



King's Research Portal

DOI:

[10.1016/j.jad.2017.07.046](https://doi.org/10.1016/j.jad.2017.07.046)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Baron, E., Bass, J., Murray, S. M., Schneider, M., & Lund, C. (2017). A systematic review of growth curve mixture modelling literature investigating trajectories of perinatal depressive symptoms and associated risk factors. *Journal of Affective Disorders*, 223, 194-208. <https://doi.org/10.1016/j.jad.2017.07.046>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

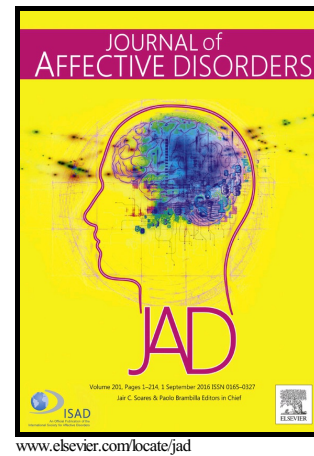
Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Author's Accepted Manuscript

A systematic review of growth curve mixture modelling literature investigating trajectories of perinatal depressive symptoms and associated risk factors

Emily Baron, Judith Bass, Sarah M Murray, Marguerite Schneider, Crick Lund



PII: S0165-0327(17)30347-6
DOI: <http://dx.doi.org/10.1016/j.jad.2017.07.046>
Reference: JAD9112

To appear in: *Journal of Affective Disorders*

Received date: 15 February 2017
Revised date: 20 July 2017
Accepted date: 24 July 2017

Cite this article as: Emily Baron, Judith Bass, Sarah M Murray, Marguerite Schneider and Crick Lund, A systematic review of growth curve mixture modelling literature investigating trajectories of perinatal depressive symptom and associated risk factors, *Journal of Affective Disorders* <http://dx.doi.org/10.1016/j.jad.2017.07.046>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and a review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain

A systematic review of growth curve mixture modelling literature investigating trajectories of perinatal depressive symptoms and associated risk factors

Emily Baron^{a*}, Dr Judith Bass^b, Dr Sarah M Murray^c, Dr Marguerite Schneider^d, Prof Crick Lund^e

^aAlan J Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, South Africa

^bDepartment of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, US

^cDepartment of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, US

^dAlan J Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, South Africa

^eAlan J Centre for Public Mental Health, Department of Psychiatry and Mental Health, University of Cape Town, South Africa; Center for Global Mental Health, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

*Corresponding author. Postal address: Office 32, Building B, 46 Sawkins Road, 7700 Rondebosch, Cape Town, South Africa. Tel. (+27) 21 650 1095. emily.baron@uct.ac.za

Abstract

Background

The aim of this study was to review the growth curve mixture modelling (GCMM) literature investigating trajectories of perinatal maternal depressive symptoms and associated risk factors. .

Methods

A systematic search of peer-reviewed articles published until November 2015 was conducted in seven databases. Articles using GCMM to identify trajectories of perinatal depressive symptoms were considered. Symptoms had to be assessed at least three times, anytime from pregnancy to two years postpartum (PROSPERO; 2016:CRD42016032600).

Results

Eleven studies met inclusion criteria. All reported a low risk trajectory, characterised by stable low depressive symptoms throughout the perinatal period. A stable moderate-high or high symptom trajectory was reported in eight of 11 studies, suggesting a high-risk group with persistent depressive

symptoms. Six studies also reported transient trajectories, with either increasing, decreasing or episodic depressive symptoms. None of the demographic, personality or clinical characteristics investigated systematically differentiated groups of women with different symptom trajectories, within or across studies. Thus, it is difficult to differentiate women at high or low risk of specific perinatal depression trajectories.

Limitations

A meta-analysis was not possible. The studies' settings and inclusion criteria limit the generalisability of the findings to low-risk, middle- to high-income women.

Conclusions

Relatively similar trajectories of perinatal depressive symptoms were identified across studies. Evidence on factors differentiating women assigned to different trajectories was inconsistent. Research with larger samples and in more diverse settings is needed to inform services and policies on how and when to effectively identify subgroups of women at high risk of perinatal depression.

Keywords

trajectory; depressive symptoms; pregnancy; postpartum; risk factors; growth mixture modelling

Introduction

The high prevalence of perinatal maternal depression is a well-documented global phenomenon. In high-income countries, common mental disorders are reported on average by 10% and 13% of pregnant and postnatal women, respectively (O'hara and Swain, 1996). A recent review of the literature suggests that in low- and middle-income countries (LMICs), approximately 16% of women experience antenatal depression and 20% postnatal depression (Fisher et al., 2012). Perinatal depression contributes to the global burden of disease, both directly, given that depression accounts for over 40% of disability adjusted life years caused by mental disorders (Whiteford et al., 2013), and indirectly, through associations with suicidal behaviour (Rahman et al., 2013; World Health Organization, 2008). Untreated perinatal depression also has detrimental effects on birth outcomes (Lusskin et al., 2007), as well as on children's health and socio-emotional development (Hayes and Sharif, 2009; Wachs et al., 2009).

Effective prevention of perinatal depression and associated poor maternal and child health outcomes requires understanding when women are most at risk and what factors are associated with the disorder's onset, severity and chronicity. To achieve this aim, longitudinal mixed-effects and latent growth curve models are commonly used to assess the progression of depressive symptoms during the perinatal period. Though these methods allow for individual variability, they assess the average pattern of change in symptoms over time and assume individuals belong to the same underlying population, represented by a single growth curve. Yet, existing evidence suggests heterogeneity in time of onset and progression of perinatal depressive symptoms. While some studies have identified antenatal depression as a major risk factor for postpartum depression (Robertson et al., 2004), others have shown a natural decline in depressive symptoms during pregnancy and the postpartum period, or symptoms developing only after giving birth (Gavin et al., 2005; Stowe and Nemeroff, 1995). These methods' assumptions therefore risk oversimplifying the complex process involved in the development and progression of perinatal depression.

An emerging, alternative method which addresses this limitation is a person-centred, latent class approach, which allows researchers to identify and describe underlying subgroups or classes within a population, based on different patterns of symptom change, or trajectories (Leiby, 2012; Ram and Grimm, 2009). Within this approach, latent growth curve models, often referred to as growth curve mixture models (GCMM) (Leiby, 2012), are a flexible subtype of models that do not require the researcher to predefine the number of trajectories being identified. This is an advantage, particularly given that predefining the number of trajectories is likely to increase the likelihood of poor model fit (Ram and Grimm, 2009).

When GCMM is used, several models are generated. In each model, parameters of growth trajectories and inter-individual variation are estimated for each latent class or trajectory. The intercept is the initial level of symptom, and the slope is the rate in change of symptom level over time. In addition, posterior probability estimates in each model indicate the probability that an individual belongs to each trajectory. The optimal model of trajectories is selected using a range of fit statistics, including model fit indices, estimated posterior probabilities, and likelihood ratio tests. Post-hoc tests, such as multinomial regressions, are often performed to compare baseline characteristics or specific

outcomes of individuals classified into the different trajectories. These analyses can also help assess whether the latent trajectories identified make pragmatic sense.

GCMM has been used in the analysis of mental health-related outcomes, including binge drinking (Tucker et al., 2003), psychosocial wellbeing (Zammit et al., 2012), and anxiety and mood disorders (Nandi et al., 2009). It has also increasingly been used to explore trajectories of depressive symptoms among women during the perinatal period (Kuo et al., 2014; Mora et al., 2009; Sutter-Dallay et al., 2012). To our knowledge, the findings of these studies have not yet been systematically synthesised. An overview of these studies would help identify how and when trajectories of perinatal depressive symptoms differ, and whether this is consistent across populations. Findings could also have implications for identifying optimal timing of screening for perinatal depression and for the content or focus of screening required to differentiate women with chronic symptoms from those with transient levels. Therefore, the aim of this study was to systematically review the growth curve mixture modelling literature investigating the trajectories and associated risk factors of maternal depressive symptoms during the perinatal period.

Methods

The review protocol was registered with PROSPERO (2016:CRD42016032600) and was developed and reported according to the MOOSE guidelines (Stroup et al., 2000).

Search strategy

A systematic search of peer-reviewed articles was conducted in the following seven databases: MEDLINE, Embase (via Scopus), the Cochrane Library (Cochrane Database of Systematic Reviews), Web of Science, PsychINFO, CINAHL and Africa Wide. A range of keywords and database subject headings were used to capture four key concepts combined using the Boolean term 'AND': (1) depressive symptoms during the perinatal period, (2) perinatal depressive symptoms trajectories, (3) factors associated with symptom trajectories, and (4) longitudinal designs using latent variable modelling approaches (see Table 1).

There were no publication date or language restrictions. Articles considered for review were those which reported the use of GCMM to identify trajectories of perinatal depressive symptoms and associated risk factors. These could be based on primary data from cohort studies, or based on secondary data from randomised controlled trials (RCTs), if no statistical differences in depressive symptoms were reported between the control and intervention arms. If a study reported only on trajectories and outcomes, rather than risk factors, results related to trajectories were still included in this review.

Depressive symptoms were defined as any sub-clinical (distress) or clinical depressive symptomatology, assessed on a longitudinal scale, either using a validated screening tool or a diagnostic assessment based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Statistical Classification of Diseases and Related Health Problems (ICD) criteria. However, studies were excluded if depressive symptoms were investigated in the context of a comorbid primary mental disorder (e.g. anxiety, bipolar, schizophrenia, psychosis). This criterion was put in place to exclude studies investigating perinatal women without a primary presentation of depression. Psychological morbidities would likely influence the severity and course of depressive symptoms, and it would be difficult to distinguish actual change in depressive symptoms over time from change in depressive symptom as a function of the comorbid primary diagnosis. Trajectories were conceptualised as the change in depressive symptoms during the perinatal period, defined as pregnancy and up to two years after birth. Assessments could be conducted during pregnancy, during the postpartum period, or during pregnancy and the postpartum period. A minimum of four assessment points has been recommended when performing GCMM, as fewer assessments limit the functions that can be modelled, and therefore the type and number of trajectories that can be generated (Berlin et al., 2014; Johnson et al., 2007). Given the limited number of studies generated in a pilot search with this criterion, the authors decided to include studies with a minimum of three depression symptom assessments. In cases where additional assessments were conducted outside of the two year postpartum period, the study was excluded as this would shape the overall estimates of symptom trajectory. Risk factors were defined as any clinical, socio-demographic or socio-economic factors measured during the first two assessments.

Abstracts generated from the search were recorded and transferred to Endnote, where duplicates were identified and deleted. After irrelevant titles were excluded by one reviewer (ECB), two pairs of two independent reviewers (ECB and JB, or ECB and SM) screened the remaining abstracts and full texts. Articles selected by both reviewers were automatically included in the next review step. When reviewers did not agree, a third reviewer (JB or SM) made a final decision. The number of articles selected at each step of the review process were captured, as well as the main reason for exclusion (Figure 1). A single reviewer (ECB) extracted the following data from each full-text article: study setting, participant characteristics, the number and timing of assessments, instruments used to measure depressive symptoms, statistical analyses performed, and covariates included in the analyses. The number and nature of trajectories identified, as well as any risk factors associated with these trajectories, were also recorded. Identification of studies' biases and limitations were based on the Newcastle-Ottawa quality assessment scale for cohort studies (Wells et al., 2012). This section covered information related to the size and representativeness of the sample, the number of assessment points, attrition rate and statistical methods used to deal with missing data, as well as the number and nature of considerations taken to select the optimal model from the GCMM. Based on these criteria, a score out of 12 was generated to rate the quality of each study; each study was then classified as either poor (+), average (++) or good (+++).

The search was conducted in November 2015, the screening and review processes were finalised in January 2016, and data extraction was completed in February 2016.

Data analysis

Given the lack of model parameters (i.e. slopes, intercepts, and variances of the latent trajectories) reported in the identified studies, a meta-analysis could not be conducted. Instead, a qualitative synthesis of findings are presented, highlighting the most common trajectories and risk factors reported across the studies identified. To help visualise and compare the different trajectories identified across the studies in the review, and because different instruments were used to measure depressive symptoms, the average depression score at each time point for each trajectory was standardised in relation to the severity cut-off of the scale indicated by authors. Standardization was completed by transforming the difference between the average depressive score at each time of

assessment and the severity cut-off score of the scale indicated by the authors into a percentage of the overall score of the scale used. A positive percentage therefore indicated a score above the cut-off and a negative percentage indicated a score below the cut-off. When a severity cut-off was not indicated by authors, the recommended cut-off for the instrument was used.

Results

Study selection

The search terms identified 5388 articles, of which 789 abstracts were screened for eligibility (Figure 1). In total, 55 articles were selected for full-text review; 19 of these did not clearly state the method of analysis used in the reviewed abstract. The majority of articles (95%) were written in English. The three non-English full-text articles (two in Korean and one in Japanese) were excluded after a translation of the analysis section revealed that a growth curve modelling approach was not used. One article did not strictly fit the two-year postpartum period criterion (Mora et al., 2009), with a final assessment conducted at 25 months postpartum. As this criterion was used to exclude articles with assessments occurring later in childhood or during adolescence, and given the article's relevance to the topic of the review, the authors opted to include this article.

Eleven articles were identified for final review (Table 2). All studies reported in these articles were conducted in high-income countries: two in Taiwan (Kuo et al., 2014; Kuo et al., 2012), one in France (Sutter-Dallay et al., 2012), one in Finland (Vänskä et al., 2011), and the remaining seven in the United States (Christensen et al., 2011; Glasheen et al., 2013; Lee et al., 2014; Marcus et al., 2011; Mora et al., 2009; Parade et al., 2014; Ramos-Marcuse et al., 2010). All studies were published between 2009 and 2014. Ten of the 11 studies investigated perinatal depressive symptoms among adult women; the remaining study was with adolescent women in the postpartum period (Ramos-Marcuse et al., 2010). The majority of studies (n=6) investigated risk factors only in relation to trajectories, three studies investigated risk factors and outcomes simultaneously (Christensen et al., 2011; Glasheen et al., 2013; Vänskä et al., 2011), and two studies investigated outcomes associated with trajectories (Marcus et al., 2011; Ramos-Marcuse et al., 2010).

[Include Figure 1 here]

Study characteristics and quality

Tables 2 and 3 provide an overview of the recruitment, assessment and statistical methods employed across the different studies identified in this review. The study's quality ratings are also reported in Table 3. Three studies were considered of good quality (score 8-12) (Glasheen et al., 2013; Kuo et al., 2014; Mora et al., 2009), five as average (score 5-7) (Kuo et al., 2012; Lee et al., 2014; Ramos-Marcuse et al., 2010; Sutter-Dallay et al., 2012; Vänskä et al., 2011), and three as relatively poor (score 1-4) (Christensen et al., 2011; Marcus et al., 2011; Parade et al., 2014). The main reason for the low score for the Marcus et al (2011) study was poor statistical methods or reporting, whereas poor design (such as sample size and attrition) was the main reason for the low scores in the Parade et al (2014) and Christensen et al (2011) studies.

Sample sizes varied widely across studies (range: $n=98$ (Parade et al., 2014) to $n=1735$ (Mora et al., 2009)), with the majority reporting samples between 120 and 600. A small sample was acknowledged as a limitation by the authors in several of the studies (Kuo et al., 2014; Parade et al., 2014; Sutter-Dallay et al., 2012). Five of the 11 studies reported analyses on secondary data (Christensen et al., 2011; Lee et al., 2014; Mora et al., 2009; Ramos-Marcuse et al., 2010; Sutter-Dallay et al., 2012). Of these, three drew data from randomized controlled trials (RCTs). These assessed an intervention for perinatal depression among Hispanic women who reported sub-clinical depressive symptomatology but did not have a diagnosis of major depression (Christensen et al., 2011); a weight loss program among overweight or obese women (Lee et al., 2014); and a parenting and adolescent development promotion program for low-income families qualifying for government grants (Ramos-Marcuse et al., 2010). Two of these studies reported that depressive symptoms were not significantly different between the control and intervention arms at any assessment point (Christensen et al., 2011) (Ramos-Marcuse et al., 2010), and one reported that there were no differences in intercept, slope or curvature between the two arms in the overall symptom trajectory generated through latent growth modelling (Lee et al., 2014). This meant that the intervention did not significantly differentiate the two arms and thus should not have biased the trajectories of depressive symptoms generated by the GCMM..

The majority of studies included in this review excluded women on the basis of other mental health or substance abuse problems (Christensen et al., 2011; Marcus et al., 2011; Sutter-Dallay et al., 2012), chronic illnesses (Kuo et al., 2014; Kuo et al., 2012; Ramos-Marcuse et al., 2010), and whether the pregnancy was planned or unplanned (Kuo et al., 2014; Sutter-Dallay et al., 2012; Vänskä et al., 2011). The limited representativeness of the samples was acknowledged by most authors (Christensen et al., 2011; Kuo et al., 2014; Kuo et al., 2012; Ramos-Marcuse et al., 2010; Sutter-Dallay et al., 2012; Vänskä et al., 2011). Both Christensen et al. (2011) and Marcus et al. (2011) investigated depressive symptoms among women at high risk of depression but without a diagnosis. Three studies were more inclusive, and only restricted the criteria to age (usually above 18 years old) and language (speaking local language) (Glasheen et al., 2013; Mora et al., 2009; Parade et al., 2014). However, the inclusion criteria were unclear in Parade et al.'s study (2014), and the data used in the Glasheen et al.'s study (2013) dated from 1982 and 1985 and may be less representative now.

Seven studies conducted more than the required three assessments of depression for inclusion in the review (range: four (Kuo et al., 2012; Mora et al., 2009; Parade et al., 2014) to eight (Sutter-Dallay et al., 2012)). The length of the trajectories modelled varied extensively. The longest trajectories were reported by Mora et al. (2009), where women were followed from their first trimester through two years postpartum. The shortest study period investigated was 9 weeks in Marcus et al.'s study (2011), and was the only study to be conducted solely during pregnancy. Two studies focused on postpartum depressive symptoms only (Lee et al., 2014; Ramos-Marcuse et al., 2010), while the remaining eight studies investigated depressive symptoms both during pregnancy and the postpartum period (Christensen et al., 2011; Glasheen et al., 2013; Kuo et al., 2014; Kuo et al., 2012; Mora et al., 2009; Parade et al., 2014; Sutter-Dallay et al., 2012; Vänskä et al., 2011).

All studies used validated depressive screening tools to assess depressive symptoms. Tools used included versions of the Centre for Epidemiological Scale – Depression (CES-D-20; (Radloff, 1977); n=4), the Beck Depression Inventory (BDI; (Beck et al., 1961), n=4), or the Edinburgh Postnatal Depression Scale (EPDS; (Cox et al., 1987), n=3).

GCMM was conducted in Mplus (Muthén and Muthén) or SAS (Jones et al., 2001; SAS Institute Inc, 2006). The majority of studies reported conducting growth mixture modelling (Christensen et al., 2011;

Glasheen et al., 2013; Kuo et al., 2012; Marcus et al., 2011; Mora et al., 2009; Vänskä et al., 2011), which allows for individual variability within each trajectory. Other studies used latent class growth modelling (Lee et al., 2014), group-based trajectory modelling (Kuo et al., 2014), semi parametric mixture models (Ramos-Marcuse et al., 2010; Sutter-Dallay et al., 2012) or unconditional latent class growth analysis (Parade et al., 2014). Nearly all studies reported using a combination of information criteria fit indices and estimated posterior probabilities to identify the optimal model (Christensen et al., 2011; Glasheen et al., 2013; Kuo et al., 2014; Lee et al., 2014; Mora et al., 2009; Ramos-Marcuse et al., 2010; Sutter-Dallay et al., 2012; Vänskä et al., 2011). Likelihood ratio tests were also utilised in some of the studies (Christensen et al., 2011; Glasheen et al., 2013; Mora et al., 2009; Parade et al., 2014). Two studies reported using the Bayesian information criterion (BIC) only (Kuo et al., 2012; Marcus et al., 2011). However, all studies also applied some level of theoretical interpretability to identifying the optimal model, especially when fit statistics indicated that multiple models fit equally well.

Attrition rates, which were reported for eight of 11 studies, are indicated in Table 4 (Glasheen et al., 2013; Kuo et al., 2014; Kuo et al., 2012; Lee et al., 2014; Marcus et al., 2011; Mora et al., 2009; Ramos-Marcuse et al., 2010; Vänskä et al., 2011). The other three studies provided no information on the average number of assessments conducted per participant, or the number of participants assessed at each time point (Christensen et al., 2011; Parade et al., 2014; Sutter-Dallay et al., 2012). Attrition or average number of assessments per participant were reported differently across the studies: four studies reported that at least 70% of the sample received all assessments (Glasheen et al., 2013; Kuo et al., 2014; Kuo et al., 2012; Ramos-Marcuse et al., 2010), one study reported completion of all assessments by 66% of the sample (Vänskä et al., 2011), and one reported that only half completed all assessments (Mora et al., 2009). Marcus et al. (2011) did not provide information on the proportion of the sample completing all assessments, but did indicate that 91% of the sample completed at least two out of three assessments. Similarly, Lee et al. (2014) reported an average of 2.5 and 1.9 assessments completed by participants in the two RCTs used in their analyses. The authors of this study did acknowledge attrition as a limitation, but argued that the expectation maximisation algorithm employed through Mplus dealt with missing data limited attrition bias.

Out of the eight studies reporting attrition, two did not provide any information on how attrition was dealt with (Glasheen et al., 2013; Ramos-Marcuse et al., 2010). One study imputed multiple missing depression measures using PROC MI in SAS (Marcus et al., 2011). The others used full-information maximum likelihood estimation methods that are robust to data missing at random, either in SAS (Kuo et al., 2014; Kuo et al., 2012) or Mplus (Lee et al., 2014; Mora et al., 2009; Vänskä et al., 2011). These estimations allow all available depression score data of an individual to be used in estimation and thus is an optimal estimator for GCM (Little et al., 2014).

Number and shape of trajectories

The number and shape of trajectories reported in each study are summarised in Table 4. The figures illustrating standardised depression scores over time for all trajectories for each study are also presented in Figures 2a and 2b. The horizontal full line in the figures indicates the severity cut-off, while the vertical full line marks the time of birth. Where studies report secondary data as part of an RCT (Christensen et al., 2011; Lee et al., 2014; Ramos-Marcuse et al., 2010), the timing of the intervention in relation to the assessments is indicated with a horizontal arrow. In the case of very short trajectories (Kuo et al., 2014; Kuo et al., 2012; Marcus et al., 2011), the time axis was adjusted (Figure 2b). For all other studies, the time scale ranged from the start of pregnancy to 25 months postpartum (34 months after the start of pregnancy) (Figure 2a).

The severity cut-off on the instrument measuring depressive symptoms was not indicated in three studies (Glasheen et al., 2013; Marcus et al., 2011; Parade et al., 2014). For this reason, a cut-off of 16 was used for the CES-D-20 in Glasheen et al. (2013), as recommended by Weissman et al. (1977). In Parade et al. (2014), where five items were dropped from the original CES-D-20 scale, the cut-off was readjusted from 16 to 12. Finally, the recommended cut-off of 16 for the BDI during pregnancy (Holcomb Jr et al., 1996) was used for standardizing the Marcus et al. (2011) findings. For comparison purposes, this cut-off was also used to calculate standardised BDI scores in Ramos-Marcuse et al (2010)'s study, instead of the cut-off of 9 suggested and used by the authors.

The number of trajectories identified and reported ranged from two to five, with the most common number of trajectories being 3 (Christensen et al., 2011; Kuo et al., 2014; Lee et al., 2014; Marcus et

al., 2011; Ramos-Marcuse et al., 2010). Five studies reported that all identified trajectories were relatively stable over time, with average depression scores remaining either above or below the given symptom severity cut-point (Glasheen et al., 2013; Kuo et al., 2014; Kuo et al., 2012; Marcus et al., 2011; Ramos-Marcuse et al., 2010). The remaining six studies reported a combination of stable and non-stable trajectories over time (Christensen et al., 2011; Lee et al., 2014; Mora et al., 2009; Parade et al., 2014; Sutter-Dallay et al., 2012; Vänskä et al., 2011), also described as transient in Mora et al. (2009).

All studies reported either a low and/or a moderate-low stable symptom trajectory, and five studies reported both (Kuo et al., 2014; Kuo et al., 2012; Marcus et al., 2011; Ramos-Marcuse et al., 2010; Vänskä et al., 2011). The low stable symptom trajectory is characterised by very low symptoms levels throughout the perinatal period, while the moderate-low stable trajectory is characterised by a higher level of symptoms that approaches but remains under the severity cut-off. Of the studies reporting either a low or a moderate-low stable trajectory (n=6), this trajectory represented the majority of the sample in all studies (range 71.0% (Mora et al., 2009) to 82.5% (Lee et al., 2014)), but one (Glasheen et al., 2013), where this trajectory represented a minority of the total sample (16.5%). The trajectory's sample size was not reported in Parade et al. (2014). Of the studies which reported both types of trajectories, the moderate-low stable trajectory tended to have a higher sample size than the low stable trajectory (Kuo et al., 2014; Kuo et al., 2012; Marcus et al., 2011; Ramos-Marcuse et al., 2010). An exception was the Vänskä et al. (2011) study, in which 75.5% of women were classified in the stable low trajectory, while only 8.7% were classified in the moderate-low stable trajectory.

A stable moderate-high and/or high symptom trajectory group was also reported by eight studies. Three studies reported identifying a trajectory with very high stable symptoms that represented a minority of the sample in all three studies: 8.3% of the sample in Kuo et al. (2012), 7% in Mora et al. (2009), and 3% in Sutter-Dallay et al. (2012). Seven studies, including Kuo et al. (2012) and Sutter-Dallay et al. (2012) reported a stable moderate-high symptom group, in which symptom levels hovered close to but above the severity cut-off, suggesting persistent but relatively less severe symptoms of depression. This trajectory represented the majority of the sample (83.5%) in Glasheen et al. (2013), but the minority of the sample in the remainder of the studies (range 4.2% (Vänskä et al.,

2011) to 27.3% (Kuo et al., 2014)). In Vänskä et al. (2011), this trajectory exhibited substantial heterogeneity.

Six studies reported transient trajectories that can be grouped into three categories: increasing, decreasing or episodic. Four reported a decreasing trajectory, of which three were characterised by high symptoms during pregnancy, followed by a decline to low or mild levels in the postpartum period (Christensen et al., 2011; Mora et al., 2009; Vänskä et al., 2011). The decreasing trajectory reported by Lee et al. (2014) was characterized by a steady decrease from the first six months postpartum to two years postpartum. This trajectory represented less than 10% of the samples in all four studies.

The increasing trajectory pattern, characterised by initially low depressive symptoms that increase to a level above the severity cut-off in the postpartum period, was reported by three studies (Lee et al., 2014; Mora et al., 2009; Vänskä et al., 2011). A similar trajectory was reported by Sutter-Dallay et al. (2012), but symptom levels ultimately remained under the depression severity cut-off. This trajectory also represented a minority of the total samples (range 4.0% (Sutter-Dallay et al., 2012) to 10.2% (Lee et al., 2014)).

The third type of trajectory reported by three studies is characterised by episodic increases and decreases in depressive symptoms. In Christensen et al. (2011) and Mora et al. (2009), this trajectory begins with symptoms just above the cut-off that increase rapidly just after birth, but then return to levels either just above (Christensen et al., 2011) or below the cut-off (Mora et al., 2009). Nine percent (Mora et al., 2009) and 10% (Christensen et al., 2011) of women were classified in this trajectory. In Parade et al. (2014), the opposite episodic trajectory is reported: the trajectory begins above the severity cut-off in the third trimester of pregnancy, temporarily abates two weeks before birth, but increases again four weeks after birth and reaches the cut-off again by two months postpartum.

Factors associated with the trajectories

A total of nine studies investigated the association of baseline characteristics with membership in different trajectories of perinatal depressive symptoms. All used the likelihood of being classified in the stable low or moderate-low depressive symptom trajectory (low-risk group) as the reference against which to compare the likelihood of being classified into other trajectories (Table 3). In Kuo et al. (2012), the stable low and moderate-low trajectories were combined into one trajectory, and the stable high and moderate-high were combined into a second trajectory, so that only these two groups were compared. For ease of reporting, these combined trajectories are considered stable low and stable high symptom trajectories, respectively.

Of the two studies that both reported low and moderate-low stable trajectories and investigated risk factors, one study found no differences between women assigned to either trajectory (Kuo et al., 2014). The other only found that a greater proportion of the women assigned to the stable moderate-low trajectory were multiparous, compared to those assigned to the stable low trajectory (Vänskä et al., 2011).

The likelihood of belonging to a stable moderate-high symptom group (high, but close to cut-off) was higher among women who smoked more than a pack of cigarettes per day (Glasheen et al., 2013), reported sleep difficulty in the third trimester and used patient-controlled analgesics after a caesarean section (Kuo et al., 2014). Glasheen et al. (2013) also reported that women with high social support were 51% less likely to be in this group compared to those with low social support. In Sutter-Dallay et al. (2012), the likelihood of being classified in this trajectory was greater for women who were older, nulliparous, reported a lower salary and had higher levels of trait anxiety. Vänskä et al. (2011), however, reported no demographic differences between the stable moderate-high and stable low trajectories during the perinatal period..

The likelihood of belonging to the stable high symptom trajectory compared to a stable low or moderate-low trajectory was higher for those who reported sleep difficulty in the third trimester (Kuo et al., 2012). Mora et al. (2009) found a higher likelihood of belonging to this group among women who reported being white (vs. black), were multiparous, had fair or poor emotional health, were anxious

about the pregnancy and showed moderate or high objective stress. Sutter-Dallay et al. (Sutter-Dallay et al., 2012) also reported a higher likelihood of belonging to that group among women who had higher trait anxiety levels. .

In Mora et al. (2009), the likelihood of belonging to the transient, decreasing trajectory was greater for women who reported not being born in the country (US), self-identified as white (as opposed to Latina/Hispanic), rated their emotional health as fair or poor, had recently consumed alcohol, were anxious about the pregnancy, and reported a high level of objective stress (Mora et al., 2009). Multiparous women (Vänskä et al., 2011) and women with lower social support (Christensen et al., 2011) also had a greater likelihood of belonging to this trajectory. Lee et al (2014), who identified a similar trajectory but only in the postpartum period, did not find any differences between women in this group versus those with stable low level symptoms.

Lee et al (2014) did find differences for women who were classified in the transient, increasing trajectory compared to those in the stable low trajectory: postpartum women with increasing symptoms were less likely to report good physical health in the third trimester of pregnancy (Lee et al., 2014). Mora et al. (2009) reported that the likelihood of belonging to this trajectory was also higher for women who had less than high school education and had higher levels of objective stress. No differences were found in the likelihood of being classified in this trajectory in comparison to the stable low trajectory in the other two studies reporting these trajectories (Sutter-Dallay et al., 2012; Vänskä et al., 2011).

Finally, differences found between women assigned to the episodic trajectory in Mora et al. (2009) compared to those assigned to the stable moderate-low trajectory were similar to those found for women assigned to the stable high trajectory, in terms of parity, ethnicity, emotional health and objective stress. Women classified in this episodic trajectory were also more likely to report low educational attainment and having a comorbid condition, however anxiety about pregnancy was not associated with this trajectory (Mora et al., 2009). Christensen et al (2011), who also report a similar episodic trajectory, indicate that women assigned to this trajectory were more likely to not have health insurance or to report a history of abuse. In Parade et al. (2014), where the episodic trajectory shows temporary decreased symptoms just before and after birth, results suggests that women who were

classified in this trajectory were more likely to have a lower education and decreased remembered paternal care.

Discussion

The objective of this study was to systematically review the literature that has used GCMM to identify groups of women with different trajectories of depressive symptoms and associated risk factors over the perinatal period.

Summary of evidence

All 11 studies included in this review reported identifying at least one low-risk group, characterised by stable low or moderate-low symptoms of depression (associated with scores below the severity cut-off for the instrument) throughout the perinatal period. The majority of studies also reported a high risk trajectory, in which stable high depressive symptoms persisted from pregnancy throughout the postpartum period. Six of the 11 studies identified transient trajectories, with either increasing, decreasing or episodic depressive symptoms occurring in the perinatal period.

A range of risk factors or predictors of trajectories were investigated in nine studies, using the characteristics of women assigned to the stable low symptom trajectory as a reference. Very few predictors differentiated trajectories. Anxiety or stress were reported by two studies as increasing the likelihood of belonging to the stable high symptom trajectory (Mora et al., 2009; Sutter-Dallay et al., 2012), and lower education as increasing the likelihood of having episodic depressive symptoms (Mora et al., 2009; Parade et al., 2014). Results also suggest that, within the same study, many of the same factors were identified as increasing the chances of belonging to stable or transient trajectories, in comparison to stable low trajectories. This suggests that predictors could not necessarily differentiate women with persistent low risk of depression from those who might eventually experience severe symptoms in the perinatal period, or differentiating women whose symptoms might abate naturally from those who will continue to have severe symptoms throughout the perinatal period.

While studies were consistent in identifying a low risk trajectory, the size of the stable low or moderate-low symptom trajectory varied by study. Out of the 10 studies who provided sample sizes, these trajectories comprised the majority of the sample in nine studies. The fact that the stable moderate-low trajectory was a minority in Glasheen et al. (2013) is particularly striking, given that the only other trajectory identified in their model represented stable high symptoms. The inclusion criteria were relatively broad in this study, and the sample was not a high-risk group. The authors do indicate, however, that the women recruited were predominantly of low socio-economic status. The fact that none of the socio-demographic predictors differentiated the two trajectories suggest that the sample was homogeneous, and supports the idea that socio-economic status may explain the very high proportion of women being classified in the stable high trajectory. This supports previous evidence suggesting that income or socio-economic level is associated with a greater risk for depression (Lund et al., 2010). This finding was not replicated by Ramos-Marcuse et al. (2010), however, where over 40% and 45% of low-income women were assigned to the low and moderate-low symptom level trajectories, respectively.

Three studies did not report a stable high or moderate-high symptom trajectory (Christensen et al., 2011; Lee et al., 2014; Parade et al., 2014). In Christensen et al. (2011), this may have been due to the inclusion criterion which excluded all women who were already diagnosed with depression. This was however not seen in Marcus et al. (2011), who also excluded women with depression but still report a stable moderate-high trajectory. Also, symptoms were only assessed for several weeks during pregnancy, and a longer assessment period may have revealed trajectories similar to those found by Christensen et al. (2011).

The second study not reporting a stable high trajectory was Parade et al's (2014). This study was one of the three that were rated as 'low quality'; contributing to the low score was the vague inclusion criteria, the extremely small sample size, and the lack of reporting on the proportion of the sample classified in either trajectory. The small sample size may have limited the ability to identify a robust third trajectory class through GCM. The third study without a stable high or moderate-high symptom trajectory was Lee et al. (2014), which investigated postnatal depressive symptoms among overweight or obese women. This is an interesting finding, given that the evidence suggests that obesity is a risk factor for postpartum depression (Marchi et al., 2015; Milgrom et al., 2012). However,

the lowest number of assessments per women was reported in this study, which is likely to have influenced the trajectories identified through GCMM.

The fact that depressive symptoms were investigated for various lengths of time pre- and/or postnatally may explain the variety of trajectories reported, and make it difficult to truly compare trajectories across studies. Indeed, many of the studies which investigated both antenatal and postnatal symptoms only did so for a few weeks during pregnancy (Kuo et al., 2014; Kuo et al., 2012; Parade et al., 2014; Sutter-Dallay et al., 2012) and/or up to six or 12 months postpartum (Christensen et al., 2011; Kuo et al., 2014; Kuo et al., 2012; Vänskä et al., 2011). Yet, studies with longer follow-up periods show that symptoms can still increase or decrease substantially after the first year postpartum (Lee et al., 2014; Mora et al., 2009; Sutter-Dallay et al., 2012). Investigating depressive symptoms only during pregnancy or in the postpartum period also gives a partial picture of the change in mood from pregnancy to motherhood, especially given that both periods are hormonally and psychologically different. For example, in Lee's study (2014), the decreasing symptom trajectory from birth to two years postpartum may be similar to the 'postpartum' trajectory in Christensen et al. (2011) or Mora et al. (2009), where symptoms increased periodically around birth. Similarly, while Ramos-Marcuse et al. (2010) report only stable trajectories, the stable moderate-low trajectory identified could represent the postnatal part of the transient decreasing trajectory reported in another three studies (Christensen et al., 2011; Mora et al., 2009; Vänskä et al., 2011). In other words, women assigned to low-risk trajectories in the postpartum period may have been at risk for depression antenatally, and vice versa.

Regular symptom assessments over a shorter period of time also have advantages, mainly in allowing more nuanced description of changes in symptom experiences. This is what Marcus et al. (2011) and Kuo et al. (2012) did, by examining depressive symptoms over two months during pregnancy and several days after birth, respectively. Both studies, however, report relatively stable symptoms, even in the first week after giving birth (Kuo et al., 2012). It is common for women to develop depressive symptoms during the first two weeks after giving birth, often referred to as the 'baby blues', but this phenomenon was not reflected in Kuo et al. (2012) or Parade et al. (2014)'s study. In fact, Parade et al. (2014) report an atypical trajectory of high levels of depressive symptoms during pregnancy and postnatally, with a temporary decrease immediately after birth. The authors referred to this as the 'honeymoon period'. None of the other studies included in this review examined

depressive symptoms close enough to the birth of the baby to corroborate these findings. The existence of a 'honeymoon' period would have important implications on timing of screening and detection of 'postpartum blues' or postpartum depression.

Findings from studies which report transient trajectories also have implications for timing of screening. While transient trajectories only represented a minority of the sample in each study, they often overlapped with stable low or high trajectories for several months (Christensen et al., 2011; Mora et al., 2009; Sutter-Dallay et al., 2012; Vänskä et al., 2011). Further complicating the identification of women at risk of perinatal depression, results from this review indicate that few socio-economic predictors differentiate women who have chronically high symptom levels from those who have decreasing levels of symptoms (Mora et al., 2009; Vänskä et al., 2011). There is also mixed evidence on predictors differentiating women who have stable low levels of symptoms from those whose symptoms increase in the postpartum period (Mora et al., 2009; Vänskä et al., 2011). Indeed, in Mora et al. (2009) for example, while each trajectory could be differentiated from the stable moderate-low trajectory on a range of demographic and psychosocial factors, these predictors were similar across trajectories. Perhaps important distinctions were missed by conducting post-hoc multinomial analyses, which require that all trajectories be compared to one reference trajectory only, when comparisons of two transient trajectories, or comparison of stable high-symptom trajectory with a transient one may be more useful in this context.

The lack of demographic or socio-economic differences identified across different trajectories may be due to the homogeneity of the samples recruited, as a result of the strict inclusion criteria implemented in most studies in this review. For example, the fact that no differences were found between the stable low and stable high symptom trajectory classes in Glasheen et al's study (2013) may be due to the low variability of the socio-economic status of the women recruited.

The lack of differences across trajectories may also be due to the relatively small samples reported in the studies reviewed. There are usually no restrictions on sample sizes for conducting GCMM; they usually depend on the number of parameters in the model, attrition and missing data, the instruments used, as well as on the distribution of variables. However, small samples can lead to an inability to identify smaller but meaningful latent trajectories (Berlin et al., 2014), or to generate smaller

trajectories, comprising of a sample too small to detect statistical differences in predictors between the trajectories.

Another possible reason for the lack of differences in characteristics across trajectories is the type of risk factors investigated in each study. Besides demographic factors that were investigated in most studies, risk factors examined mostly pertained to the emotional state of women (trait anxiety, stress, emotional health), their ambivalence about pregnancy, their self-esteem, or general health (physical health and sleep difficulty). Other common risk factors for perinatal depression, such as partner and social support, domestic violence, obstetric complications, as well as anxiety and traumatic life events (Lancaster et al., 2010; Robertson et al., 2004; Verreault et al., 2014), should be investigated further in the context of GCM analyses.. Researchers should also be cognisant of the different effects risk factors have on depressive symptoms when investigating differences between trajectories. Indeed, one could argue that transient stressors, such as a traumatic life event, may have a more episodic effect on depressive symptoms over time, compared to persistent risk factors, such as community violence or poverty, which may have more permanent repercussions on depressive mood.

Limitations

A few limitations of the review should be noted. First, a small number of studies were identified, and none from low- and middle-income countries, despite the absence of publication date or language restrictions. This may be due to the fact that only studies using GCM were included, and not all researchers investigating perinatal symptoms over multiple time points may have had the statistical expertise to use this analytical approach. However, only a minority of articles were excluded based on this criterion alone. Second, though one study investigated perinatal depressive symptoms among low-income women (Ramos-Marcuse et al., 2010), most studies were conducted in urban settings, none were based in low or low-middle income countries, and, in general, studies' inclusion criteria were quite strict. Altogether, these factors may therefore limit the generalisability of the review's findings to low-risk women in middle- high income countries. Third, important distinctions between trajectories may also have been missed by focusing on baseline characteristics only. In addition to predictors, investigating outcomes may give a better indication of whether different trajectories identify 'real' subgroups of women. Though this was beyond the current study's scope, it represents an

important future direction of research. Finally, none of the articles reported the parameters of the model identified, such as the slopes, intercepts, variances and co-variances of latent trajectories, and instead reported fit statistics used in the identification of the optimal model and number of latent trajectories. For this reason, a meta-analysis of the articles was not possible.

Limitations relating to latent growth mixture models and their interpretation must also be noted. First, the observed data (i.e. baseline risk factors or characteristics of women) are a function of the probability of belonging to certain trajectories in the growth curve mixture model, so any differences identified between trajectories in subsequent phases have a level of uncertainty, and do not capture the individual variability within each class. For these reasons, findings should be interpreted with caution, especially when, with attrition, less than three assessments on average are completed (Lee et al., 2014) or when few fit statistics are used to select the optimal model (Kuo et al., 2012; Marcus et al., 2011). Two studies used only the BIC as the fit statistic to select the optimal model generated from the GCMM. While some researchers report that BIC is the most reliable criteria for identifying the optimal number of trajectories (Nylund et al., 2007), others report that entropy, which indicates the extent to which an individual is classified into one latent group with confidence, is usually favoured over other fit indices (Leiby, 2012; Ram and Grimm, 2009). Also, the use of prospective data, rather than retrospective (secondary) data, such as is the case for several studies included in this review, also limits the interpretation of the findings since data are constrained by the original recruitment and data collection methods. Instead, prospective studies can be tailored to GCMM, by recruiting bigger samples, planning for at least 5 assessments, and having a clear plan about how to deal with missing data.

Others have also warned about creating trajectories when there is only one 'real' trajectory (Bauer and Curran, 2003). Indeed, a growth curve mixture model will generate a multiple-trajectory model, regardless of whether these trajectories have realistic benefits or implications for understanding the aetiology or progression of perinatal depression. Unnecessary trajectories could be generated due to non-normally distributed outcomes, inappropriate screening or measurement tools or even overly large samples (Ram and Grimm, 2009). Moreover, by using GCMM, researchers may run the risk of creating overly complex models that are difficult to interpret, or of creating overly simplistic models that ignore heterogeneity in the actual progression of symptoms among different subgroups of women.

While this possibility cannot be ignored, the majority of studies included in this review did use extensively locally validated screening tools, and most often selected models with parsimony, all fit statistics being equal. Given the exploratory nature of GCMM, Ram and Grimm (2009) also suggest basing the identification of the optimal model on past research and theory as much as possible. Authors of studies included in this review do acknowledge using theoretical and pragmatic interpretability to select optimal models, though they are not very explicit in stating which theories or interpretations these are. Future studies using this method could benefit from clearly stating the theoretical considerations taken into account, besides the fit statistics, when selecting optimal models.

Despite these limitations, the use of GCMM over more common longitudinal data analysis methods, allows for the identification of subgroups of women with different profiles of depression, health risks and treatment needs. GCMM research could help tailor the timing and content of screening, improving the efficiency of identification, referral, and treatment strategies, which is especially necessary in a context of limited mental health resources. The use of GCMM in the context of randomised controlled trials has also been advocated to identify whether response to the treatment under investigation differs across subgroups of women with perinatal depression. Bearing in mind the diverse trajectories of depressive symptoms during the perinatal period, evidence can be built for how, when and for whom psychosocial interventions are most likely to be effective.

Conclusions

Bearing in mind the constraints of GCMM, this method has allowed researchers to identify heterogeneity in the course of perinatal depressive symptoms within populations. The studies included in this review report relatively similar types of depressive symptom trajectories during the perinatal period. The stable high symptom group consistently reported suggest that there is clearly an at-risk population of women, with strikingly persistent severe symptoms throughout the perinatal period. It is important for policy makers and providers to realise that severe symptoms may not necessarily abate on their own, and that not identifying and treating these women early is likely to result in a long period of distress and potential health risks for the mother and the child. This review also suggests that there is little information on how groups of women with transient or stable levels of depressive symptoms differ. It is unclear whether this finding is due to the relatively small samples

recruited for this type of analysis, whether the trajectories generated are actually not meaningfully different, or whether they are but they differ in ways that have not yet been measured. It is important that more high quality GCM studies are completed, particularly with bigger samples, risk factor selection guided by a theoretical framework, and be clear reporting of all theoretical and practical steps taken. More consistency in assessment schedules throughout the perinatal period would also allow greater comparison of findings across studies. Only then will we be able to draw conclusions on the meaningfulness and applicability of trajectories identified through GCM, and potentially improve the identification systems of at-risk perinatal women. More research should also focus on women living in LMIC. Identifying high risk groups in settings where mental health services and resources are limited would allow screening and interventions to be focused on women with greater needs, thereby reducing the burden of service delivery.

Acknowledgements

The authors would like to acknowledge the support Alfred Musekiwa and Professor Ding-Geng Chen provided on finalising the review's analysis plan.

Funding

This work was supported by the National Institute of Mental Health of the National Institutes of Health [grant number U19MH095699] and by the National Institute of Mental Health Global Mental Health training grant [grant number MH103210 to S.M.M.]. The funders did not have any role in the study design, data collection, analysis or interpretation of the data, or in the writing of the report.

Conflicts of interest

None.

Contributors

ECB, CL and MS conceptualised the manuscript. ECB conducted the search, and together with JB and SMM, screened and reviewed the titles, abstracts and full texts generated through search. ECB

drafted the manuscript, which was reviewed and edited by JB, SMM, CL and MS. All authors read and agreed on the final version of the manuscript.

References

- Bauer, D.J., Curran, P.J., 2003. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychol. Methods* 8, 338-363.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561-571.
- Berlin, K.S., Parra, G.R., Williams, N.A., 2014. An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *J. Pediatr. Psychol.* 39, 188-203.
- Christensen, A.L., Stuart, E.A., Perry, D.F., Le, H.-N., 2011. Unintended pregnancy and perinatal depression trajectories in low-income, high-risk Hispanic immigrants. *Prevention science* 12, 289-299.
- Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry* 150, 782-786.
- Fisher, J., Mello, M.C.d., Patel, V., Rahman, A., Tran, T., Holton, S., Holmes, W., 2012. Prevalence and determinants of common perinatal mental disorders in women in low-and lower-middle-income countries: a systematic review. *Bull. World Health Organ.* 90, 139-149.
- Gavin, N.I., Gaynes, B.N., Lohr, K.N., Meltzer-Brody, S., Gartlehner, G., Swinson, T., 2005. Perinatal depression: a systematic review of prevalence and incidence. *Obstet. Gynecol.* 106, 1071-1083.
- Glasheen, C., Richardson, G.A., Kim, K.H., Larkby, C.A., Swartz, H.A., Day, N.L., 2013. Exposure to maternal pre- and postnatal depression and anxiety symptoms: risk for major depression, anxiety disorders, and conduct disorder in adolescent offspring. *Dev. Psychopathol.* 25, 1045-1063.
- Hayes, B., Sharif, F., 2009. Behavioural and emotional outcome of very low birth weight infants—literature review. *The journal of maternal-fetal & neonatal medicine* 22, 849-856.
- Holcomb Jr, W.L., Stone, L.S., Lustman, P.J., Gavard, J.A., Mostello, D.J., 1996. Screening for depression in pregnancy: characteristics of the Beck Depression Inventory. *Obstet. Gynecol.* 88, 1021-1025.
- Johnson, W., Hicks, B.M., McGue, M., Iacono, W.G., 2007. Most of the girls are alright, but some aren't: Personality trajectory groups from ages 14 to 24 and some associations with outcomes. *J. Pers. Soc. Psychol.* 93, 266-284.
- Jones, B.L., Nagin, D.S., Roeder, K., 2001. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological methods & research* 29, 374-393.
- Kuo, S.Y., Chen, S.R., Tzeng, Y.L., 2014. Depression and anxiety trajectories among women who undergo an elective cesarean section. *PLoS One* 9, e86653.
- Kuo, S.Y., Yang, Y.L., Kuo, P.C., Tseng, C.M., Tzeng, Y.L., 2012. Trajectories of depressive symptoms and fatigue among postpartum women. *J. Obstet. Gynecol. Neonatal Nurs.* 41, 216-226.
- Lancaster, C.A., Gold, K.J., Flynn, H.A., Yoo, H., Marcus, S.M., Davis, M.M., 2010. Risk factors for depressive symptoms during pregnancy: a systematic review. *American journal of obstetrics and gynecology* 202, 5-14.
- Lee, C.T., Stroo, M., Fuemmeler, B., Malhotra, R., Ostbye, T., 2014. Trajectories of depressive symptoms over 2 years postpartum among overweight or obese women. *Womens Health Issues* 24, 559-566.
- Leiby, B.E., 2012. Growth curve mixture models. *Shanghai Archives of Psychiatry* 24, 355.
- Little, T.D., Jorgensen, T.D., Lang, K.M., Moore, E.W.G., 2014. On the joys of missing data. *J. Pediatr. Psychol.* 39, 151-162.

- Lund, C., Breen, A., Flisher, A.J., Kakuma, R., Corrigall, J., Joska, J.A., Swartz, L., Patel, V., 2010. Poverty and common mental disorders in low and middle income countries: A systematic review. *Soc. Sci. Med.* 71, 517-528.
- Lusskin, S.I., Pundiak, T.M., Habib, S.M., 2007. Perinatal depression: hiding in plain sight. *The Canadian journal of psychiatry* 52, 479-488.
- Marchi, J., Berg, M., Dencker, A., Olander, E., Begley, C., 2015. Risks associated with obesity in pregnancy, for the mother and baby: a systematic review of reviews. *Obes. Rev.* 16, 621-638.
- Marcus, S., Lopez, J.F., McDonough, S., Mackenzie, M.J., Flynn, H., Neal, C.R., Jr., Gahagan, S., Volling, B., Kaciroti, N., Vazquez, D.M., 2011. Depressive symptoms during pregnancy: impact on neuroendocrine and neonatal outcomes. *Infant Behav Dev* 34, 26-34.
- Milgrom, J., Skouteris, H., Worotniuk, T., Henwood, A., Bruce, L., 2012. The association between ante-and postnatal depressive symptoms and obesity in both mother and child: a systematic review of the literature. *Womens Health Issues* 22, e319-e328.
- Mora, P.A., Bennett, I.M., Elo, I.T., Mathew, L., Coyne, J.C., Culhane, J.F., 2009. Distinct trajectories of perinatal depressive symptomatology: evidence from growth mixture modeling. *Am. J. Epidemiol.* 169, 24-32.
- Muthén, L., Muthén, B., (1998-2007). *Mplus User's Guide: Statistical Analysis with Latent Variables*. Fifth Ed. Los Angeles, CA: Muthén & Muthén.
- Nandi, A., Beard, J.R., Galea, S., 2009. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. *BMC Psychiatry* 9, 31.
- Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling* 14, 535-569.
- O'hara, M.W., Swain, A.M., 1996. Rates and risk of postpartum depression-a meta-analysis. *Int. Rev. Psychiatry* 8, 37-54.
- Parade, S.H., Nayena Blankson, A., Leerkes, E.M., Crockenberg, S.C., Faldowski, R., 2014. Close relationships predict curvilinear trajectories of maternal depressive symptoms over the transition to parenthood. *Family Relations* 63, 206-218.
- Radloff, L.S., 1977. The CES-D scale a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385-401.
- Rahman, A., Surkan, P.J., Cayetano, C.E., Rwagatare, P., Dickson, K.E., 2013. Grand challenges: integrating maternal mental health into maternal and child health programmes. *PLoS Med.* 10, e1001442.
- Ram, N., Grimm, K.J., 2009. Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development* 33, 565-576.
- Ramos-Marcuse, F., Oberlander, S.E., 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.
- Robertson, E., Grace, S., Wallington, T., Stewart, D.E., 2004. Antenatal risk factors for postpartum depression: a synthesis of recent literature. *General hospital psychiatry* 26, 289-295.
- SAS Institute Inc, 2006. *SAS Software, Version 9.1.*, Cary, NC, USA.
- Stowe, Z.N., Nemeroff, C.B., 1995. Women at risk for postpartum-onset major depression. *Am. J. Obstet. Gynecol.* 173, 639-645.
- Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 283, 2008-2012.
- Sutter-Dallay, A.L., Cosnefroy, O., Glatigny-Dallay, E., Verdoux, H., Rasclé, N., 2012. Evolution of perinatal depressive symptoms from pregnancy to two years postpartum in a low-risk sample: the MATQUID cohort. *J. Affect. Disord.* 139, 23-29.
- Tucker, J.S., Orlando, M., Ellickson, P.L., 2003. Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychol.* 22, 79-87.

Vänskä, M., Punamäki, R.L., Tolvanen, A., Lindblom, J., Flykt, M., Unkila-Kallio, L., Tiitinen, A., Repokari, L., Sinkkonen, J., Tulppala, M., 2011. Maternal pre- and postnatal mental health trajectories and child mental health and development: Prospective study in a normative and formerly infertile sample. *International Journal of Behavioral Development* 35, 517-531.

Verreault, N., Da Costa, D., Marchand, A., Ireland, K., Dritsa, M., Khalifé, S., 2014. Rates and risk factors associated with depressive symptoms during pregnancy and with postpartum onset. *Journal of psychosomatic obstetrics & gynecology* 35, 84-91.

Wachs, T.D., Black, M.M., Engle, P.L., 2009. Maternal depression: a global threat to children's health, development, and behavior and to human rights. *Child Development Perspectives* 3, 51-59.

Weissman, M.M., Sholomskas, D., Pottenger, M., Prusoff, B.A., Locke, B.Z., 1977. Assessing depressive symptoms in five psychiatric populations: a validation study. *Am. J. Epidemiol.* 106, 203-214.

Wells, G., Shea, B., O'Connell, D., 2012. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies

in meta-analyses.

Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* 382, 1575-1586.

World Health Organization, 2008. Maternal mental health and child health and development in low and middle income countries: report of the meeting, Geneva, Switzerland, 30 January-1 February., Geneva.

Zammit, A.R., Starr, J.M., Johnson, W., Deary, I.J., 2012. Profiles of physical, emotional and psychosocial wellbeing in the Lothian birth cohort 1936. *BMC Geriatr.* 12, 1.

Highlights

- Two to five different trajectories were identified in each study
- Most studies report both chronic and transient symptom level trajectories
- Most studies report one trajectory with chronic high levels of depressive symptoms
- Chronic and transient trajectories could not be distinguished based on women's features

Figure 1. Search process for studies investigating trajectories of perinatal depression using growth curve mixture modelling

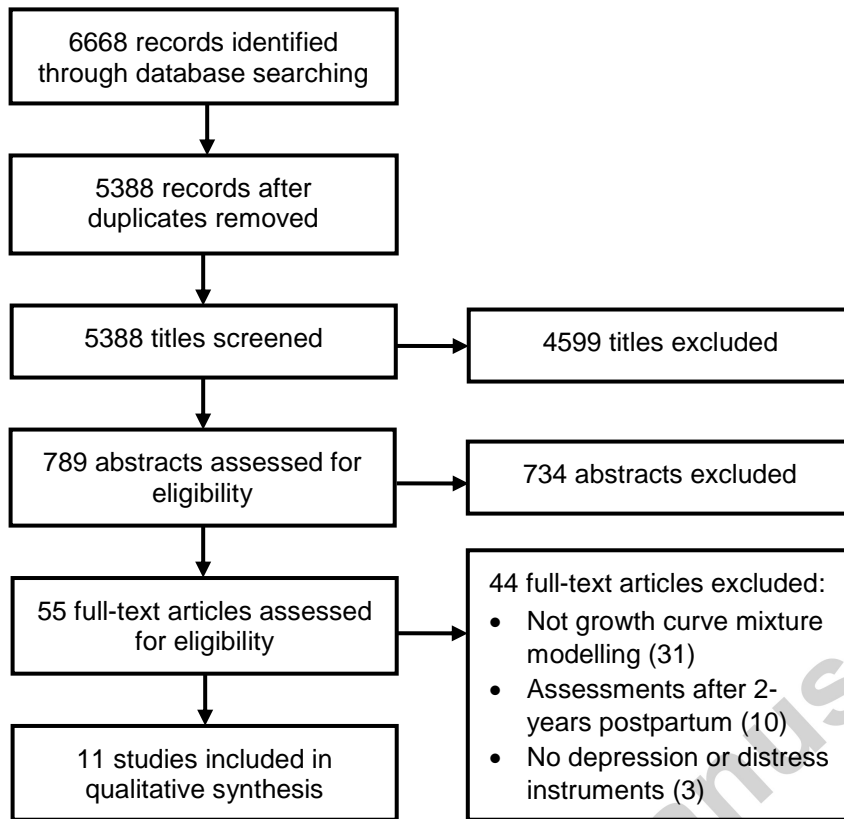


Figure 2a. Standardised depressive symptoms over time for each trajectory, for studies with follow-up time beyond 6-month postpartum, Standardised depressive symptoms over time for each trajectory, for studies with a maximum follow-up time of 6-month postpartum

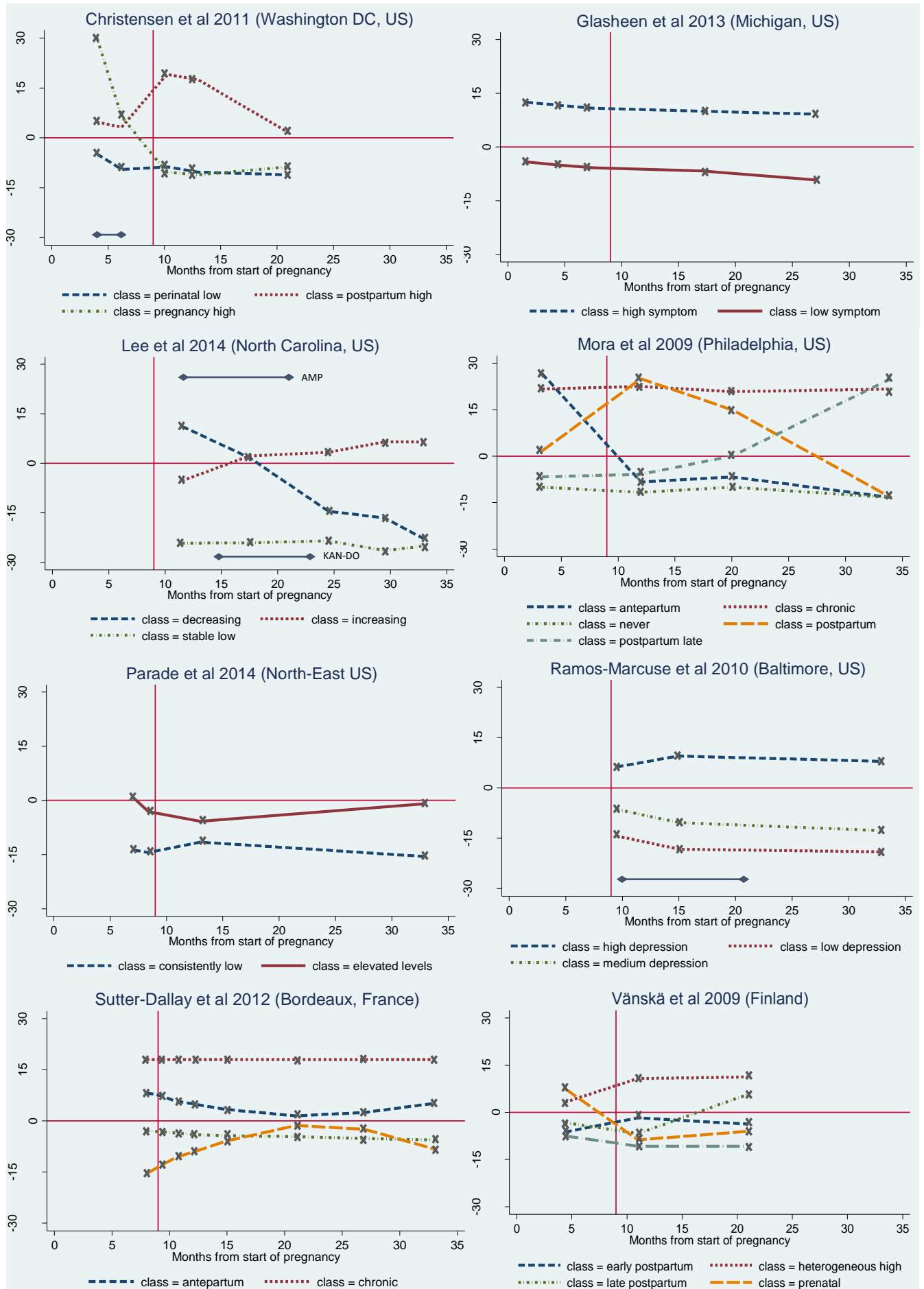


Figure 2b.

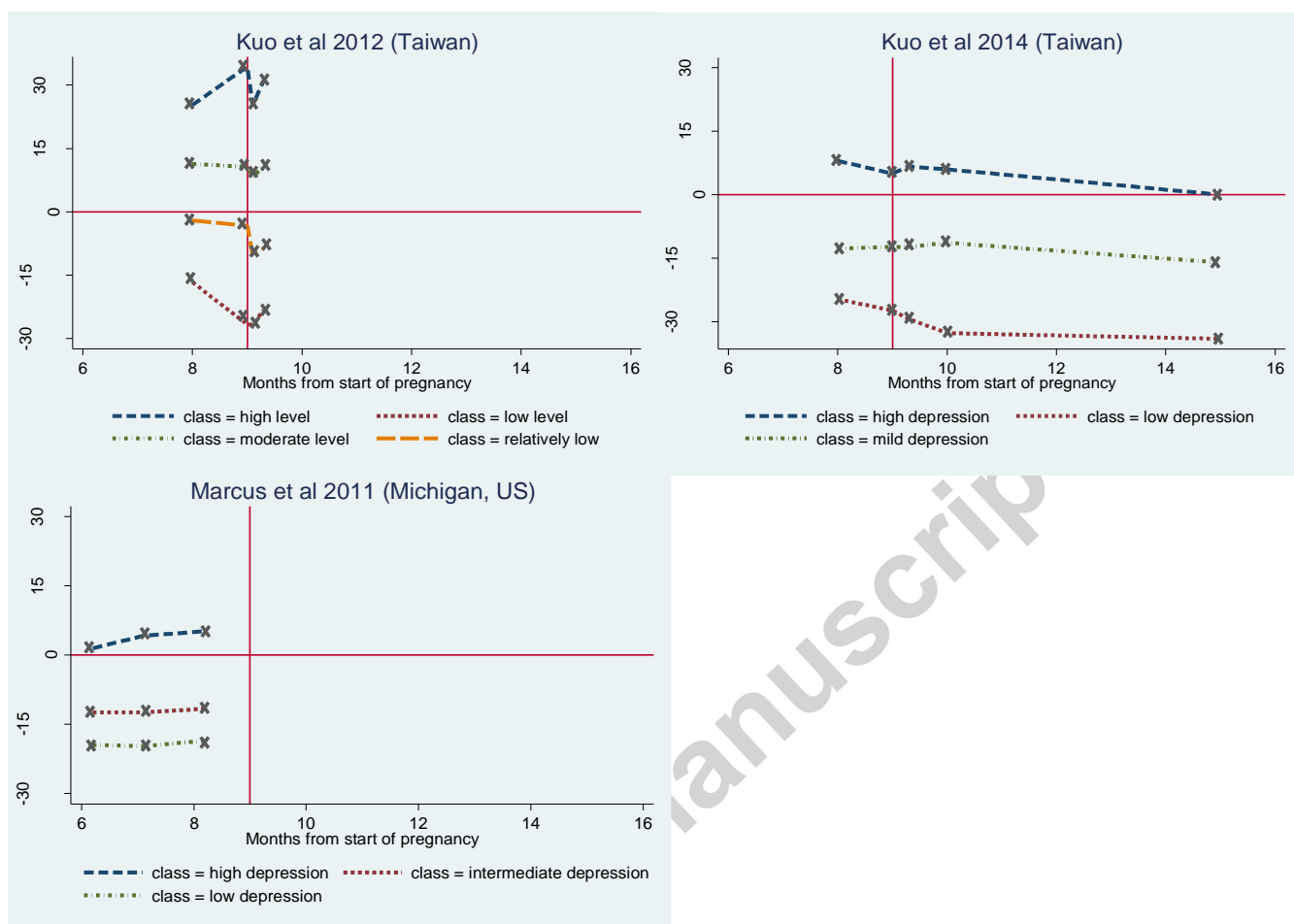


Table 1. Database search strategy for the systematic review

Concept	Search terms
1. perinatal depressive symptoms	(depression OR depressive symptoms OR mood OR dysthymia OR distress OR mental health) AND (perinatal OR antenatal OR prenatal OR pregnancy OR pregnant OR birth OR postnatal OR postpartum OR maternal)
2. trajectories	trajectory OR trajectories OR evolution OR evolutionary OR progress OR progression OR development OR growth OR prognosis OR remission OR epidemiology OR persistence OR chronic OR change
3. factors associated with trajectories	profiles OR "risk factors" OR symptoms OR "socio-economic" OR socioeconomic OR "psychosocial factors" OR correlates OR "prognostic factors" OR predictors
4. longitudinal design using latent variable modelling	cohort OR prospective OR longitudinal OR modelling OR modeling OR follow-up OR latent classes OR "growth mixture modelling"

Table 2. Design, recruitment and data collection methods of the 11 studies included in the systematic review

Study	Location	Sample size	Population	Inclusion/exclusion criteria	Time of recruitment	Outcome of interest	Assessments
Christensen et al, 2011 (29)	Washington DC, US (urban)	215	Pregnant women undergoing preventive intervention for perinatal depression; Mean age 25.4	<i>Inclusion criteria:</i> 18 to 35 years; Hispanic; less than 25 weeks pregnant; CES-D-20 score >15 but not currently depressed. <i>Exclusion criteria:</i> report smoking, drinking alcohol or using illicit drugs, or meet criteria for other major mental disorder or significant psychosocial problems	Antenatal (average 18 weeks gestation)	BDI-II: cut-off of 16 suggesting depressive symptomatology	5 assessments: First assessment: 18 weeks gestation; Last assessment: 12 months postpartum
Glasheen et al, 2013 (21)	Michigan, US (urban)	577	General perinatal population; Mean age 22.9	<i>Inclusion criteria:</i> at least 3 of 5 assessments; mother-child pair available at 16 year follow-up <i>No other exclusion criteria</i>	Antenatal, first trimester	CES-D-20 (no cut-off indicated)	5 assessments First assessment: first trimester of pregnancy; Last assessment: 18 months postpartum
Kuo et al, 2012 (26)	Central Taiwan	121	General perinatal population; Mean age 33.4	<i>Inclusion criteria:</i> 20 years or older; singleton pregnancy <i>Exclusion criteria:</i> perinatal complications, chronic medical history	Antenatal, third trimester	EPDS: cut-off of 10 suggesting high level of symptoms	4 assessments First assessment: third trimester; Last assessment: 7 days postpartum
Kuo et al, 2014 (25)	Central Taiwan	139	General perinatal population; Mean age 31.2	<i>Inclusion criteria:</i> 20 years or older; considered for caesarean section, <i>Exclusion criteria:</i> other perinatal complications; chronic medical illness	Antenatal, third trimester (average 37 weeks gestation)	EPDS: cut-off of 13 suggesting probable case of depression	5 assessments First assessment: third trimester; Last assessment: 6 months postpartum
Lee et al, 2014 (30)	North Carolina, US	844 (from two RCTs, AMP and KAN-DO)	Overweight or obese perinatal population Mean age 31.7	<i>Inclusion criteria:</i> BMI>25 before pregnancy; 18 years or older; English-speaking AMP: able to walk a mile KAN-DO: maximum 6 months postpartum at recruitment; pre-schooler in the	Postnatal; AMP: 2 months postpartum KAN-DO: 6 months postpartum	EPDS: cut-off of 13 suggesting postpartum depression	3 assessments per RCT (one assessment in common), 5 altogether First assessment: 1-7 months postpartum; Last assessment: 24 months

				home; no medical complications <i>Inclusion criteria:</i> 21 years or older, EPDS score >10 but not currently depressed; fluent in English <i>Exclusion criteria:</i> adoption plan; chronic medical condition or use of medication that impact LHPA; treated with psychotropic medication; substance abuse; eating disorder; bipolar illness			postpartum
Marcus et al, 2011 (22)	Michigan, US	154	General perinatal population ; 48% between 20-30 years 50% between 31-40 years	<i>Inclusion criteria:</i> English- or Spanish-speaking; singleton intrauterine pregnancy; at least one postpartum interview; live birth <i>No other exclusion criteria</i>	Antenatal, between 8 and 28 weeks gestation	BDI: cut-off of 20 criteria for further assessment with SCID (no cut-off indicated for severity)	3 assessments First assessment: 28 weeks; Last assessment: 37 weeks
Mora et al, 2009 (24)	Philadelphia, US (urban)	1735	General perinatal population ; Mean age 23.9	<i>Inclusion criteria:</i> English- or Spanish-speaking; singleton intrauterine pregnancy; at least one postpartum interview; live birth <i>No other exclusion criteria</i>	Antenatal, first antenatal care visit (average 15 weeks gestation)	CES-D-20: cut-off of 16 indicating significant levels of symptoms	4 assessments First assessment: at first antenatal visit (approximately 15 weeks gestation); Last assessment: 25 months postpartum
Parade et al, 2014 (23)	North-East US (state not specified)	98	General perinatal population ; Mean age 29.0	<i>Inclusion criteria:</i> primiparous, 20 years or older <i>No other exclusion criteria</i>	Antenatal, 8 weeks before expected date of delivery (EDD)	CES-D-20 (15 items - excluding 5 items relating to somatic symptoms) Not cut-off indicated	4 assessments First assessment: 8 weeks prior to EDD; Last assessment: 24 weeks postpartum
Ramos-Marcuse et al, 2010 (31)	Baltimore, Maryland, US (urban)	181	Low-income general perinatal population ; Mean age 16.3	<i>Inclusion criteria:</i> 17 years or younger, first time delivery, African American, low-income (eligible for WIC, family income under 185% of poverty level); living with their mother <i>Exclusion criteria:</i> chronic physical illnesses	Postnatal, shortly after delivery	BDI: cut-off of 9 suggesting risk for depression	3 assessments First assessment: within 3 weeks of delivery; Last assessment: 24 months postpartum
Sutter-Dallay et al, 2012 (27)	Bordeaux, France (urban)	579	General perinatal population ; Mean age 29.4	<i>Inclusion criteria:</i> French-speaking, living in catchment area of hospital, less than one week of hospitalisation for pregnancy complications <i>Exclusion criteria:</i>	Antenatal, third trimester	CES-D-20: cut-off of 16 indicating clinically significant level of depressive symptoms	8 assessments First assessment: 8 months gestation; Last assessment: 24 months postpartum

Vänkä et al, 2011 (28)	Finland (city not specified)	805	Pregnant women who have undergone successful infertility treatment; Mean age 33.1	planned or unplanned CS delivery; personal history of chronic severe mental illness; multiple pregnancy or in vitro fertilisation for current pregnancy; premature birth	<i>Inclusion criteria:</i> Finnish-speaking Intervention group: successful singleton pregnancy after infertility treatment Control group: no infertility history <i>No other exclusion criteria</i>	Antenatal, 18-20 weeks gestation	GHQ-36: cut-off of 9 clinical criterion for psychiatric disorder; BDI-13: cut-off of 5 suggesting mild depression (after recoding items)	3 assessments First assessment: 18-20 weeks gestation; Last assessment: 12 months postpartum
------------------------	------------------------------	-----	---	--	---	----------------------------------	--	--

AMP: Active Mothers Postpartum; BDI: Beck Depression Inventory; BMI: Body mass index; CES-D-20: Centre for Epidemiological Studies – Depression scale; EDD: expected date of delivery; EPDS: Edinburgh Postnatal Depression Scale; GHQ-36: General Health Questionnaire – 36 items; KAN-DO: Kids and Adults Now! – Defeat Obesity; LHPA: limbic hypothalamic pituitary axis; WIC: Women, Infants and Children (financial assistance for pregnant, postpartum and breastfeeding women); US: United States of America.

Table 3. Statistical methods of the 11 studies included in the systematic review

Study	Data used	Intervention	Aim of analysis	Analytical approach	Variables controlled for	Method used for missing data	Considerations in model selection	Study quality
Christensen et al, 2011 (29)	Secondary (RCT)	Preventive intervention for perinatal depression	Trajectories, risk profiles and outcomes	Trajectories: growth mixture modelling (Mplus); Risk factors: multinomial logistic regression	Demographic, psychosocial characteristics and randomization status for adjusted logistic regression models	Not specified	Information criteria fit indices: BIC and ABIC Estimated posterior probabilities: entropy, sample size of latent trajectories Likelihood ratio tests: BLRT	+
Glasheen et al, 2013 (21)	Secondary	None	Trajectories, risk profiles and outcomes	Trajectories: growth mixture modelling (Mplus); Risk factors: logistic and multinomial regression analyses	Demographic characteristics, social support, pregnancy, labour and delivery complications, substance use for regression models	Not specified	Information criteria fit indices: BIC, AIC, Estimated posterior probabilities: entropy Likelihood ratio tests: LMR*	+++
Kuo et al, 2012	Primary	None	Trajectories and	Trajectories: growth	Parity, education,	PROC TRAJ	Information criteria fit	++

(26)			risk profiles	mixture modelling (SAS); Risk factors: multinomial logistic regression	prenatal employment, prenatal exercise, mode of birth, sleep quality for regression models Age, parity, education, pregnancy BMI, use of patient controlled analgesics (PCAs) and sleep quality for regression models	(SAS): maximum likelihood estimation	indices: BIC	
Kuo et al, 2014 (25)	Primary	None	Trajectories and risk profiles	Trajectories: group-based trajectory modelling (SAS); Risk factors: logistic regressions		PROC TRAJ (SAS): maximum likelihood estimation	Information criteria fit indices: BIC Sample size of latent trajectories	+++
Lee et al, 2014 (30)	Secondary (RCT)	Weight loss (diet and exercise)	Trajectories and risk profiles	Trajectories: latent growth modelling and latent class growth analysis (Mplus); Risk factors: multinomial logistic regression	Maternal BMI, parity, study (KAN-DO vs. AMP) and arm (control vs. intervention) for LCGA and regression models	Full information maximum likelihood OR expectation maximisation algorithm	Information criteria fit indices: BIC Estimated posterior probabilities: entropy Sample size of latent trajectories	++
Marcus et al, 2011 (22)	Primary	None	Trajectories and outcomes	Trajectories: mixture growth curve approach (SAS)	n/a	Imputed missing values using Proc MI (SAS)	Information criteria fit indices: BIC	+
Mora et al, 2009 (24)	Secondary (Cohort)	None	Trajectories and risk profiles	Trajectories: growth mixture modelling (MPlus); Risk factors: bivariate analyses (chi square and analysis of variance), multinomial regression	Adjusting for all covariates (maternal characteristics) in regression models	Expectation-maximization algorithm	Estimated posterior probabilities: entropy Information criteria fit indices: BIC, AIC, ABIC Likelihood ratio tests: BLRT Sample size of latent trajectories	+++
Parade et al, 2014 (23)	Primary	None	Trajectories and risk profiles	Trajectories: unconditional latent class growth analysis (using Mplus); Risk factors: analyses not specified	Education in latent class growth analysis; education, family income, romantic relationship length and type in regression models	Not specified	Information criteria fit indices: BIC Likelihood ratio tests: BLRT	+
Ramos-	Second	Promoting	Trajectories	Trajectories:	RCT arm	Not	Information	++

Marcuse et al, 2010 (31)	Primary (RCT)	parenting and adolescent development	Outcomes	group-based modelling (semiparametric - using PROC TRAJ, in SAS); Risk factors: polynomial function, analysis of variance and pairwise comparisons	not controlled for (no difference in depressive symptoms); arm allocation in post-hoc analyses	specified	criteria fit indices: BIC Sample size of latent trajectories	
Sutter-Dallay et al, 2012 (27)	Secondary (Cohort)	None	Trajectories and risk profiles	Trajectories: semiparametric mixture models using PROC TRAJ (SAS); Risk factors: multinomial logistic regression	Adjusting for all covariates in the regression model; education excluded from regression due to collinearity	Not specified	Information criteria fit indices: BIC Average posterior probability Sample size of latent trajectories	++
Vänskä et al, 2011 (28)	Primary	None	Trajectories, risk profiles and outcomes	Trajectories: mixture modelling (Mplus); Risk factors: ANCOVA	Current psychological distress (based on GHQ36) and parity at recruitment in ANCOVA	Missing-data method (Mplus)	Information criteria fit indices: BIC, AIC, ABIC, Likelihood ratio tests: VLMR, LMR and BLRT Individual and average posterior probabilities	++

ABIC sample size adjusted Bayesian information criterion; AIC: Akaike information criterion; BIC: Bayesian information criterion; BLRT: bootstrap likelihood ratio test; CBT: cognitive behavioural therapy; LMR: Lo-Mendell-Rubin adjusted likelihood ratio test; RCT: randomised controlled trial; VLMR: Vuong-Lo-Mendell-Rubin likelihood ratio test.

Table 4. Main findings of the 11 studies included in the systematic review

Study	Attrition rate	Classes	Trajectory labels and size
Christensen et al, 2011 (29)	Not specified	3	1. Pregnancy high – high symptom levels antenatally, drops postnatally below risk cut-off (9.8%) 2. Postpartum high – near cut-off antenatally, marked increase postnatally, decrease to initial levels at 12-month postpartum (10.2%) 3. Perinatal low – never exceeds cut-off during pregnancy and postpartum period (80.0%)
Glasheen et al, 2013 (21)	Range of follow-up rate: 76-82%	2	1. Low symptom group – low symptom levels, stable but small decrease over time (16.5%) 2. High symptom group – higher scores, stable but small decrease over time (83.5%)
Kuo et al, 2012 (26)	One participant missing at first assessment,	4	1. Low levels of depressive symptoms antenatally, slight decrease in first three days after birth, slight increase one week after birth (23.1%) 2. Relatively low antenatally, slight decrease one day after birth and stabilises (43.0%)

	7 missing at last assessment		3. Moderate stable levels antenatally and postnatally (25.6%) 4. High scores antenatally, increases one day after birth, decreases 3 days after birth, slight increase to original levels at one week postpartum (8.3%)	High risk group
Kuo et al, 2014 (25)	All participants completed 4 assessments, 102 (73%) completed 5 assessments	3	1. Low depression – low levels stable over postpartum period (30.9%) 2. Mild depression – mild levels stable over postpartum period (41.7%) 3. High depression – high levels antenatally, slight decrease in first month postpartum, then stable (27.3%)	
Lee et al, 2014 (30)	Mean assessments : AMP: 2.5; KAN-DO: 1.86	3	1. Stable-low symptoms throughout postpartum period (82.5%) 2. Decreasing symptoms throughout postpartum period (7.3%) 3. Increasing symptoms throughout postpartum period (10.2%)	
Marcus et al, 2011 (22)	140 (91%) participants had at least 2 assessments	3	1. Low depression - low stable non-depressive during pregnancy (36.0%) 2. Intermediate depression - intermediate-stable depressive during pregnancy (56.0%) 3. High depression - high-elevated depressive during pregnancy (8.0%)	
Mora et al, 2009 (24)	More than 85% completed at least 2 assessments, 48% completed 4 assessments	5	1. Chronic – persistently high level of depressive symptoms antenatally and postnatally (7.0%) 2. Antepartum - depressive symptomatology present only at first antenatal visit (6.0%) 3. Postpartum – depressive symptoms present within 6 weeks of delivery, subsides over time (9.0%) 4. Late – low levels of depressive symptoms antenatally, increase in second year postpartum (7.0%) 5. Never – continuous low levels of depressive symptoms (71.0%)	
Parade et al, 2014 (23)	Not specified	2	1. Consistently low levels of depressive symptoms antenatally and postnatally 2. Elevated levels of depressive symptoms during pregnancy, temporary decline around birth, elevated again at 6-month postpartum	
Ramos-Marcuse et al, 2010 (31)	82% did 2 assessments, 70% did all 3 assessments	3	1. Low depressive symptoms – stable low symptom levels (40.9%) 2. Medium depressive symptoms - just below cut-off after birth, decrease over postpartum period, remain below cut-off at 6 month postpartum (45.3%) 3. High depressive symptoms – symptom levels above cut-off which increase over the postpartum period (13.8%)	
Sutter-Dallay et al, 2012 (27)	Not specified	4	1. Postpartum - lowest levels of the sample in third trimester, increase rapidly to reach maximum level at one year postpartum (4.0%) 2. Never * – below cut-off with low decrease over postpartum period (72.0%) 3. Antepartum – high average levels during pregnancy, decrease until one year postpartum (still above cut-off), increase slightly after that (21.0%) 4. Chronic – stable high symptom levels from the end of pregnancy to 2 years postpartum (3.0%)	
Vänskä et al, 2011 (28)	788 (98%) completed first assessment, 81.6% the second and, 67.7% the third; 65.7% completed all assessments	8 (4 + 4 combined in 1)	1. Stable low levels of mental health problems antenatally and postnatally (75.7%) 2. Prenatal mental health problems (5.8%) 3. Early postpartum mental health problems (8.7%) 4. Late postpartum mental health problems (5.6%) 5. Heterogeneous high levels of mental health problems (combined group of 4 classes) (4.2%)	

Table 5. Factors associated with the trajectories reported in the 11 studies included in the review

Study ^a	Stable			Transient			
	Low	Moderate low	Moderate high	High	Increasing	Decreasing	Episodic
1		Reference group				Lower social support	No health insurance History of abuse
2		Reference group	Lower social support Cigarette smoking (>1 pack a day)				
3	Reference group			Sleep difficulty in third trimester			
4	Reference group	No associations	Sleep difficulty in third trimester No patient-controlled analgesics				
5	Reference group				Poorer self-reported physical health	No associations	
6 ^b	n/a	n/a	n/a				
7		Reference group		Ethnicity (white vs. Black non-Hispanic) Non-nulliparous Poor/fair self-rated emotional health Anxious about pregnancy Moderate/high objective stress	Education (<high school) Moderate/high objective stress	Not US-born Ethnicity (white vs Latina/Hispanic) Poor/fair self-rated emotional health Recent alcohol use Anxious about pregnancy High objective stress	Ethnicity (white vs. Black non-Hispanic) Parity (1-2 children) Education (≤ high school) Poor/fair self-rated emotional health Comorbid disorder Moderate/high objective stress
8	Reference group						Lower education Decreased remembered paternal care Conflict resolution strategies

9 ^b	n/a	n/a	n/a			
10		Reference group	Older age Higher trait anxiety Lower income Nulliparous	Higher trait anxiety	No associations	
11	Reference group	Multiparity	No associations		No associations	Multiparity

^a 1 Christensen et al. (2011); 2 Glasheen et al. (2013); 3 Kuo et al. (2012); 4 Kuo et al. (2014); 5 Lee et al. (2014); 6 Marcus et al. (2011); 7 Mora et al. (2009); 8 Parade et al. (2014); 9 Ramos-Marcuse et al. (2010); 10 Sutter-Dallay et al. (2012); 11 Vänskä et al. (2009). ^b Risk factors not investigated in the study.

Accepted manuscript