Constatou-se que 5,6% dos estudos aplicaram entrevista clínica a todas as participantes e os outros 20,8% fizeram o rastreamento de possíveis casos de DPP usando escalas de autoavaliação e aplicando entrevista clínica apenas às mulheres que tiveram resultado positivo (ver Tabela 10).

3. Planejamento do tamanho amostral

A maioria dos estudos (85,5%) não declarou ter planejado o tamanho da amostra com base em análise de poder. Trabalhos de países em desenvolvimento declararam ter planejado a amostra com maior frequência (21,5%) que os de países desenvolvidos (11,2%), uma diferença estatisticamente significativa ($\chi^2_{(1)}$=5,890, p=0,015).

4. Apresentação dos critérios de inclusão e/ou exclusão de participantes

Os critérios de inclusão e/ou exclusão de participantes foram apresentados pela maior parte dos estudos incluídos (75,4%), sem diferença estatisticamente significativa entre países desenvolvidos e em desenvolvimento.

6.3.4 Prevalência de DPP

Dentre os 337 estudos incluídos, 102 (30,3%) avaliaram DPP mais de uma vez na mesma amostra, por duas razões: 1. foi realizado seguimento longitudinal das participantes; 2. foi usado mais de um instrumento para identificar DPP. Nesses casos, foi preciso selecionar os valores de prevalência que seriam usados nas análises, de acordo com as regras descritas na seção "5.6.1 Definição dos dados a serem usados na análise estatística". Assim, a cada estudo incluído correspondeu um único valor de prevalência no banco de dados, eliminando-se a ocorrência de medidas repetidas.

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%). Esse valor é superior ao que havia sido previsto pela Hipótese 1 do presente trabalho.

As médias ponderadas das estimativas pontuais de prevalência em cada país e cidade foram mapeadas nas Figuras 8 e 9, respectivamente, e apresentadas no Apêndice 3. Os resultados por país também foram representados em um gráfico de floresta, conforme a Figura 10. Da mesma forma que a média ponderada das prevalências dos estudos, essas figuras têm caráter ilustrativo e não devem ser tomadas como base para nortear decisões referentes à gestão em saúde ou à prática profissional, visto que os estudos foram heterogêneos e outras variáveis precisam ser consideradas.

Na Figura 8, observa-se que as prevalências de DPP foram geralmente elevadas em países latino-americanos, asiáticos e do Leste Europeu. Poucos

países da África apresentaram estudos sobre prevalência de DPP. As prevalências médias nesse continente foram bastante heterogêneas, apresentando-se elevadas na África do Sul, Zimbabwe e Egito, e menores na Nigéria e Zâmbia. Prevalências menores de DPP foram encontradas na Austrália, no Canadá e em alguns países da Europa Ocidental, como Portugal, França e Alemanha, mas Estados Unidos da América, que também é um país desenvolvido, apresentou prevalência média elevada.

Na Figura 9, percebe-se que estudos sobre DPP estão mais concentrados em cidades da Europa, leste dos Estados Unidos da América e Sul e Sudeste do Brasil. Poucas cidades da África estiveram representadas nos estudos de prevalência de DPP, e a maior parte encontrou prevalência alta. Comparando a Figura 8 com a Figura 9, observa-se que a síntese de prevalências em estimativa única por país não se aproxima, necessariamente, das prevalências das cidades, que são heterogêneas. Mesmo em cidades próximas, as prevalências de DPP foram heterogêneas e podem ser encontradas representações sobrepostas de círculos maiores e menores.

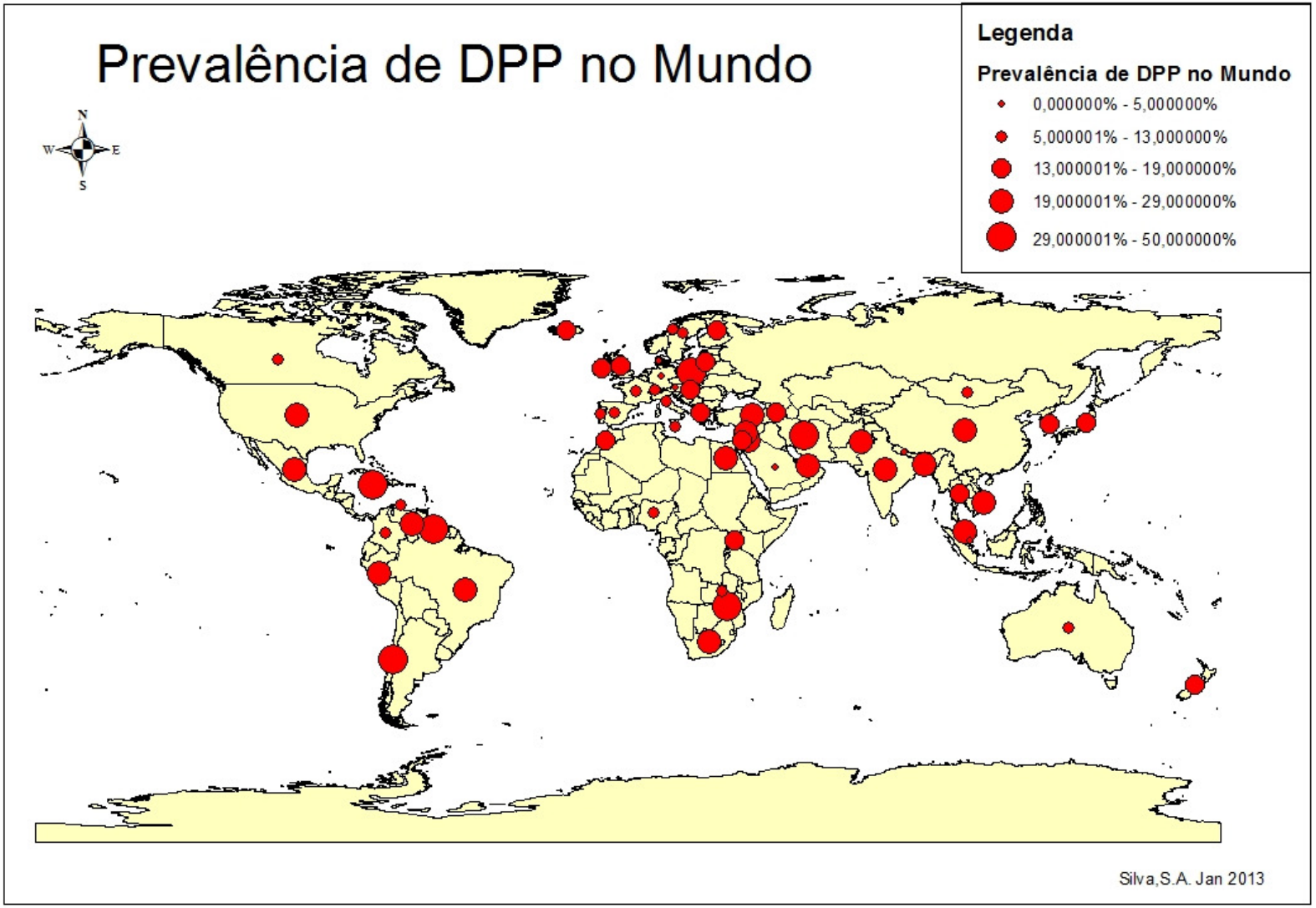

Figura 8 – Mapa indicando as médias das estimativas pontuais de prevalência de DPP por país.

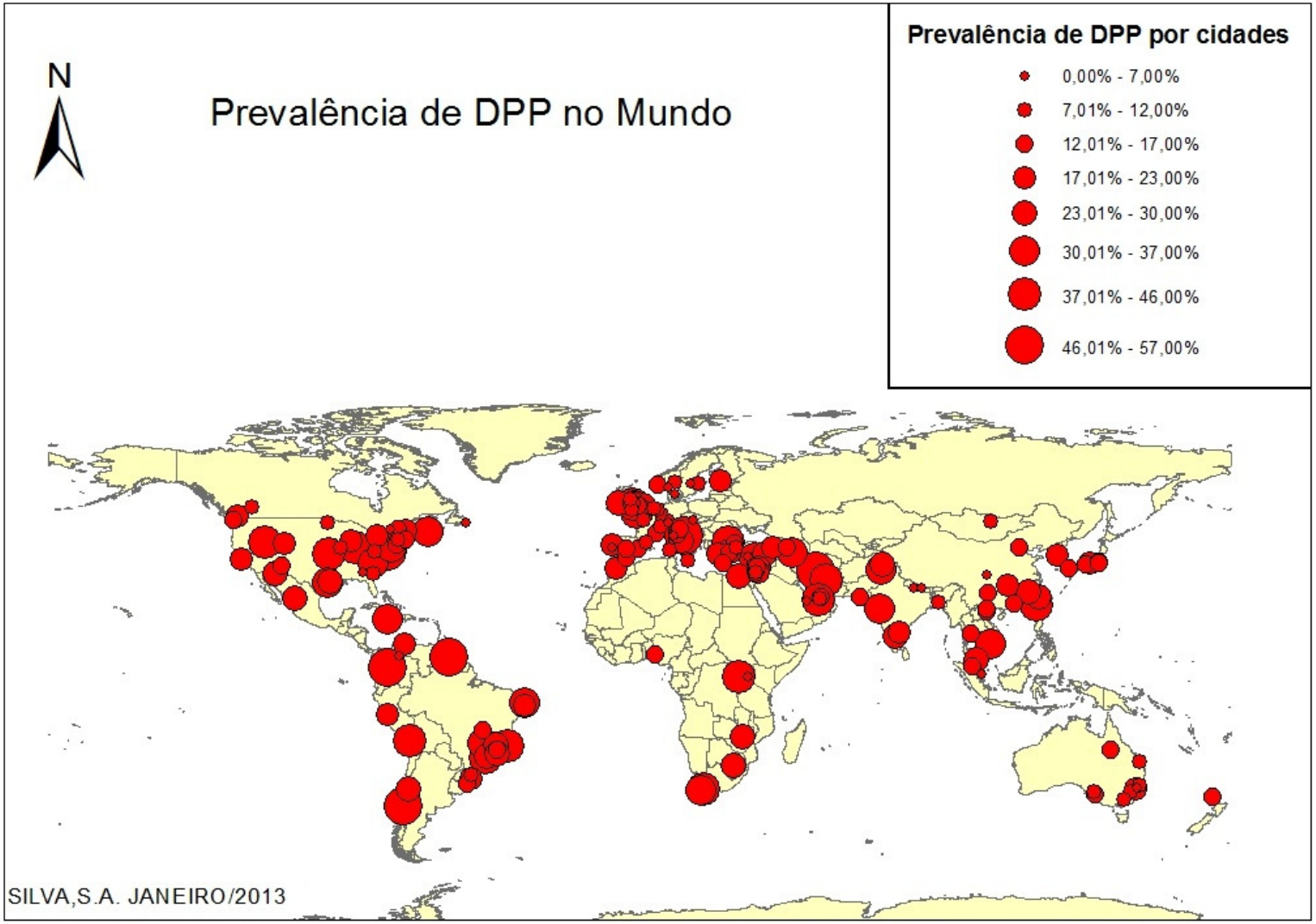

Figura 9 – Mapa indicando as médias das estimativas pontuais de prevalência de DPP por cidade.

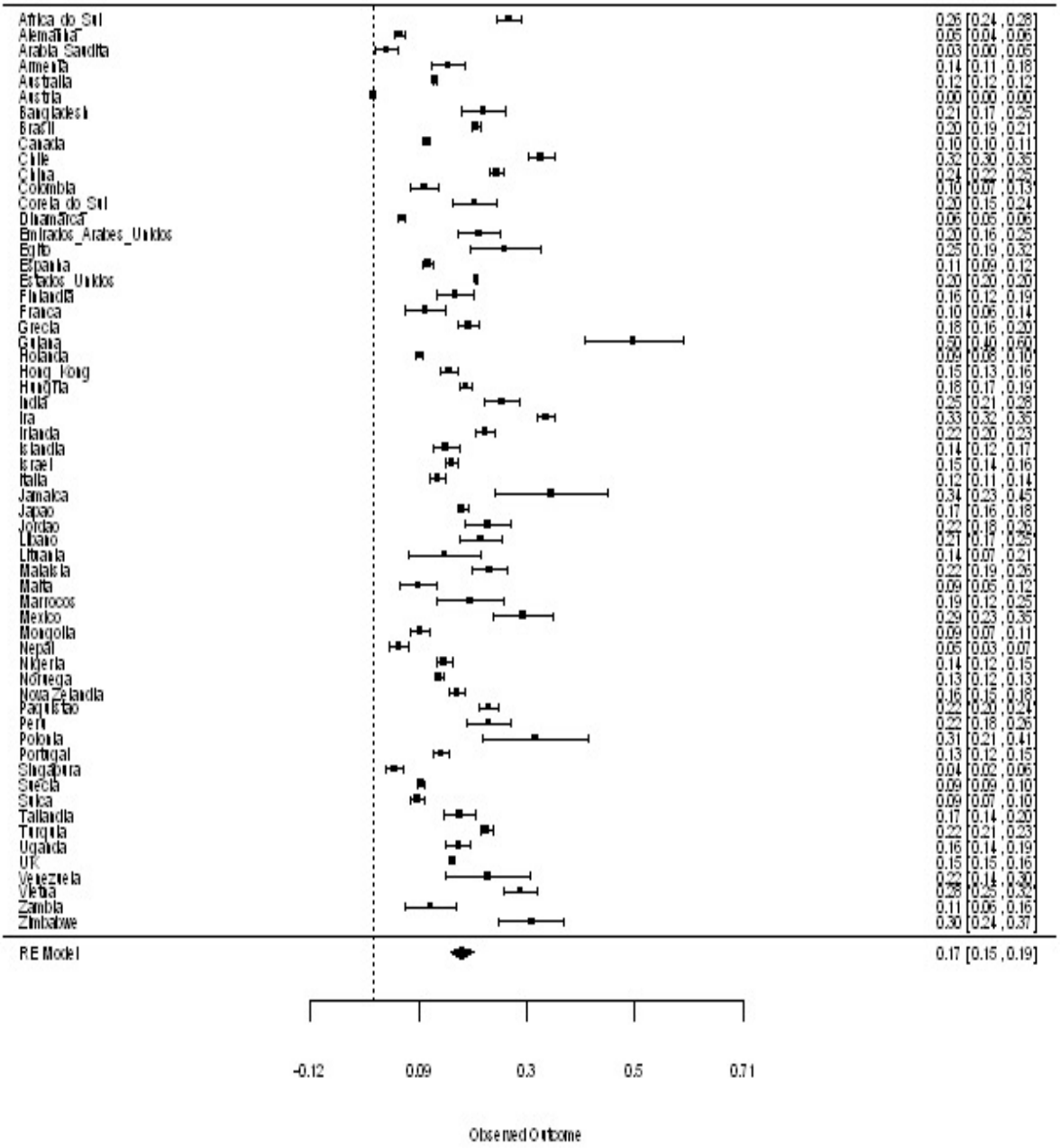

Figura 10 – Gráfico de floresta indicando a prevalência média pontual e os intervalos de confiança por país.

6.4 Análises de heterogeneidade

Gráficos de funil (Figura 11) foram elaborados para verificar o nível de heterogeneidade entre os estudos, usando-se o pacote *Metafor* do *software* R. O Gráfico A apresenta as prevalências brutas de DPP nos trabalhos incluídos. Nessa representação, usou-se a prevalência dos estudos em relação ao erro padrão. Visualmente, identificou-se alta heterogeneidade, pois os pontos se distanciaram do formato de funil. O teste de qui-quadrado resultou significativo (χ^2_{331}=19280.2214, p<0,0001) e o valor de I^2 foi de 99,06%, considerado superior ao aceitável (50%). Os Gráficos B e C apresentam os modelos em que país e cidade, respectivamente, foram usados como moderadores. Nesses gráficos, os resíduos dos estudos foram representados em relação ao erro padrão. Visualmente, observou-se menor heterogeneidade nos modelos que usaram como moderadores variáveis geográficas (países ou cidades). O modelo usando as cidades foi o que mais se aproximou do formato de funil.

As amostras dos modelos que usaram país ou cidade como moderadores foram menores, devido aos dados faltantes. Em cinco estudos (Affonso et al., 2000; Fuggle, Glover, Khan, & Haydon, 2002; Gorman et al., 2004; Vik et al., 2009; Cohen, 2010), foram registrados valores faltantes (*missing values*) na coluna correspondente ao país, porque os dados foram coletados em mais de um país. Codificar mais de um país por estudo implicaria em considerar múltiplos resultados de prevalência (um para cada país), além de diferentes IDHs e classificações quanto ao grau de desenvolvimento. As redundâncias criadas no banco de dados não seriam compatíveis com o modelo de metarregressão usado no presente trabalho. Da mesma maneira, códigos de valores faltantes foram inseridos na variável cidade, em 103 estudos, devido a duas situações: 1. o estudo declarou ter realizado a coleta de dados em mais de uma cidade; 2. o estudo não declarou em qual cidade a coleta de dados foi realizada.

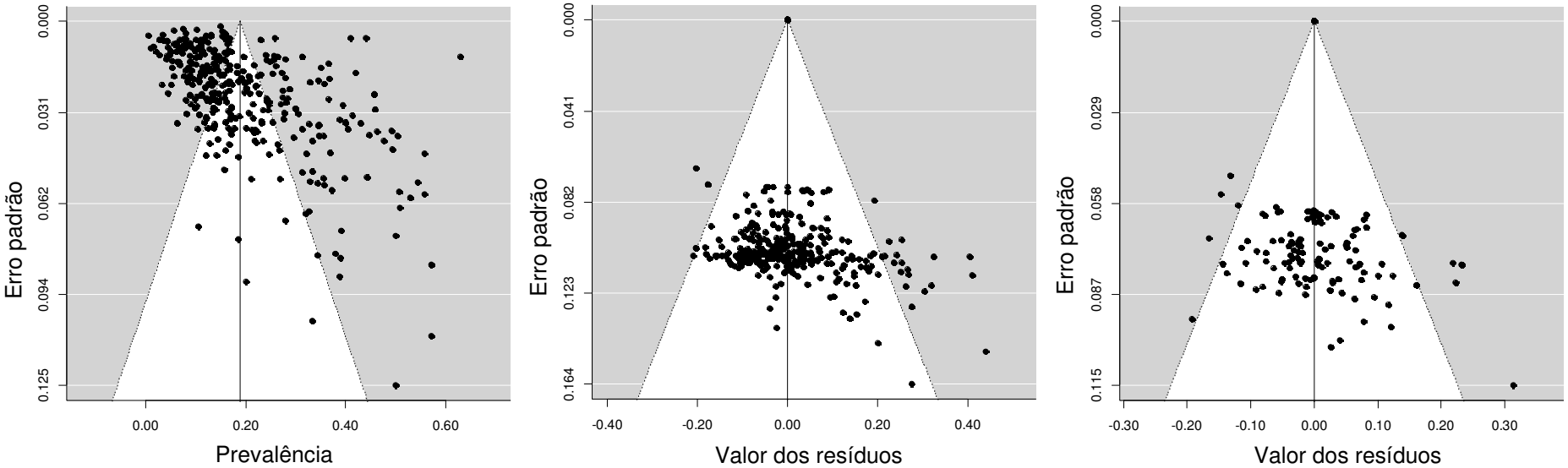

Gráfico A – Prevalência geral, sem moderadores. N = 337.

Gráfico B – País como moderador. N = 332.

Gráfico C – Cidade como moderador. N = 234.

Figura 11 – Gráficos de funil para avaliar heterogeneidade dos estudos incluídos. Gráfico A: prevalências dos estudos incluídos em relação ao erro padrão; Gráfico B: resíduos em relação ao erro padrão, usando país como moderador; Gráfico C: resíduos em relação ao erro padrão, usando cidade como moderador.

6.5 Análises de subgrupos

A heterogeneidade entre estudos incluídos foi considerada alta, indicando que variáveis metodológicas prejudicaram a comparabilidade entre eles, e/ou que ocorreu viés de publicação. Assim, a estimativa geral de prevalência, buscando sintetizar os resultados de todos os estudos em um único valor, mostrou-se inadequada e distante da realidade. É necessário, portanto, considerar as variáveis metodológicas que poderiam ser responsáveis pela heterogeneidade, realizando análises de subgrupos, que serão apresentadas a seguir:

1. Instrumento para identificar DPP

A prevalência média de DPP dos 257 estudos que usaram instrumentos de autoavaliação foi de 17,30% (IC 95% 17,17%-17,44%). Esse valor foi superior à média de prevalência dos 80 estudos que usaram entrevista clínica, que foi de 10,89% (IC 95% 10,55%-11,23%). Os intervalos de confiança não apresentaram sobreposição, conforme se pode observar na Figura 12. Esse resultado está de acordo com a Hipótese 3B do presente estudo.

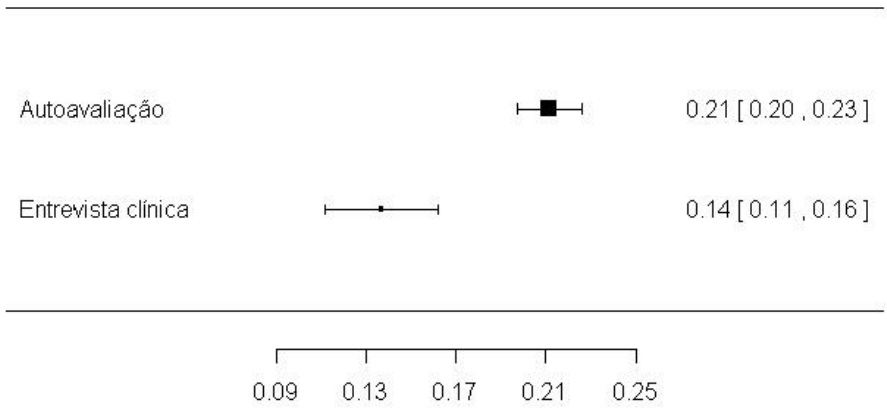

Figura 12 – Prevalência média de DPP por tipo de instrumento usado para identificar o transtorno.

2. Grau de desenvolvimento do país onde 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%). Os intervalos de confiança não apresentaram sobreposição, conforme se pode observar na Figura 13. Usando o IDH como variável contínua, teste de correlação de Pearson bivariado mostrou associação inversa e fraca entre IDH e prevalência de DPP (R= -0,165, p= 0,003). Esses resultados estão de acordo com a Hipótese 2A do presente trabalho.

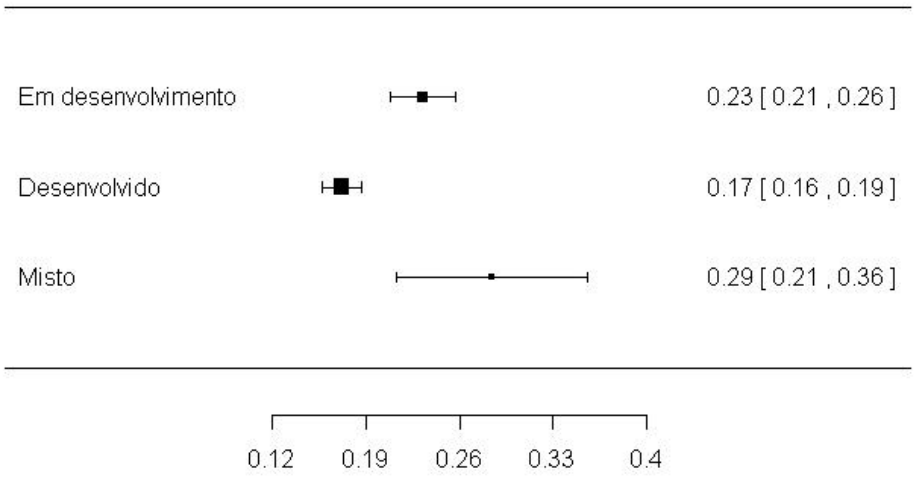

Figura 13 – Prevalência média de DPP por grau de desenvolvimento do país onde o estudo primário foi realizado.

As diferenças nas prevalências em relação ao grau de desenvolvimento do país se mantiveram quando considerada a metodologia usada para identificar as mães com DPP. A Tabela 16 mostra a prevalência em países desenvolvidos e em desenvolvimento, de acordo com o tipo de instrumento.

Tabela 16 – Prevalência média de DPP de acordo com país e tipo de instrumento usado para identificar DPP.

Grau de desenvolvimento do país / Método para identificar DPP	Prevalência média %	IC 95% limite inferior %	IC 95% limite superior %	Número de estudos %
Países desenvolvidos / Autoavaliação	16,66	16,52	16,81	185
Países desenvolvidos / Entrevista clínica	9,87	9,51	10,24	55
Países em desenvolvimento / Autoavaliação	22,96	22,45	23,47	69
Países em desenvolvimento / Entrevista clínica	14,88	14,04	15,73	25

3. Período após o parto

O período após o parto em que as coletas de dados foram realizadas apresentou ampla variação entre os estudos. Em muitos casos, não foi possível classificar o período em uma das faixas planejadas (primeiro, segundo ou terceiro quadrimestre após o parto), porque o intervalo em que os dados foram coletados correspondeu a duas ou três faixas. Por exemplo, estudos coletaram dados de cinco a nove meses após o parto, compreendendo, portanto, o segundo e o terceiro quadrimestres (Coates, Schaefer, & Alexander, 2004); ou do primeiro ao décimo segundo mês de vida do bebê, compreendendo os três quadrimestres (Escobar, Pacora, Custodio, & Villar, 2009).

Outro fator que dificultou a síntese dos dados de prevalência de acordo com os períodos após o parto é que muitos estudos não fizeram distinção entre prevalência pontual (proporção de mulheres com depressão no momento do estudo) e de período (proporção de mulheres com depressão desde o nascimento do bebê até o momento do estudo), conforme já havia sido observado por Halbreich e Karkun (2006).

Também foram frequentes os trabalhos que indicaram um momento aproximado em que as avaliações foram feitas, sem especificar o período mínimo e máximo. Por exemplo, estudos indicaram que o rastreamento de DPP foi feito “próximo à sexta semana após o parto” (Georgiopoulos et al., 1999, p. 654) ou que usaram a EPDS “nas consultas do quarto mês após o parto” (Milgrom, Ericksen, Negri, & Gemmill, 2005, p. 834). No primeiro caso, não se sabe qual foi o intervalo considerado próximo de seis semanas após o parto; e no segundo exemplo, as consultas dos quatro meses pós-parto provavelmente não foram realizadas exatamente nas datas em que os bebês completaram 120 dias. Nesses casos, por indisponibilidade de informações, os dados foram tabulados como se tivessem sido coletados em um único dia, isto é, aos 42 dias após o parto no estudo de Georgiopoulos et al. (1999) e aos 120 dias no trabalho de Milgrom et al. (2005), consistindo em meras estimativas.

Pelas razões expostas, não foi feita análise de subgrupos por período após o parto, considerando-se que as informações coletadas nos trabalhos não foram suficientemente precisas para dar embasamento a essa análise. Portanto, a Hipótese 3A do presente trabalho não foi testada.

4. Delineamento

A média da prevalência dos 157 estudos transversais foi de 17,19% (IC 95% 17,01%-17,37%) e a dos estudos de coorte longitudinal foi de 15,95% (IC 95% 15,75%-16,13%). Os intervalos de confiança não apresentaram sobreposição, conforme se pode observar na Figura 14. Esse resultado está de acordo com a Hipótese 3C do presente trabalho.

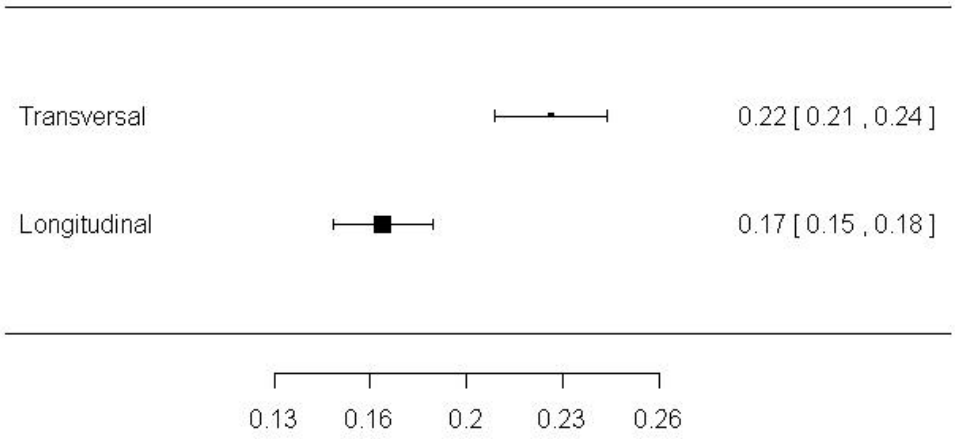

Figura 14 – Prevalência média de DPP por delineamento de pesquisa.

5. Técnica de amostragem

O item 1 da Escala de Loney et al. (1998) classificou a técnica de amostragem em duas categorias: adequada (aleatória ou censo, recebendo pontuação nesse item da escala) e inadequada (não aleatória ou não está clara, não recebendo pontuação nesse item da escala). Os 31 estudos com amostra considerada adequada tiveram prevalência média de DPP de 15,77% (IC 95% 15,51%-16,03%), enquanto os demais, que tiveram amostra considerada inadequada, tiveram prevalência média de 16,92% (IC95% 16,77%-17,07%). Os intervalos de confiança não apresentaram sobreposição.

As prevalências médias para as categorias relativas às técnicas de amostragem foram apresentadas na Tabela 17.

O único intervalo de confiança que não apresentou sobreposição com os demais foi o dos estudos em que a técnica de amostragem não estava clara, categoria que apresentou prevalência média superior às das demais.

Tabela 17 – Prevalência média de DPP de acordo com país e técnica de amostragem.

Técnica de amostragem	Prevalência média %	IC 95% limite inferior %	IC 95% limite superior %	Número de estudos %
Aleatória	15,64	15,34	15,94	17
Censo	16,19	15,67	16,72	14
Não aleatória	16,06	15,86	16,26	181
Não está clara	17,94	17,72	18,17	125

Assim, considera-se que a Hipótese 3D foi corroborada, pois ao testar a técnica de amostragem de forma binária, comparando estudos que usaram estratégias adequadas ou inadequadas, foi encontrada diferença. Análise mais detalhada acrescentou a compreensão de que essa diferença se deu devido aos estudos cujos relatos não permitiram identificar a técnica de amostragem.

6. Tamanho da amostra

Usando-se o item 3 da Escala de Loney et al. (1998), foram comparadas as prevalências médias de DPP de estudos que usaram amostras menores que 470 participantes e aqueles que usaram amostras maiores ou iguais a 470 participantes. Estudos com amostras menores que 470 tiveram prevalência média de 18,80% (221 estudos, IC 95% 18,44%-19,17%), enquanto os que tiveram amostras maiores ou iguais a 470 participantes tiveram prevalência média de 16,28% (116 estudos, IC 95% 16,14%-16,42%).

Teste de correlação de Pearson bivariado apontou que não houve associação linear significativa entre tamanho de amostra, inserido como variável contínua, e prevalência de DPP nos estudos incluídos (R= -0,065, p= 0,235). No entanto, esse resultado deve ser interpretado com cautela, devido aos *outliers* (estudos com amostras muito superiores às dos demais), que podem alterar fortemente os resultados de testes lineares de correlação.

Refazendo-se o teste com os 277 estudos com amostra de até 1.000 participantes, encontrou-se associação inversa moderada, porém significativa entre tamanho de amostra e prevalência de DPP (R= -0,300, p<0,001).

Considera-se, portanto, que a Hipótese 3E do presente estudo, que previa que não haveria associação entre tamanho de amostra e prevalência de DPP, foi refutada.

7. Planejamento do tamanho da amostra

A média da prevalência de DPP em estudos que planejaram o tamanho amostral com base em análise de poder foi de 19,19% (48 estudos, IC 95% 18,84%-19,53%), enquanto a média dos que não planejaram a amostra foi de 16,16% (289 estudos, IC 95% 16,02%-16,29%). Esse resultado está de acordo com a Hipótese 3G do presente trabalho.

8. Base amostral

O item 2 da Escala de Loney et al. (1998) atribuiu um ponto aos estudos que usaram base amostral adequada, isto é, representativa da população geral. A prevalência média de DPP foi de 16,01% (IC 95% 15,76%-16,26%) para os 23 estudos que usaram base amostral adequada e 16,87% (IC 95% 16,71%-17,02%) para os 314 que não usaram. Não ocorreu sobreposição dos intervalos de confiança. Esse resultado está de acordo com o que foi previsto pela Hipótese 3F da presente revisão sistemática.

9. Contexto de pesquisa

As prevalências médias de DPP para os diferentes contextos de pesquisa foram apresentadas na Tabela 18. A maior prevalência média foi encontrada em estudos nacionais. Embora tenha ocorrido sobreposição de alguns intervalos de confiança (ex.: estudos realizados em instalações médicas e os realizados em províncias, estados ou regiões), outros não apresentaram sobreposição (ex.: estudos realizados em cidades tiveram maior prevalência média que os realizados em instalações médicas, províncias, estados ou

regiões). Assim, considera-se que a Hipótese 3H do presente trabalho foi refutada.

Tabela 18 – Prevalência média de DPP de acordo com o contexto de pesquisa.

Contexto	Prevalência média %	IC 95% limite inferior %	IC 95% limite superior %	Número de estudos %
Nacional	19,65	19,41	19,88	14
Província/estado/região	14,93	14,68	15,18	54
Cidade	15,79	15,49	16,09	111
Instalação médica	14,53	14,21	14,85	114
Não está claro	14,37	13,81	14,93	29
Outros	16,18	15,59	16,78	15

10. Ponto de corte da Escala de Depressão Pós-parto de Edimburgo (EDPE)

A EDPE foi usada por 228 trabalhos (67,7% dos incluídos), sendo o instrumento mais frequente para identificar DPP em estudos científicos. A relação entre o ponto de corte desse instrumento e a prevalência de DPP foi testada em uma subamostra de 198 trabalhos que atenderam a três critérios: 1. usaram a EPDE; 2. relataram de forma clara o ponto de corte; e 3. não realizaram entrevista clínica com as mães identificadas acima do ponto de corte. Não foram incluídos na subamostra trabalhos cujo critério para identificar DPP consistiu em uma combinação de resultados de instrumentos, como por exemplo: EDPE maior ou igual a nove e *General Health Questionnaire* (GHQ) maior ou igual a quatro (Grussu & Quatraro, 2009); EDPE maior ou igual a 10 e *Lubin Depression Adjective Check-List* (DACL) maior ou igual a 14 ou EDPE maior ou igual a 14 e DACL maior ou igual a 11 (Costa, Larouche, Dritsa, & Brender, 2000). Também não foram incluídos estudos que não relataram o ponto de corte da EDPE. Os resultados foram apresentados na Tabela 19.

Tabela 19 – Prevalência média de DPP de acordo com o ponto de corte da EDPE.

Ponto de corte	Prevalência média	IC 95% limite inferior	IC 95% limite superior	Número de estudos
7	6,83	3,87	9,80	1
8	14,68	13,75	15,61	3
9	14,83	13,97	15,68	15
10	15,64	15,23	16,05	45
11	20,38	17,02	23,74	5
12	13,57	13,18	13,96	44
13	13,22	13,00	13,44	83
14	33,03	28,73	37,32	2

De acordo com os resultados apresentados na Tabela 19, não parece ter ocorrido associação entre o ponto de corte da EDPE e a prevalência de DPP. Esse resultado foi reforçado pelo resultado do teste de correlação de Pearson bivariada entre o valor do ponto de corte da EDPE e a prevalência de DPP, que não se mostrou significativo (r=-0,008, p=0,915). Assim, foi refutada a Hipótese 3I do presente estudo.

6.6 Impacto da metodologia sobre os resultados de prevalência: metarregressão

Devido a teórica existência de dependência dos estudos de um mesmo país e de uma mesma cidade, inicialmente foi ajustado um modelo de efeitos mistos hierárquico de três níveis (nível 1: estudo, nível 2: cidade e nível 3: país) e a variável resposta (proporção de mulheres identificadas como deprimidas em cada estudo), que se assumiu ter distribuição Normal. Foi usada a função de ligação logarítmica, já que a variável resposta assume valores positivos.

Considerou-se, ainda, que a estrutura de correlação dos efeitos aleatórios é estimada por componente de variância.

As variáveis de efeitos fixos inseridas no modelo inicial do MLMG foram: tempo desde a publicação do estudo, em anos (ou seja, subtraiu-se o ano de publicação de 2013); tipo de publicação (três categorias: 1 = artigo, 2 = dissertação, 3 = tese); IDH (variável contínua); tipo de instrumento (duas categorias: 0 = autoavaliação ou 1 = entrevista clínica); delineamento (duas categorias: 0 = transversal ou 1 = coorte longitudinal); planejamento da amostra com base em análise de poder (duas categorias: 0 = não ou 1 = sim); contexto de pesquisa (seis categorias: 1 = nacional, 2 = província/estado/região, 3 = cidade, 4 = instalação médica, 5 = não está claro, 6 = outros); se o estudo declarou que teve como objetivo avaliar a prevalência de DPP (0 = não ou 1 = sim); itens 1, 2, 4, 5, 6, 7 e 8 da Escala de Loney et al. (1998) (0 = inadequado, 1 = adequado); e tamanho amostral (variável contínua)[38]. Foram testados os efeitos principais e de interação de primeira ordem.

Através de teste para verificar se o nível de "país" foi significativo para o modelo, observou-se que não havia indícios para rejeitar a hipótese de que existe dependência dos estudos dentro de um dado país. Assim, foi usado o modelo de 2 níveis (cidade e estudo), em que o teste de dependência dos estudos de uma mesma cidade foi significativo ($Z = 4{,}268$, $p<0{,}001$), apresentando valor do intercepto para a cidade igual a 0,116.

Já em relação aos efeitos fixos, o modelo escolhido por fazer sentido prático e também com o auxílio do critério de Akaike, foi composto pelas covariáveis tipo de instrumento usado para identificar DPP na amostra ($F_{1,229}=39{,}299$, $p<0{,}001$), delineamento ($F_{1,229}=28{,}220$, $p<0{,}001$), tamanho

[38] A variável grau de desenvolvimento do país (duas categorias: desenvolvido ou em desenvolvimento) não foi inserida por apresentar redundância com a variável IDH (contínua); o item 3 da Escala de Loney et. al (1998) (duas categorias: menos de 470 participantes ou número maior ou igual a 470 participantes) não foi inserido devido à redundância com a variável tamanho amostral (contínua); a variável técnica de amostragem (quatro categorias: aleatória, censo, não aleatória, não está claro) não foi inserida devido à redundância com o item 1 da Escala de Loney et al. (1998) (duas categorias: 0 = amostragem inadequada, 1 = amostragem adequada). O critério heurístico adotado para definir, em cada par de variáveis redundantes, qual seria inserida no modelo foi o menor valor de AIC, conforme análises preliminares testando combinações entre variáveis.

amostral ($F_{1,229}$=13,619, p<0,001) e IDH do país ($F_{1,229}$=5,916, p<0,016). A Tabela 20 mostra as estimativas dos coeficientes do modelo, assim como o seu valor p. Observa-se que, quanto maior o IDH, menor a prevalência de DPP e o mesmo acontece quanto maior for o tamanho da amostra, dado as demais covariáveis fixas. Em relação ao tipo de instrumento, a auto-avaliação resultou em maior prevalência de DPP que a entrevista clínica e o delineamento transversal resultou em maior prevalência de DPP que o longitudinal, dado as demais covariáveis com valores e status fixos. O AIC do modelo inicial, com todas as variáveis de efeitos fixos, foi de 454,701, e o do modelo final, formado apenas pelas variáveis significativas, foi de 418,719.

Tabela 20 – Ajuste final do Modelo Linear Misto Generalizado tendo DPP como variável resposta e características metodológicas dos estudos primários como preditores.

	Coeficiente	p-valor
Intercepto	-1,495	<0,001
IDH do país	-0,736	0,016
Tipo de instrumento = 0 (autoavaliação)	0,557	<0,001
Tipo de instrumento = 1 (entrevista clínica)	0,000*	-
Delineamento = 0 (transversal)	0,353	<0,001
Delineamento = 1 (longitudinal)	0,000*	-
Tamanho amostral	-0,000	<0,001

* Coeficientes iguais a zero devido a redundância.

A Figura 15 mostra o valor predito da prevalência de DPP, no eixo Y, em relação ao valor observado, representado no eixo X. Os dados se aproximaram

de uma reta, com maiores valores preditos correspondendo aos maiores valores observados, o que indica boa qualidade de ajuste do modelo.

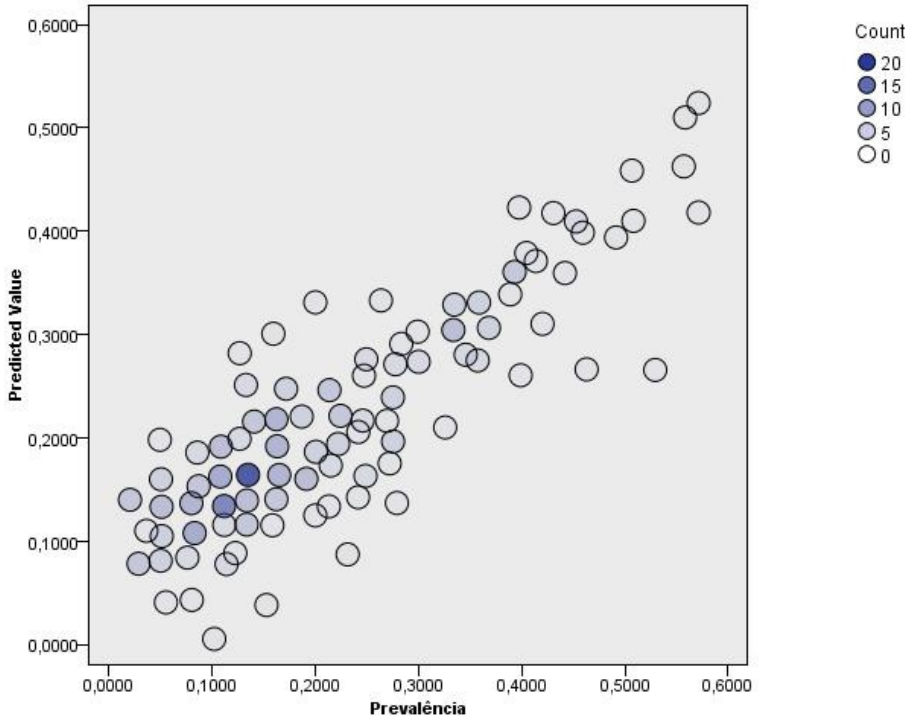

Figura 15 – Gráfico mostrando o valor predito da prevalência de DPP em relação ao valor observado.

Portanto, foi possível construir o modelo explicativo da prevalência de DPP a partir de variáveis preditivas. Considera-se que os resultados estão parcialmente de acordo com a Hipótese 5 do presente estudo, pois a técnica de amostragem não se manteve no modelo final, conforme havia sido previsto. Como o IDH foi um preditor significativo, que se manteve no modelo final, foi refutada a Hipótese 2B do presente trabalho, que previa que a diferença entre países desenvolvidos e em desenvolvimento desapareceria se fossem controladas as variáveis metodológicas. Também foi refutada a Hipótese 4C, que previa que a qualidade do estudo, medida pela Escala de Loney et al. (1998), teria associação com a prevalência de DPP. Todos os itens da Escala de Loney et al. (1998) foram eliminados do modelo de Metarregressão. O único preditor do Modelo Final que apresentou associação com a escala de

qualidade foi o tamanho amostral, inserido no modelo como variável contínua, mas que apresenta relação direta com o item 3.

6.7 Análises de sensibilidade

Análises de sensibilidade foram realizadas para verificar se eventuais mudanças nas caraterísticas metodológicas da presente revisão sistemática poderiam causar alterações expressivas nos resultados. Haviam sido planejadas análises omitindo: 1. teses e dissertações; 2. estudos publicados até 31 de dezembro de 1989; 3. estudos que não declararam ter como objetivo avaliar a prevalência de DPP; 4. estudos que tiveram pontuação abaixo da mediana na Escala de Loney et al. (1998). Como foi verificado que apenas quatro estudos incluídos foram publicados antes de 1990, uma segunda análise do ano de publicação foi realizada, omitindo-se estudos publicados até 31/12/1999. Para cada uma das análises de sensibilidade, foi verificada a prevalência média geral, em países desenvolvidos e em países em desenvolvimento, conforme apresentado na Tabela 21.

As prevalências médias das análises de sensibilidade, em geral, foram semelhantes às encontradas na amostra original, e os intervalos de confiança se sobrepuseram. Em três situações, ocorreram diferenças nas prevalências médias dos países desenvolvidos e, consequentemente, na prevalência média global, visto que maior parte dos estudos incluídos foi realizada em países desenvolvidos. Ao omitir teses e dissertações, a prevalência média em países desenvolvidos diminuiu cerca de um ponto percentual. Omitindo estudos com pontuação na Escala de Loney et al. (1998) inferior à mediana, a redução na prevalência média de países desenvolvidos foi de cerca de 1,5 ponto percentual. E omitindo estudos que não declararam ter como objetivo investigar a prevalência de DPP, a prevalência média em países desenvolvidos aumentou cerca de dois pontos percentuais.

Tabela 21 – Resultados das análises de sensibilidade.

Análise de sensibilidade	**Global**		**Países desenvolvidos**		**Países em desenvolvimento**	
	Prevalência média (IC) %	N	Prevalência média (IC) %	N	Prevalência média (IC) %	N
Prevalência média na amostra original (sem omissões)	**16,63 (16,50-16,76)**	**337**	**16,02 (15,89-16,16)**	**240**	**21,31 (20,87-21,75)**	**94**
Omitindo teses e dissertações	15,83 (15,70-15,96)*	305	15,17 (15,04-15,31)*	219	21,46 (21,00-21,91)	86
Omitindo trabalhos publicados até 31/12/1989	16,63 (16,51-16,77)	333	16,03 (15,90-16,17)	236	21,31 (20,87-21,75)	94
Omitindo trabalhos publicados até 31/12/1999	16,86 (16,73-17,00)	293	16,24 (16,10-16,38)	199	21,31 (20,86-21,75)	91
Omitindo estudos que não declararam ter como objetivo avaliar prevalência de DPP	17,81 (17,60-18,02)*	163	17,03 (16,81-17,26)*	98	20,46 (19,94-20,97)	65
Omitindo estudos com pontuação na Escala de Loney et al. (1998) abaixo da mediana	15,55 (15,38-15,72)*	85	14,87 (14,69-15,05)*	59	21,44 (20,78-22,11)	24

* Intervalo de confiança não apresenta sobreposição com o da prevalência média na amostra original.

IC = intervalo de confiança.
N = número de estudos incluídos na análise.

7 Discussão

Considera-se que a presente revisão sistemática foi a mais abrangente dentre as que objetivaram estudar a prevalência de DPP. A busca bibliográfica envolveu 14 bases de dados eletrônicas e incluiu 337 trabalhos (entre artigos em periódicos científicos, teses e dissertações), totalizando 293.743 participantes. Até o presente momento, a maior revisão sistemática sobre prevalência de DPP (Halbreich & Karkun, 2006) havia incluído cerca de 140 estudos, correspondendo a menos da metade do número de trabalhos incluídos na presente revisão, e consultou somente quatro bases de dados. A revisão sistemática mais recente sobre prevalência de DPP (Lobato et al., 2012) buscou apenas estudos brasileiros e incluiu 14, enquanto a presente revisão incluiu 26 trabalhos publicados no Brasil.

Nas próximas seções, serão discutidas forças e limitações da presente revisão sistemática. A abrangência da estratégia de busca, a qualidade metodológica dos estudos incluídos, as hipóteses de pesquisa à luz dos resultados, os resultados da metarregressão e a comparação da prevalência de DPP em países desenvolvidos e em desenvolvimento serão assuntos abordados. Para finalizar, serão discutidas as implicações dos resultados para a prática profissional e para pesquisas primárias ou secundárias sobre prevalência de DPP.

7.1 Forças e limitações da estratégia de busca

Um diferencial da presente revisão foi a ampliação da busca em bases de dados latino-americanas, totalizando sete: BDTD, Portal Nacional BVS Brasil em Saúde, ERIC, LILACS, Psicodoc, Redalyc e SciELO. Duas revisões sistemáticas realizadas por autores brasileiros (Melo Junior, 2011; Lobato et al., 2012) realizaram buscas na SciELO e na LILACS, mas as revisões feitas por autores de outros países (O'Hara & Swain, 1996; Gavin et al., 2005; Halbreich

& Karkun, 2006; Sawyer et al., 2010; Vigod et al., 2010; Villegas et al., 2011) ignoraram completamente as bases de dados lationoamericanas.

Quantitativamente, acredita-se que a não avaliação de estudos devido ao idioma e à dificuldade de localizar o texto completo teve um impacto desconsiderável nos resultados da metanálise, por duas razões: 1. dentre os registros selecionados por título e resumo, 87,1% foram avaliados pelo texto completo. 2. dentre os estudos que não foram avaliados, nem todos atenderiam aos critérios de inclusão. A taxa de inclusão de estudos na presente revisão foi de 43,4% (338 de 778 avaliados). Os estudos excluídos devido ao idioma de publicação corresponderam a 4,4% dos trabalhos, e os que não puderam ser recuperados corresponderam a 6,0%, totalizando 10,4% de perda. Se 43,4% desses 10,4% fosse incluído na revisão, o acréscimo no número de trabalhos seria de 4,5%. Essa quantidade de estudos provavelmente não traria alterações relevantes aos resultados encontrados, devido ao grande número de trabalhos (338) e de participantes (293.743) incluídos, que permitiram análises estatísticas com alto poder.

Acredita-se, portanto, que o alto número de estudos e de participantes incluídos – que é o maior dentre as revisões sistemáticas de prevalência de DPP a que se teve acesso – tenha minimizado o impacto das limitações da estratégia de busca: 1. não foram consultados especialistas na área para procurar identificar estudos em andamento ou não publicados, 2. não foram consultadas as listas de referências bibliográficas dos trabalhos incluídos e 3. foram excluídos estudos em idiomas diferentes de inglês, espanhol e português.

A limitação de idiomas também ocorreu em outras revisões sistemáticas sobre prevalência de DPP, que declararam ter incluído apenas pesquisas publicadas em inglês (Gavin et al., 2005; Sawyer et al., 2010), ou em inglês e francês (Vigod et al.; Villegas et al., 2011). Uma revisão sistemática sobre fatores de risco para DPP, realizada por autora latino-americana, limitou o idioma de publicação a inglês, português e espanhol (Aliane et al., 2011). Na presente revisão, caso a avaliação dos trabalhos não fosse limitada pelo idioma de publicação, o tempo necessário e o alto custo das traduções provavelmente resultaria em uma relação de custo-benefício pouco satisfatória, devido ao

grande número de trabalhos localizados e à variedade de idiomas: além de inglês, espanhol e português, foram encontrados estudos em alemão, chinês, coreano, curdo, francês, holandês, húngaro, italiano, persa, polonês e turco.

De fato, revisões sistemáticas de estudos observacionais tendem a encontrar um número grande de registros (Gülmezoglu et al., 2004). A variedade de idiomas é compreensível, pois diferentemente das pesquisas experimentais, os estudos observacionais sobre saúde coletiva são altamente dependentes do contexto local e, portanto, têm maior tendência de serem publicados em periódicos nacionais (Victora & Moreira, 2006). A publicação de artigos sobre prevalência em periódicos nacionais e escritos em idioma local é importante para aumentar a divulgação e a aplicação dos resultados na prática profissional, ao facilitar o acesso de profissionais de saúde que não atuam em meio acadêmico. Em geral, os periódicos nacionais também costumam estar disponíveis com maior frequência em acesso aberto (Victora & Moreira, 2006). Uma solução para assegurar a divulgação em âmbito nacional e, ao mesmo tempo, facilitar o acesso a pesquisadores e revisores seria a publicação bilíngue, que já é adotada por alguns periódicos, como por exemplo, a Revista de Psiquiatria Clínica (a publicação bilíngue em português e inglês é facultada aos autores). Com a possibilidade de publicação em meio digital, manter versões do mesmo artigo em diferentes idiomas não é mais uma opção demasiadamente custosa para os periódicos científicos.

Outra limitação do presente estudo se refere aos 47 trabalhos que não foram avaliados porque o texto completo não pôde ser obtido. Essa situação reforça a importância do Movimento de Acesso Aberto, que defende a ampliação e a popularização do acesso gratuito aos periódicos científicos (Sampaio & Serradas, 2009). Esforços foram empreendidos para acessar o texto completo dos artigos de interesse, e 122 foram obtidos via serviços de comutação bibliográfica, entretanto o custo alto e a dificuldade de obtenção dos trabalhos foram fatores que dificultaram essa tarefa. Os estudos não localizados, em geral, podem ser considerados de menor relevância, por serem mais antigos.

7.2 Viés de publicação contra o Sul

Os resultados da busca sistemática corroboraram a visão de Victora & Moreira (2006) sobre viés de publicação contra o Sul, isto é, desproporção entre a produção científica de países de renda alta ("Norte") e os de renda média ou baixa ("Sul"). Dentre os estudos incluídos, 71,5% coletaram dados em países desenvolvidos. Não foram encontradas diferenças estatisticamente significativas entre estudos incluídos e excluídos quanto ao nível de desenvolvimento socioeconômico do país, afastando-se a possibilidade de que a desproporção tenha ocorrido por razões metodológicas da presente revisão. A diferença foi maior considerando-se o número de participantes incluídas na revisão: 262.706 eram de países desenvolvidos, totalizando 89,4% do total.

Essa discrepância pode ter sido causada por diversos fatores: 1. menor número de pesquisas em países em desenvolvimento, devido a menores investimentos financeiros na área; 2. "racismo editorial", ou menor chance de publicação de artigos de países em desenvolvimento nos periódicos de maior visibilidade no meio científico, devido a preconceito dos editores e pareceristas (Victora & Moreira, 2006); 3. menor chance de indexação de periódicos científicos nacionais de países em desenvolvimento nas bases de dados internacionais.

Os resultados referentes ao número de estudos incluídos por país na presente revisão sistemática, de forma geral, estão de acordo com o quadro dos *Top-20*, ou seja, 20 maiores produtores mundiais de artigos científicos na área de saúde entre 1992 e 2001 (Paraje et al., 2005). Estados Unidos da América, país que ocupou o primeiro lugar na lista dos *Top-20* e responsável por 36,68% da produção mundial de artigos na área de saúde, também encabeçou a lista dos países com maior número de estudos sobre prevalência de DPP (81 trabalhos, 24,04% do total). Outros países que, no presente estudo, tiveram o maior número de trabalhos sobre prevalência de DPP também faziam parte da lista dos *top-20*, como: Reino Unido, Canadá, China, Itália e Suécia. Algumas exceções foram: Turquia (11 trabalhos incluídos na presente revisão, 3,2%) e Chile (10 trabalhos incluídos na presente revisão,

3,0%), que não fazem parte da lista dos *Top-20*; e Brasil que, empatado com a Austrália, teve o segundo maior número de estudos incluídos na presente revisão sobre prevalência de DPP (26 estudos, 7,7%), apesar de ter ocupado o 19º lugar na lista dos *Top-20*, sendo responsável por 0,73% da produção mundial de artigos na área de saúde.

A presença do Brasil, Turquia e Chile dentre os países que mais tiveram trabalhos incluídos na presente revisão sistemática pode estar relacionada com um crescente aumento das pesquisas realizadas em países em desenvolvimento, que pôde ser constatada no presente trabalho: estudos de países em desenvolvimento foram mais recentes que os de países desenvolvidos e muitos estudos realizados em países latino-americanos foram publicados após o período contemplado pela revisão bibliométrica de Paraje et al., (2005). Também pode refletir uma maior preocupação desses países em investigar a DPP como um problema de saúde coletiva. A ampla busca em bases de dados latino-americanas, que não foram incluídas no estudo sobre os produtores *Top-20* de artigos científicos na área de saúde (Paraje et al., 2005), pode ter aumentado o número de estudos brasileiros e chilenos na presente revisão. A maior disponibilidade de trabalhos científicos nacionais em acesso aberto (Victora & Moreira, 2006) ou via serviços de comutação bibliográfica deve ser considerada ao interpretar a posição do Brasil como o segundo país com maior número de trabalhos incluídos.

Foram incluídos 43 estudos que coletaram dados na América Latina, que corresponderam a 12,8% do total. Em revisões sistemáticas anteriores, a participação de trabalhos latino-americanos foi menor: 10 estudos na revisão de Halbreich e Karkun (2006), seis na de Melo Junior (2011) e nenhum em outras revisões (O'Hara & Swain, 1996; Villegas et al., 2011). A baixa presença de estudos realizados na América Latina nas revisões sistemáticas internacionais pode estar relacionada com restrições idiomáticas e com a ausência de buscas nas bases de dados latino-americanas, indicando que o viés “contra o Sul” se deu também nas estratégias de busca das revisões sistemáticas. O artigo de Victora e Moreira (2006) sobre preconceito editorial propôs importantes medidas para melhorar a qualidade dos artigos latino-americanos, aumentando as chances de que sejam publicados em periódicos

internacionais “do Norte”. Paralelamente, também seria interessante incentivar os autores “do Norte” a consultar as bases eletrônicas “do Sul”, sobretudo quando realizam revisões sistemáticas.

Além das buscas em bases de dados latino-americanas, a possibilidade de incluir estudos em espanhol e português aumentou a participação de estudos da América Latina na presente revisão sistemática, visto que apenas 40,8% deles apresentaram texto completo em inglês. Seria pertinente argumentar que as buscas em sete bases de dados latino-americanas e a possibilidade de incluir estudos publicados em português e espanhol podem ter favorecido estudos realizados na América Latina, em detrimento dos realizados em outros países em desenvolvimento, sobretudo na África e na Ásia. O presente estudo, portanto, teria um viés de seleção latino-americano. Entretanto, essa característica pode ser considerada uma força da presente Tese, visto que se trata de uma ampliação da abrangência das revisões sistemáticas sobre prevalência de DPP, que historicamente, tiveram pouca ou nenhuma participação de trabalhos da América Latina. Além disso, foram incluídos 60 trabalhos de outros países em desenvolvimento, número elevado em relação às revisões anteriores. A necessidade de inclusão de mais estudos de países em desenvolvimento já havia sido apontada por Mann et al. (2010).

7.3 Concordância entre revisoras

Um ponto forte da presente revisão sistemática foi ter verificado a confiabilidade dos instrumentos empregados para seleção de estudos para a revisão e para avaliação da qualidade metodológica dos mesmos, através de análises independentes realizadas por duas revisoras, que foram comparadas para estabelecer uma taxa de concordância. Esse procedimento, que é recomendado pela Colaboração Cochrane (Higgins & Green, 2011), não tem sido usado de forma adequada nas revisões sistemáticas de estudos observacionais. Por exemplo, O'Hara e Swain (1996) não declararam ter realizado avaliações independentes por dois revisores; Lobato et al. (2012) declararam que os estudos primários foram avaliados quanto aos critérios de

inclusão duas vezes pelo mesmo autor, para verificar possíveis divergências; Gavin et al. (2005) e Melo Junior (2011) relataram que dois revisores avaliaram os estudos quanto aos critérios de inclusão de forma independente e que divergências foram solucionadas por consenso, mas não relataram as taxas de concordância.

A concordância entre as revisoras na seleção dos estudos de acordo com os critérios de inclusão estabelecidos foi considerada satisfatória em todas as fases: seleção por títulos, por resumos e pelos textos completos. Isso indica que os critérios definidos estavam claros e que a estratégia de seleção é replicável.

A concordância entre as revisoras na avaliação da qualidade metodológica pela Escala de Loney et al. (1998), por sua vez, foi satisfatória em sete dos oito itens. O item 6 (apresentação de intervalos de confiança para a estimativa de prevalência e análises de subgrupos) foi insatisfatório, obtendo concordância entre avaliadoras próxima à que seria esperada ao acaso. A concordância baixa se deveu a uma divergência na compreensão das revisoras. Enquanto uma delas atribuiu pontuação ao item apenas quando foi apresentado o intervalo de confiança da prevalência (interpretação considerada correta), a outra atribuiu pontuação caso fosse apresentado qualquer intervalo de confiança. Diversos estudos não apresentaram o intervalo de confiança da prevalência, e sim o de outros parâmetros, como risco relativo, razão de chance, etc. Essa situação mostra a importância de se fazer treinamentos e atingir boa compreensão dos itens avaliados por todos os revisores. A concordância entre revisoras quanto à pontuação total na escala foi moderada, de acordo com o coeficiente de correlação intraclasse. Para contornar possíveis vieses na pontuação total da escala, em vez de incluir a soma das pontuações na Escala de Loney et al. (1998) na metarregressão, foram incluídos os itens um a um, visto que a concordância em cada item (com exceção do de número 7) foi considerada adequada.

7.4 Análise crítica da qualidade metodológica dos estudos incluídos

Testes estatísticos comparando os resultados da Escala de Loney et al. (1998) mostraram que não houve diferença na qualidade de estudos realizados em países desenvolvidos e em desenvolvimento. Esse resultado, que corroborou a Hipótese 4A do presente trabalho, contrariou o senso comum do meio acadêmico de países da América Latina. Ao realizar pesquisa com autores latino-americanos que publicaram trabalhos em periódicos científicos, questionando sobre a possível existência de preconceito editorial, Victora e Moreira (2006) relataram:

> Durante o inquérito, vários colegas expressaram sua opinião sobre a questão. Os principais comentários foram do tipo "concordo que haja algum grau de preconceito, mas nossos artigos também apresentam problemas e peculiaridades que os tornam pouco atrativos". Os problemas mais frequentemente mencionados incluem deficiências na redação em inglês e a má qualidade geral de alguns artigos. (p. 38)

Ao menos em estudos sobre prevalência de DPP, a impressão de que trabalhos de países em desenvolvimento teriam qualidade inferior à dos realizados em países desenvolvidos não pôde ser sustentada pelos dados. Portanto, é possível questionar se autores de países "do Sul" estariam perpetuando o preconceito, ao considerar que os trabalhos realizados em países em desenvolvimento teriam menor qualidade, o que os levaria a citá-los com menor frequência. Pesquisadores de países em desenvolvimento estariam vivenciando um autopreconceito?

Infelizmente, a notícia de que trabalhos científicos de países em desenvolvimento tiveram qualidade semelhante à de países desenvolvidos não pode ser considerada positiva, visto que a avaliação pela Escala de Loney et al. (1998) mostrou que ambos os grupos apresentaram qualidade fraca. Assim, apesar de existir grande número de trabalhos sobre prevalência de DPP, a maior parte não usou metodologia adequada e/ou apresentou problemas no relato das informações sobre métodos e resultados.

O único item que obteve pontuação em mais de 90% dos trabalhos avaliados foi o número 8, ou seja: apresentação do perfil das participantes. Os demais não foram observados nem por metade dos estudos incluídos. Algumas questões, como por exemplo, apresentação do intervalo de confiança e detalhamento por subgrupos de análise; apresentação da confiabilidade e validade do instrumento; apresentação da taxa de resposta e das análises sobre perfil de respondentes e não respondentes; e apresentação do perfil das participantes poderiam ser solucionadas pela simples introdução dessas informações no relato, indicando que muitas falhas resultaram de ignorância e/ou negligência de autores e pareceristas de periódicos científicos.

A Escala de Loney et al. (1998) foi alvo de críticas: considerou-se que os itens não fornecem muitos detalhes sobre as propriedades estatísticas do trabalho avaliado (Giannakopoulos et al., 2012); e que o resultado de zero a oito pontos não diferencia os três domínios que a escala se propõe a avaliar, isto é, validade dos métodos, interpretação e aplicabilidade dos resultados (Shamliyan et al., 2010). No momento em que o método da presente revisão estava em preparação, a Escala de Loney et al. (1998) era um dos poucos instrumentos desenvolvidos especificamente para avaliação de estudos observacionais, e apresentava a vantagem de fornecer um resultado numérico e passível de ser usado em análises estatísticas. Posteriormente, foi publicado o instrumento de Giannakopoulos et al. (2012). Considera-se que itens desse questionário que não estavam presentes na Escala de Loney et al. (1998) foram contemplados em outras questões do formulário de coleta de dados usado no presente estudo, como por exemplo, relato de critérios de inclusão e exclusão.

A Escala de Loney et al. (1998) também apresenta a desvantagem de não diferenciar entre qualidade do método e qualidade do relato do estudo científico. Os itens 1 (amostragem), 2 (base amostral), 3 (tamanho da amostra), 5 (medida do desfecho com baixo nível de viés) e 6 (taxa de resposta adequada) se referem à qualidade do método e precisam ser pensados desde o planejamento da pesquisa. Já os itens 4 (relatar confiabilidade de validade do instrumento de medida), 7 (relatar intervalo de confiança e fazer análises de subgrupos) e 8 (relatar perfil dos participantes) se referem ao relato e podem

ser corrigidos se houver necessidade (por exemplo, por solicitação de um parecerista). Considera-se, portanto, que para avaliar a qualidade dos estudos para uma revisão sistemática com metanálise, os itens 1, 2, 3, 5 e 6 são mais importantes.

Uma crítica possível à presente revisão seria ter usado demasiado rigor na avaliação dos trabalhos. Por exemplo, apenas estudos que usaram amostra aleatorizada receberam pontuação no item 1, sendo que outros tipos de amostra poderiam resultar em um bom custo-benefício. A amostra sistemática, em que um número da lista de possíveis participantes é escolhido aleatoriamente e, a partir dele, um intervalo numérico é usado para selecionar os demais participantes; e a amostra consecutiva, em que todos os participantes acessíveis em um determinado período são incluídos, podem ser consideradas alternativas viáveis à amostra aleatória, por apresentarem menores chances de viés que a amostra de conveniência (Lunsford & Lunsford, 1995). Optou-se por maior rigor nas avaliações, visto que havia interesse em identificar estudos que ocupassem os níveis mais elevados de evidência científica, conforme recomendações da Colaboração Cochrane (Higgins & Green, 2011) e da OCEBM (OCEBM Levels of Evidence Working Group, 2011).

7.5 Prevalência de DPP no mundo

A revisão sistemática resultou na inclusão de 337 trabalhos sobre prevalência de DPP, realizados em 61 países e publicados nos últimos 32 anos em 147 periódicos científicos. As prevalências variaram entre 0,5% (IC 95% 0,00%-1,48%) (Kok, Chan, & Ratnam, 1994) e 62,8% (IC 95% 60,44%-65,23%) (Beck, Gable, Sakala, & Declercq, 2011). A prevalência média global de DPP foi de 16,63% (IC 95% 16,50%-16,76%). No entanto, a síntese global das prevalências em uma metanálise não pode ser considerada representativa da prevalência de DPP no mundo, devido à alta variabilidade entre os estudos primários. Essa variação nos resultados de prevalência de DPP pode se dever a, pelo menos, três grupos de fatores: diferenças metodológicas entre estudos incluídos, diferenças no perfil biopsicossocial e econômico das participantes dos estudos, e diferenças culturais entre as regiões. As diferenças culturais

podem afetar a forma como os sintomas de DPP são percebidos e interpretados, de forma que a tradução de um instrumento para outro idioma, sem levar em consideração as características do novo contexto em que será aplicado, pode prejudicar a identificação de transtornos mentais (Halbreich & Karkun, 2006). Uma metanálise da validade da EDPE, que é o instrumento mais usado do mundo para identificar DPP, indicou que não há evidências suficientes para assegurar a validade dessa escala em diferentes grupos culturais (Gibson et al., 2009).

Ao interpretar a prevalência de DPP praticamente nula em estudos realizados em culturas orientais, Halbreich e Karkun (2006) consideraram que estaria relacionada com variações na percepção de normalidade e dos sintomas de DPP, que levaria a diferenças de interpretação de instrumentos ocidentais, como a EDPE. No entanto, conforme dados da presente revisão, embora a menor prevalência de DPP tenha sido encontrada em estudo realizado em Singapura (prevalência = 0,50%, IC 95% 0,00-1,48; Kok, Chan, & Ratnam, 1994), outras duas pesquisas que encontraram prevalências excepcionalmente baixas foram realizadas em países de cultura ocidental, como Noruega (prevalência = 0,92%, IC 95% 0,00-2,71; Skari et al., 2002) e Estados Unidos da América (prevalência = 1,49%, IC 95% 0,00-3,15; Affonso, Lovett, Paul, & Sheptak, 1990). Esses dois artigos não foram incluídos na revisão de Halbreich e Karkun (2006), mostrando a importância de se realizar buscas abrangentes em revisões sistemáticas.

Examinando-se em detalhe os três estudos com as menores prevalências de DPP na presente revisão sistemática, não foi encontrado um fator único que pudesse explicar a ocorrência de taxas tão baixas, e sim, características específicas de cada trabalho. Skari et al. (2002) usaram o General Health Questionaire (GHQ) como instrumento para avaliar DPP e argumentaram que a subescala de depressão desse instrumento identifica sintomas de transtorno depressivo maior, e que estudos anteriores podem ter considerado, de forma errônea, sintomas de depressão menor ou outros

sintomas de angústia[39] após o parto como indicativos de depressão clínica, por terem usado instrumentos de autoavaliação menos específicos, como a EDPE. Affonso et al. (1990) usaram amostra de participantes cujo perfil predispunha a baixo risco de DPP: critérios de inclusão determinaram que as mães deveriam ser primíparas, com feto viável, casadas ou vivendo com o pai do bebê e sem episódio de depressão nos últimos cinco anos. Além disso, as mulheres tiveram média de idade de 30 anos e alta escolaridade, a maioria era caucasiana e havia planejado a gestação. O critério para identificar DPP foi um protocolo de diagnóstico baseado em entrevista clínica e adaptado para evitar a confusão de sintomas comuns da gestação e do puerpério com sintomas depressivos. Kok et al. (1994) realizaram estudo em puérperas de Singapura, de forma que as diferenças interculturais são importantes na interpretação do resultado. Foi usada a EDPE com ponto de corte 16, considerado elevado em relação a outros estudos. Não foi apresentado o perfil das participantes, que também pode ter contribuído para a baixa prevalência, caso a maioria apresentasse poucos fatores de risco.

7.6 Comparação da prevalência de DPP em países desenvolvidos e em desenvolvimento

Husain et al. (2006) observou que os resultados relativos à prevalência de DPP em estudos realizados em países desenvolvidos e em desenvolvimento apresentaram muitas divergências, de forma que não foi possível verificar, em revisão narrativa, se o grau de desenvolvimento do país teria relação com o desfecho. No entanto, ao realizar a presente revisão sistemática com metanálise, foi possível concluir que, de fato, países em desenvolvimento apresentaram, em média, maior prevalência de DPP que os países desenvolvidos.

Análises descritivas indicaram que a prevalência de DPP em países em desenvolvimento foi de 21,31% (IC 95% 20,87%-21,85%), valor semelhante ao resultado obtido por Melo Junior (2011): 20,09% (IC 99% 19,28%-20,90%). O

---

[39] No original, "*distress*".

valor encontrado para países desenvolvidos, de 16,02% (IC 95% 15,89%-16,16%), foi superior ao obtido por Melo Junior (2011): 11,22% (IC 99% 10,97%-11,47%). Essa diferença provavelmente ocorreu pela abrangência da presente revisão, que incluiu maior quantidade de estudos primários. Além disso, Melo Junior (2011) usou critérios de seleção que determinaram a exclusão de estudos com amostras inferiores a 150 participantes, e foi verificado, na presente revisão, que amostras menores estiveram associadas com maiores prevalências de DPP. Por fim, também podem ter ocorrido diferenças nos critérios usados para definir países desenvolvidos ou em desenvolvimento, pois Melo Junior (2011) não mencionou como os países foram classificados.

A Hipótese 2B do presente estudo propôs que, controlando-se variáveis metodológicas, a diferença entre médias ponderadas das prevalências de DPP de países desenvolvidos e em desenvolvimento desapareceria. Os resultados do modelo de metarregressão refutaram essa hipótese, mostrando que, mesmo mantendo controle das variáveis metodológicas, a prevalência de DPP foi inversamente proporcional ao IDH do país. É preciso analisar, portanto, por que ocorreu essa diferença.

A interpretação dos resultados deve levar em consideração, novamente, a grande heterogeneidade entre os trabalhos. Alguns estudos realizados em países em desenvolvimento apresentaram prevalências compatíveis com os resultados médios de países desenvolvidos, como por exemplo: 2,6% (IC 95% 0,4%-4,9%) na Arábia Saudita (Amr, Hassan, & Balaha, 2010), 4,0% (IC 95% 1,4%-6,6%) na cidade de Recife, Brasil (Cantilino et al., 2010) e 4,2% (IC 95% 1,9%-6,5%) na Colômbia (Latorre-Latorre, Contreras-Pezzotti, García-Rueda, & Arteaga-Medina, 2006); e o estudo com maior prevalência (62,8%, IC 95% 60,4%-65,2%) na presente revisão sistemática foi realizado nos Estados Unidos da América, que é um país desenvolvido (Beck et al., 2011). Prevalências bastante distintas puderam ser encontradas em estudos realizados dentro de um mesmo país, como por exemplo, no Brasil: 4,0% (IC 95% 1,4%-6,6%) em Recife (Cantilino et al., 2010) e 39,4% (IC 95% 33,8%-45,0%) em Vitória (Ruschi et al., 2009); ou nos Estados Unidos da América: 1,5% (IC 95% 0,00%-3,2%) em San Francisco (Affonso et al., 1990) e 62,8%

(IC 95% 60,4%-65,2%) em estudo realizado pela internet (Beck et al., 2011). Esses resultados são reforçados pelo dado encontrado na metarregressão, indicando que a cidade, e não o país, foi um nível significativo de agrupamento de estudos.

Mas até mesmo dentro da mesma cidade, as prevalências foram bastante diferentes: em Santiago (Chile), variaram entre 10,2% (IC 95% 4,5%-15,9%; Jadresic, Araya, & Jara, 1995) e 34,5% (IC 95% 30,5%-38,5%; Jadresic & Araya, 1995). Esses resultados mostraram que a heterogeneidade da prevalência nos estudos incluídos não esteve associada, somente, ao local onde os dados foram coletados. Outras variáveis devem ter influenciado os resultados.

O modelo de metarregressão avaliou o efeito multivariado de diversas variáveis metodológicas sobre a prevalência de DPP e encontrou preditores significativos, mas devido à baixa quantidade e precisão de alguns dados nos estudos primários avaliados, não foi possível testar algumas variáveis que poderiam ser importantes no modelo, como por exemplo: período após o parto e perfil das participantes incluídas. Renda e escolaridade são itens que compõem o IDH (UNDP, 2011) e que não puderam ser testados na metarregressão. Baixa renda ou nível socioeconômico e baixa escolaridade foram fatores de risco para a ocorrência de DPP (O'Hara & Swain, 1996; Moraes et al., 2006). Portanto, uma das explicações para a maior prevalência em países em desenvolvimento poderia ser a maior quantidade de mães com baixa renda e escolaridade nesses países, e a consequente maior participação de mulheres com esse perfil nas pesquisas realizadas nesses locais.

Ainda assim, não é possível determinar quais seriam os fatores que aumentariam a predisposição de mulheres pobres e com poucos anos de educação formal a apresentar sintomas depressivos no puerpério. As pesquisas quantitativas têm constatado que essas variáveis apresentaram uma associação, cujo mecanismo ainda não foi elucidado. Até o momento, não se conhecem pesquisas qualitativas que tenham investigado a relação entre renda, escolaridade e DPP. Sugere-se a realização de pesquisas usando técnicas de grupos focais ou entrevistas semi-estruturadas, com mulheres de baixo e alto nível socioeconômico, com e sem DPP.

7.7 Influência de fatores metodológicos sobre resultados de estudos de prevalência de DPP

A síntese dos resultados dos estudos incluídos em estimativa única não pôde ser considerada representativa da prevalência mundial de DPP, devido à alta heterogeneidade, constatada no teste de qui-quadrado, no índice I^2 e nas observações de gráficos de funil. A ausência de análises de heterogeneidade em revisões sistemáticas de prevalência de DPP, portanto, permite questionamento sobre a validade dos resultados, que podem ter apresentado o problema conhecido como "misturar maçãs e laranjas" (Higgins & Green, 2011), isto é: combinar estudos com metodologias demasiadamente distintas e cuja síntese em valor único teria significado duvidoso do ponto de vista prático. Duas metanálise anteriores (Gavin et al., 2005; Villegas et al., 2011) verificaram heterogeneidade significativa entre os estudos, mas não buscaram explicar as possíveis causas.

Uma força do presente trabalho foi ter explorado a heterogeneidade, investigando os fatores metodológicos que seriam responsáveis pela variabilidade nos resultados dos estudos de prevalência em análises de subgrupos e usando metarregressão. Dessa maneira, foi possível testar hipóteses referentes ao efeito de variáveis metodológicas e regionais sobre os resultados de prevalência.

7.7.1 Discussão das hipóteses de pesquisa

A Hipótese 1 do presente trabalho estimou, com base em metanálises anteriores (O'Hara & Swain, 1996; Melo Junior, 2011), que a prevalência média global de DPP seria de 12% a 14% e foi refutada, pois a prevalência média global foi de 16,63% (IC 95% 16,50-16,76). Esse valor foi superior ao previsto pela Hipótese 1 e também ao encontrado em três metanálises anteriores que tiveram pouca ou nenhuma participação de trabalhos realizados em países em desenvolvimento (O'Hara & Swain, 1996; Gavin et al., 2011). No entanto, foi semelhante ao encontrado em metanálise de estudos realizados na África (Sawyer et al., 2010) e menor que o encontrado em metanálise de estudos

realizados em zonas rurais de países desenvolvidos e em desenvolvimento (Villegas et al., 2011). Assim, permanece a crítica de que a prevalência de 10% a 15%, usualmente citada na literatura sobre DPP, não é representativa do universo de estudos sobre o tema (Halbreich & Karkun, 2006; Lobato et al., 2012).

A Hipótese 2A, que previu que o valor da prevalência média em estudos realizados em países desenvolvidos seria menor que o valor da prevalência média em estudos realizados em países em desenvolvimento, foi corroborada pelos resultados. Até o momento, as únicas metanálises conhecidas que haviam comparado países desenvolvidos e em desenvolvimento quanto à prevalência de DPP foram a de Melo Junior (2011) e a de Villegas et al. (2011). Ambas apresentaram conclusão semelhante à da presente revisão, isto é, de que a prevalência média de DPP de países em desenvolvimento foi superior à de países desenvolvidos. No entanto, não procuraram verificar se fatores metodológicos atuariam como variáveis de confusão. Essa hipótese foi testada e refutada na presente revisão (Hipótese 2B), pois mesmo controlando variáveis metodológicas dos estudos primários, o IDH do país foi um preditor significativo da prevalência de DPP, conforme resultado do modelo de metarregressão. Portanto, é provável que fatores econômicos e psicossociais tenham influência sobre a ocorrência de DPP.

A influência de fatores metodológicos sobre os resultados de prevalência de DPP foi testada nas Hipóteses 3B a 3I. A Hipótese 3A não pôde ser testada, devido a deficiências dos relatos dos estudos primários em relação ao tempo exato após o parto em que os dados foram coletados.

A Hipótese 3B, que previa que o grupo formado por estudos que usaram entrevista clínica como método para avaliar DPP teria prevalência média menor que o grupo formado por estudos que usaram escalas de autoavaliação, foi corroborada. O'Hara e Swain (1996) encontraram diferentes prevalências de acordo com o instrumento usado para identificar DPP, mas aparentemente, essas diferenças não estiveram relacionadas com o tipo de avaliação, isto é, se foram usados instrumentos de autoavaliação ou entrevista clínica. Alguns exemplos de resultados obtidos nessa metanálise foram: prevalência de 12,0% (IC 95% 10,9%-13,1%) usando a EDPE (autoavaliação); 18,0% (IC 95%

16,1%-19,9%) usando o *Center for Epidemiologic Studies Depression Scale* (CES-D – autoavaliação); 10,5% (IC 95% 9,7%-11,3%) usando o *Research Diagnostic Criteria* (RDC – entrevista clínica); e 17,8% (IC 95% 15,8%-19,8%) usando o critério de Goldberg (entrevista clínica). A metanálise de Melo Junior (2011) indicou não haver diferenças entre instrumentos auto-administrados (15,58%; IC 99% 15,12%-16,04%), entrevistas presenciais (14,52%; IC 99% 13,87%-15,17%), entrevistas por telefone (15,24%; IC 99% 14,49%-15,99%) ou envio de questionários pelo correio (12,74%; IC 99% 12,30%-13,18%).

Na presente metanálise, as diferenças de prevalência média associadas ao tipo de instrumento usado para identificar DPP foram relevantes e ocorreram em países desenvolvidos e em desenvolvimento. É possível inferir, da comparação entre o presente estudo e as metanálises anteriores, que o principal fator que afetou a prevalência não foi a forma de administrar o instrumento (presencial, por telefone, por correio, etc.), como foi testado por Melo Junior (2011), e sim, quem foi o avaliador: a própria puérpera (autoavaliação) ou um profissional de saúde (heteroavaliação). Esse resultado tem fundamento, visto que escalas de autoavaliação costumam usar pontos de corte que assegurem maior sensibilidade do que especificidade, pois é preferível encaminhar mais puérperas que o necessário para diagnóstico por profissionais de saúde mental do que deixar de acompanhar puérperas que poderiam apresentar DPP e não foram corretamente identificadas pelo instrumento.

A Hipótese 3C, referente ao delineamento de pesquisa, foi corroborada pelos resultados, o que pode ser explicado pelo viés de prevalência: as mães que manifestaram DPP por um maior período de tempo tiveram maior chance de ser incluídas nos estudos transversais, de forma que não se pode diferenciar entre o surgimento de novos casos e a duração dos mesmos (Bastos & Duquia, 2007). É importante, portanto, que sejam feitas pesquisas longitudinais, com boa distinção entre incidência e prevalência de DPP.

A Hipótese 3D, referente à técnica de amostragem, foi corroborada, pois estudos que usaram técnicas de amostragem consideradas inadequadas tiveram prevalência maior que os que usaram técnicas consideradas adequadas. Estudos que não apresentaram informações suficientes para que

se compreendesse o procedimento de amostragem tiveram prevalência de DPP superior à dos estudos que usaram amostra aleatória, não aleatória ou censo. Isso indica que, provavelmente, esses trabalhos usaram amostras de conveniência em que as puérperas com DPP estiveram mais disponíveis que as sem DPP. É possível que as mães com dificuldades para lidar com o ajustamento exigido no puerpério tenham participado com maior frequência das pesquisas, em busca de atenção, acolhimento e escuta oferecidos pelo pesquisador, ou de encaminhamento para tratamento.

O tamanho amostral apresentou associação com a prevalência de DPP, contrariando a Hipótese 3E e concordou com o resultado verificado na metanálise de Melo Junior (2011). Se as participantes das pesquisas sobre DPP fossem selecionadas de forma a dar a cada mulher da população em estudo chance igual de ser incluída na amostra, esperar-se-ia que o tamanho amostral não teria associação com o desfecho. A presença de associação indicou que as amostras dos estudos primários apresentaram vieses, suposição reforçada pelo resultado da prevalência de DPP de acordo com a técnica de amostragem, que indicou maior prevalência de DPP quando o procedimento de amostragem estava inadequado. Pode-se inferir que a prevalência de DPP em estudos com amostras maiores é mais confiável, visto que se aproximam mais da realidade do universo. Por razão desconhecida, mulheres com DPP parecem ter apresentado maior chance de serem incluídas nos trabalhos com amostras menores. Novamente, constatou-se uma tendência a superestimar a prevalência de DPP devido a deficiências metodológicas.

Constatou-se efeito do uso de base amostral adequada, isto é, amostra que representou toda a população de uma cidade ou região, sobre os resultados de prevalência de DPP, corroborando a Hipótese 3F. Mais uma vez, o uso de método adequado resultou em menor prevalência de DPP. Esse dado merece interpretação cautelosa, visto que poucos trabalhos usaram bases amostrais adequadas. O uso de base amostral adequada associado a censo em pequenas vilas ou cidades (ex.: Cooper et al., 1999; Ali, Ali, & Azam, 2009) foi comum, o que implicou na inclusão de populações com características diferenciadas.

O contexto de pesquisa teve efeito sobre a prevalência média de DPP, resultado que contrariou a Hipótese 3H do presente estudo. As razões para as diferenças entre estudos nacionais, regionais, realizados em uma única cidade ou instalação médica não estão claras e demandam investigação.

Foi refutada, também, a Hipótese 3I do presente trabalho, que previu que o ponto de corte da EDPE seria inversamente proporcional às prevalências de DPP encontradas nos estudos primários, conforme resultado encontrado na revisão sistemática de Melo Junior (2011). Uma das razões para essa ocorrência pode ser a heterogeneidade metodológica, visto que os 198 trabalhos que foram comparados apresentaram diferenças em outras variáveis que foram consideradas preditores significativos da prevalência de DPP, como delineamento, cidade onde os dados foram coletados, tamanho amostral e IDH do país. Também é possível que o uso de validações regionais para a EDPE tenha adaptado os pontos de corte da escala às questões idiomáticas e culturais, evitando alterações expressivas nos resultados de prevalência.

As Hipóteses 4A a 4C dizem respeito à qualidade metodológica dos estudos incluídos, avaliada pela Escala de Loney et al. (1998). A qualidade metodológica não diferiu entre estudos de países desenvolvidos ou em desenvolvimento (corroborando a Hipótese 4A), não apresentou associação com o ano de publicação (refutando a Hipótese 4B) e não teve efeito sobre a prevalência média de DPP (refutando a Hipótese 4C). Portanto, a qualidade metodológica medida pela pontuação na Escala de Loney et al. (1998) não apresentou associação importante com nenhuma das variáveis analisadas. A ausência de associação entre a prevalência de DPP e a qualidade metodológica de estudos, avaliada de forma quantitativa por instrumento padronizado, também foi constatada em revisão sistemática de estudos africanos (Sawyer et al., 2010).

No entanto, considera-se que a qualidade metodológica apresentou efeito sobre os resultados dos estudos epidemiológicos sobre DPP, pois itens como a técnica de amostragem, o tamanho amostral e o tipo de instrumento usado para identificar DPP são fatores de qualidade importantes e apresentaram associação com a prevalência. A explicação mais provável é que a pontuação da Escala de Loney et al. (1998) não se mostrou adequada para o

propósito de avaliar a qualidade metodológica na presente revisão sistemática. Algumas razões podem ter sido a não diferenciação entre qualidade do relato e qualidade da metodologia empregada e a baixa concordância entre revisoras no item 7 da escala, o que interferiu na pontuação como um todo. Outras críticas à Escala de Loney et al. (1998) podem ser encontradas na Seção "7.4 Análise crítica da qualidade metodológica dos estudos incluídos" da presente Tese.

Por fim, a Hipótese 5 se referiu ao modelo de metarregressão, que será comentado na seção "7.7.2 Discussão dos resultados da metarregressão", a seguir.

7.7.2 Discussão dos resultados da metarregressão

Os resultados da metarregressão indicaram que os principais fatores metodológicos relacionados com as prevalências de DPP em diferentes estudos foram o instrumento usado para identificar DPP, o delineamento do estudo, o tamanho amostral e o IDH, além da cidade em que a coleta de dados foi realizada.

Em relação às variáveis regionais, o país em que o estudo foi realizado não foi significativo como fator explicativo da prevalência de DPP na presente metarregressão, assim como não havia sido em estudo anterior (O'Hara & Swain, 1996). No entanto, uma informação que não foi investigada em revisões anteriores foi significativa: a cidade em que a coleta de dados foi realizada, que foi inserida no modelo como cluster. Portanto, diferenças culturais e sociais dentro da mesma nação parecem exercer papel importante sobre a prevalência de DPP. As disparidades dentro do mesmo país e o agrupamento dos estudos por cidades são bastante plausíveis, sobretudo, quando são analisados países de grandes dimensões e com variações regionais, como por exemplo, o Brasil. Em estudo de validação da PDSS no Brasil, afirmou-se que:

> O Brasil tem sido considerado um país multifacetado, em que estão reunidos diversos grupos étnicos. Uma combinação de fatores culturais, como idioma, tradições, crenças, valores e situação socioeconômica interagem regionalmente e nacionalmente como as

> forças sociais que definem a sociedade brasileira. As discrepâncias observadas entre o Nordeste – uma região afetada por limitações de alimento e água e deficiências nos sistemas educacional e de saúde – e a região Sul, que é mais desenvolvida, formam uma das disparidades mais conspícuas do Brasil. Essas variações regionais são de tal magnitude, que foi considerado 'discutível' se instrumentos de rastreamento para transtornos mentais 'podem ser aplicados em diferentes e regiões e cidades do país sem interferência da cultura local' (Carvalho, Silva, Grisi & Escobar, 2008, apud Zubaran et al., 2009). De fato, dados diferentes de prevalência de depressão pós-parto têm sido documentados em regiões distintas do Brasil – mesmo quando o mesmo instrumento de rastreamento foi usado.[40] (Zubaran et al., 2009, p. 245).

Foi encontrada maior prevalência média associada ao uso de escalas de autoavaliação, tanto em países desenvolvidos, quanto em desenvolvimento. Como escalas de autoavaliação não devem ser consideradas um diagnóstico oficial, esse resultado está de acordo com afirmação de que estariam ocorrendo excessos ao identificar patologias no período do puerpério (Mathey, 2010).

Também se constatou que o tamanho amostral e o delineamento dos estudos foram preditores significativos no modelo de metarregressão. Esses resultados, que corroboraram a Hipótese 3C e refutaram a 3E do presente estudo, já foram discutidos na Seção "7.7.1 Discussão das hipóteses de pesquisa". Até onde se tem conhecimento, o delineamento de pesquisa não foi investigado em revisões sistemáticas anteriores como um fator que poderia ter impacto sobre os resultados dos estudos sobre epidemiologia da DPP.

Na metanálise de O'Hara e Swain (1996), a duração do período após o parto em que foi feita a avaliação de DPP também foi identificada como fator preditivo da prevalência de DPP. As informações sobre período após o parto

---

[40] No original: Brazil has been considered a multifaceted country in which several ethnic groups congregate. A combination of cultural factors such language, traditions, beliefs, values, and socio-economic status interact regionally and nationally as the defining social forces in Brazilian society. The discrepancies observed between the Northeast – a region affected by limited food and water resources and deficient educational and health care systems – and the more developed Southern region form one of the most conspicuous national disparities in the Brazil. These regional variances are of such magnitude that it has been considered ''disputable'' whether screening instruments for mental disorders ''can be applied in different regions and cities of the country without the interference by local culture''. In fact, dissimilar prevalence figures of postnatal depression have been documented in distinct regions in Brazil – even when the same screening tool was used […].

não foram testadas na presente revisão, devido às lacunas constatadas nos relatos dos estudos incluídos, que resultariam em número excessivo de dados faltantes (*missing values*) nas análises.

As análises de sensibilidade, que testaram os efeitos de possíveis alterações na metodologia sobre os resultados do presente estudo, mostraram que os resultados não sofreriam alteração expressiva caso fossem inseridas restrições na revisão sistemática quanto aos tipos de trabalho incluídos (excluindo-se teses e dissertações; ou excluindo-se estudos que não declararam que avaliar prevalência de DPP foi um dos objetivos), aos anos de publicação (excluindo-se estudos mais antigos) ou à qualidade mínima dos trabalhos (excluindo estudos com qualidade igual ou inferior a dois pontos na Escala de Loney et al., 1998). Esses resultados indicaram que as evidências relativas à prevalência de DPP, encontradas no presente trabalho, são consistentes.

7.8 Implicações para a prática profissional

Estudos primários sobre prevalência de DPP podem ser facilmente encontrados por profissionais de saúde, visto que 337 foram incluídos na presente revisão. No entanto, a qualidade fraca desses trabalhos tem comprometido a interpretação e aplicabilidade dos resultados. Os dados mais confiáveis seriam provenientes dos trabalhos considerados evidências nível 1 em estudos de prevalência, ou seja, levantamentos realizados em amostras aleatórias de participantes nos locais de interesse (OCEBM Levels of Evidence Working Group, 2011). No caso da DPP, sugere-se que sejam considerados nível 1 de evidência apenas trabalhos que usaram entrevista clínica, visto que pontuações de instrumentos de autoavaliação não podem ser considerados como único critério para um diagnóstico em saúde mental. No entanto, o número de trabalhos com características metodológicas adequadas foi pequeno, de forma que os profissionais de saúde raramente podem contar com boas evidências sobre prevalência de DPP para nortear as suas práticas.

O presente trabalho mostrou que os resultados relativos à prevalência de DPP foram fortemente influenciados por fatores metodológicos. Assim, é

importante que os profissionais de saúde compreendam e estejam familiarizados com conceitos de metodologia de pesquisa, de forma que possam avaliar a validade dos estudos epidemiológicos e até que ponto eles podem ser tomados como base para identificar a prevalência de DPP na região e no período de interesse. Constatou-se que o critério usado para identificar as mães que estariam com DPP foi uma variável que contribuiu para a heterogeneidade entre estudos, pois instrumentos de autoavaliação resultaram em prevalência mais alta que entrevistas clínicas. Também foram fatores importantes o tamanho amostral e o delineamento do trabalho. Outras variáveis, que não puderam ser investigadas no presente estudo pela insuficiência de dados, devem ser levadas em conta pelo profissional, como por exemplo, o perfil socioeconômico e obstétrico das participantes.

A discrepância entre resultados obtidos por meio de entrevista clínica e de instrumentos de autoavaliação mostra que o método de rastreamento e diagnóstico de DPP deve ser planejado de forma cuidadosa, visto que pode influenciar significativamente os resultados. Sugere-se que, em instituições de grande porte, a identificação de mulheres com DPP seja feita através de método misto, combinando rastreamento por escalas de autoavaliação com avaliação clínica feita por profissionais de saúde mental (psiquiatras ou psicólogos) para as mães que apresentarem pontuação acima do ponto de corte. Esse método apresenta vantagens econômicas, pois a aplicação de escalas de autoavaliação é menos custosa que a avaliação clínica. O instrumento de rastreamento deve ter propriedades psicométricas aceitáveis, apresentadas em estudos de validação, e a entrevista clínica jamais deverá ser dispensada para as mulheres que apresentarem pontuações acima do ponto de corte da escala. Os profissionais de saúde devem ser instruídos e treinados para realização de diagnóstico, evitando que situações comuns no puerpério, como perda de peso, alterações no sono e fadiga, sejam confundidas com sintomas de DPP (Mathey, 2010). Considerando-se que a DPP pode surgir desde as primeiras semanas após o parto até cerca de um ano após o nascimento do bebê, é recomendável que a escala de rastreamento seja aplicada em diversos momentos, iniciando-se no período gestacional, quando já existe vulnerabilidade para sintomas depressivos (Bennett et al., 2004).

Nos estudos que usaram instrumentos de autoavaliação, 17,30% (IC 95% 17,17%-17,44%) das participantes tiveram pontuação indicativa de risco de DPP, ou seja, elas mesmas declararam que apresentaram sofrimento psíquico após o nascimento do bebê. Esses resultados indicaram que o primeiro ano após o nascimento do bebê é um período em que muitas mães estão sujeitas a sentimentos vivenciados como negativos, provavelmente decorrentes da adaptação necessária à nova dinâmica familiar e ao novo papel assumido pela mulher. Ainda que parte das participantes que tiveram altas pontuações em instrumentos de rastreamento não estivesse clinicamente deprimida, provavelmente necessitavam de apoio para enfrentar a transição para a maternidade. As mudanças relacionadas ao nascimento dos filhos, inevitavelmente, ocorrem em todas as culturas:

> O nascimento de um novo bebê, especialmente o primeiro, é um evento de vida significante em qualquer cultura. Está associado com uma das situações mais dramáticas de múltiplas quedas hormonais – pareada com uma mudança significante na situação socioeconômica, estilo de vida e carga de trabalho da mãe e de sua família imediata. Se um modelo compreensivo biopsicossocioeconômico for aplicado ao período pós-parto (como deveria), então é difícil explicar como um evento significante de vida como esse não resultaria em angústia em pelo menos algumas mães em qualquer cultura.[41] (Halbreich & Karkun, 2006,109).

Assim, o planejamento de intervenções em saúde mental deve considerar não apenas o tratamento das mães consideradas deprimidas de acordo com critérios clínicos, mas também o apoio às mulheres em período perinatal, independentemente de apresentarem qualquer patologia. Essas intervenções podem consistir em grupos de apoio, plantão psicológico, visitas domiciliares, oferecimento de atividades que propiciem interação social e lazer às gestantes e puérperas, ou outra opção que as equipes de saúde e os

[41] No original: The delivery of a new baby, especially the first one, is a significant life event in any culture. It is associated with one of the most dramatic multiple hormonal withdrawal situations — coupled with a significant change in the socioeconomic, life style and work load of the mother and her immediate family. If a comprehensive bio-psychosocio-economic model is applied to the postpartum period (as it should), then it is difficult to explain how such a significant life event would not result in distress in at least some mothers in any culture.

usuários considerem adequada. A avaliação da efetividade dos programas através de pesquisas de boa qualidade é fundamental para o aprimoramento, bem como para a ampliação dessas atividades, quando efetivas.

As estimativas de prevalência de DPP, mesmo considerando apenas os estudos que usaram entrevista clínica como critério para diagnóstico, foram relativamente elevadas, atingindo, em média, 9,87% (IC 95% 9,51%-10,24%) das mulheres em países desenvolvidos e 14,88% (IC 95% 14,04%-15,73%) em países em desenvolvimento. A alta frequência do transtorno indicou a necessidade de criação de programas não apenas de tratamento, mas também de prevenção e, principalmente, promoção da saúde.

Se a DPP for, como parece ser, um transtorno cuja prevalência varia de acordo com a renda das participantes, ou com o nível de desenvolvimento dos países, deve ser considerada um fenômeno não apenas psicológico ou biológico, mas também social. Desse ponto de vista, seriam necessárias políticas de promoção da saúde, isto é, que desenvolvam padrões de vida mais saudáveis, com maior participação dos indivíduos na sociedade. Não basta ampliar o acesso a serviços de saúde mental, embora, obviamente, essa ação também seja necessária. Situações como más condições de trabalho, baixa escolaridade, renda insuficiente, habitações precárias, discriminação, segregação, ausência de oportunidades culturais e educacionais são a base de diversos transtornos mentais. O enfrentamento dessas questões sociais deve ser feito de forma intersetorial, envolvendo ações coordenadas dos governos, setores de saúde, organizações não governamentais, autoridades locais, indústria e mídia, além da participação de indivíduos, famílias e comunidades (Terris, 1990).

O desafio dos clínicos é planejar intervenções que extrapolem os dados estatísticos, levando em conta as singularidades das comunidades onde intervêm e a individualidade de cada mulher. Afinal, é da interação entre boas evidências científicas, experiência profissional e preferências e características dos pacientes que se constrói a Prática da Psicologia Baseada em Evidências.

7.9 Implicações para pesquisas

Se fosse considerado apenas o número de trabalhos incluídos na presente revisão sistemática sobre prevalência de DPP, seria possível afirmar que novos estudos sobre essa temática não seriam necessários. No entanto, dentre os 337 trabalhos incluídos, poucos podem ser considerados nível 1 de evidência científica de acordo com a hierarquia de evidências científicas de estudos de prevalência sugerida pelo OCEBM Levels of Evidence Working Group (2011). Além disso, a qualidade dos trabalhos incluídos, avaliada usando-se a Escala de Loney et al. (1998), foi considerada fraca. Assim, é possível afirmar que faltam pesquisas que possibilitem gerar evidências de boa qualidade sobre prevalência de DPP.

No meio científico, parece ser generalizada a impressão (muitas vezes falsa) de que estudos observacionais são simples e baratos em comparação com os experimentais. Muitos pesquisadores parecem desconhecer a complexidade de se fazer um estudo sobre prevalência que agregue informações precisas e, portanto, úteis para os clínicos e gestores de saúde. Além disso, conforme alertou Cantilino (2010), falta consenso metodológico entre pesquisadores, para que os dados sejam coletados de forma comparável. O presente trabalho levantou questões metodológicas fundamentais, que devem ser observadas em pesquisas sobre a prevalência de DPP:

1. **Técnica de amostragem:** em estudos epidemiológicos, é preciso realizar censo ou usar amostra probabilística, para evitar viés de seleção de participantes (Lunsford & Lunsford, 1995; Boyle, 1998; Loney et al., 1998). Esse fator é fundamental para caracterizar estudos nível 1 de evidência em pesquisas sobre prevalência e incidência (OCEBM Levels of Evidence Working Group, 2011).

2. **Base amostral adequada:** ao estimar a prevalência de DPP, é preciso que a amostra seja representativa da população estudada. Assim, respostas de uma amostra selecionada com base apenas nas mulheres que frequentam uma determinada instituição de saúde, ou em lista

telefônica, ou por meio de respostas a questionário disponível na internet, não podem ser generalizadas para a população de uma cidade ou país. Um exemplo de base amostral adequada para estimar a prevalência de DPP em uma cidade de grande porte pode ser encontrado em um estudo realizado em Porto Alegre, Brasil (Tannous, Gigante, Fuchs, & Busnello, 2008). Foi feito sorteio a partir do Registro de Nascidos Vivos, um banco de dados oficial do governo brasileiro, que armazena o nome e endereço de todas as mulheres que deram à luz a recém-nascidos vivos, a partir de dados fornecidos por hospitais e cartórios. As mulheres selecionadas foram visitadas em seus locais de moradia, que poderiam incluir residências comuns, hotéis, pensões e penitenciárias. Dessa maneira, assegurou-se que amostra selecionada foi representativa da população estudada.

3. **Cálculo do tamanho amostral**: é preciso planejar o tamanho da amostra, pois pode ocorrer desperdício de recursos quando são usadas amostras demasiadamente pequenas, que não têm capacidade de produzir resultados úteis; e também quando são usadas amostras exageradamente grandes, pois além de usar mais recursos que o necessário, podem fazer com que um efeito de pouca importância científica seja detectado como estatisticamente significativo (Lenth, 2001). O tamanho da amostra não apresenta qualquer relação com o tamanho do universo, e sim com o poder que se deseja alcançar na análise estatística, ou seja, a capacidade do teste de identificar diferenças significativas entre grupos de participantes. Uma calculadora de tamanho amostral pode ser encontrada na página: http://www.lee.dante.br/pesquisa/amostragem/calculo_amostra.html (acessado em 26/12/2012). No entanto, é preciso levar em consideração outros fatores, como os recursos disponíveis, a taxa esperada de perda de participantes e questões éticas. Recomenda-se consultar um Bacharel em Estatística, preferencialmente com experiência em estudos epidemiológicos.

4. **Registro das taxas de perda**: a taxa de perda de participantes deve ser declarada no estudo de prevalência, porque indica o quanto a amostra é representativa da população. Frequentemente, podem ocorrer vieses nos estudos devido às recusas e desistências. Por exemplo, seria possível que mães com DPP recusassem com maior frequência participar do estudo, por sentirem-se desinteressadas; ou que aceitassem com maior frequência, por sentirem necessidade de apoio. Assim, deve-se manter registro de quantos participantes seriam elegíveis para participar da pesquisa; quantos foram, de fato, convidados; quantos aceitaram participar; e, nos estudos longitudinais, quantos foram perdidos ao longo do tempo. Também é importante registrar características de perfil de respondentes e não respondentes, que devem ser comparadas estatisticamente e declaradas na publicação da pesquisa (Boyle, 1998; Loney et al., 1998; Giannakopoulos et al., 2012).

5. **Cuidados no diagnóstico**: os resultados da presente revisão sistemática mostraram que o uso de escalas de autoavaliação produziu prevalências de DPP maiores que o uso de critérios de diagnóstico baseados em entrevista clínica. De fato, críticas aos estudos epidemiológicos de DPP têm alertado para a tendência a superestimar a prevalência do transtorno, devido ao uso de escalas de autoavaliação; uso de pontos de corte inadequados para os instrumentos; e confusão de situações comuns no puerpério, como perda de peso, alterações no sono e fadiga, com sintomas de DPP (Mathey, 2010). Portanto, é necessário que todas as puérperas sejam avaliadas por meio de entrevista clínica realizada por profissionais de saúde mental instruídos e treinados. Ainda assim, não é possível confiar na avaliação feita por apenas um profissional, visto que a identificação de transtornos mentais passa por critérios subjetivos. Cada mulher deve ser avaliada por, no mínimo, dois profissionais, que devem estar cegos para qualquer outra avaliação, de forma a permitir calcular a taxa de concordância (Loney et al., 1998).

6. **Padronização do período após o parto**: de acordo com revisões sistemáticas anteriores, o período após o parto no qual a avaliação de DPP é realizada tem impacto sobre os resultados de prevalência (O'Hara & Swain, 1996; Halbreich & Karkun, 2006). As prevalências encontradas, em geral, foram maiores, quanto mais distante do parto foi realizada a avaliação de DPP; além disso, a duração do intervalo de tempo em que são realizadas as avaliações foi diretamente proporcional à prevalência. No presente estudo, não foi testado o efeito do período após o parto sobre os resultados de prevalência, devido à ausência de informações suficientes em grande parte dos estudos incluídos. Assim, é importante que futuros estudos declarem com maior precisão o intervalo de tempo após o parto em que as avaliações de DPP foram realizadas, incluindo parâmetros como: mínimo, máximo, média, desvio padrão, mediana. Também é interessante escolher um período de tempo que permita a comparação com outros trabalhos similares, a partir de revisão da literatura.

7. **Distinção entre incidência e prevalência**: é necessário que as pesquisas sobre epidemiologia da DPP façam distinção clara entre incidência e prevalência do transtorno (Halbreich & Karkun, 2006). A prevalência indica quantas puérperas estão deprimidas em um dado instante, enquanto a incidência indica o número de novos casos de depressão. Assim, trabalhos que excluem mulheres avaliadas como deprimidas na gestação são apropriados para avaliar incidência e inadequados para avaliar prevalência. Os conceitos epidemiológicos devem ser compreendidos e aplicados corretamente.

8. **Registro de características de perfil das participantes incluídas**: a maior parte dos trabalhos declarou características de perfil das participantes incluídas, mas ocorreu grande variação em relação a quais características foram mantidas em registro e como esses dados foram apresentados. A escala de qualidade de Giannakopoulos et al. (2012)

considerou que as características mais importantes a serem avaliadas em estudos de prevalência são: idade; sexo; condições de trabalho ou hobbies; classe social, educacional ou financeira; etnia; região de residência (urbana, suburbana ou rural); e dados relevantes sobre a saúde dos participantes, registrados em anamnese. No caso específico da DPP, sugere-se que, além desses aspectos, sejam registradas variáveis que foram consideradas significativas em estudos sobre fatores de risco, como por exemplo: histórico de depressão ou outro transtorno psiquiátrico (O'Hara & Swain, 1996; Beck, 1996; Beck, 2001; Robertson et al., 2004), depressão durante a gestação (Beck, 1996; Beck, 2001) e estado civil (Beck, 2001).

9. **Registro de locais e datas em que os dados foram coletados:** informações de local e data da coleta de dados estiveram ausentes em grande parte dos estudos incluídos sobre prevalência de DPP. Evidências epidemiológicas de nível 1 são levantamentos locais e atuais, com amostras aleatorizadas ou censitárias (OCEBM Levels of Evidence Working Group, 2011). As palavras “locais” e “atuais” mostram a importância de se situar a pesquisa quanto ao espaço e tempo em que os dados foram coletados. Além disso, o modelo de metarregressão indicou que as cidades diferiram quanto aos resultados de prevalência de DPP. Assim, é importante incluir não apenas o registro do país em que a coleta foi feita, mas também da cidade. Essa informação deixou de ser mencionada em diversos estudos, principalmente naqueles que foram realizados nos Estados Unidos da América.

Pode-se argumentar que pesquisas nos moldes sugeridos seriam demasiadamente custosas em relação ao tempo e aos recursos financeiros e humanos necessários. No entanto, pelo menos 337 estudos sobre prevalência de DPP já foram publicados, envolvendo 296.139 participantes, e que não foi possível chegar a uma conclusão clara sobre a prevalência de DPP, devido à baixa qualidade metodológica. Tendo em vista esses parâmetros, o custo-

benefício de se fazer pesquisas de boa qualidade parece razoável. Investigações sobre a prevalência de DPP com metodologia adequada e relatadas de forma clara e completa precisam ser feitas não apenas nos grandes centros urbanos de países desenvolvidos, mas principalmente, nos países em desenvolvimento, onde há menos estudos publicados sobre a prevalência de DPP. Também é importante priorizar cidades e países onde não há, até o momento, pesquisas sobre esse tema.

Para facilitar a realização de revisões sistemáticas, seria interessante que os autores observassem algumas recomendações quanto ao relato do trabalho científico. O revisor se depara com uma quantidade grande de trabalhos para avaliar, e o primeiro elemento com o qual entra em contato é o título. É importante, portanto, que o título seja claro e explicativo, remetendo o leitor para a pergunta de pesquisa e o desfecho investigado.

Recomenda-se, ainda, o uso de resumos estruturados, ou seja: aqueles que nomeiam cada seção do texto (introdução, objetivos, método, resultados, discussão ou conclusão), facilitando ao leitor encontrar os elementos de interesse (Sabadini, Sampaio, & Koller, 2009). Não foi por acaso que se optou pelo uso de resumo estruturado na presente Tese. Ao lidar com o grande volume de material científico gerado pelas buscas sistemáticas, esse detalhe pode poupar trabalho e tempo do revisor.

No texto completo, algumas informações precisam estar presentes, como por exemplo: o nome do instrumento usado e suas características psicométricas, indicando a confiabilidade e a validade, preferencialmente em população comparável à de estudo; o período após o parto em que foi feita a avaliação de DPP; a estimativa intervalar de prevalência, isto é, o intervalo de confiança; e registros do local e data da coleta de dados.

Muitas vezes, uma mesma pesquisa científica pode gerar diversas publicações, com recortes diferentes dos dados. Na presente metanálise, foram encontrados 80 trabalhos que duplicaram ou apresentaram sobreposição parcial de amostra com outro estudo incluído. A busca por duplicidades foi feita manualmente, comparando-se dados como autoria, local e data da coleta de dados, tamanho amostral e prevalência encontrada. Nem sempre foi simples

perceber as duplicações, visto que: 1. o mesmo autor pode publicar estudos com amostras diferentes; 2. a mesma amostra pode ser usada em estudos de diferentes autores, no caso de um projeto mais amplo; 3. o local e a data da coleta de dados nem sempre foram disponibilizados nos relatos; 4. o tamanho da amostra e a prevalência poderiam apresentar diferenças entre os estudos, devido a dados faltantes (*missing values*) e aos diferentes recortes possíveis de uma mesma amostra. Considera-se que não há problema na prática de publicar diferentes artigos usando os mesmos dados, sobretudo quando muitas informações estiverem disponíveis, possibilitando diversas conclusões sobre as relações entre variáveis e desfechos. Mas seria importante declarar explicitamente que os mesmos dados foram usados em outros trabalhos, evitando que os mesmos participantes sejam inseridos diversas vezes nas metanálises, gerando medidas repetidas que podem enviesar os resultados.

7.10 Recomendações para revisões sistemáticas de estudos de prevalência

Ao realizar a presente revisão sistemática, foram superados alguns desafios metodológicos, observando-se questões que podem ser compartilhadas, buscando colaborar com futuras revisões de estudos observacionais de prevalência. A primeira delas foi a ausência de diretrizes estabelecidas a partir de discussões entre a comunidade científica para a metodologia de revisões sistemáticas de estudos observacionais. A maior parte dos esforços dos pesquisadores na área de saúde tem sido empreendida para a sistematização dos resultados de trabalhos delineados para avaliar eficácia ou efetividade de intervenções (Gülmezoglu et al., 2004). Apenas uma recomendação de hierarquia de evidências científicas na área de prevalência foi encontrada (OCEBM Levels of Evidence Working Group, 2011), e foram observados poucos trabalhos que buscaram elaborar métodos sistemáticos para a avaliação de qualidade de estudos de prevalência ou incidência (Shamliyan et al., 2010; Giannakopoulos et al., 2012). Dessa ausência de

diretrizes decorreu a ampla variação de metodologias adotadas em revisões sistemáticas sobre prevalência, não havendo consenso a respeito de elaboração de estratégias de busca, critérios de inclusão, adoção de escalas de qualidade e análise estatística, colocando em dúvida a qualidade da evidência gerada pelas revisões. Na ausência de diretrizes consolidadas, recomenda-se aos pesquisadores interessados em fazer revisão sistemática de estudos de prevalência que busquem outros trabalhos realizados em sua área de interesse, avaliando-os de forma crítica. Paralelamente, considera-se que a criação de diretrizes para revisões sistemáticas de estudos observacionais de prevalência é uma recomendação para futuros trabalhos acadêmicos.

O número de registros localizados (1.894) e trabalhos incluídos (337) na presente revisão foi muito superior ao geralmente descrito em revisões da Cochrane sobre eficácia de intervenções na área de saúde mental (exemplos: 28 estudos incluídos sobre psicoterapia no tratamento de personalidade borderline em Stoffers et al., 2012; 11 estudos incluídos sobre intervenções psicossociais para tratamento de disfunção erétil em Melnik, Soares, & Nasello, 2007). De fato, pesquisas observacionais são mais frequentes que ensaios clínicos controlados, e trabalhar com número grande de estudos exige maior investimento de tempo e de recursos financeiros, o que deve ser observado na fase de planejamento das revisões sistemáticas.

Recomenda-se o uso de recursos tecnológicos para facilitar o armazenamento e as análises dos dados, como os *softwares* para gerenciar referências bibliográficas, tabular dados e fazer análises estatísticas. Já existem *softwares* livres especializados em revisões sistemáticas de estudos sobre eficácia ou efetividade de intervenções, como o *RevMan*, criado e distribuído pela Colaboração Cochrane. No entanto, até o momento, não há *softwares* conhecidos para auxiliar na elaboração de revisões sistemáticas de estudos observacionais, o que configura mais uma sugestão para futuros trabalhos acadêmicos.

Devido à tendência de publicação de artigos relatando estudos observacionais em periódicos científicos nacionais (Victora & Moreira, 2006), sugere-se que revisões sistemáticas de trabalhos epidemiológicos procurem incluir o maior número de bases de dados eletrônicas possíveis, inclusive as

que indexam artigos de periódicos nacionais ou regionais, como as bases de dados latino-americanas, africanas e asiáticas. Se possível, é interessante contar com recursos suficientes para realizar traduções, evitando a limitação idiomática.

A estratégia de busca deve ser elaborada com auxílio de um Bacharel em Biblioteconomia, preferencialmente especializado em revisões sistemáticas. O apoio desse profissional também será fundamental para localizar os textos completos dos artigos, que podem não estar disponíveis em acesso aberto.

A avaliação da qualidade metodológica dos estudos incluídos é fundamental em revisões sistemáticas, visto que a qualidade das estimativas e seu potencial de uso para fundamentar as ações em saúde dependem da qualidade dos estudos primários. Os resultados da presente revisão indicaram que foi mais interessante verificar a qualidade através de variáveis metodológicas dos estudos primários, observando-se seus efeitos sobre o desfecho, do pontuação obtida em instrumento de avaliação da qualidade. Caso seja usada uma escala, recomenda-se testar o instrumento de Giannakopoulos et al. (2012), que foi publicado recentemente e é mais detalhado que a Escala de Loney et al. (1998).

Em revisões sistemáticas, é importante que seja verificada a confiabilidade dos critérios de seleção e de avaliação da qualidade metodológica dos estudos primários incluídos. Para isso, é necessária a participação de, no mínimo, dois revisores, que devem trabalhar de forma independente. Essa prática não tem sido observada em revisões de estudos observacionais. O ideal é que os revisores avaliem todos os estudos para estabelecer o nível de confiança, e que os casos de divergência sejam resolvidos por consenso ou pela avaliação de um terceiro revisor. No entanto, em revisões com grande número de trabalhos incluídos, como tendem a ser as de estudos observacionais, a confiabilidade pode ser estabelecida através de amostra aleatória dos estudos incluídos, como foi feito no presente trabalho.

8 Conclusões

Concluiu-se que os resultados relativos à prevalência de DPP, relatados nos estudos primários, são influenciados pela metodologia adotada. Foram encontradas prevalências maiores de DPP nas pesquisas transversais, em comparação com as longitudinais; nas pesquisas que usaram instrumentos de autoavaliação, em comparação com as que usaram entrevista clínica; nas pesquisas com menor tamanho amostral; e nas pesquisas realizadas em países com menores IDHs. Assim, também foi possível concluir que a DPP foi mais prevalente em países em desenvolvimento do que nos países desenvolvidos, mesmo levando em conta características metodológicas dos estudos incluídos. No entanto, as conclusões do presente trabalho são preliminares e não podem ser consideradas definitivas, devido à qualidade fraca dos estudos primários, que pode ter gerado vieses nos resultados.

Referências

Affonso, D., De, A., Horowitz, J., & Mayberry, L. (2000). An international study exploring levels of postpartum depressive symptomatology. *Journal of Psychosomatic Research, 49*, 207-216.

Affonso, D., Lovett, S., Paul, S., & Sheptak, S. (1990). A standardized interview that differentiates pregnancy and postpartum symptoms from perinatal clinical depression. *BIRTH, 17*, 121-130.

Ahlund, S., Clarke, P., Hill, J., & Thalange, N. K. (2009). Post-traumatic stress symptoms in mothers of very low birth weight infants 2-3 years post-partum. *Archives of Women's Mental Health, 12*, 261-264.

Alhabib, S., Nur, U., & Jones, R. (2010). Domestic violence against women: Systematic review of prevalence studies. *Journal of Family Violence, 25*, 369–382.

Ali, N., Ali, B., & Azam, I. (2009). Post partum anxiety and depression in peri-urban communities of Karachi, Pakistan: A quasi-experimental study. *BMC Public Health, 9*, 384.

Aliane, P., Mamede, M., & Furtado, E. (2011). Revisão sistemática sobre fatores de risco associados à depressão pós-parto. *Psicologia em Pesquisa, 5*, 146-155.

American Psychiatric Association. (1994). *DSM-IV. Manual de Diagnóstico e Estatística dos Distúrbios Mentais* (4a ed.). Lisboa: Climepsi.

American Psychological Association. (2005). *Evidence-Based Practice in Psychology.* Acesso em 12 de 01 de 2011, disponível em American Psychological Association: http://www.apa.org/practice/resources/evidence/evidence-based-statement.pdf

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-Based Practice in Psychology. *American Psychologist, 61*, 271-285.

American Psychological Association Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically validated psychological treatments. *The Clinical Psychologist, 48*, 3-23.

American Psychological Association Task Force on Psychological Intervention Guidelines. (1995). *Template for developing guidelines: Interventions for mental disorders and psychological Association.* Washington DC: American Psychological Association.

Amr, M. A. -M., & Balaha, M. H. H. (2010). Minor psychiatric morbidity in young Saudi mothers using Mini International Neuropsychiatric Interview (MINI). *Journal of the College of Physicians and Surgeons Pakistan, 20*(10), 680-684.

Anzures-Cabrera, J., & Higgins, J. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods, 1*, 66-80.

Asunción Lara, M., Navarro, C., Navarrete, L., & Le, H.-N. (2010). Retention rates and potential predictors in a longitudinal randomized control trial to prevent postpartum depression. *Salud Mental, 33*, 429-436.

Austin, M.-P. (2004). Antenatal screening and early intervention for ‘‘perinatal’’ distress, depression and anxiety: where to from here? *Archives of Women's Mental Health, 7*, 1-6.

Ballestrem, C. -L. v., Strauß, M., & Kächele, H. (2005). Contribution to the epidemiology of postnatal depression in Germany – Implications for the utilization of treatment. *Archives of Women's Mental Health, 8*, 29-35.

Bantia, S., Mauria, M., Oppoa, A., Borria, C., Rambellia, C., Ramacciottia, D., ..., Cassanoa, G. B. (2011). Prevalence, incidence, recurrence, and new onset of depression: Results from the Perinatal Depression - Research & Screening Unit study. *Comprehensive Psychiatry, 52*, 343-351.

Bastos, J., & Duquia, R. (2007). Um dos delineamentos mais empregados em epidemiologia: Estudo transversal. *Scientia Medica, 17*, 229-232.

Beck, C. (1996). A meta-analysis of predictors of postpartum depression. *Nursing Research, 45*(5), 297-303.

Beck, C. (1998). The effects of postpartum depression on child development: A meta-analysis. *Archives of Psychiatric Nursing, 12*, 12-20.

Beck, C. (1999). Maternal depression and child behavior problems: A meta-analysis. *Journal of Advanced Nursing, 29*, 623-629.

Beck, C. (2001). Predictors of postpartum depression: An update. *Nursing Research, 50*, 275-285.

Beck, C., Gable, R., Sakala, C., & Declercq, E. (2011). Postpartum depressive symptomatology: Results from a two-stage US national survey. *Journal of Midwifery &Women’s Health, 56*, 427-435.

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.

Beydoun, H., Beydoun, M., Kaufman, J., Lo, B., & Zonderman, A. (2012). Intimate partner violence against adult women and its association with major depressive disorder, depressive symptoms and postpartum depression: A systematic review and meta-analysis. *Social Science & Medicine, 75*, 959-975.

Biruel, E., & Pinto, R. (2011). Bibliotecário: Um profissional a serviço da pesquisa. *Anais do XXIV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação.* Maceió, Alagoas.

Bledsoe, S., & Grote, N. (2006). Treating depression during pregnancy and the postpartum: A preliminary meta-analysis. *Research on Social Work Practice, 16*, 109-120.

Bowen, A., Stewart, N., Baetz, M., & Muhajarine, N. (2009). Antenatal depression in socially high-risk women in Canada. *Journal of Epidemiology and Community Health, 63*, 414-416.

Boyle, M. (1998). Guidelines for evaluating prevalence studies. *Evidence Based Mental Health, 1*, 37-39.

Breakwell, G. M., Fife-Schaw, C., Hammond, S., & Smith, J. (2010). *Métodos de pesquisa em Psicologia.* Porto Alegre: Artmed.

Brent, L., Koban, T., & Ramirez, S. (2002). Abnormal, abusive, and stress-related behaviors in baboon mothers. *Biological Psychiatry, 52*, 1047-1056.

Bromet, E., Andrade, L., Hwang, I., Sampson, N., Alonso, J., de Girolamo, G., …, Kessler, R. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine, 9*, 90.

Cantilino, A. (2010). Por um consenso conceitual e metodológico nas pesquisas em depressão pós-parto. *Jornal Brasileiro de Psiquiatria, 59*, 257.

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.

Carter, F., Frampton, C., & Mulder, R. (2006). Cesarean section and postpartum depression: A review of the evidence examining the link. *Psychosomatic Medicine, 68*, 321-330.

Castro, A., Saconato, H., Guidugli, F., & Clark, O. (n.d.). *Curso de revisão sistemática e metanálise [online]*. Acesso em 05 de 06 de 2010, disponível em LED-DIS/UNIFESP: http://www.virtual.epm.br/cursos/metanalise/

Centers for Disease Control and Prevention. (2008). Prevalence of self-reported postpartum depressive symptoms — 17 states, 2004–2005. *Morbidity and Mortality Weekly Report, 57*, 361-392.

Chambless, D., & Ollendick, T. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685-716.

Coates, A., Schaefer, C., & Alexander, J. (2004). Detection of postpartum depression and anxiety in a large health plan. *Journal of Behavioral Health Services & Research, 31*, 117-133.

Cohen, J. S. (2010). *Mindfulness and self-compassion in the transition to motherhood: A prospective study of postnatal mood and attachment.* Tese de Doutorado, Columbia University, New York.

Concato, J. (2004). Observational versus experimental studies: What's the evidence for a hierarchy? *The Journal of the American Society for Experimental NeuroTherapeutics, 1*, 341-347.

Cook, D., Mulrow, C., & Haynes, R. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine, 126*, 376-380.

Cooper, P., Tomlinson, M., Swartz, L., Woolgar, M., Murray, L., & Molteno, C. (1999). Post-partum depression and the mother-infant relationship in a South African peri-urban settlement. *British Journal of Psychiatry, 175*, 554-558.

Costa, D., Larouche, J., Dritsa, M., & Brender, W. (2000). Psychosocial correlates of prepartum and postpartum depressed mood. *Journal of Affective Disorders, 59*, 31-40.

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

Cox, J., Holden, J., & Sagovsky, R. (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry, 150*, 782–786.

Cozby, P. C. (2009). *Métodos de pesquisa em ciências do comportamento* (1a ed.). (P. I. Gomide, E. Otta, & J. O. Siqueira, Trads.) São Paulo: Atlas.

Cruz, E., Simões, G., & 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*, 181-188.

Csada, R., James, P., & Espie, R. (1996). The "file drawer problem" of non-significant results: Does it apply to biological research? *Oikos, 76*, 591-593.

Cuijpers, P., Brännmark, J., & van Straten, A. (2008). Psychological treatment of postpartum depression: A meta-analysis. *Journal of Clinical Psychology, 64*, 103-118.

Cummings, N. (1977). About this special issue. *American Psychologist, 32*, 695-696.

Cunha, M., & Cavalcanti, C. (2008). *Dicionário de Biblioteconomia e Arquivologia.* Brasília: Briquet de Lemos.

Daley, A., Jolly, K., & MacArthur, C. (2009). The effectiveness of exercise in the management of post-natal depression: Systematic review and meta-analysis. *Family Practice, 26*, 154-162.

Da-Silva, V., Moraes-Santos, A., Carvalho, M., Martins, M., & Teixeira, N. (1998). Prenatal and postnatal depression among low income Brazilian women. *Brazilian Journal of Medical and Biological Research, 31*, 799-804.

Delgado-Rodríguez, M. (2001). Glossary on meta-analysis. *Journal of Epidemiology and Community Health, 55*, 534–536.

Dennis, C., & Allen, K. (2008). Interventions (other than pharmacological, psychosocial or psychological) for treating antenatal depression. *Cochrane Database of Systematic Reviews*, CD006795.

Dennis, C., & Creedy, D. (2004). Psychosocial and psychological interventions for preventing postpartum depression. *Cochrane Database of Systematic Reviews, 4*, CD001134.

Dennis, C., & Hodnett, E. (2007). Psychosocial and psychological interventions for treating postpartum depression. *Cochrane Database of Systematic Reviews*, CD006309.

Diagnostic Test Accuracy Working Group. (2011). Acesso em 12/10/2012, disponível em http://srdta.cochrane.org/cochrane-diagnostic-test-accuracy-reviews

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634.

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*, 115-118.

Eugenio, B. (2000). On the usage of Kappa to evaluate agreement on coding tasks. *Second International Conference on Language Resources and Evaluation.* Athens, Greece.

Evans, J., Heron, J., Francomb, H., Oke, S., & Golding, J. (2001). Cohort study of depressed mood during pregnancy and after childbirth. *British Medical Journal, 323*, 257-260.

Faisal-Cury, A., Tedesco, J., Kahhale, S., Menezes, P., & Zugaib, M. (2004). Postpartum depression: In relation to life events and patterns of coping. *Archieves of Women's Mental Health, 7*, 123-131.

Figueira, P., Corrêa, H., Malloy-Diniz, L., & Romano-Silva, M. (2009). Escala de Depressão Pós-natal de Edimburgo para triagem no sistema público de saúde. *Revista de Saúde Pública, 43*, 79-84.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*, 451-474.

Fuggle, P., Glover, L., Khan, F., & Haydon, K. (2002). Screening for postnatal depression in Bengali women: Preliminary observations from using a translated version of the Edinburgh Postnatal Depression Scale (EPDS). *Journal of Reproductive and Infant Psychology, 20*, 72-82.

Gavin, N., Gaynes, N., Kathleen, N., Meltzer-Brody, S., Gartlehner, G., & Swinson, T. (2005). Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics and Gynecology, 5*, 1071-1083.

Georgiopoulos, A., Bryan, T., Yawn, B., Houston, M., Rummans, T., & Therneau, T. (1999). Population-based screening for postpartum depression. *Obstetrics & Gynecology, 93*, 653-657.

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.

Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., & Gray, R. (2009). A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. *Acta Psychiatrica Scandinavica, 119*, 150-364.

Gilbert, B., Shulman, H., Fischer, L., & Rogers, M. (1999). The Pregnancy Risk Assessment Monitoring System (PRAMS): Methods and 1996 response rates from 11 states. *Maternal and Child Health Journal, 3*, 199-209.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Gorman, L., O'Hara, M., Figueiredo, B., Hayes, S., Jacquemain, F., Kammerer, M., …, Sutter-Dallay, A.-L. (2004). Adaptation of the Structured Clinical Interview for DSM-IV Disorders for assessing depression in women during pregnancy and post-partum across countries and cultures. *British Journal of Psychiatry, 184*, 17-23.

Grace, S., Evindar, A., & Stewart, D. (2003). The effect of postpartum depression on child cognitive development: A review and critical analysis of the literature. *Archives of Women's Mental Health, 6*, 263-274.

Greenberg, P., Stiglin, L., Finkelstein, S., & Berndt, E. (1993). The economic burden of depression in 1990. *Journal of Clinical Psychiatry, 54*, 405-418.

Greenhalgh, T. (2008). *Como ler artigos científicos: Fundamentos da medicina baseada em evidências.* (A. P. Fajardo, Trad.) Porto Alegre: Artmed.

Grussu, P., & Quatraro, R. (2009). Prevalence and risk factors for a high level of postnatal depression symptomatology in Italian women: A sample drawn from ante-natal classes. *European Psychiatry, 24*, 327-333.

Guajardo, V., Rojas C., M., Fritsch, R., Jadresic M., E., Díaz, V., & Araya, R. (2007). Efectividad de un programa de tratamiento de la depresión posparto en la atención primaria a tres y seis meses de seguimiento. *Revista de Psiquiatría Clínica (Santiago de Chile), 44*, 70-88.

Gülmezoglu, A. M., Say, L., Betrán, A. P., Villar, J., & Piaggio, G. (2004). WHO systematic review of maternal mortality and morbidity: Methodological issues and challenges. *BMC Medical Research Methodology*.

Halbreich, U., & Karkun, S. (2006). Cross-cultural and social diversity of prevalence of postpartum. *Journal of Affective Disorders, 91*, 97-111.

Hasbún H., J., Risco N., L., Jadresic M., E., Galleguillo U., T., Gonzáles A., M., & Garay S., J. (1999). Depresión postparto: Prevalencia y factores de riesgo. *Revista Chilena de Obstetricia y Ginecología, 64*, 466-470.

Henshaw, C. (2003). Mood disturbance in the early puerperium: A review. *Archives of Women's Mental Health, 6*, s33-s42.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [atualizado em Março de 2011]. The Cochrane Collaboration, 2011. Disponível em: www.cochrane-handbook.org.

Higgins, J., & Thompson, S. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.

Ho, P., Peterson, P., & Masoudi, F. (2008). Evaluating the evidence: Is there a rigid hierarchy? *Circulation, 118*, 1675-1684.

Hofbrand, S., Howard, L., & Crawley, H. (2001). Antidepressant treatment for post-natal depression. *Cochrane Database of Systematic Reviews*, CD002018.

Howard, L., Hoffbrand, S., Henshaw, C., Boath, L., & Bradley, E. (2009). Antidepressant prevention of postnatal depression. *Cochrane Database of Systematic Reviews*, CD004363.

Husain, N., Bevc, I., Husain, M., Chaudhry, I., Atif, N., & Rahman, A. (2006). Prevalence and social correlates of postnatal depression in a low income country. *Archives of Women's Mental Health, 9*, 197-202.

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.

Josefsson, A., Berg, G., Nordin, C., & Sydsjö, G. (2001). Prevalence of depressive sumptoms in late pregnancy and postpartum. *Acta Obstetrica et Gynecologica Scandinavica, 80*, 251-255.

Kersten-Alvarez, L., Hosman, C., Riksen-Walraven, J., van Doesum, K., & Hoefnagels, C. (2011). Which preventive interventions effectively enhance depressed mothers' sensitivity? A meta-analysis. *Infant Mental Health Journal, 32*, 362-376.

Kok, L., Chan, P., & Ratnam, S. (1994). Postnatal Depression in Singapore women. *Singapore Medical Journal, 35*, 33-35.

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.

Lenth, R. (2001). Some practical guidelines for effective sample size determination. *American Statistical Association, 55*, 187-193.

Lewis, R. (2011). Epidemiology of sexual dysfunction in Asia compared to the rest of the world. *Asian Journal of Andrology, 13*, 152-158.

Lewis, S., & Clarke, M. (2001). Forest plot: Trying to see the wood and the trees. *British Medical Journal, 322*, 1479-1480.

Lwanga, S.K., & Lemeshow, S. (1991). *Sample size determination in health studies: a practical manual*. Geneva, World Health Organization.

Lobato, G., Moraes, C., & Reichenheim, M. (2012). Magnitude da depressão pós-parto no Brasil: Uma revisão sistemática. *Revista Brasileira de Saúde Materno Infantil, 11*, 369-179.

Loney, P., Chambers, L., Bennett, K., Roberts, J., & Stratford, P. (1998). Critical appraisal of the health research literature: Prevalence or incidence of a health problem. *Chronic Diseases in Canada, 19*, 170-176.

Lumley, J., Austin, M.-P., & Mitchell, C. (2004). Intervening to reduce depression after birth: A systematic review of the randomized trials. *International Journal of Technology Assessment in Health Care, 20*, 128-144.

Lunsford, T., & Lunsford, B. (1995). Research Forum: The research sample, Part 1: Sampling. *Journal of Prosthetics and Orthotics, 7*, 105-112.

Macaskill, P., Gatsonis, C., Deeks, J. J., Harbord, R. M., Takwoingi, Y. Chapter 10: Analysing and Presenting Results. Em: Deeks, J. J., Bossuyt, P. M., Gatsonis, C. (editores). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. Disponível em: http://srdta.cochrane.org/.

Maija Reblin, M., & Uchino, B. (2008). Social and emocional support and its implication for health. *Current Opinion in Psychiatry, 21*, 201-205.

Mann, R., Gilbody, S., & Adamson, J. (2010). Prevalence and incidence of postnatal depression: What can systematic reviews tell us? *Archives of Women's Mental Health, 13*, 295–305.

Martins, C., & Gaffan, E. (2000). Effects of early maternal depression on patterns of infant-mother attachment: A meta-analytic investigation. *Journal of Child Psychology and Psychiatry, 41*, 737-746.

Martins, C., Siqueira, K., Tyrrell, M., Barbosa, M., Carvalho, S., & Santos, L. (2008). Dinâmica familiar em situação de nascimento e puerpério. *Revista Eletrônica de Enfermagem, 10*, 1015-1025.

Mathey, S. (2010). Are we overpathologising motherhood? *Journal of Affective Disorders, 120*, 263–266.

McGraw, K., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

Melnik, T., & Atallah, A. (2011). *Psicologia baseada em evidências: Provas científicas da efetividade da psicoterapia.* São Paulo: Santos.

Melnik, T., Soares, B., & Nasello, A. (2007). Psychosocial interventions for erectile dysfunction. *Cochrane Database of Systematic Reviews*, Art. No.: CD004825.

Melo Junior, E. (2011). Prevalência de depressão perinatal e fatores associados. Tese de Doutorado apresentada à Universidade Estadual de Campinas. Campinas, SP, Brasil.

Milgrom, J., Ericksen, J., Negri, L., & Gemmill, A. (2005). Screening for postnatal depression in routine primary care: properties of the Edinburgh Postnatal Depression Scale in an Australian sample. *Australian and New Zealand Journal of Psychiatry, 39*, 833–839.

Moraes, I., Pinheiro, R., Silva, R., Horta, B., Sousa, P., & Faria, A. (2006). Prevalência da depressão pós-parto e fatores associados. *Revista de Saúde Pública, 40*, 65-70.

Murray, C. J. L., & Lopez, A. D. (1996). Evidence-based health policy – Lessons from the Global Burden of Disease Study. *Science, 274*, 740-743.

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*, 1498–1504.

Murray, C. J. L., & Cooper, P. (1997). *Postpartum depression and child development.* New York: The Guilford Press.

OCEBM Levels of Evidence Working Group. (2011). *The Oxford 2011 Levels of Evidence.* Acesso em 25 de 09 de 2012, disponível em Oxford Centre for Evidence-Based Medicine: http://www.cebm.net/index.aspx?o=5653

O'Hara, M., & Swain, A. (1996). Rates and risk of postpartum depression - A meta-analysis. *International Review of Psychiatry, 8*, 37-54.

O'Hara, M.W., Zekoski, E.M., Phillips, L.H. & Wright, E.J. (1990). Controlled prospective study of postpartum mood disorders: Comparison of childbearing and nonchildbearing women. *Journal of Abnormal Psychology, 99,* 3-15.

Paraje, G., Sadana, R., & Karam, G. (2005). Increasing international gaps in health-related publications. *Science, 308*, 959-960.

Pearlstein, T., Howard, M., Salisbury, A., & Zlotnick, C. (2009). Postpartum depression. *American Journal of Obstetrics & Gynecology*, 357-364.

Pereira, J., Paes, A., & Okano, V. (2000). Espaço aberto: Questões comuns sobre epidemiologia, estatística e informática. *Revista do IDPC, 7*, 12-17.

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*, 422-432.

Righetti-Veltema, M., Conne-Perréard, E., Bousquet, A., & Manzano, J. (2001). Postpartum depression and mother–infant relationship at 3 months old. *Journal of Affective Disorders, 70*: 291-306.

Righetti-Veltema, M., Bousquet, A., & Manzano, J. (2003). Impact of postpartum depressive symptoms on mother and her 18-month-old infant. *European Child & Adolescent Psychiatry, 12*:75-83.

Robertson, E., Grace, S., Wallington, T., & Stewart, D. (2004). Antenatal risk factors for postpartum depression: A synthesis of recent literature. *General Hospital Psychiatry, 26*, 289–295.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Ruschi, G., Chambô-Filho, A., Lima, V., Yazaki-Sun, S., Zandonade, E., & Mattar, R. (2009). Alteração tireoidiana: Um fator de risco associado à depressão pós-parto? *Revista Brasileira de Saúde Materno-Infantil, 9*, 207-213.

Ruschi, G., Sun, S., Mattar, R., Chambô-Filho, A., Zandonade, E., & Lima, V. (2007). Aspectos epidemiológicos da depressão pós-parto em amostra brasileira. *Revista de Psiquiatria do Rio Grande do Sul, 29*, 274-280.

Sabadini, A., Sampaio, M., & Koller, S. (2009). Preparando um artigo científico. In: A. Sabadini, M. Sampaio, & S. Koller, *Publicar em psicologia: Um enfoque para a revista científica* (p. 117-154). São Paulo: Associação Brasileira de Editores Científicos de Psicologia / Instituto de Psicologia da Universidade de São Paulo.

Sanchez-Meca, J., & Botella, J. (2010). Revisiones sistemáticas y meta-análisis: Herramientas para la práctica profesional. *Papeles del Psicólogo, 31*(1):7-17.

Sampaio, M., & Serradas, A. (2009). O Movimento de Acesso Aberto, os repositórios e as revistas científicas. In A. Sabadini, M. Sampaio & S. Koller, *Publicar em psicologia: Um enfoque para a revista científica* (p. 77-88). São Paulo: Associação Brasileira de Editores Científicos de Psicologia.

Sánchez-Meca, J., Boruch, R., Petrosino, A., & Rosa, A. (2002). La Colaboración Campbell y la práctica basada en la evidencia. *Papeles del Psicólogo, 83*, 44-48.

Santos, I., Matijasevich, A., Tavares, B., Barros, A., Botelho, I., Lapolli, C., …, Barros, F. (2007). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of mothers from the 2004 Pelotas Birth Cohort Study. *Cadernos de Saúde Pública, 23*, 2577-2588.

Santos, M., Martins, F., & Pasquali, L. (1999). Escalas de auto-avaliação de depressão pós-parto: Estudo no Brasil. *Revista de Psiquiatria Clínica, 26*, 32-40.

Sawyer, A., Ayers, S., & Smith, H. (2010). Pre- and postnatal psychological wellbeing in Africa: A systematic review. *Journal of Affective Disorders, 123*, 17-29.

Schmidt, E., Piccoloto, N., & Müller, M. (2005). Depressão pós-parto: fatores de risco e repercussões no desenvolvimento infantil. *Psico-USF, 10*, 61-68.

Schram, M., Tedja, A., Spijker, R., Bos, J., Williams, H., & Spuls, P. (2010). Is there a rural/urban gradient in the prevalence of eczema? A systematic review. *British Journal of Dermatology, 162*, 964–973.

Schwengber, D.D.S., & Piccinini, C.A. (2003). O impacto da depressão pós-parto para a interação mãe-bebê. *Estudos de psicologia, 8*(3):403-411.

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*, e17591.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Silva, G. A. (2008). Estudo longitudinal sobre prevalência e fatores de risco para depressão pós-parto em mães de baixa renda. Dissertação de Mestrado, Instituto de Psicologia da Universidade de São Paulo, São Paulo, SP.

Silva, G. A., Sampaio, M., & Otta, E. (2011). A efetividade das intervenções psicoterápicas no tratamento da depressão durante a gestação e no puerpério. In T. Melnik & A. Attalah, *A efetividade das intervenções psicoterápicas no tratamento de transtornos psiquiátricos: Classificação das evidências e do grau de recomendação.* São Paulo: Gen.

Skari, H., Skreden, M., Malt, U., Dalholt, M., Ostensen, A., 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 Gynecology, 109*, 1154–1163.

Smith, M., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

Sobocki, P., Jönsson, B., Angst, J., & Rehnberg, C. (2006). Cost of depression in Europe. *Journal of Mental Health Policy and Economics, 9*, 87-98.

Sockol, L., Epperson, C., & Barber, J. (2011). A meta-analysis of treatments for perinatal depression. *Clinical Psychology Review, 31*, 839–849.

Stewart, R., Kauye, F., Umar, E., Vokhiwa, M., Bunn, J., Fitzgerald, M., …, Creed, F. (2009). Validation of a Chichewa version of the Self-Reporting Questionnaire (SRQ) as a brief screening measure for maternal depressive disorder in Malawi, Africa. *Journal of Affective Disorders, 112*, 126-134.

Stoffers, J., Völlm, B., Rücker, G., Timmer, A., Huband, N., & Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, Art. No.: CD005652.

Tannous, L., Gigante, L., Fuchs, S., & Busnello, E. (2008). Postnatal depression in Southern Brazil: prevalence and its demographic and socioeconomic determinants. *BMC Psychiatry, 8*, 1.

Terris, M. (1990). Public health policy for the 1900s. *Annual Review of Public Health, 11*, 39-51.

The Cochrane Collaboration. (2005). *Glossary of Terms in The Cochrane Collaboration, Version 4.2.5.* Acesso em 28 de 11 de 2012, disponível em Centro Cochrane do Brasil: http://www.cochrane.org/sites/default/files/uploads/glossary.pdf

Thompson, S., & Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559-1573.

Thompson, S., & Sharp, S. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18*, 2693-2708.

United Nations Development Programme. (02 de 11 de 2011). Acesso em 30 de 11 de 2012, disponível em Human Development Reports: http://hdr.undp.org/en/media/HDR_2011_EN_Table1.pdf

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*, 102-112.

Victora, C., & Moreira, C. (2006). Publicações científicas e as relações Norte-Sul: Racismo editorial? *Revista de Saúde Pública, 40*, 36-42.

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. *British Journal of Obstetrics and Gynecology, 117*, 540-550.

Vik, T., Grote, V., Escribano, J., Socha, J., Verduci, E., Fritsch, M., …, Koletzko, B. (2009). Infantile colic, prolonged crying and maternal postnatal depression. *Acta Pædiatrica, 98*, 1344-1348.

Villegas, L., McKay, K., Dennis, C. L., & Ross, L. E. (2011). Postpartum depression among rural women from developed and developing countries: A systematic review. *Journal of Rural Health, 27*, 278-288.

Walker, S., Wachs, T., Gardner, J., Lozoff, B., Wasserman, G., Pollitt, E., …, International Child Development Steering Group. (2007). Child development: Risk factors for adverse outcomes in developing countries. *Lancet, 369*, 145-157.

Waraich, P., Goldner, E., Somers, J., & Hsu, L. (2004). Prevalence and incidence studies of mood disorders: A systematic review of the literature. *Canadian Journal of Psychiatry, 49*, 124-138.

Wong, J., & Fisher, J. (2009). The role of traditional confinement practices in determining postpartum depression inwomen in Chinese cultures: A systematic review of the English language evidence. *Journal of Affective Disorders, 116*, 161–169.

Wu, Q., Chen, H.-L., & Xu, X.-J. (2012). Violence as a risk factor for postpartum depression in mothers: A meta-analysis. *Archives of Women's Mental Health, 15*, 107-114.

Zaconeta, A., Motta, L., & França, P. (2004). Depresión postparto: Prevalencia de test de rastreo positivo en puérperas del Hospital Universitário de Brasília, Brasil. *Revista Chilena de Obstetricia y Ginecologia, 69*, 209-213.

Zagoneli, I., Martins, M., Pereira, K., & Athayde, J. (2003). O cuidado humano diante da transição ao papel materno: Vivências no puerpério. *Revista Eletrônica de Enfermagem, 5*, 24-32.

Zubaran, C., Foresti, K., Schumacher, M., Amoretti, A., Müller, L., Thorell, M., … Madi, J. (2009). Validation of a screening instrument for postpartum depression in Southern Brazil. *Journal of Psychosomatic Obstetrics & Gynecology, 30*, pp. 244-254.

Apêndice 1 – Protocolo de avaliação de estudos por título e resumo.

TÍTULO

Excluir estudo apenas quando estiver claro que não atende aos critérios.

Quando um critério claramente não for atendido, colocar uma letra “F” (de “falso”). Caso contrário, deixar a célula vazia.

Ex.:

1. Não há evidências de que o estudo não atende algum critério.

A	B	C	D	E	F	G	Status do título
							VERDADEIRO

2. Há evidências de que o estudo não atende o critério A.

A	B	C	D	E	F	G	Status do título
F							FALSO

3. Há evidências de que o estudo não atende os critérios B, F e G.

A	B	C	D	E	F	G	Status do título
	F				F	F	FALSO

RESUMO

Avaliar somente os registros que tiverem “Verdadeiro” no status do título (conforme avaliação de títulos).

Avaliar critério a critério. Nenhuma célula poderá ficar vazia (a não ser que o trabalho não tenha resumo. Nesse caso, manter todas as células vazias e direcioná-lo à avaliação do texto completo).

Usar a seguinte legenda:

1. Verdadeiro – está claro que o estudo atende ao critério avaliado.
2. Falso – está claro que o estudo não atende ao critério avaliado.
3. Não está claro – não está claro se o estudo atende ao critério avaliado.
4. Não se aplica – não é possível avaliar esse critério, devido à natureza do estudo. Exemplo: se o estudo não for sobre DPP, não faz sentido avaliar os critérios D, E, F, G (todos serão preenchidos com 4 – não se aplica). Se for uma revisão de literatura, não faz sentido avaliar os critérios B, C, D, E, F, G (todos serão preenchidos com 4 – não se aplica).

Exemplos:

1. O critério C foi julgado falso de acordo com o título. Não é preciso ler o resumo. Portanto, as células de avaliação do resumo ficam em branco.

Títulos								Resumos							
A	B	C	D	E	F	G	Status do título	A	B	C	D	E	F	G	Status do resumo
		F					FALSO								FALSO

2. Nenhum critério foi julgado falso de acordo com o título. De acordo com o resumo, pode-se compreender claramente que todos os critérios são atendidos.

Títulos								Resumos							
A	B	C	D	E	F	G	Status do título	A	B	C	D	E	F	G	Status do resumo
							VERDADEIRO	1	1	1	1	1	1	1	VERDADEIRO

3. Nenhum critério foi julgado falso de acordo com o título. De acordo com o resumo, observa-se que os critérios A, B, C, D e G são atendidos claramente. Mas não está claro se o trabalho atende os critérios E e F. Ou seja, será necessário ler o texto completo para esclarecer.

Títulos								Resumos							
A	B	C	D	E	F	G	Status do título	A	B	C	D	E	F	G	Status do resumo
							VERDADEIRO	1	1	1	1	3	3	1	VERDADEIRO

4. Nenhum critério foi julgado falso de acordo com o título. Lendo o resumo, percebeu-se que todos os critérios são claramente atendidos,

com exceção do critério F, que é claramente falso. Portanto, não será avaliado o texto completo.

Títulos								Resumos							
A	B	C	D	E	F	G	Status do título	A	B	C	D	E	F	G	Status do resumo
							VERDADEIRO	1	1	1	1	1	2	1	FALSO

5. Nenhum critério foi julgado falso de acordo com o título. Lendo o resumo, percebeu-se que se trata de revisão de literatura. Portanto, em "Resumos", o critério A é falso (2) e todos os outros não se aplicam (4).

Títulos								Resumos							
A	B	C	D	E	F	G	Status do título	A	B	C	D	E	F	G	Status do resumo
							VERDADEIRO	2	4	4	4	4	4	4	FALSO

OBSERVAÇÕES

Depressão pós-parto é um transtorno de humor (*mood disorder*). Se o título ou o resumo mencionar que investigou transtornos de humor após o parto, considerar que pode ter avaliado DPP (ou seja, o critério B será marcado com 3 – não está claro). Fazer o mesmo com o termo "*postpartum distress*", que costuma incluir ansiedade e depressão.

Categorias que certamente são distintas de DPP (critério B será marcado com 2 – falso): transtorno de stress pós-traumático, transtorno obsessivo compulsivo, psicose pós-parto, melancolia pós-parto (*baby blues, maternity blues*), ansiedade pós-parto, disforia pós-parto.

Artigos sobre depressão em pais: não excluir pelo título, pois podem comparar prevalência em pais e mães. Pelo resumo, incluir apenas quando houver evidências de que também relataram prevalência nas mães.

Apêndice 2 – Referências dos estudos incluídos

Abbaszadeh, A., Kermani, F. P., Safizadeh, H., & Nakhee, N. (2011). Violence during pregnancy and postpartum depression. *Pakistan Journal of Medical Sciences, 27*(1), 177-181.

Abbott, M. W., & Williams, M. M. (2006). Postnatal depressive symptoms among Pacific mothers in Auckland: Prevalence and risk factors. *Australian and New Zealand Journal of Psychiatry, 40*, 230-238.

Abou Nazel, M. W., & Nosseir, S. A. (1994). Antepartum and postpartum depression and infant feeding pattern: A prospective study. *The Journal of the Egyptian Public Health Association, 69*(5), 397-424.

Abraham, A. J.. (2003). *Correlating health locus of control and risk for postpartum depression.* Dissertação de Mestrado, University of New Hampshire, New Hampshire.

Adewuya, A. O. (2006). Early postpartum mood as a risk factor for postnatal depression in Nigerian women. *American Journal of Psychiatry, 163*, 1435-1437.

Adewuya, A. O., & Afolabi, O. T. (2005). The course of anxiety and depressive symptoms in Nigerian postpartum women. *Archives of Women's Mental Health, 8*, 257-259.

Adewuya, A. O., Eegunranti, A. B., & Lawal, A. M. (2005). Prevalence of postnatal depression in Western Nigerian women: A controlled study. *International Journal of Psychiatry in Clinical Practice, 9*(1), 60-64.

Affonso, D. D., De, A. K., Horowitz, J. A., & Mayberry, L. J. (2000). An international study exploring levels of postpartum depressive symptomatology. *Journal of Psychosomatic Research, 49*, 207-216.

Affonso, D. D., Lovett, S., Paul, S. M., & Sheptak, S. (1990). A standardized interview that differentiates pregnancy and postpartum symptoms from perinatal clinical depression. *Birth-Issues in Perinatal Care, 17*(3), 121-130.

Agoub, M., Moussaoui, D., & Battas, O. (2005). Prevalence of postpartum depression in a Moroccan sample. *Archives of Women's Mental Health, 8*, 37-43.

Akman, C., Uguz, F., & Kaya, N. (2007). Postpartum-onset major depression is associated with personality disorders. *Comprehensive Psychiatry, 48*, 343-347.

Alcorn, K. L., O'Donovan, A., Patrick, J. C., Creedy, D., & Devilly, G. J. (2010). A prospective longitudinal study of the prevalence of post-traumatic stress disorder resulting from childbirth events. *Psychological Medicine, 40*, 1849-1859.

Ali, N. S., Ali, B. S., & Azam, I. S. (2009). Post partum anxiety and depression in peri-urban communities of Karachi, Pakistan: A quasi-experimental study. *BMC Public Health, 9*, 384-393.

Aliane, P. P.. (2008). *Alcohol use in pregnancy and its relationship with postpartum depressive symptoms.* Dissertação de Mestrado, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto.

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., Hernándes-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), 0.

Ammerman, R. T., Putnam, F. W., Altaye, M., Chen, L., Holleb, L. J., Stevens, J., ... Van Ginkel, J. B. (2009). Changes in depressive symptoms in first time mothers in home visitation. *Child Abuse & Neglect, 33*, 127-138.

Amr, M. A. -M., & Balaha, M. H. H. (2010). Minor psychiatric morbidity in young Saudi mothers using Mini International Neuropsychiatric Interview (MINI). *Journal of the College of Physicians and Surgeons Pakistan, 20*(10), 680-684.

Anderson, L. N., Campbell, M. C., daSilva, 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.

Andretto, D. A.. (2010). *Transtorno de estresse pós-traumático pós-parto e depressão pós-parto: Prevalência e fatores associados em puérperas do setor público e privado da zona leste de São Paulo.* Dissertação de Mestrado, Universidade de São Paulo, São Paulo.

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.

Astbury, J., Brown, S., Lumley, J., & Small, R. (1994). Birth events, birth experiences and social differences in postnatal depression. *Australian Journal of Public Health, 18*(2), 176-184.

Augusto, A., Kumar, R., Calheiros, J. M., Matos, E., & Figueiredo, E. (1996). Post-natal depression in an urban area of Portugal: Comparison of childbearing women and matched controls. *Psychological Medicine, 26*, 135-141.

Austin, M. -P., Hadzi-Pavlovic, D., Priest, S. R., Reilly, N., Wilhelm, K., Saint, K., & Parker, G. (2005). Depressive and anxiety disorders in the postpartum period: how prevalent are they and can we improve their detection?

Archives of Women's Mental Health, 13, 395-401.

Avan, B., Richter, L. M., Ramchandani, P. G., Norris, S. A., & Stein, A. (2010). Maternal postnatal depression and children's growth and behaviour during the early years of life: Exploring the interaction between physical and mental health. *Archives of Disease in Childhood, 95*, 690-695.

Aydin, N., Inandi, T., & Karabulut, N. (2005). Depression and associated factors among women within their first postnatal year in Erzurum province in eastern Turkey. *Women & Health, 41*(2), 1-12.

Ayela, D. U.. (2008). *Incidence of postpartum depression in African-Canadian women attending a prenatal program in Toronto, Canada.* Dissertação de Mestrado, Faculty of D'Youville College, Buffalo, NY.

Bagedahl-Strindlund, M., & Monsen Borjesson, K. (1998). Postnatal depression: A hidden illness. *Acta Psychiatrica Scandinavica, 98*, 272-275.

Baghianimoghadam, M. H., Shodjaee zadeh, D., & Aminian, A. H. (2009). Caesarean section, vaginal delivery and post natal depression. *Iranian Journal of Public Health, 38*(3), 118-122.

Baird, J., Hill, C. M., Kendrick, T., & Inskip, H. M. (2009). Infant sleep disturbance is associated with preconceptional psychological distress: Findings from the Southampton Women's Survey. *SLEEP, 32*(4), 566-568.

Baker, L., & Oswalt, K. (2008). Screening for postpartum depression in a rural community. *Community Mental Health Journal, 44*, 171-180.

Baker, L., Cross, S., Greaver, L., Wei, G., Lewis, R., & Helth Start, Corps. (2005). Prevalence of postpartum depression in a native American population. *Maternal and Child Health Journal, 9*(1), 21-25.

Ballard, C. G., Davis, R., Cullen, P. C., Mohan, R. N., & Dean, C. (1994). Prevalence of postnatal psychiatric morbidity in mothers and fathers. *British Journal of Psychiatry, 164*, 782-788.

Ballestrem, C. -L. v., Strauß, M., & Kächele, H. (2005). Contribution to the epidemiology of postnatal depression in Germany – implications for the utilization of treatment. *Archives of Women's Mental Health, 8*, 29-35.

Bandyopadhyay, M., Small, R., Watson, L. F., & Brown, S. (2010). Life with a new baby: How do immigrant and Australian-born women's experiences compare? *Australian and New Zealand Journal of Public Health, 34*(4), 412-421.

Banti, S., Mauri, M., Oppo, A., Borri, C., Rambelli, C., Ramacciotti, D., ... Cassano, G. B. (2011). From the third month of pregnancy to 1 year postpartum. Prevalence, incidence, recurrence, and new onset of depression. Results from the Perinatal Depression–Research & Screening Unit study. *Comprehensive Psychiatry, 52*, 343-351.

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.

Barbosa, R. C. A.. (2008). *Prevalência e incidência de depressão pós-parto e sua associação com o apoio social.* Dissertação de Mestrado, Universidade Federal de Pernambuco, Recife.

Beck, C. T., & Gable, R. K. (2001). Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nursing Research, 50*(4), 242-250.

Beck, C. T., Gable, R. K., Sakala, C., & Declercq, E. R. (2011). Postpartum depression symptomatology: Results from a two-stage US national survey. *Journal of Midwifery & Women's Health, 56*, 427-435.

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.

Beeghly, M., Weinberg, M. K., Olson, K. L., Kernan, H., Riley, J., & Tronick, E. Z. (2002). Stability and change in level of maternal depressive symptomatology during the first postpartum year. *Journal of Affective Disorders, 71*, 169-180.

Bennett, I. M., Coco, A., Coyne, J. C., Mitchell, A. J., Nicholson, J., Johnson, E., ... Ratcliffe, S. (2008). Efficiency of a two-item pre-screen to reduce the burden of depression screening in pregnancy and postpartum: An IMPLICIT network study. *The Journal of the American Board of Family Medicine, 21*, 317-325.

Benoit, C., Westfall, R., Treloar, A. E. B., Phillips, R., & Jansson, M. (2007). Social factors linked to postpartum depression: A mixed-methods longitudinal study. *Journal of Mental Health, 16*(6), 719-730.

Benvenuti, P., Ferrara, M., Niccolai, C., Valoriani, V., & Cox, J. L. (1999). The Edinburgh Postnatal Depression Scale: Validation for an Italian sample. *Journal of Affective Disorders, 53*, 137-141.

Berle, J. Ø., Aarre, T. F., Mykletun, A., Dahl, A. A., & Holsten, F. (2003). Screening for postnatal depression Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. *Journal of Affective Disorders, 76*, 151-156.

Bielawska-Batorowicz, E., & Kossakowska-Petrycka, K. (2006). Depressive mood in men after the birth of their offspring in relation to a partner's depression, social support, fathers' personality and prenatal expectations. *Journal of Reproductive and Infant Psychology, 24*(1), 21-29.

Bielinski-Blattmann, D., Lemola, S., Jaussi, C., Stadlmayr, W., & Grob, A. (2009). Postpartum depressive symptoms in the first 17 months after childbirth: The impact of an emotionally supportive partnership. *International Journal of Public Health, 54*, 333-339.

Bilszta, J. L. C., Gu, Y. Z., Meyer, D., & Buist, A. E. (2008). A geographic comparison of the prevalence and risk factors for postnatal depression in an Australian population. *Australian and New Zealand Journal of Public Health, 32*(5), 424-430.

Bina, R.. (2011). *Seeking for help for postpartum depression: The behavioral model of health service use as a framework for predicting treatment utilization for postpartum depression.* Tese de Doutorado, University of Maryland Baltimore, Baltimore.

Binder, E. B., Newport, D. J., Zach, E. B., Smith, A. K., Deveau, T. C., Altshuler, L. L., ... Cubells, J. F. (2010). A serotonin transporter gene polymorphism predicts peripartum depressive symptoms in an at-risk psychiatric cohort. *Journal of Psychiatric Research, 44*, 640-646.

Bjerke, S. E. Y., Vangen, S., Nordhagen, R., Ytterdahl, T., Magnus, P., & Stray-Pedersen, B. (2008). Postpartum depression among Pakistani women in Norway: Prevalence and risk factors. *The Journal of Maternal-Fetal and Neonatal Medicine, 21*(12), 889-894.

Blom, E. A., Jansen, P. W., Verhulst, F. C., Hofman, A., Raat, H., Jaddoe, V. W. V., ... Tiemeier, H. (2010). Perinatal complications increase the risk of postpartum depression. The Generation R Study. *British Journal of Obstetrics and Gynaecology (BJOG), 117*, 1390-1398.

Boyce, P. M., Johnstone, S. J., Hickey, A. R., Morris-Yates, A., Harris, M. G., & Stracham, T. (2000). Functioning and well-being at 24 weeks postpartum of women with postnatal depression. *Archives of Women's Mental Health, 3*, 91-97.

Boyce, P., & Hickey, A. (2005). Psychosocial risk factors to major depression after childbirth. *Social Psychiatry and Psychiatric Epidemiology, 40*, 605-612.

Boyce, P., Parker, G., Barnett, B., Cooney, M., & Smith, F. (1991). Personality as a vulnerability factor to depression. *British Journal of Psychiatry, 159*, 106-114.

Bozoky, I., & Corwin, E. J. (2002). Fatigue as a predictor of postpartum depression. *Journal of Obstetric Gynecologic and Neonatal Nursing, 31*(4), 436-443.

Brown, S., & Lumley, J. (1993). Maternal health after childbirth: Results of an Australian population based survey. *British Journal of Obstetrics and Gynaecology (BJOG), 105*, 156-161.

Bugdayci, R., Sasmaz, C. T., Tezcan, H., Kurt, A. O., & Öner, S. (2004). A cross-sectional prevalence study of depression at various times after delivery in Mersin Province in Turkey. *Journal of Women's Health, 13*(1), 63-68.

Buist, A. E., Austin, M. -P. V., Hayes, B. A., Speelman, C., Bilszta, J. L. C., Gemmill, A. W., ... Milgrom, J. (2008). Postnatal mental health of women giving birth in Australia 2002-2004: Findings from the beyondblue National Postnatal Depression Program. *Australian and New Zealand Journal of Psychiatry, 42*, 66-73.

Bunevičius, A., Kusminskas, L., & Bunevičius, R. (2009). Validation of the Lithuanian version of the Edinburgh Postnatal Depression Scale. *Medicina (Kaunas), 45*(7), 544-548.

Campbell, S. B., Cohn, J. F., & Meyers, T. (1995). Depression in first-time mothers: Mother-Infant Interaction and depression chronicity. *Developmental Psychology, 31*(3), 349-357.

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 fo the Edinburgh Postnatal Depression Scale to detect postpartum depression. *Revista Médica de Chile, 136*, 851-858.

Centers for Disease Control. (2008). Prevalence of Self-Reported Postpartum Depressive Symptoms — 17 States, 2004–2005. *Morbidity and Mortality Weekly Report, 57*(14), 361-392.

Chaaya, M., Campbell, O. M. R., El Kak, F., Shaar, D., Harb, H., & Kaddour, A. (2002). Postpartum depression: Prevalence and determinants in Lebanon. *Archives of Women's Mental Health, 5*, 65-72.

Chandran, M., Tharyan, P., Muliyil, J., & Abraham, S. (2002). Post-partum depression in a cohort of women from a rural area of Tamil Nadu, India. *British Journal of Psychiatry, 181*, 499-504.

Chatzi, L., Melaki, V., Sarri, K., Apostolaki, I., Roumeliotaki, T., Georgiou, V., ... Kogevinas, M. (2011). Dietary patterns during pregnancy and the risk of postpartum depression: The mother-child 'Rhea' cohort in Crete, Greece. *Public Health Nutrition, 14*(9), 1663-1670.

Chaudron, L. H., Klein, M. H., Remington, P., Palta, M., Allen, C., & Essex, M. J. (2001). Predictors, prodromes and incidence of postpartum depression. *Journal of Psychosomatic Obstetrics and Gynecology, 22*, 103-112.

Chaudron, L. H., Szilagyi, P. G., Kitzman, H. J., Wadkins, H. I. M., & Conwell, Y. (2004). Detection of postpartum depressive symptoms by screening at well-child visits. *Pediatrics, 113*, 551-558.

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.

Chen, C. -H. (2001). Association of work status and mental well-being in new mothers. *Kaohsiung Journal of Medical Sciences, 17*(11), 570-575.

Chibanda, D., Mangezi, W., Tshimanga, M., Woelk, G., Rusakaniko, S., Stranix-Chibanda,., ... Shetty, A. K. (2010). Postnatal depression by HIV status among women in Zimbabwe. *Journal of Women's Health, 19*(11), 2071-2077.

Chien, L. -Y., Tai, C. -J., Hwang, F. -M., & Huang, C. -M. (2009). Postpartum physical symptoms and depressive symptomatology at 1 month and 1 year after delivery: A longitudinal questionnaire survey. *International Journal of Nursing Studies, 46*, 1201-1208.

Choi, Y., Bishai, D., & Minkovitz, C. (2009). Multiple births are a risk factor for postpartum maternal depressive symptoms. *Pediatrics, 123*, 1147-1154.

Clarke, P. J. (2008). Validation of two postpartum depression screening scales with a sample of First Nations and Métis women. *Canadian Journal of Nursing Research, 40*(1), 112-125.

Coates, A. O., Schaefer, C. A., & Alexander, J. L. (2004). Detection of postpartum depression and anxiety in a large health plan. *Journal of Behavioral Health Services & Research, 31*(2), 117-133.

Coburn, B. J.. (2010). *Postpartum depression: A comparison of military and civilian population.* Dissertação de Mestrado, Montana State University, Bozeman, Montana.

Cohen, J. S. (2010). *Mindfulness and self-compassion in the transition to motherhood: A prospective study of postnatal mood and attachment.* Tese de Doutorado, Columbia University, New York.

Cooper, P. J., Murray, L., Hooper, R., & West, A. (1996). The development and validation of a predictive index for postpartum depression. *Psychological Medicine, 26*, 627-634.

Cooper, P. J., Tomlinson, M., Swartz, L., Woolgar, M., Murray, L., & Molteno, C. (1999). Postpartum depression and the mother-infant relationship in a South African peri-urban settlement. *British Journal of Psychiatry, 175*, 554-558.

Costa, R., Figueiredo, B., Pacheco, A., Marques, A., & Pais, A. (2005). Anticipation of Childbirth Questionnaire (ACQ). *Psychologica, 38*, 265-295.

Costa, R., Pacheco, A., & Figueiredo, B. (2007). Prevalence and predictors of depressive symptoms after childbirth. *Revista de Psiquiatria Clínica, 34*(4), 157-165.

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 Adolescênc, 10*(2), 63-71.

Cowley-Malcolm, E. T., Fairbairn-Dunlop, T. P., Paterson, J., Gao, W., & Williams, M. (2009). Child discipline and nurturing practices among a cohort

of Pacific mothers living in New Zealand. *Pacific Health Dialog, 15*(1), 36-45.

Cox, J. L., Murray, D., & Chapman, G. (1993). A controlled study of the onset, duration and prevalence of postnatal depression. *British Journal of Psychiatry, 163*, 27-31.

Crossett, S. E.. (2009). *Interpersonal and cognitive risk factors for postparum depression.* Tese de Doutorado, Binghamton University, State Univers, Binghamton.

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.

Cryan, E., Keogh, F., Connolly, E., Cody, S., Quinlan, A., & Daly, I. (1993). Depression among postnatal women in an urban Irish community. *Irish Journal of Psychological Medicine, 18*(1), 5-10.

Cury, A. F.. (2001). *Depressão puerperal: Relação com eventos vitais estressantes, modo de enfrentamento e níveis do ácido 5 hidroxiindolacético.* Tese de Doutorado, Universidade de São Paulo, São Paulo.

Czarnocka, J., & Slade, P. (2000). Prevalence and predictors of post-traumatic stress symptoms following childbirth. *British Journal of Clinical Psychology, 39*, 35-51.

Da Costa, D., Larouche, J., Dritsa, M., & Brender, W. (2000). Psychosocial correlates of prepartum and postpartum depressed mood. *Journal of Affective Disorders, 59*, 31-40.

Dankner, R., Goldberg, R. P., Fisch, R. Z., & Crum, R. M. (2000). Cultural elements of postpartum depression - A study of 327 Jewish Jerusalem women. *Journal of Reproductive Medicine, 45*, 97-104.

Davey, H. L.. (2006). *Epidemiology of postpartum depression: A prospective study.* Dissertação de Mestrado, University of Calgary, Calgary.

Davis, S., Cross, J., & Lind, B. K. (2008). Exploring the Postpartum Adjustment Questionnaire as a predictor of postpartum depression. *Journal of Obstetric Gynecologic and Neonatal Nursing, 37*, 622-630.

De Tychey, C., Briançon, S., Lighezzolo, J., Spitz, E., Kabuth, B., Luigi, V., ... Vincent, S. (2008). Quality of life, postnatal depression and baby gender. *Journal of Clinical Nursing, 17*, 312-322.

Demissie, Z.. (2010). *The associations between physical activity and antepartum and postpartum depression.* Tese de Doutorado, Chapel Hill, Chapel Hill.

Dennis, C. -L. (2004). Can we identify mothers at risk for postpartum depression

in the immediate postpartum period using the Edinburgh Postnatal Depression Scale? *Journal of Affective Disorders, 78*, 163-169.

Dørheim, S. K., Bondevik, G. T., Eberhard-Gran, M., & Bjorvatn, B. (2009). Sleep and depression in postpartum women: A population-based study. *SLEEP, 32*(7), 847-855.

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.

Eberhard-Gran, M., Eskild, A., Tambs, K., Samuelsen, S. O., & Opjordsmoen, S. (2002). Depression in postpartum and non-postpartum women: Prevalence and risk factors. *Acta Psychiatrica Scandinavica, 106*, 426-433.

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.

Edwards, B., Galletly, C., Semmler-Booth, T., & Dekker, G. (2008). Does antenatal screening for psychosocial risk factors predict postnatal depression? A follow-up study of 154 women in Adelaide, South Australia. *Australian and New Zealand Journal of Psychiatry, 42*(1), 51-55.

Ege, E., Timur, S., Zincir, H., Geçkil, E., & Sunar-Reeder, B. (2008). Social support and symptoms of postpartum depression among new mothers in Eastern Turkey. *Journal of Obstetrics and Gynaecology Research, 34*(4), 585-593.

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). Post partum depression: is it associated to gender based violence? *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 realidad en el sistema publico de atencion de salud. *Revista Chilena de Obstetricia y Ginecología, 68*(6), 491-494.

Evans, J., Heron, J., Francomb, H., Oke, S., Golding, J., & Avon Longitudinal. (2001). Cohort study of depressed mood during pregnancy and after childbirth. *British Medical Journal, 323*(4), 257-260.

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.

Felice, E., Saliba, J., Grech, V., & Cox, J. (2004). Prevalence rates and psychosocial characteristics associated with depression in pregnancy and postpartum in Maltese women. *Journal of Affective Disorders, 82*, 297-301.

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.

Figueiredo, B., Pacheco, A., & Costa, R. (2007). Depression during pregnancy and the postpartum period in adolescent and adult Portuguese mothers. *Archives of Women's Mental Health, 10*, 103-109.

Fisher, J,., Tran, T., La, B. t., Kriitmaa, K., Rosenthal, D., & Tran, T. (2010). Common perinatal mental disorders in northern Viet Nam: Community prevalence and health care use. *Bulletin of the World Health Organization, 88*, 737-745.

Fisher, J. R. W., Morrow, M. M., Ngoc, N. T. N., & Anh, L. T. H. (2004). Prevalence, nature, severity and correlates of postpartum depressive symptoms in Vietnam. *British Journal of Obstetrics and Gynaecology (BJOG), 111*, 1353-1360.

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.

Fonseca, V. R. J. R. M., Silva, G. A., & Otta, E. (2010). The relationship between postpartum depression and maternal emotional availability. *Cadernos de Saúde Pública, 26*(4), 738-746.

Forman, D. N., Videbech, P., Hedegaard, M., Salving, J. D., & Secher, N. J. (2000). Postpartum depression: Identification of women at risk. *British Journal of Obstetrics and Gynaecology (BJOG), 107*, 1210-1217.

Freeman, M. P., Wright, R., Watchman, M., Wahl, R. A., Sisk, D. J., Fraleigh, L., & Weibrecht, J. M. (2005). Postpartum depression assessments at well-baby visits: Screening feasibility, prevalence, and risk factors. *Journal of Women's Health, 14*(10), 929-935.

Friedman, K. B.. (2008). *Are symptoms of postpartum depression associated with deficits in facial and auditory emotional recognition?* Tese de Doutorado, Drexel University, Philadelphia.

Fuggle, P., & Haydon, K. (2000). Estimating resources for setting up an HV service for postnatal depression. *British Journal of Community Nursing, 5*(7), 348-351.

Fuggle, P., Glover, L., Khan, F., & Haydon, K. (2002). Screening for postnatal depression in Bengali women: Preliminary observations from using a translated version of the Edinburgh Postnatal Depression Scale (EPDS). *Journal of Reproductive and Infant Psychology, 20*(2), 71-82.

Gao, L. L., Chan, W-C., & Mao, Q. (2009). Depression, perceived stress, and social support among first-time Chinese mothers and fathers in the postpartum period. *Research in Nursing & Health, 32*, 50-58.

Garcia-Esteve, L., Acaso, C., Ojuel, J., & Navarro, P. (2003). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *Journal of Affective Disorders, 75*, 71-76.

Gausia, K., Fisher, C., Ali, M., & Oosthuizen, J. (2009). Magnitude and contributory factors of postnatal depression: a community-based cohort study from a rural subdistrict of Bangladesh. *Psychological Medicine, 39*, 999-1007.

Georgiopoulos, A. M., Bryan, T. L., Yawn, B. P., Houston, M. S., Rummans, T. A., & Therneau, T. M. (1999). Population-based screening for postpartum depression. *Obstetrics and Gynecology, 93*(5), 653-657.

Ghubash, R., & Abou-Saleh, M. T. (1997). Postpartum psychiatric illnessin Arab culture: Prevalence and psychosocial correlates. *The British Journal of Psychiatry, 171*, 65-68.

Giakoumaki, O., Vasilaki, K., Skouroliakou, M., & Liosis, G. (2009). The role of maternal anxiety in the early postpartum period: Screening for anxiety and depressive symptomatology in Greece. *Journal of Psychosomatic Obstetrics & Gynecology, 30*(1), 21-28.

Gilligan, C. E.. (2010). *Screening for postpartum depression and posttraumatic stress disorder in the immediate postpartum.* Tese de Doutorado, Antioch University New England, Keene.

Gjerdingen, D., Crow, S., McGovern, P., Miner, M., & Center, B. (2011). Changes in depressive symptoms over 0–9 months postpartum. *Journal of Women's Health, 20*(3), 381-386.

Glasser, S., Barell, V., Boyko, V., Ziv, A., Lusky, A., Shoham, A., & Hart, S. (2000). Postpartum depression in an Israeli cohort: Demographic, psychosocial and medical risk factors. *Journal of Psychosomatic Obstetrics*

& Gynecology, 21, 99-108.

Glasser, S., Tanous, M., Shihab, S., Goldman, N., Ziv, A., & Kaplan, G. (2011). Perinatal depressive symptoms among Arab women in Northern Israel. *Maternal and Child Health Journal, 15*, 0.

Glavin, K., Smith, L., & Sørum, R. (2009). Prevalence of postpartum depression in two municipalities in Norway. *Scandinavian Journal of Caring Sciences, 23*, 705-710.

Gonidakis, F., Rabavilas, A. D., Varsou, E., Kreatsas, G., & Christodoulou, G. N. (2008). A 6-month study of postpartum depression and related factors in Athens Greece. *Comprehensive Psychiatry, 49*, 275-282.

Goodman, J. H., & Tyer-Viola, L. (2010). Detection, treatment, and referral of perinatal depression and anxiety by obstetrical providers. *Journal of Women's Health, 19*(3), 477-490.

Gorman, L. L., O'Hara, M. W., Figueiredo, B., Hayes, S., Jacquemain, F., Kammerer, M. H., ... Sutter-Dallay, A. -L. (2004). Adaptation of the Structured Clinical Interview for DSM-IV Disorders for assessing depression in women during pregnancy and post-partum across countries and cultures. *The British Journal of Psychiatry, 184*, 17-23.

Gotlib, I. H., Whiffen, V. E., Mount, J. H., Milne, K., & Cordy, N. I. (1989). Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. *Journal of Consulling and Clinical Psychology, 57*(2), 269-274.

Goyal, D., Gay, C., & Lee, K. A. (2010). How much does low socioeconomic status increase the risk of prenatal and postpartum depressive symptoms in first-time mothers? *Women's Health Issues, 20*, 96-104.

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.

Green, J. M. (1998). Postnatal depression or perinatal dysphoria? Findings from a longitudinal community-based study using the Edinburgh Postnatal Depression Scale. *Journal of Reproductive & Infant Psychology, 16*(2), 143-155.

Green, K., Broome, H., & Mirabella, J. (2006). Postnatal depression among mothers in the United Arab Emirates: Socio-cultural and physical factors. *Psychology, Health & Medicine, 11*(4), 425-431.

Gremigni, P., Mariani, L., Marracino, V., Tranquilli, A. L., & Turi, A. (2011). Partner support and postpartum depressive symptoms. *Journal of Psychosomatic Obstetrics & Gynecology, 32*(3), 135-140.

Gress-Smith, J. L., Luecken, L. J., Lemery-Chalfant, K., & Howe, R. (2012). Postpartum depression prevalence and impact on infant health, weight, and sleep in low-income and ethnic minority women and infants. *Maternal and Child Health Journal, 16*, 887-893.

Grote, N. K., & Bledsoe, S. E. (2007). Predicting postpartum depressive

symptoms in new mothers: The role of optimism and stress frequency during pregnancy. *Health & Social Work, 32*(2), 107-118.

Grussu, P., & Quatraro, R. M. (2009). Prevalence and risk factors for a high level of postnatal depression symptomatology in Italian women: A sample drawn from ante-natal classes. *European Psychiatry, 24*, 327-333.

Guest, M., & Stamp, G. (2009). South Australian rural women's views of their pregnancy, birthing and postnatal care. *Rural and Remote Health, 9*(3), 1101.

Gulseren, L., Erol, A., Gulseren, S., Kuey, L., Kilic, B., & Ergor, G. (2006). From antepartum to postpartum: A prospective study on the prevalence of peripartum depression in a semiurban Turkish community. *Journal of Reproductive Medicine, 51*, 955-960.

Hamdan, A., & Tamim, H. (2011). Psychosocial risk and protective factors for postpartum depression in the United Arab Emirates. *Archives of Women's Mental Health, 14*, 125-133.

Hannah, P., Adams, D., Lee, A., Glover, V., & Sandler, M. (1992). Links between early postpartum mood and postnatal depression. *British Journa! of Psychiatry, 160*, 777-780.

Hanusa, B. H., Scholle, S. H., Haskett, R. F., Spadaro, K., & Wisner, K. L. (2008). Screening for depression in the postpartum period: A comparison of three instruments. *Journal of Women's Health, 17*(4), 585-596.

Hasbún H. , J., Risco N. , L., Jadresic M. , E., Galleguillo U. ,., 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, D. M., Pakenham, K. I., & Smith, A. (2006). Social support and postpartum depressive symptomatology: The mediating role of maternal self-efficacy. *Infant Mental Health Journal, 27*(3), 276-291.

Haywood, A., Slade, P., & King, H. (2007). Is there evidence of an association between postnatal distress and premenstrual symptoms? *Journal of Affective Disorders, 99*, 241-245.

Heron, J., Haque, S., Oyebode, F., Craddock, N., & Jones, I. (2009). A longitudinal study of hypomania and depression symptoms in pregnancy and the postpartum period. *Bipolar Disorders, 11*, 410-417.

Hobfoll, S. E., Ritter, C., Lavin, J., Hulsizer, M. R., & Cameron, R. P. (1995). Depression prevalence and incidence among inner-city pregnant and postpartum women. *Journal of Consulting and Clinical Psychology, 63*(3), 445-453.

Horowitz, J. A. (2006). Community-based postpartum depression screening within the first month after delivery. *Contemporary Nurse, 21*(1), 85.

Horowitz, J. A., Murphy, C. A., Gregory, K. E., & Wojcik, J. (2009). Community-based postpartum depression screening: Results from the CARE study. *Psychiatric Services, 60*(11), 1432-1434.

House, J. D., Iriarte, R. I., & Burns, E. A. (1986). Stressful life events and depressive symptomatology in obstetric patients: A pilot study in a family practice setting. *Family Practice Research Journal, 6*(2), 98-105.

Howard, L. M., Flach, C., Mehay, A., Sharp, D., & Tylee, A. (2011). The prevalence of suicidal ideation identified by the Edinburgh Postnatal Depression Scale in postpartum women in primary care: Findings from the RESPOND trial. *BMC Pregnancy and Childbirth, 11*, 57.

Ho-Yen, S., Bondevik, G. T., Eberhard-Gran, M., & Bjorvatn, B. (2006). The prevalence of depressive symptoms in the postnatal period in Lalitpur district, Nepal. *Acta Obstetricia et Gynecologica, 85*, 1186-1192.

Huang, C. -M., Carter, P. A., & Guo, J. L. (2004). A comparison of sleep and daytime sleepiness in depressed and non-depressed mothers during the early postpartum period. *Journal of Nursing Research, 12*(4), 287-295.

Huang, Z. J., Wong, F. Y., Ronzio, C. R., & Yu, S. M. (2007). Depressive symptomatology and mental health help-seeking patterns of US- and foreign-born mothers. *Maternal and Child Health Journal, 11*, 257-267.

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.

Husain, N., Beve, I., Husain, M., Chaudhry, I. B., Atif, N., & Rahman, A. (2006). Prevalence and social correlates of postnatal depression in a low income country. *Archives of Women's Mental Health, 9*, 197-202.

Husain, N., Parveen, A., Husain, M., Saeed, Q., Jafri, F., Rahman, R., ... Chaudhry, I. B. (2011). Prevalence and psychosocial correlates of perinatal depression: A cohort study from urban Pakistan. *Archives of Women's Mental Health, 13*, 395-403.

Hutto, H. F., Kim-Godwin, Y. S., Pollard, D., & Kemppainen, J. (2011). Postpartum depression among white, african american, and hispanic low-income mothers in rural southeastern North Carolina. *Journal of Community Health Nursing, 28*(1), 41-53.

Inandi, T., Elci, O. C., Ozturk, A., Egri, M., Polat, A., & Sahin, T. K. (2002). Risk factors for depression in postnatal first year, in eastern Turkey. *International Journal of Epidemiology, 31*, 1201-1207.

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.

Josefsson, A., Berg, G., Nordin, C., & Sydsjö, G. (2001). Prevalence of depressive symptoms in late pregnancy and postpartum. *Acta Obstetricia et Gynecologica Scandinavica, 80*, 251-255.

Kadir, A. A., Daud, M. N. M., Yaacob, M. J., & Hussain, N. H. N. (2009). Relationship between obstetric risk factors and postnatal depression in Malaysian women. *International Medical Journal, 16*(2), 101-106.

Kakyo, T. A., Muliira, J. K., Mbalinda, S. N., Kizza, I. B., & Muliira, R. S. (2012). Factors associated with depressive symptoms among postpartum mothers in a rural district in Uganda. *Midwifery, 28*, 374-379.

Kaminsky, L. M., Carlo, J., Muench, M. V., Nath, C., Harrigan, J. T., & Canterino, J. (2008). Screening for postpartum depression with the Edinburgh Postnatal Depression Scale in an indigent population: Does a directed interview improve detection rates compared with the standard self-completed q. *The Journal of Maternal-Fetal and Neonatal Medicine, 21*(5), 321-325.

Kammerer, M., Marks, M. N., Pinard, C., Taylor, A., Castelberg, B. v., Künzli, H., & Glover, V. (2009). Symptoms associated with the DSM IV diagnosis of depression in pregnancy and post partum. *Archives of Women's Mental Health, 12*, 135-141.

Kara, B., Ünalan, P., Çifçili, S., Cebeci, D. S., & Sarper, N. (2008). Is there a role for the family and close community to help reduce the risk of postpartum depression in new mothers? A cross-sectional study of Turkish women. *Maternal and Child Health Journal, 12*, 155-161.

Karkum, S.. (2005). *Occupational role performance and post-partum depression: A pilot exploratory study.* Dissertação de Mestrado, State University of New York, Buffalo.

Kent, V. A.. (2004). *A survey of postpartum depression in Southeastern Montana.* Dissertação de Mestrado, Montana State University, Bozeman.

Kheirabadi, G. R., & Maracy, M. R. (2010). Perinatal depression in a cohort study on Iranian women. *Journal of Research in Medical Sciences, 15*(1), 41-49.

Kim, J. J., Gordon, T. E. J., Porte, L. M. L., Adams, M., Kuendig, J. M., & Silver, R. K. (2008). The utility of maternal depression screening in the third trimester. *American Journal of Obstetrics and Gynecology, 199*, 509.

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.

Kirpinar, I., Gözüm, S., & Pasinlioglu, T. (2010). Prospective study of postpartum depression in eastern Turkey prevalence, socio-demographic and obstetric correlates, prenatal anxiety and early awareness. *Journal of Clinical Nursing, 19*, 422-431.

Kogima, E. O.. (2010). *Postpartum depression in adolescents enrolled at Family Health Strategy in the city of Embu Guaçu.* Tese de Doutorado, Universidade de São Paulo, São Paulo.

Kok, L. P., Chan, P. S. L., & Ratnam, S. S. (1994). Postnatal depression in

Singapore women. *Singapore Medical Journal, 35*, 33-35.

Koo, V., Lynch, J., & Cooper, S. (2003). Risk of postnatal depression after emergency delivery. *Journal of Obstetrics and Gynaecology Research, 29*(4), 246-250.

Kozinszky, Z., Dudas, R. B., Csatordai, S., Devosa, I., Tóth, E., Szabó, D., ... Pál, A. (2011). Social dynamics of postpartum depression: a population-based screening in South-Eastern Hungary. *Social Psychiatry and Psychiatric Epidemiology, 46*, 413-423.

Kron, T., & Brosh, A. (2003). Can dreams during pregnancy predict postpartum depression? *Dreaming, 13*(2), 67-81.

Lane, A., Keville, R., Morris, M., Kinsella, A., Turner, M., & Barry, S. (1997). Postnatal depressionand elation among mothers and their partners: Prevalence and predictors. *The British Journal of Psychiatry, 171*, 550-555.

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.

Lau, Y., Wang, Y., Yin, L., Chan, K. S., & Guo, X. (2010). Validation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale in Chengdu mothers. *International Journal of Nursing Studies, 47*, 1139-1151.

Leahy-Warren, P., McCarthy, G., & Corcoran, P. (2011). Postnatal depression in first-time mothers: Prevalence and relationships between functional and structural social support at 6 and 12 weeks postpartum. *Archives of Psychiatric Nursing, 25*(3), 174-184.

Lee, A,M., Lam, S. K., Lau, M. S. M., Chong, C. S. Y., Chui, H. W., & Fong, D. Y. T. (2007). Prevalence, course, and risk factors for antenatal anxiety and depression. *Obstetrics and Gynecology, 110*(5), 1102-1112.

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.

Lee, D. T. S., Yip, S. K., Chiu, H. F. K., Leung, T. S. Y., Chan, K. P. M., Chau, I. O. L., ... Chung, T. K. H. (1998). Detecting postnatal depression in Chinese women: Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry, 172*, 433-437.

Lee, E. A.. (2009). *Achieving semantic equivalence between the Chinese and English versions of the Postpartum Depression Screening Scale.* Tese de Doutorado, University of Connecticut, Connecticut.

Leigh, B., & Milgrom, J. (2008). Risk factors for antenatal depression, postnatal depression and parenting stress. *BMC Psychiatry, 8*, 24.

Leonardou, A. A., Zervas, Y. M., Papageorgiou, C. C., Marks, M. N., Tsartsara, E. C., Antsaklis, A., ... Soldatos, C. R. (2009). Validation of the Edinburgh Postnatal Depression Scale and prevalence of postnatal depression at two months postpartum in a sample of Greek mothers. *Journal of Reproductive*

and Infant Psychology, 27(1), 28-39.

Leung, S. S. K., Martinson, I. M., & Arthur, D. (2005). Postpartum depression and related psychosocial variables in Hong Kong Chinese women: Findings from a prospective study. *Research in Nursing & Health, 28*, 27-38.

Leung, W. C., Kung, F., Lam, J., Leung, T. W., & Ho, P. C. (2002). Domestic violence and postnatal depression in a Chinese community. *International Journal of Gynecology and Obstetrics, 79*, 159-166.

Leverton, T. J., & Elliott, S. A. (2000). Is the EPDS a magic wand?: 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. *Journal of Reproductive and Infant Psychology, 18*(4), 279-296.

Liberto, T. L.. (2011). *Attitudes toward help seeking among depressed and non-depressed postpartum women.* Tese de Doutorado, West Virginia University, Morgantown, West Virginia.

Limlomwongse, N., & Liabsuetrakul, T. (2006). Cohort study of depressive moods in Thai women during late pregnancy and 6–8 weeks of postpartum using the Edinburgh Postnatal Depression Scale (EPDS). *Archives of Women's Mental Health, 9*, 131-138.

Lincoln, M. S. W., Feyerharm, R., Damran, P., DeVault, M., Lorenz, D., & Dooley, S. (2008). Maternal depression after delivery in Oklahoma. *The Journal of the Oklahoma State Medical Association, 101*(12), 307-311.

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.

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.

Magistris, A., Coni, E., Puddu, M., Zonza, M., & Fanos, V. (2010). Screening of postpartum depression: Comparison between mothers in the neonatal intensive care unit and in the neonatal section. *The Journal of Maternal-Fetal and Neonatal Medicine, 23*, 101-103.

Mancini, F., Carlson, C., & Albers, L. (2007). Use of the Postpartum Depression Screening Scale in a collaborative obstetric practice. *Journal of Midwifery & Women's Health, 52*(5), 429-434.

Mandy, A., Gard, P. R., Ross, K., & Valentine, B. H. (1998). Psychological sequelae in women following either parturition or non-gynaecological surgery. *Journal of Reproductive and Infant Psychology, 16*(2), 133-141.

Mann, J. R., McKeown, R. E., Bacon, J., Vesselinov, R., & Bush, F. (2008). Do antenatal religious and spiritual factors impact the risk of postpartum depressive symptoms? *Journal of Women's Health, 17*(5), 745-755.

Mao, Q., Zhu, L. X., & Su, X. -Y. (2011). A comparison of postnatal depression

and related factors between Chinese new mothers and fathers. *Journal of Clinical Nursing, 20*, 645-652.

Mariam, K. A., & Srinivasan, K. (2009). Antenatal psychological distress and postnatal depression: A prospective study from an urban clinic. *Asian Journal of Psychiatry, 2*, 71-73.

Marques, M., Bos, S., Soares, M. J., Maia, B., Pereira, A. T., Valente, J., ... Azevedo, M. H. (2011). Is insomnia in late pregnancy a risk factor for postpartum depression/depressive symptomatology? *Psychiatry Research, 186*, 272-280.

Martínez, E., Toledo, M., Pineda, M., Monleón, J., Ferrero, J., & Barreto, P. (2001). La depresión postparto: Un estudio exploratorio con una muestra de mujeres valencianas. *Revista de Psicopatología y Psicología Clínica, 6*(1), 37-48.

Mason, Z. S.. (2009). *The role of maternal attachment: Its effects on postpartum depression and infant social-emotional development.* Tese de Doutorado, Yeshiva University, New York.

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.

Matthey, S., Barnett, B., Kavanagh, D. J., & Howie, P. (2001). Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. *Journal of Affective Disorders, 64*, 175-184.

Mauri, M., Oppo, A., Montagnani, M. S., Borri, C., Banti, S., Camilleri, V., ... Cassano, G. B. (2010). Beyond “postpartum depressions”: Specific anxiety diagnoses during pregnancy predict different outcomes - Results from PND-ReScU. *Journal of Affective Disorders, 17*, 177-184.

Mayberry, L. J., Horowitz, J. A., & Declercq, E. (2007). Depression symptom prevalence and demographic risk factors among US women during the first 2 years postpartum. *Journal of Obstetric Gynecologic and Neonatal Nursing, 36*, 542-549.

McCoy, S. J. B., Beal, J. M., Shipman, S. B. M., Payton, M. E., & Watson, G. H. (2006). Risk factors for postpartum depression: A retrospective investigation at 4-weeks postnatal and a review of the literature. *The Journal of the American Osteopathic Association, 106*, 193-198.

McLearn, K. T., Minkovitz, C. S., Strobino, D. M., Marks, E., & Hou, W. (2006). The timing of maternal depressive symptoms and mothers' parenting practices with young children: Implications for pediatric practice. *Pediatrics,*

118(1), 174-182.

Mechakra-Tahiri, S., Zunzunegui, M. V., & Seguin, L. (2007). Self-rated health and postnatal depressive symptoms among immigrant mothers in Quebec. *Women & Health, 45*(4), 1-17.

Meltzer-Brody, S., Stuebe, A., Dole, N., Savitz, D., Rubinow, D., & Thorp, J. (2011). Elevated Corticotropin Releasing Hormone (CRH) during Pregnancy and Risk of Postpartum Depression (PPD). *Journal of Clinical Endocrinology & Metabolism, 96*(1), 40-47.

Milgrom, J., Ericksen, J., Negri, L., & Gemmill, A. W. (2005). Screening for postnatal depression in routine primary care: Properties of the Edinburgh Postnatal Depression Scale in an Australian sample. *Australian and New Zealand Journal of Psychiatry, 39*, 833-839.

Miller, R. L., Pallant, J. F., & Negri, L. M. (2006). Anxiety and stress in the postpartum: Is there more to postnatal distress than depression? *BMC Psychiatry, 6*, 12.

Mishina, H., Hayashino, Y., Takayama, J. I., Kasahara, M., & Fukuhara, S. (2010). Can pediatricians accurately identify maternal depression at well-child visits? *Pediatrics International, 52*, 284-289.

Miyake, Y., Tanaka, K., Sasaki, S., & Hirota, Y. (2011). Employment, income, and education and risk of postpartum depression: The Osaka Maternal and Child Health Study. *Journal of Affective Disorders, 130*, 133-137.

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.

Mori, T., Tsuchiya, K. J., Matsumoto, K., Suzuki, K., Mori, N., Takei, N., & The HBC Study Team. (2011). Psychosocial risk factors for postpartum depression and their relation to timing of onset: The Hamamatsu Birth Cohort (HBC) Study. *Journal of Affective Disorders, 135*, 341-346.

Morof, D., Barrett, G., Peacock, J., Victor, C. R., & Manyonda, I. (2003). Postnatal depression and sexual health after childbirth. *The American College of Obstetricians and Gynecologists, 102*(6), 1318-1325.

Morris-Rush, J. K., Freda, M. C., & Bernstein, P. S. (2003). Screening for postpartum depression in an inner-city population. *American Journal of Obstetrics and Gynecology, 188*(5), 1217-1219.

Murphy, P. K.. (2009). *Postpartum depression and vitamin D.* Tese de Doutorado, Medical University of South Carolina, Charleston.

Nagy, E., Molnar, P., Pal, A., & Orvos, H. (2011). Prevalence rates and socioeconomic characteristics of post-partum depression in Hungary. *Psychiatry Research, 185*, 113-120.

Nakku, J. E. M., Nakasi, G., & Mirembe, F. (2006). Postpartum major depression at six weeks in primary health care: prevalence and associated factors. *African Health Sciences, 6*(4), 207-214.

Navarro, G. M., Calvente, M. M. G., & Parra, A. F. (2005). Estado de ánimo al mes postparto evaluado con la Escala de Depresión Postnatal de Edimburgo. *Psicología Conductual, 13*(2), 231-241.

Navarro, P., García-Esteve, L., Ascaso, C., Aguado, J., Gelabert, E., & Martín-Santos, R. (2008). Non-psychotic psychiatric disorders after childbirth: Prevalence and comorbidity in a community sample. *Journal of Affective Disorders, 109*, 171-176.

Ndokera, R., & MacArthur, C. (2011). The relationship between maternal depression and adverse infant health outcomes in Zambia: a cross-sectional feasibility study. *Child: Care, Health and Development, 37*(1), 74-81.

Nishizono-Maher, A., Kishimoto, J., Yoshida, H., Urayama, K., Miyato, M., Otsuka, Y., & Matsui, H. (2004). The role of self-report questionnaire in the screening of postnatal depression: A community sample survey in central Tokyo. *Social Psychiatry and Psychiatric Epidemiology, 39*, 185-190.

O., Neunaber, D. J., & Zekoskd, E. M. (1984). Prospective study of postpartum depression: Prevalence, course, and predictive factors. *Journal of Abnormal Psychology, 93*(2), 158-171.

Özbaoaran, F., Çoban, A., & Kucuk, M. (2011). Prevalence and risk factors concerning postpartum depression among women within early postnatal periods in Turkey. *Archives of Gynecology and Obstetrics, 283*, 483-490.

Ozdemir, H., Ergin, N., Selimoglu, K., & Bilgel, N. (2005). Postnatal depressive mood in Turkish women. *Psychology, Health & Medicine, 10*(1), 96-107.

Page, M. A.. (2008). *The context and experience of new mothers: Postpartum depression, family relationships, knowledge of infant development, and infant outcomes.* Tese de Doutorado, University of Arizona, Tucson.

Parish, C. J.. (2005). *Evaluation of two depression screening tools for postpartum mothers at a rural WIC program.* Dissertação de Mestrado, University of Wyoming, Laramie, Wyoming (EUA).

Pawlby, S., Sharp, D., Hay, D., & O'Keane, V. (2008). Postnatal depression and child outcome at 11 years: The importance of accurate diagnosis. *Journal of Affective Disorders, 107*, 241-245.

Paykel, E. S., Emms, E. M., Fletcher, J., & Rassaby, E. S. (1980). Life events and social support in puerperal depression. *British Journal of Psychiatry, 136*, 339-346.

Pedrós Roselló, A., Tomás Pérez, A., Gómez Beneyto, M., & García Iturrospe, K. (1998). Child rearing patterns in mothers with postpartum depression.

Actas Luso-Espanolas de Neurologia Psiquiatria y Ciencias Afines, 26(2), 104-110.

Pereira, A. T., Bos, S., Marques, M., Maia, R. B., Soares, M. J., Valente, J., ... Azevedo, M. H. (2010). The Portuguese version of the Postpartum Depression Screening Scale. *Journal of Psychosomatic Obstetrics & Gynecology, 31*(2), 90-100.

Petrosyan, D., Armenian, H. K., & Arzoumanian, K. (2011). Interaction of maternal age and mode of delivery in the development of postpartum depression in Yerevan, Armenia. *Journal of Affective Disorders, 135*, 77-81.

Phillips, J., Sharpe, L., Matthey, S., & Charles, M. (2010). Subtypes of postnatal depression? A comparison of women with recurrent and de novo postnatal depression. *Journal of Affective Disorders, 120*, 67-75.

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.

Pollock, J. I., Manaseki-Holland, S., & Patel, V. (2009). Depression in Mongolian women over the first 2 months after childbirth: Prevalence and risk factors. *Journal of Affective Disorders, 116*, 126-133.

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.

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.

Posmontier, B. (2008). Functional status outcomes in mothers with and without postpartum depression. *Journal of Midwifery & Women's Health, 53*, 310-318.

Priel, B., & Besser, A. (1999). Vulnerability to postpartum depressive symptomatology: Dependency, self-criticism and the moderating role of antenatal attachment. *Journal of Social and Clinical Psychology, 18*(2), 240-253.

Quelopana, A. M., Champion, J. D., & Reyes-Rubilar, T. (2011). Factors associated with postpartum depression in Chilean women. *Health Care for Women International, 32*, 939-949.

Rahman, A., Iqbal, Z., & Harrington, R. (2003). Life events, social support and depression in childbirth: Perspectives from a rural community in the developing world. *Psychological Medicine, 33*, 1161-1167.

Ramos-Marcuse, F., Oberlander, S., Papas, M. A., McNary, S. W., Hurley, K. M., & Black, M. M. (2010). Stability of maternal depressive symptoms among urban, low-income, African American adolescent mothers. *Journal of*

Affective Disorders, 122, 68-75.

Reading, R., & Reynolds, S. (2001). Debt, social disadvantage and maternal depression. *Social Science & Medicine, 53*, 441-453.

Reck, C., Struben, K., Backenstrass, M., Stefenelli, U., Reinig, K., Fuchs, T., ... Mundt, C. (2008). Prevalence, onset and comorbidity of postpartum anxiety and depressive disorders. *Acta Psychiatrica Scandinavica, 118*, 459-468.

Regmi, S., Sligl, W., Carter, D., Grut, W., & Seear, M. (2002). A controlled study of postpartum depression among Nepalese women: Validation of the Edinburgh Postpartum Depression Scale in Kathmandu. *Tropical Medicine and International Health, 7*(4), 378-382.

Reiha, J.. (2004). *An examination of the relationship between perceived spousal social support and mood state of primiparous postpartum women, and the subsequent effect on maternal role identity.* Dissertação de Mestrado, University of Windsor, Windsor, Ontario, Canada.

Rhinehart, P. C.. (2008). *Self-perception of postpartum depression.* Tese de Doutorado, University of Houston, Houston, United States of America.

Rich-Edwards, J. W., Kleinman, K., Abrams, A., Harlow, B. L., McLaughlin, T. J., Joffe, H., & Gillman, M. W. (2006). Sociodemographic predictors of antenatal and postpartum depressive symptoms among women in a medical group practice. *Journal of Epidemiology and Community Health, 60*, 221-227.

Righetti-Veltema, M., Conne-Perréard, E., Bousquet, A., & Manzano, J. (1998). Risk factors and predictive signs of postpartum depression. *Journal of Affective Disorders, 49*, 167-180.

Ross, L. E., Sellers, E. M., Gilbert Evans, S. E., & Romach, M. K. (2004). Mood changes during pregnancy and the postpartum period: Development of a biopsychosocial model.. *Acta Psychiatrica Scandinavica, 109*, 457-466.

Ross, L. E., Villegas, L., Dennis, C. -L., Bourgeault, I. L., Cairney, J., Grigoriadis, S., ... Yudin, M. H. (2011). Rural residence and risk for perinatal depression: A Canadian pilot study. *Archives of Women's Mental Health, 14*, 175-185.

Rouhi, M., Mohammedalizadeh, C. S., Usefi, H., & Rouhi, N. (2011). Postpartum morbidity and help-seeking behaviours in Iran. *British Journal of Midwifery, 19*(3), 178-184.

Rubertsson, C., Wickberg, B., Gustavsson, P., & Radestad, I. (2005). Depressive symptoms in early pregnancy, two months and one year postpartum-prevalence and psychosocial risk factors in a national Swedish sample. *Archives of Women's Mental Health, 8*, 97-104.

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 post-partum depression? *Revista Brasileira de Saúde Materno Infantil, 9*(2), 207-213.

Russel, E. J.. (2009). *Demographic, reproductive, and psychosocial predictors*

of mood change in the postpartum period. Dissertação de Mestrado, Lakehead University, Thunderbay.

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.

Sarton, C.. (2006). *Postpartum depression: Timing, location of residence, and perceived stress.* Tese de Doutorado, State University of New York, Binghamton.

Sato, Y., Kato, T., & Kakee, N. (2008). A six-month follow-up study of maternal anxiety and depressive symptoms among Japanese. *Journal of Epidemiology, 18*(2), 84-87.

Schmidt, R. M., Wiemann, C. M., Rickert, V. I., & Smith, E. O. (2006). Moderate to severe depressive symptoms among adolescent mothers followed four years postpartum. *Journal of Adolescent Health, 38*, 712-718.

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.

Segre, L. S., O'Hara, M. W., Arndt, S., & Stuart, S. (2007). The prevalence of postpartum depression: The relative significance of three social status indices. *Social Psychiatry and Psychiatric Epidemiology, 42*, 316-321.

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.

Sheeder, J., Kabir, K., & Stafford, B. (2009). Screening for postpartum depression at well-child visits: Is once enough during the first 6 months of life? *Pediatrics, 123*, 982-988.

Silverman, M. E., & Loudon, H. (2010). Antenatal reports of pre-pregnancy abuse is associated with symptoms of depression in the postpartum period. *Archives of Women's Mental Health, 13*, 411-415.

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.

Skazufka, E. T.. (2000). *Prevalência da depressão pós-parto e fatores associados em mães que deram à luz em um hospital universitário do município de São Paulo.* Dissertação de Mestrado, Universidade de São

Paulo, São Paulo.

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.

Sorenson, D. S., & Tschetter, L. (2009). Prevalence of negative birth perception, disaffirmation, perinatal trauma symptoms, and depression among postpartum women. *Perspectives in Psychiatric Care, 46*(1), 14-25.

Stamp, G. E., & Crowther, C. A. (1994). Postnatal depression: A South Australian prospective survey. *Australian & New Zealand Journal of Obstetrics & Gynaecology, 34*(2), 164-167.

Stuart, S., Couser, G. B. A., Schilder, K. B. A., O'Hara, M. W., & Gorman, L. (1998). Postpartum anxiety and depression: Onset and comorbidity in a community sample. *Journal of Nervous and Mental Disease, 186*(7), 420-424.

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.

Sylvén, S. M., Papadopoulos, F. C., Mpazakidis, V., Ekselius, L., Sundström-Poromaa., & Skalkidou, A. (2011). Newborn gender as a predictor of postpartum mood disturbances in a sample of Swedish women. *Archives of Women's Mental Health, 14*, 195-201.

Tamaki, R., Murata, M., & Okano, T. (1997). Risk factors for postpartum depression in Japan. *Psychiatry and Clinical Neurosciences, 51*, 93-98.

Tammentie, T., Tarkka, M. -T., Åstedt-Kurki, P., & Paavilainen, E. (2002). Sociodemographic factors of families related to postnatal depressive symptoms of mothers. *International Journal of Nursing Practice, 8*, 240-246.

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.

Tashakori, A., Shanesaz, A., & Rezapour, A. (2009). Assessment of some potential risk factors of postpartum depression. *Pakistan Journal of Medical Sciences, 25*(2), 261-264.

Teng, H. -W., Hsu, C. -S., Shih, S. -M., Lu, M. -L., Pan, J. -J., & Shen, W. W. (2005). Screening postpartum depression with the Taiwanese version of the Edinburgh Postnatal Depression Scale. *Comprehensive Psychiatry, 46*, 261-265.

Terry, D. J., Mayocchi, L., & Hynes, G. J. (1996). Depressive symptomatology in new mothers: A stress and coping perspective. *Journal of Abnormal Psychology, 105*(2), 220-231.

Thio, I. M., Browne, M. A. O., Coverdale, J. H., & Argyle, N. (2006). Postnatal depressive symptoms go largely untreated: A probability study in urban New Zealand. *Social Psychiatry and Psychiatric Epidemiology, 41*, 814-818.

Thome, M. (2000). Predictors of postpartum depressive symptoms in Icelandic women. *Archives of Women's Mental Health, 3*, 7-14.

Thompson, J. F., Roberts, C. L., Currie, M., & Ellwood, D. A. (2002). Prevalence and persistence of health problems after childbirth: Associations with parity and method of birth. *Birth-Issues in Perinatal Care, 29*(2), 83-94.

Tomasic, M. M.. (2005). *A prospective study of prevalence and risk factors related to postpartum depression.* Dissertação de Mestrado, Memorial University of Newfoundland, St. John's.

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.

van de Pol, G., van Brummen, H. J., Bruinse, H. W., Heintz, A. P. M., & van der Vaart, C. H. (2007). Is there an association between depressive and urinary symptoms during and after pregnancy? *International Urogynecology Journal, 18*, 1409-1415.

Verkerk, G. J. M., Denollet, J., Van Heck, G. L., Van Son, M. J. M., & Pop, V. J. M. (2005). Personality factors as determinants of depression in postpartum women: A prospective 1-year follow-up study. *Psychosomatic Medicine, 67*, 632-637.

Vik, T., Grote, V., Escribano, J., Socha, J., Verduci, E., Fritsch, M., ... European Childhood Obesity Trial Study Group. (2009). Infantile colic, prolonged crying and maternal postnatal depression. *Acta Pædiatrica, 98*, 1344-1348.

Wan, E. Y., Moyer, C. A., Harlow, S. D., Fan, Z., Jie, Y., & Yang, H. (2009). Postpartum depression and traditional postpartum care in China: Role of Zuoyuezi. *International Journal of Gynecology and Obstetrics, 104*, 209-213.

Wang, L., Wu, T., Anderson, J. L., & Florence, J. E. (2011). Prevalence and risk factors of maternal depression during the first three years of child rearing. *Journal of Women's Health, 20*(5), 711-718.

Wang, S. -Y., & Chen, C. -H. (2006). Psychosocial health of Taiwanese postnatal husbands and wives. *Journal of Psychosomatic Research, 60*, 303-307.

Wang, S. -Y., Jiang, X. -Y., Jan, W. -C., & Chen, C. -H. (2003). A comparative study of postnatal depression and its predictors in Taiwan and mainland China. *American Journal of Obstetrics and Gynecology, 189*, 1407-1412.

Warner, R., Appleby, L., Whitton, A., & Faragher, B. (1996). Demographic and obstetric risk factors for postnatal psychiatric morbidity. *British Journal of Psychiatry, 168*, 607-611.

Watanabe, M., Wada, K., Sakata, Y., Aratake, Y., Kato, N., Ohta, H., & Tanaka, K. (2008). Maternity blues as predictor of postpartum depression: A prospective cohort study among Japanese women. *Journal of Psychosomatic Obstetrics & Gynecology, 29*(3), 206-212.

Watt, S., Sword, W., Krueger, P., & Sheehan, D. (2002). A cross-sectional study of early identification of postpartum depression: Implications for primary care providers from The Ontario Mother & Infant Survey. *BMC Family Practice, 3*, 5.

Wei, G., Greaver, L. B., Marson, S. M., Herndon, C. H., Rogers, J., & Robeson Healthcare. (2008). Postpartum depression: Racial differences and ethnic disparities in a tri-racial and bi-ethnic population. *Maternal and Child Health Journal, 12*, 699-707.

Wenzel, A., Haugen, E. N., Jackson, L. C., & Brendle, J. R. (2005). Anxiety symptoms and disorders at eight weeks postpartum. *Anxiety Disorders, 19*, 295-311.

Westwood, B. A.. (2006). *Comparing prevalence rates of depressive symptoms in postpartum and nonpostpartum samples in a low-income community.* Tese de Doutorado, University of Stellenbosch, Stellenbosch.

White, T., Matthey, S., Boyd, K., & Barnett, B. (2006). Postnatal depression and post-traumatic stress after childbirth: Prevalence, course and co-occurrence. *Journal of Reproductive and Infant Psychology, 24*(2), 107-120.

Wickberg, B., & Hwang, C. P. (1996). Counselling of postnatal depression: A controlled study on a population based Swedish sample. *Journal of Affective Disorders, 39*, 209-216.

Wissart, J., Parshad, O., & Kulkarni, S. (2005). Prevalence of pre- and postpartum depression in Jamaican women. *BMC Pregnancy and Childbirth, 5*, 15.

Wojcicki, J. M., Holbrook, K., Lustig, R. H., Epel, E., Caughey, A. B., Muñoz, R. F., ... Heyman, M. B. (2011). Chronic maternal depression is associated with reduced weight gain in Latino infants from birth to 2 years of age. *Plos One, 6*(2), 0.

Xie, R. -h., He, G., Koszycki, D., Walker, M., & Wen, S. W. (2009). Fetal sex, social support, and postpartum depression. *La Revue canadienne de psychiatrie, 54*(11), 750-756.

Xie, R. -h., He, G., Liu, A., Bradwejn, J., Walker, M., & Wen, S. W. (2007). Fetal gender and postpartum depression in a cohort of Chinese women. *Social Science & Medicine, 65*, 680-684.

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.

Yelland, J., Sutherland, G., & Brown, S. J. (2010). Postpartum anxiety, depression and social health: findings from a population-based survey of Australian women. *BMC Public Health, 10*, 771.

Yonkers, K. A., Ramin, S. M., Rush, A. J., Navarrete, C. A., Carmody, T., March, D., ... Leveno, K. J. (2001). Onset and persistence of postpartum depression in an inner-city maternal health clinic system. *American Journal*

of Psychiatry, 158(11), 1856-1863.

Yoshida, K., Marks, M. N., Kibe, N., Kumar, R., Nakano, H., & Tashiro, N. (1997). Postnatal depression in Japanese women who have given birth in England. *Journal of Affective Disorders, 43*, 69-77.

Zaconeta, A. M., da Motta, L. D. C., & França, P. S. (2004). Depresión postparto: Prevalencia de test de rastreo positivo en puérperas del hospital universitario de Brasilia, Brasil. *Revista Chilena de Obstetricia y Ginecología, 69*(3), 209-213.

Zaers, S., Waschke, M., & Ehlert, U. (2008). Depressive symptoms and symptoms of post-traumatic stress disorder in women after childbirth. *Journal of Psychosomatic Obstetrics & Gynecology, 29*(1), 61-71.

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.

Apêndice 3 – Prevalência média de DPP por cidade

Tabela 22 – Prevalência média de DPP por cidade nos estudos incluídos (continua)

País	Cidade	Prevalência média (%)	IC – limite inferior (%)	IC – limite superior (%)
Africa do Sul	Cidade do Cabo	34,69	27,00	42,39
Africa do Sul	Kylemore	44,44	28,21	60,68
Africa do Sul	Soweto	24,15	21,55	26,76
Alemanha	Stuttgart	3,63	2,31	4,95
Alemanha	Trier	12,00	2,99	21,01
Arabia Saudita	Al Ahsa	2,63	0,36	4,91
Armenia	Yerevan	14,42	11,12	17,71
Australia	Adelaide	14,36	10,87	17,86
Australia	Brisbane	8,42	6,93	9,92
Australia	Campbelltown	20,07	15,56	24,57
Australia	Canberra	10,55	8,83	12,27
Australia	Melbourne	8,35	6,90	9,80
Australia	Queensland	15,69	8,63	22,74
Australia	Randwick	10,78	9,09	12,48
Australia	Sidney	7,26	4,84	9,68
Australia	Wodonga	8,51	6,83	10,19
Austria	Viena	0,00	0,00	0,00
Bangladesh	Dhaka	11,54	0,00	23,82
Bangladesh	Matlab	21,97	17,60	26,33
Brasil	Belo Horizonte	26,94	21,38	32,49
Brasil	Brasilia	14,48	10,89	18,08
Brasil	Caxias do Sul	8,91	3,35	14,47
Brasil	Embu Guacu	39,13	25,03	53,23
Brasil	Joao Pessoa	33,33	23,25	43,41
Brasil	Juiz de Fora	14,69	9,47	19,90
Brasil	Pelotas	13,15	11,58	14,72
Brasil	Porto Alegre	20,66	15,84	25,48
Brasil	Recife	20,01	18,50	21,51
Brasil	Rio de Janeiro	24,55	18,92	30,19
Brasil	Sao Paulo	24,92	22,15	27,69
Brasil	Uberaba	34,29	18,56	50,01
Brasil	Vitoria	39,38	33,78	44,99
Canada	Calgary	11,05	9,41	12,69
Canada	Montreal	16,67	8,40	24,94
Canada	Ontario	8,23	7,22	9,24
Canada	Ottawa	4,39	2,64	6,14

Tabela 22 – Prevalência média de DPP por cidade nos estudos incluídos (continuação)

País	Cidade	Prevalência média (%)	IC – limite inferior (%)	IC – limite superior (%)
Canada	St Johns	3,13	0,00	7,39
Canada	Toronto	18,13	12,78	23,47
Canada	Vancouver	22,99	19,43	26,56
Canada	Victoria	13,48	6,39	20,58
Canada	Windsor	33,33	13,17	53,50
Chile	Arica	44,79	37,15	52,42
Chile	Santiago	28,65	25,78	31,51
Chile	Temuco	50,68	39,22	62,15
China	Beijing	15,44	11,88	19,01
China	Changsha	18,59	15,95	21,22
China	Chengdu	4,68	2,44	6,92
China	Fuzhou	25,00	18,94	31,06
China	Guangzhou	13,85	7,91	19,78
China	Kaohsiung	44,40	40,30	48,51
China	Pingtung	59,60	49,93	69,26
China	Taipei	19,75	15,95	23,55
Colombia	Bucaramanga	4,20	1,87	6,52
Colombia	Cali	57,14	40,75	73,54
Coreia do Sul	Seul	19,52	15,43	23,61
Dinamarca	Aarhus	5,52	4,89	6,15
E. A. U.	Abu Dhabi	44,19	33,69	54,68
E. A. U.	Dubai	13,68	6,77	20,60
E. A. U.	Sharjah	10,22	5,15	15,29
Egito	Alexandria	25,32	18,54	32,10
Espanha	Barcelona	8,53	7,16	9,90
Espanha	Costa	14,47	6,56	22,38
Espanha	Toledo	13,16	8,35	17,96
Espanha	Valencia	14,18	11,61	16,75
EUA	Albuquerque	15,95	13,31	18,58
EUA	Asheville	35,44	24,89	45,99
EUA	Baltimore	49,15	41,79	56,52
EUA	Boisi	40,00	33,21	46,79
EUA	Boston	13,56	12,27	14,84
EUA	Buffalo	21,78	13,73	29,83
EUA	Charleston	11,63	4,85	18,40
EUA	Dallas	6,48	4,78	8,19
EUA	Galveston	36,77	32,92	40,63
EUA	Grand Fork	12,24	6,95	17,54
EUA	Hartford	18,67	12,43	24,90

Tabela 22 – Prevalência média de DPP por cidade nos estudos incluídos (continuação)

País	Cidade	Prevalência média (%)	IC – limite inferior (%)	IC – limite superior (%)
EUA	Houston	26,87	16,25	37,48
EUA	Iowa	10,78	6,14	15,41
EUA	Keene	10,64	4,41	16,87
EUA	Manchester	11,79	10,63	12,94
EUA	Minneapolis	20,47	16,80	24,15
EUA	New Hampshire	15,94	7,30	24,58
EUA	Nova York	14,23	11,88	16,59
EUA	Philadelphia	32,00	19,07	44,93
EUA	Pittsburgh	11,62	10,11	13,13
EUA	Rochester	45,45	35,46	55,45
EUA	San Francisco	18,92	16,04	21,80
EUA	Tucson	24,72	15,76	33,68
EUA	Tulsa	38,76	32,15	45,36
EUA	Wyoming	21,05	10,47	31,64
Finlandia	Helsinki	22,41	11,68	33,15
Franca	Bordeaux	15,00	-0,65	30,65
Franca	Paris	11,76	-3,55	27,08
Grecia	Atenas	21,90	18,90	24,91
Grecia	Heraklion	13,99	11,03	16,94
Guiana	Georgetown	50,00	40,48	59,52
Holanda	Eindhoven	9,22	5,90	12,53
Holanda	Rotterdam	8,01	7,26	8,77
Hong Kong	Hong Kong	14,61	12,87	16,35
India	Bangalore	29,55	21,76	37,33
India	Calcutta	35,45	26,51	44,39
India	Vellore	19,78	15,66	23,90
Ira	Kerman	45,75	40,87	50,63
Ira	Tabriz	36,46	33,65	39,27
Ira	Yazd	55,83	46,95	64,72
Irlanda	Dublin	24,30	21,95	26,65
Israel	Acre	16,25	14,75	17,75
Israel	Afula	9,93	7,48	12,38
Israel	Jerusalem	11,48	9,63	13,34
Italia	Ancona	55,71	44,08	67,35
Italia	Cagliari	8,00	2,68	13,32
Italia	Emilia Romagna	13,25	8,90	17,59
Italia	Florenca	13,43	7,74	19,12
Italia	Pisa	5,97	4,51	7,42
Italia	Roma	27,08	18,19	35,97

Tabela 22 – Prevalência média de DPP por cidade nos estudos incluídos (continuação)

País	Cidade	Prevalência média (%)	IC – limite inferior (%)	IC – limite superior (%)
Italia	Veneza	13,13	9,29	16,97
Jamaica	Mona	34,25	23,36	45,13
Japao	Fukuoka	13,64	6,47	20,81
Japao	Hamamatsu	14,81	12,13	17,49
Japao	Kanagawa	12,77	8,50	17,03
Japao	Neyagawa	13,75	11,32	16,18
Japao	Suzuka	18,18	15,16	21,20
Japao	Tokyo	13,99	12,12	15,87
Jordao	Irbid	22,10	17,77	26,42
Libano	Beirut	15,94	10,96	20,93
Libano	Bekaa Valley	25,93	19,68	32,17
Malaisia	Kuala Lumpur	16,67	12,01	21,32
Malaisia	Kubang Kerian	27,30	22,20	32,41
Malta	Tal Pieta	8,73	5,08	12,39
Marrocos	Casablanca	18,75	12,37	25,13
Mexico	Durango	28,82	23,03	34,62
Mongolia	Ulaanbaatar	9,10	7,35	10,84
Nepal	Kathmandu	5,00	0,73	9,27
Nepal	Patan	4,93	2,87	6,99
Nigeria	Ilesa	13,66	12,15	15,17
Noruega	Arendal	0,92	0,00	2,71
Noruega	Oslo	7,61	3,91	11,32
Noruega	Stavanger	16,47	15,10	17,83
Nova Zelandia	Auckland	15,86	13,96	17,76
Paquistao	Karachi	17,38	15,11	19,65
Paquistao	Rawalpindi	29,57	26,17	32,96
Peru	Lima	22,16	17,89	26,43
Portugal	Coimbra	4,86	2,88	6,84
Portugal	Porto	18,27	15,64	20,90
Singapura	Singapura	4,18	2,41	5,96
Suecia	Estocolmo	10,53	8,97	12,09
Suecia	Gutemburgo	3,44	2,57	4,32
Suecia	Linköping	5,62	4,12	7,11
Suica	Genova	8,77	6,45	11,09
Suica	Zurique	4,10	2,84	5,36
Tailandia	Bangkok	16,72	13,76	19,68
Taiwan	Taipei	73,74	65,07	82,41
Turquia	Bursa	9,76	7,83	11,68
Turquia	Edirne	40,45	33,24	47,66

Tabela 22 – Prevalência média de DPP por cidade nos estudos incluídos (conclusão)

País	Cidade	Prevalência média (%)	IC – limite inferior (%)	IC – limite superior (%)
Turquia	Erzurum	18,15	15,05	21,25
Turquia	Istambul	17,18	11,39	22,97
Turquia	Izmir	16,80	10,25	23,35
Turquia	Konya	6,29	3,55	9,03
Turquia	Malatya	29,91	24,05	35,78
Turquia	Manisa	28,33	23,17	33,49
Turquia	Mersin	36,59	31,37	41,80
Uganda	Kabarole	43,07	36,24	49,90
Uganda	Kampala	6,12	4,06	8,17
UK	Avon	10,20	9,65	10,75
UK	Birmingham	14,63	9,80	19,47
UK	Cambridge	15,26	14,26	16,26
UK	Coventry	27,53	20,97	34,09
UK	Londres	13,07	11,75	14,39
UK	Manchester	24,50	18,54	30,46
UK	Norwich	24,69	19,27	30,11
UK	Penrith	8,94	6,23	11,65
UK	Sheffield	15,86	11,11	20,60
UK	Southampton	42,01	38,54	45,48
UK	Stoke on Trent	41,38	35,04	47,72
Venezuela	Maracaibo	22,00	13,88	30,12
Vietna	Hanam	15,00	8,00	22,00
Vietna	Hanoi	13,85	5,45	22,24
Vietna	Ho Chi Minh	32,81	28,72	36,90
Zimbabwe	Harare	30,48	24,25	36,70