Detection of differential depressive symptom patterns in a cohort of perinatal women: an exploratory factor analysis using a robust statistics approach

Kenneth J. Nieser,[a] Zachary N. Stowe,[b] D. Jeffrey Newport,[c,d] Jessica L. Coker,[e,f] and Amy L. Cochran[a,g,*]

[a]Department of Population Health Sciences, University of Wisconsin–Madison, United States
[b]Department of Psychiatry, University of Wisconsin–Madison, United States
[c]Department of Psychiatry & Behavioral Sciences, University of Texas at Austin Dell Medical School, United States
[d]Department of Women's Health, University of Texas at Austin Dell Medical School, United States
[e]Department of Psychiatry, University of Arkansas for Medical Sciences, United States
[f]Department of Obstetrics & Gynecology, University of Arkansas for Medical Sciences, United States
[g]Department of Mathematics, University of Wisconsin–Madison, United States

eClinicalMedicine
2023;57: 101830

Published Online 1 February 2023
https://doi.org/10.1016/j.eclinm.2023.101830

Summary

Background Postpartum depression can take many forms. Different symptom patterns could have divergent implications for how we screen, diagnose, and treat postpartum depression. We sought to utilise a recently developed robust estimation algorithm to automatically identify differential patterns in depressive symptoms and subsequently characterise the individuals who exhibit different patterns.

Methods Depressive symptom data (N = 548) were collected from women with neuropsychiatric illnesses at two U.S. urban sites participating in a longitudinal observational study of stress across the perinatal period. Data were collected from Emory University between 1994 and 2012 and from the University of Arkansas for Medical Sciences between 2012 and 2017. We conducted an exploratory factor analysis of Beck Depression Inventory (BDI) items using a robust expectation-maximization algorithm, rather than a conventional expectation-maximization algorithm. This recently developed method enabled automatic detection of differential symptom patterns. We described differences in symptom patterns and conducted unadjusted and adjusted analyses of associations of symptom patterns with demographics and psychiatric histories.

Findings 53 (9.7%) participants were identified by the algorithm as having a different pattern of reported symptoms compared to other participants. This group had more severe symptoms across all items—especially items related to thoughts of self-harm and self-judgement—and differed in how their symptoms related to underlying psychological constructs. History of social anxiety disorder (OR: 4.0; 95% CI [1.9, 8.1]) and history of childhood trauma (for each 5-point increase, OR: 1.2; 95% CI [1.1, 1.3]) were significantly associated with this differential pattern after adjustment for other covariates.

Interpretation Social anxiety disorder and childhood trauma are associated with differential patterns of severe postpartum depressive symptoms, which might warrant tailored strategies for screening, diagnosis, and treatment to address these comorbid conditions.

Funding There are no funding sources to declare.

Keywords: Postpartum depression; Factor analysis; Self-injurious behavior; Adverse childhood experiences; Social phobia; Symptom heterogeneity; Robust statistics

*Corresponding author. Department of Population Health Sciences, University of Wisconsin–Madison, 610 Walnut Street, Madison, WI, 53726, United States.
E-mail address: cochran4@wisc.edu (A.L. Cochran).

Research in context

Evidence before this study

Previous research has pointed out the heterogeneity of postpartum depressive symptoms; yet clinical decisions are guided largely by whether a sum-score exceeds a threshold. We searched PubMed and PsycInfo for articles from inception through November 11, 2021 using the advanced search logic: (depression) AND ((perinatal) OR (postpartum)) AND ((phenotypes) OR (heterogeneity) OR (measurement invariance) OR (model fit)). We identified 30 original articles from PubMed and an additional 6 from PsycInfo that were relevant to our study. Previous studies have suggested data-driven clinical subtypes of perinatal depression which were characterised by symptom severity, timing of onset of depressive symptoms, comorbid anxiety, and thoughts of self-harm. Gaps in previous research include the lack of analysis of how factor structures vary across subgroups and in the limited set of symptoms included in analyses. Subtypes might be more well-characterised by variables that were not included in previous analyses, such as somatic symptoms.

Added value of this study

We conducted an exploratory factor analysis of depressive symptom data from the Beck Depression Inventory, using an analytic method that can automatically detect differences in factor model parameters. We found that a homogeneous two-factor model did not fit the entire sample well; about 10% of our sample had symptoms that differed in their factor structure. This subgroup differed in how their depressive symptoms related to underlying psychological constructs and had more severe responses for all items—particularly for items related to thoughts of self-harm and negative self-judgement. History of childhood trauma and history of social anxiety disorder were strongly associated with this differential symptom pattern.

Implications of all the available evidence

Postpartum depressive symptom patterns vary across the population. Timing of onset of depressive symptoms, psychiatric comorbidities, and trauma history all can affect how an individual experiences and reports postpartum depressive symptoms. Screening tools and treatment strategies should be developed to accommodate the varied needs of perinatal women with depressive symptoms.

Introduction

Depression during pregnancy and the postpartum period merits special attention.[1] Perinatal depression impacts maternal health, impairs maternal-child bonding and caregiving, and has been associated with psychological and developmental disorders in offspring.[2,3] Approximately 9–10% of postpartum women in high income countries and 16–19% of postpartum women in low- and middle-income countries experience major depressive disorder according to diagnostic criteria.[4] Screening for depression during the perinatal period identifies individuals who can benefit from intervention and has been recommended by the US Preventive Services Task Force,[5] the American College of Obstetricians and Gynecologists,[6] the National Institute for Health and Care Excellence in the U.K.,[7] and the Centre of Perinatal Excellence in Australia.[8]

Evidence suggests that perinatal depression is heterogeneous[9]; varying symptom profiles potentially reflect distinct clinical phenotypes.[10–13] Like depression outside of the perinatal period, different clinical phenotypes might be associated with distinct risk factors, pathophysiology, and treatment response. Using data from an international perinatal psychiatry consortium—the Postpartum Depression: Action Towards Causes and Treatment (PACT) Consortium—two previous studies applied clustering methods to participant responses from the Edinburgh Postnatal Depression Scale (EPDS), a widely used screening instrument for perinatal depression.[14] One of the studies used latent class analysis (LCA) to detect three latent classes of increasing symptom severity[10] and the other used k-means clustering of EPDS latent factors (depression, anxiety, and anhedonia) to detect five latent clusters, which they described as severe anxious depression, moderate anxious depression, resolved depression, anxious anhedonia, and pure anhedonia.[11] In a study conducted with perinatal women in China, an LCA of the Patient Health Questionnaire-9 (PHQ-9) found five latent classes distinguished by symptom severity and presence of somatic symptoms.[12] Taken together, previous studies seem to suggest that symptom severity, comorbid anxiety, timing of symptom onset, and suicidal ideation are important distinguishing features of perinatal depression.[10,11] In addition, depressive symptom patterns can be dynamic throughout the perinatal period,[12] with the time around delivery being particularly salient.[15] Vital groundwork has been laid in this area, but further characterization of symptom patterns, baseline characteristics, and comorbid conditions could help clinicians personalise treatments.

Different methodological approaches might pick up on distinct aspects of diversity in symptom patterns. Depression inventories are not perfectly interchangeable with one another; each measures slightly different aspects of psychopathology.[16] The reporting of symptoms in the EPDS varies across sociocultural contexts[17,18] and the instrument excludes common somatic symptoms of depression with the justification that the physiological changes that women experience throughout the perinatal period might confound the interpretation of individual items. Despite their ambiguity, somatic

symptoms can still be pertinent indicators of postpartum depression[19] and might be expressed more often by some populations than others.[20] To investigate symptoms outside of the EPDS, we analysed data from the Beck Depression Inventory (BDI),[21] a common general depressive symptom inventory which has also been shown to be highly predictive of major depressive episodes throughout the perinatal period.[22]

The choice of analytic method could also influence which subgroups are uncovered. When subgroups, such as ethnic groups, are pre-specified, results are limited by this choice. Approaches such as latent class analysis and factor mixture analysis address this by allowing subgroups to empirically arise from the data, although inferences can be limited by various modelling assumptions. For example, latent class analyses often assume that items are uncorrelated within each class. Factor mixture models offer a rich analysis of data[23] but can yield inferences that are sensitive to the chosen number of factors or classes. As an alternative, we employed a robust expectation-maximization (REM) algorithm[24] in an otherwise typical exploratory factor analysis, which enabled automatic detection of differences in factor structure across the sample and yielded inferences about how symptoms were correlated within latent groups. This approach can be used as a diagnostic step in data analysis that finds discrepancies in model fit across the sample, prompting model re-specification or prioritization of population subgroups for future data collection.

The primary goals of our study were to: (1) detect if significant differential patterning existed in the depressive symptom patterns in our sample and if so, (2) describe symptom patterns and (3) examine associations with demographics and psychiatric histories. A secondary goal was to demonstrate how robust estimation, such as the REM algorithm, can be employed as a diagnostic tool for automatically identifying heterogeneity in symptom patterns.

Methods

Participants

The sample consisted of perinatal women participating in a longitudinal observational study of neuropsychiatric illnesses across the perinatal period through the Women's Mental Health Program at Emory University between 1994 and 2012 (referred to as Emory site) and at the University of Arkansas for Medical Sciences between 2012 and 2017 (referred to as UAMS site). Women were referred to tertiary referral centres specializing in perinatal psychiatric evaluation and treatment and were enrolled prior to delivery. The studies were approved by Institutional Review Boards at Emory and UAMS. All participants provided informed consent.

In this paper, we conducted secondary analyses which focused on the first 13 weeks following delivery. The two sites were analysed together to increase sociodemographic diversity and sample size. Given the difficulties of measurement of depression during the perinatal period, we included participants with a lifetime diagnosis of major depressive disorder based on the Structured Clinical Interview for the DSM-IV[25] (SCID), which was administered at study entry. Women with a history of depressive symptoms are at higher risk for depression during the perinatal period,[26] and so we chose to focus on this subset of perinatal women, rather than only those with a current clinical diagnosis, to include a broad range of depressive symptom patterns. This enhances the representativeness of the sample beyond women with a clinical diagnosis during the perinatal period to a broader population of perinatal women suffering from depressive symptoms throughout the life course. Participants were excluded if they had a lifetime diagnosis of bipolar disorder (I, II, or other), schizophrenia, or schizophreniform; did not have a viable birth; or were missing one or more BDI items. In addition, participants with twin births were excluded because we thought that their experiences might be sufficiently different to warrant a separate analysis, and participants with less than 7 years of education were excluded to ensure participants had sufficient reading comprehension.

Depressive symptoms

Depressive symptoms were self-reported using the BDI original version,[21] which is predictive of major depressive episodes throughout the perinatal period.[22] The BDI consists of 21 items that can be self-rated on a 4-point scale of increasing severity from 0 to 3. Commonly, the BDI decomposes into three symptom clusters: negative attitude towards self and negative affect (self-hate, sense of failure, guilty feeling, self-accusation, sense of punishment, thoughts of self-harm, pessimism, body image, sadness, lack of satisfaction, crying spells); performance impairment or anhedonia (fatigue, difficulty working, social withdrawal, irritability, somatic concern, libido loss, indecisiveness); and somatic symptoms (appetite change, weight loss, sleep disturbance).[27,28] If a participant completed the BDI more than one time in the study period, we used data only from one time point for each participant, chosen uniformly at random. Given its widespread use, we also measured depressive symptoms using the EPDS and compared findings.

Independent variables

The following variables were collected at baseline: maternal age at delivery, years of education, self-classified race and ethnicity, and marital status. Lifetime diagnoses of major depressive disorder, panic disorder, social anxiety disorder, generalised anxiety disorder, obsessive–compulsive disorder, and post-traumatic stress disorder were measured at baseline using the SCID.[25] Experiences of childhood trauma were

measured using the Childhood Trauma Questionnaire-Short Form (CTQ-SF), which can be decomposed into five subscales: physical abuse, sexual abuse, emotional abuse, physical neglect, and emotional neglect.[29]

Statistical analysis

We conducted exploratory factor analyses of the BDI items in two ways using custom source code in MATLAB 2021a.[30] We obtained estimates of item intercepts, factor loadings, and residual variances using the expectation-maximization (EM) algorithm[31]—a standard approach to maximum likelihood estimation in factor analysis models—and the robust expectation-maximization (REM) algorithm.[24] Technical details of these two estimation algorithms can be found in the Supplementary Appendix. We selected the hyperparameter for REM estimation using the recommended heuristic with $\delta = 0.05$. For model identification, we fixed factor means and covariances to zero and factor variances to one for both EM and REM. To select the number of factors, we considered prior subject matter knowledge, eigenvalues of the sample correlation matrix, magnitude of factor loadings (≥0.40), and two modified Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) evaluated at REM estimated values.[24] A simulation study suggests that for factor analyses with moderate loadings (0.50), 2 or 3 factors, and 8 or more indicators, a sample size of 150–200 yields minimal bias and at least 80% statistical power (alpha = 0.05) for all model parameters.[32] Parameter estimates were rescaled to standard deviation units of the items to enable comparisons between EM and REM estimates on the same scale; varimax rotations were applied to factor loadings.

The REM algorithm has been demonstrated in simulated factor analyses to be less sensitive to violations in modelling assumptions and can be used to automatically identify disparity in how well a model fits across a sample, without requiring researchers to fit mixture models.[24] The REM algorithm down-weights the contribution of data points that poorly fit the specified model with informative, probabilistic weights that can be examined a posteriori. To examine differences in symptom patterns, we compared item intercepts and factor loadings based on the EM algorithm to those based on the REM algorithm. We defined groups based on estimated REM weights and compared frequencies of severe responses (2 or 3) for each item across groups. Using R,[33] we performed unadjusted and adjusted analyses of the associations of demographics and psychiatric histories with down-weighting in the REM algorithm. We performed several additional analyses to assess consistency and generalizability of our results. We repeated analyses stratifying by site to investigate whether our results were specific to a site or generalizable across sites. We repeated analyses using data from EPDS rather than BDI, because of its prominent usage in clinical practice. Lastly, we fit exploratory factor mixture models on BDI data using Mplus Version 8.4[34] to assess whether alternative modelling strategies would reveal the same latent subgrouping.

Role of the funding source

There was no funding source for this study. All authors had access to the dataset and agreed to submit for publication.

Results

There were 681 women with a history of major depressive disorder in the sample. Based on exclusion criteria, 133 women were excluded from analysis, leaving 548 women in the analytic sample (426 from Emory and 122 from UAMS). Of the 681 women, 16 were excluded for having twin births, 38 did not have a viable birth, three had less than 7 years of education, one had less than 7 years of education and did not have a viable birth, three had excluding diagnoses, and 72 were missing data on one or more BDI items. Of the participants missing data on one or more BDI items, 68 out of 72 were missing data on all BDI items.

Baseline characteristics of participants at each site are summarised in Table 1. Compared to participants from the Emory site, participants from the UAMS site were younger on average, were less likely to be married, had fewer years of education, and had higher CTQ-SF scores. Most participants from both sites identified as white and non-Hispanic. The UAMS site had a greater proportion of participants with histories of social anxiety disorder (SAD) and posttraumatic stress disorder, while the Emory site had a greater proportion of participants with histories of panic disorder, generalised anxiety disorder, and obsessive–compulsive disorder.

The sample correlation matrix of the BDI items had three eigenvalues greater than one, suggesting a three-factor model. We fit factor models with one to six latent factors. AIC-based criteria suggested a four-factor model, while BIC-based criteria suggested a two-factor model. In the three-factor model, only one item (loss of appetite) had a substantial loading (≥0.40) on the third factor, so we ultimately decided on a two-factor model.

Out of the 548 participants, 53 (9.7%) were down-weighted by the REM algorithm. Estimated weights were effectively either zero or one, up to four decimal places. We dichotomised the sample into two groups, one with REM estimated weight approximately equal to zero and the other with REM estimated weight approximately equal to one.

Analysis of depressive symptom patterns

The down-weighted group tended to have higher BDI total scores than the group that was not down-weighted (boxplots in Fig. A1). For all items except for item 19 (weight loss), responses from the down-weighted group indicated greater symptom severity on average than

Site	Emory (n = 426)			UAMS (n = 122)		
Variable	*M*	*SD*	Missing	*M*	*SD*	Missing
Maternal age at delivery	33.6	4.7	0	28.2	5.7	0
Education, years	16.3	2.1	0	13.5	2.4	0
Weeks postpartum	6.1	3.5	0	5.8	3.8	0
CTQ Emotional Abuse	10.0	4.8	24	12.4	6.1	4
CTQ Emotional Neglect	10.6	4.6	21	12.6	5.5	3
CTQ Physical Abuse	6.9	3.2	24	9.3	5.0	3
CTQ Physical Neglect	6.6	2.6	21	8.6	3.8	4
CTQ Sexual Abuse	6.8	4.2	24	10.4	7.2	5
CTQ Total	40.7	14.6	28	53.1	21.3	7
Variable	Count	%	Missing	Count	%	Missing
Race			1			3
Asian	13	3.1		1	0.8	
Black	29	6.8		21	17.2	
Multiple	1	0.2		9	7.4	
Native American	6	1.4		0	0.0	
White	376	88.3		88	72.1	
Ethnicity			0			0
Hispanic	9	2.1		7	5.7	
Non-Hispanic	417	97.9		115	94.3	
Marital Status			0			0
Divorced	9	2.1		9	7.4	
Married	366	85.9		55	45.1	
Never Married, Lives Alone	26	6.1		37	30.3	
Never Married, Lives w/Partner	17	4.0		20	16.4	
Separated	8	1.9		1	0.8	
Widowed	0	0.0		0	0.0	
Lifetime psychiatric diagnoses[a]						
Panic disorder	102	23.9	0	20	16.4	0
Social anxiety disorder	69	16.2	0	34	27.9	0
Generalised anxiety disorder	102	23.9	0	23	18.9	0
Obsessive compulsive disorder	61	14.3	0	6	4.9	0
Posttraumatic stress disorder	74	17.4	0	56	45.9	0

[a]Percentages will not add up to 100% because some women had multiple lifetime psychiatric diagnoses.

Table 1: **Demographics and psychiatric histories by site.**

responses from the rest of the sample at a significance level of 0.01 based on Mann–Whitney U tests (Fig. A2). Comparing the down-weighted group to other participants, the greatest discrepancy in relative proportions reporting the most severe response categories (values of 2 and 3) occurred for thoughts of self-harm (15.1% vs. 0.0%), self-hate (18.9% vs 1.0%), lack of satisfaction (22.6% vs 1.8%), sense of failure (37.7% vs 3.8%), and somatic preoccupation (26.4% vs 2.8%) items, shown in Table 2. For item 9 (thoughts of self-harm), there was complete separation between the two groups; participants in the down-weighted group were exactly the participants who had reported thoughts of self-harm. We performed several sensitivity analyses to validate that our results were not entirely due to the responses on the "thoughts of self-harm" item (details in Supplementary Appendix).

For both EM and REM estimation, item intercepts were generally the same for all items, except thoughts of self-harm (Table 3). The first factor seemed to correspond to cognitive-affective symptoms (e.g., pessimism, sense of failure, self-hate) and the second factor reflected inhibition or somatic symptoms (e.g. work inhibition, fatigue, sleep disturbance). Several items loaded on different factors during the REM estimation compared to the EM estimation, indicating variation in correlations of symptoms with underlying factors (Table 3). Based on a loading cut-off of 0.40, thoughts of self-harm and body image items loaded on the first factor with EM estimation and did not load on either factor with REM; sadness and social withdrawal items cross-loaded on both factors with EM but only loaded on the second factor with REM; irritability and loss of libido items loaded on the second factor with EM estimation

Item[a]	Weight ≈ 0[b] Count (%)[c]	Weight ≈ 1[b] Count (%)[c]	Ratio[d]
Self-harm	8 (15.1)	0 (0)	–
Self-hate	10 (18.9)	5 (1)	18.7
Lack of satisfaction	12 (22.6)	9 (1.8)	12.5
Sense of failure	20 (37.7)	19 (3.8)	9.8
Somatic preoccupation	14 (26.4)	14 (2.8)	9.3
Sense of punishment	12 (22.6)	13 (2.6)	8.6
Sadness	21 (39.6)	23 (4.6)	8.5
Pessimism	16 (30.2)	19 (3.8)	7.9
Self-accusations	15 (28.3)	20 (4)	7.0
Social withdrawal	16 (30.2)	23 (4.6)	6.5
Crying	14 (26.4)	22 (4.4)	5.9
Indecisiveness	16 (30.2)	27 (5.5)	5.5
Loss of appetite	19 (35.8)	33 (6.7)	5.4
Irritability	20 (37.7)	41 (8.3)	4.6
Guilty feeling	21 (39.6)	45 (9.1)	4.4
Fatigue	25 (47.2)	58 (11.7)	4.0
Work inhibition	20 (37.7)	47 (9.5)	4.0
Body image	27 (50.9)	81 (16.4)	3.1
Sleep disturbance	18 (34)	78 (15.8)	2.2
Loss of libido	29 (54.7)	127 (25.7)	2.1
Weight loss	13 (24.5)	127 (25.7)	1.0

[a]Table is sorted descending by ratio column to highlight items with greatest discrepancies. [b]"Weight" refers to the individual estimated weights from the REM algorithm. Estimated weights were effectively either 0 or 1. These indicate how well the fitted factor model describes each individual's symptom data; lower weight indicates poorer model fit. [c]Column data includes responses of 2 or 3. [d]For each item, the ratio column shows the proportion of individuals with weight 0 who had severe symptoms (scored 2 or 3) divided by the proportion of individuals with weight 1 who had severe symptoms. For example, individuals with weight 0 were 18.7 times as likely to indicate severe responses for "self-hate" compared to individuals with weight 1.

Table 2: **Comparison of frequency of severe responses between REM groups.**

and did not load on either factor with REM; and loss of appetite and somatic preoccupation items loaded on the first factor with EM and on the second with REM.

Associations with demographics and psychiatric histories

Unadjusted, bivariate analyses of demographics and psychiatric histories between the two groups are shown in Table 4. On average, participants in the down-weighted group were younger; had fewer years of education; and had higher CTQ-SF scores on each of the five subscales. Compared to the majority, participants in the down-weighted group were more likely to be from the UAMS site; more likely to identify as Black; and more likely to be never married, living alone. Lastly, participants in the down-weighted group were more likely to have histories of SAD.

We fit an adjusted logistic regression model of group membership (Table 5). After adjustment for all other variables, we found strong evidence that greater CTQ-SF score (for each 5 point increase, OR: 1.2; 95% CI [1.1, 1.3]; p = 0.002) and a history of SAD (OR: 4.0; 95% CI [1.9, 8.1]; p = 0.0002) were associated with being in the down-weighted group. There was some evidence that each additional year of age at delivery was associated with not being down-weighted (OR: 0.9; 95% CI [0.9, 1.0]; p = 0.033).

Site-stratified analysis

Our results were consistent across both the Emory and UAMS sites, suggesting that the identified subgroup is not specific to only one of the sites in our study. After fitting exploratory two-factor models using the REM algorithm separately by site, we found that the same individuals were down-weighted as those in the combined data sample. Unadjusted and adjusted analyses of associations between group membership and demographic and psychiatric histories can be found in the Supplementary Appendix. At the Emory site, women identifying as Black were more likely to be down-weighted (OR: 3.8; 95% CI [1.3, 9.8]) to a similar degree as in the overall analysis; however, this association was weaker at the UAMS site (OR: 1.5; 95% CI [0.4, 4.6]). There was a larger increase in mean CTQ-SF scores for the down-weighted group relative to the rest of the sample at UAMS (16.1; 95% CI [4.7, 27.5]) compared with the down-weighted group relative to the rest of the sample at Emory (6.2; 95% CI [−1.5, 14.0]). In adjusted analyses, the associations with age, childhood trauma, and SAD were all stronger at the UAMS site compared to the Emory site.

Comparison with EPDS results

Analysis of EPDS data yielded very similar results in terms of which individuals were down-weighted (details in Supplementary Appendix). Unlike the BDI analysis, there were not clear differences in the factor model parameter estimates in the EPDS analysis except for the "thoughts of self-harm" item.

Comparison with factor mixture analysis

We conducted an exploratory factor mixture analysis using Mplus version 8.4 (details in Supplementary Appendix). Unlike in EM and REM estimation, we treated indicators as categorical variables rather than continuous. We found that a three-factor model had the lowest BIC out of all fitted models and Lo-Mendell–Rubin tests indicated that factor mixture models with more than one latent class did not explain the data significantly better than factor mixture models with a single latent class. This would suggest that factor mixture analyses were unnecessary. Among the fitted factor mixture models with multiple latent classes, a model with two latent classes, each with their own one-factor model, had the lowest BIC and strong classification quality (entropy = 0.92). Latent class 1 was the most likely class for 463 (84%) participants, while latent class 2 was most likely for 85 (16%) participants. There was a strong

Item	Description	EM Estimates[a]				REM Estimates[a]			
		Intercept	Cognitive/Affective	Somatic/Inhibition	Residual variance	Intercept	Cognitive/Affective	Somatic/Inhibition	Residual variance
1	Sadness	0.68	**0.66**	**0.40**	0.40	0.65	**0.61**	0.36	0.50
2	Pessimism	0.58	**0.72**	0.34	0.36	0.53	**0.69**	0.28	0.44
3	Sense of failure	0.57	**0.79**	0.24	0.32	0.51	**0.75**	0.20	0.40
4	Lack of satisfaction	0.61	**0.58**	**0.50**	0.41	0.55	**0.55**	**0.50**	0.46
5	Guilty feeling	0.61	**0.70**	0.29	0.43	0.56	**0.71**	0.22	0.45
6	Sense of punishment	0.42	**0.62**	0.17	0.58	0.36	**0.49**	0.20	0.72
7	Self-hate	0.72	**0.77**	0.22	0.36	0.68	**0.76**	0.11	0.41
8	Self-accusations	0.83	**0.71**	0.33	0.39	0.78	**0.76**	0.23	0.37
9	Thoughts of self-harm	0.30	**0.55**	0.17	0.67	0.00	0.00	0.00	1.00
10	Crying	0.69	**0.44**	0.37	0.67	0.63	**0.40**	0.34	0.72
11	Irritability	0.95	0.39	**0.40**	0.69	0.91	0.35	0.35	0.75
12	Social withdrawal	0.68	**0.46**	**0.60**	0.42	0.63	0.39	**0.59**	0.50
13	Indecisiveness	0.70	**0.53**	**0.45**	0.52	0.65	**0.50**	**0.41**	0.59
14	Body image	0.91	**0.44**	0.29	0.72	0.86	0.38	0.24	0.79
15	Work inhibition	0.85	**0.43**	**0.58**	0.48	0.80	**0.42**	**0.53**	0.54
16	Sleep disturbance	0.99	0.16	**0.53**	0.70	0.97	0.18	**0.48**	0.74
17	Fatigue	1.13	0.23	**0.69**	0.47	1.10	0.26	**0.62**	0.54
18	Loss of appetite	0.53	**0.40**	0.37	0.70	0.48	0.23	**0.45**	0.75
19	Weight loss	0.70	0.02	0.12	0.99	0.69	−0.02	0.15	0.98
20	Somatic preoccupation	0.57	**0.45**	0.37	0.66	0.53	0.24	**0.44**	0.75
21	Loss of libido	0.88	0.28	**0.41**	0.75	0.84	0.25	0.39	0.78

[a]These estimates were obtained after scaling the data to have mean 0 and variance 1. The intercept column indicates the mean response, cognitive/affective and somatic/inhibition columns indicate the factoring loadings for the respective latent factor, and the residual variance column indicates the unique variances of item responses. Factor loadings have been rotated using the varimax procedure. Factor loadings greater than or equal to 0.40 are bolded.

Table 3: **Estimated factor model parameters.**

association between membership in the latent classes and membership in the REM-defined subgroups (OR = 5.9, $p = 8.6 \times 10^{-9}$). Probabilities of endorsing more severe response categories (2 and 3) were greater for latent class 2 for all items; the greatest discrepancies between latent class 1 and 2 occurred for thoughts of self-harm (0.0% vs 9.4%), lack of satisfaction (0.2% vs 23.5%), social withdrawal (0.9% vs 41.2%), pessimism (1.5% vs 32.9%), and self-accusations (1.5% vs 32.9%) items, shown in Table A5. In an adjusted logistic regression analysis (Table A6), a history of generalised anxiety disorder (OR: 2.1; 95% CI [1.1, 3.8]; p = 0.02) and a history of posttraumatic stress disorder (OR: 2.4; 95% CI [1.3, 4.4]; p = 0.003) was significantly associated with membership in latent class 2.

Discussion

We analysed depressive symptoms collected from postpartum women at two sites by conducting an exploratory factor analysis of BDI responses using a novel robust estimation approach. We sought to detect significant differential patterning of depressive symptoms and if present, describe how symptom patterns differed and examine associations with demographics and psychiatric histories. We found that a two-factor model did not fit all individuals in the sample equally well; individuals who were down-weighted (i.e., did not fit the model well) exhibited differential symptom patterns. The down-weighted group shared a similar symptom profile to a putative subtype of perinatal depression previously reported—more severe depressive symptoms and increased thoughts of self-harm.[10] In our study, we found that this subset was associated with histories of SAD and childhood trauma. In addition, the use of the BDI made apparent the increased prevalence of negative self-judgement among this subset. These findings suggest a need for screening and treatment strategies that accommodate different patterns of postpartum depressive symptoms.

The value of valid screening instruments is crucial in the identification of illnesses. However, striking the balance between sensitivity and specificity as well as having an instrument that is not overly burdensome may preclude recognition of subtypes germane to optimizing treatment plans. A total score of 10 for one individual might have different implications than the same total score for a different individual based on the pattern of symptoms endorsed. For example, we surmise that the heightened self-judgement cognitions reported by the subgroup identified in our sample could benefit from an applicable psychotherapy, such as dialectical behaviour therapy. By contrast, individuals who are not in the identified subgroup but otherwise score highly on a screening instrument may benefit from a different psychotherapy, such as

Group	Weight ≈ 0[a]		Weight ≈ 1[a]		Mean difference [95% CI][b]
Variable	*M*	*SD*	*M*	*SD*	
Maternal age at delivery	29.4	5.7	32.7	5.3	−3.3 [-4.9, −1.6]***
Education, years	14.2	2.7	15.8	2.3	−1.6 [-2.3, −0.8]***
Weeks postpartum	6.7	4.2	6.0	3.5	0.8 [-0.4, 2.0]
CTQ Emotional Abuse	14.1	6.2	10.1	5.0	4.0 [2.2, 5.8]***
CTQ Emotional Neglect	13.4	5.4	10.8	4.7	2.6 [1.1, 4.2]**
CTQ Physical Abuse	9.8	5.5	7.2	3.6	2.7 [1.1, 4.3]**
CTQ Physical Neglect	8.6	3.9	6.9	2.9	1.6 [0.5, 2.8]**
CTQ Sexual Abuse	10.1	7.2	7.3	4.9	2.7 [0.6, 4.8]*
CTQ Total	54.9	23.0	42.3	16.0	12.7 [5.8, 19.6]***
Variable	Count	%	Count	%	OR [95% CI][c]
Site					
Emory	31	58.5	395	79.8	ref
UAMS	22	41.5	100	20.2	2.8 [1.5, 5.0]***
Race					
Asian	1	1.9	13	2.6	0.8 [0, 4.4]
Black	11	20.8	39	7.9	3.1 [1.4, 6.3]**
Multiple	1	1.9	9	1.8	1.2 [0.1, 6.7]
Native American	0	0.0	6	1.2	–
White	39	73.6	425	85.9	ref
Ethnicity					
Hispanic	2	3.8	14	2.8	1.3 [0.2, 5.0]
Non-Hispanic	51	96.2	481	97.2	ref
Marital Status					
Divorced	1	1.9	17	3.4	0.7 [0, 3.5]
Married	33	62.3	388	78.4	ref
Never Married, Lives Alone	14	26.4	49	9.9	3.4 [1.6, 6.6]***
Never Married, Lives w/Partner	5	9.4	32	6.5	1.8 [0.6, 4.7]
Separated	0	0.0	9	1.8	–
Widowed	0	0.0	0	0.0	–
Lifetime psychiatric diagnoses[d]					
Panic disorder	10	18.9	112	22.6	0.8 [0.4, 1.6]
Social anxiety disorder	24	45.3	79	16.0	4.4 [2.4, 7.9]***
Generalised anxiety disorder	17	32.1	108	21.8	1.7 [0.9, 3.1]
Obsessive compulsive disorder	5	9.4	62	12.5	0.7 [0.2, 1.7]
Posttraumatic stress disorder	18	34.0	112	22.6	1.8 [0.9, 3.2]

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. [a]"Weight" refers to the individual estimated weights from the REM algorithm. Estimated weights were effectively either 0 or 1. These indicate how well the fitted factor model describes each individual's symptom data; lower weight indicates poorer model fit. [b]Welch's t-test. [c]Wald χ^2 test. [d]Percentages will not add up to 100% because some participants had multiple lifetime psychiatric diagnoses.

Table 4: **Demographics and psychiatric histories by estimated REM weight.**

behavioural activation. Using our method, a scoring algorithm could be used to make predictions about an individual's subgroup membership. As a proof-of-concept, we created a web application using the R Shiny package that would make these predictions based on BDI item responses (publicly available at https://knieser.shinyapps.io/app_subgroupppredict/). Before such a tool could be recommended clinically, the method used in this paper should be applied to a large representative sample of perinatal women to obtain representative parameter estimates.

We found that item responses from the down-weighted group were more severe on average across the board, particularly for items such as thoughts of self-harm, self-hate, lack of satisfaction, sense of failure, and somatic preoccupation. Social withdrawal, body image, loss of appetite, and somatic preoccupation items loaded with negative attitude or affect items in the EM estimation but not in the REM estimation. This suggests that these symptoms, or at least the questions, relate differently to underlying psychological constructs for the down-weighted group.

Variable	OR[a]	95% CI[a]	p[a]
Maternal age at delivery	0.9	[0.9, 1.0]	0.033*
Education, years	0.9	[0.8, 1.1]	0.204
Weeks postpartum	1.1	[1.0, 1.2]	0.105
CTQ Total (5-point increase)	1.2	[1.1, 1.3]	0.002**
Site			
Emory	ref	ref	ref
UAMS	0.8	[0.3, 1.9]	0.587
Race			
Asian	[b]	[b]	[b]
Black	0.7	[0.2, 1.9]	0.486
Multiple	[b]	[b]	[b]
Native American	[b]	[b]	[b]
White	ref	ref	ref
Ethnicity			
Hispanic	1.5	[0.1, 10.9]	0.748
Non-Hispanic	ref	ref	ref
Marital Status			
Divorced	[b]	[b]	[b]
Married	ref	ref	ref
Never Married, Lives Alone	1.8	[0.6, 5.1]	0.251
Never Married, Lives w/Partner	1.1	[0.3, 3.5]	0.909
Separated	[b]	[b]	[b]
Widowed	[b]	[b]	[b]
Lifetime psychiatric diagnoses			
Panic disorder	0.7	[0.3, 1.7]	0.450
Social anxiety disorder	4.0	[1.9, 8.1]	0.0002***
Generalised anxiety disorder	1.7	[0.8, 3.6]	0.169
Obsessive compulsive disorder	0.8	[0.2, 2.3]	0.680
Posttraumatic stress disorder	0.7	[0.3, 1.6]	0.403

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. [a]Results adjusted for all other variables. [b]Insufficient data for adjusted estimates.

Table 5: **Adjusted odds ratio estimates of down-weighting with REM.**

In our sample of two sites, we found that participants indicating thoughts of self-harm were more likely to be younger in age, identify as Black, have fewer years of education, and to be unmarried and living alone. This finding has important implications for racial and socioeconomic disparities in screening and treatment of postpartum depression. For example, if Black women are more likely to have severe symptoms of negative self-judgement and thoughts of self-harm, this might go unrecognised if only the sum-score from screening instruments are considered. Given that our sample was mostly non-Hispanic white women, it will be crucial for future studies in this area to better engage other racial and ethnic groups. After adjustment for several demographic factors and psychiatric histories, associations of subgroup membership with site, self-identified race, education, and marital status were no longer present, suggesting that these associations could be explained by one or more of the adjustment variables. In the adjusted analysis, we found that childhood trauma and a history of SAD were strongly associated with down-weighting in REM estimation and by extension, strongly associated with thoughts of self-harm and more severe depressive symptoms. Childhood trauma, particularly emotional abuse and neglect as we found in Table 4, is associated with SAD,[35,36] and one study suggests that shame and self-criticism play a role in the impact of emotional abuse on the development of SAD.[37] This connection could explain the increased self-judgement symptoms seen in the down-weighted group (Table 2). SAD is a risk factor for depression[38]; however, to our knowledge, few studies have investigated its impact on postpartum depression.[39] SAD might inhibit seeking out social support, engender perceived burdensomeness,[40] and increase distress in relationships.[41] With that said, our analyses were exploratory and can not necessarily be interpreted causally given the complex relationships between psychiatric comorbidities and possible unmeasured confounding; rather, these findings indicate that childhood trauma and a history of SAD are predictive of being in the down-weighted subgroup after holding other factors constant (more detail in supplementary material).

Our study focused on responses from the BDI, instead of the EPDS which is more commonly used in prior studies of postpartum depression heterogeneity. The BDI allowed us to identify distinguishing symptoms, such as lack of satisfaction, self-hate, and sense of failure, not directly addressed in the EPDS. When we repeated our analysis with the EPDS, the only distinguishing symptom was thoughts of self-harm. This suggests that the BDI would be a more apt screening instrument for predicting whether an individual is a member of the subgroup we have identified, given that it includes more items that are predictive of the subgroup membership. If the subgroup we identified does require a distinct treatment plan, using a screening instrument, such as the BDI rather than the EPDS, could enhance treatment planning.

Our application of the REM algorithm offered a novel way to detect differences in how well models describe postpartum depressive symptom patterns across population subgroups. This tool, based on ideas from robust statistics, served as a diagnostic that brought attention to variation in symptom patterns not accounted for in the assumed model. In response, we fit exploratory factor mixture models, but model selection criteria did not indicate the need for modelling multiple latent groups. When fitting exploratory factor mixture models, the estimation algorithm did not terminate properly when we treated symptoms as continuous variables, so we treated them as categorical instead, which could be a reason for the disagreement between methods.

There were several limitations with our study. First, our sample is likely not representative of the general

population of postpartum women; participants in our sample had a history of major depressive disorder, were actively seeking psychiatric care, and predominately identified as white and non-Hispanic. We focused on women with a history of major depressive disorder rather than a current clinical diagnosis so that we could analyse a broad range of symptom patterns; however, this might limit the generalizability of these results to women without a history of major depressive disorder. While we were able to replicate our findings independently at each of the two sites, applying our method in a representative sample of postpartum women would allow for further validation of our findings and facilitate the development of scoring algorithms to detect this subgroup. Second, our sample was not assessed for Axis II diagnoses, such as borderline personality disorder. The subgroup down-weighted by the REM might have higher rates of comorbidity with Axis II disorders, considering they are characterised by self-judgement, self-harm, SAD, and childhood trauma. Third, this subgroup is not necessarily one homogeneous subgroup; a larger sample would be required to reapply the REM algorithm within the down-weighted group. Fourth, we cannot disentangle whether the differences in symptom expression are due to differential measurement bias across groups or due to different underlying psychopathology. Fifth, this analysis is exploratory and ideally generated hypotheses should be tested in a separate sample.

In conclusion, an individual's psychiatric history and personal experiences of trauma provide an important context for interpreting symptoms. These findings have important implications for tailoring screening and treatment strategies for postpartum depression to address these needs.

Contributors

DJN, JLC, and ZNS acquired the data. KJN and ALC conceived the study. With the supervision of ALC, KJN developed the code, performed the statistical analyses, produced the tables and figures, and wrote the first draft of the manuscript. KJN and ALC accessed and verified the underlying data. All authors contributed to revision of the manuscript, approved the final manuscript, and agreed to submit for publication.

Data sharing statement

The de-identified data analysed in this study are available from the corresponding author on reasonable request. REM code is available online at https://github.com/knieser/REM.

Declaration of interests

KJN has no conflicts of interest to declare. ALC has received research support from the National Institute of Mental Health, UW American Family Data Science Institute, UW Center for Human Genomics and Precision Medicine; has received honorarium from the University of Michigan; and served a statistical editor for the American Journal of Psychiatry. ZNS has received research support from NIH, CDC, GSK, Pfizer, Wyeth, Janssen and SAGE; has served on speaker or advisory boards for Pfizer, Eli Lilly, Wyeth, SAGE, BMS, and GSK; and has received honoraria from Eli Lilly, GSK, Pfizer, and Wyeth. DJN has received research support from Eli Lilly, Glaxo SmithKline (GSK), Janssen, the National Alliance for Research on Schizophrenia and Depression (NARSAD), the National Institutes of Health (NIH), Sage Therapeutics, Takeda Pharmaceuticals, the Texas Health & Human Services Commission, and Wyeth. He has served on speakers' bureaus and/or received honoraria from Astra-Zeneca, Eli Lilly, GSK, Pfizer and Wyeth. He has served on advisory boards for GSK, Janssen, and Sage Therapeutics. He has served as a consultant to Sage Therapeutics. Neither he nor family members have ever held equity positions in biomedical or pharmaceutical corporations. JLC has no conflicts of interest pertaining to this project.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.101830.

References

1 Meltzer-Brody S, Howard LM, Bergink V, et al. Postpartum psychiatric disorders. *Nat Rev Dis Prim*. 2018;4(1):1–18. https://doi.org/10.1038/nrdp.2018.22.

2 Becker M, Weinberger T, Chandy A, Schmukler S. Depression during pregnancy and postpartum. *Curr Psychiatr Rep*. 2016;18(3):32. https://doi.org/10.1007/s11920-016-0664-7.

3 Stein A, Pearson RM, Goodman SH, et al. Effects of perinatal mental disorders on the fetus and child. *Lancet*. 2014;384(9956):1800–1819. https://doi.org/10.1016/S0140-6736(14)61277-0.

4 Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J Affect Disord*. 2017;219:86–92. https://doi.org/10.1016/j.jad.2017.05.003.

5 O'Connor E, Rossom RC, Henninger M, Groom HC, Burda BU. Primary care screening for and treatment of depression in pregnant and postpartum women: evidence report and systematic review for the US preventive Services Task Force. *JAMA*. 2016;315(4):388. https://doi.org/10.1001/jama.2015.18948.

6 American College of Obstetricians and Gynecologists. ACOG committee opinion No. 757: screening for perinatal depression. *Obstet Gynecol*. 2018;132(5):e208–e212.

7 National Institute for Health and Care Excellence. *Recommendations | Antenatal and postnatal mental health: clinical management and service guidance | Guidance*. NICE; 2020. https://www.nice.org.uk/guidance/cg192/chapter/Recommendations. Accessed September 7, 2022.

8 Austin MP, Highnet N, the Expert Working Group. *Mental Health Care in the Perinatal Period: Australian Clinical Practice Guideline*. Melbourne: Centre of Perinatal Excellence; 2017. https://cope.org.au/wp-content/uploads/2017/10/Final-COPE-Perinatal-Mental-Health-Guideline.pdf. Accessed September 7, 2022.

9 Santos H, Tan X, Salomon R. Heterogeneity in perinatal depression: how far have we come? A systematic review. *Arch Womens Ment Health*. 2017;20(1):11–23. https://doi.org/10.1007/s00737-016-0691-8.

10 Postpartum depression: action towards Causes and treatment (PACT) consortium. Heterogeneity of postpartum depression: a latent class analysis. *Lancet Psychiatr*. 2015;2(1):59–67. https://doi.org/10.1016/S2215-0366(14)00055-8.

11 Putnam KT, Wilcox M, Robertson-Blackmore E, et al. Clinical phenotypes of perinatal depression and time of symptom onset: analysis of data from an international consortium. *Lancet Psychiatr*. 2017;4(6):477–485. https://doi.org/10.1016/S2215-0366(17)30136-0.

12 Sun J, Cao D, Li J, et al. Profiles and characteristics of clinical subtypes of perinatal depressive symptoms: a latent class analysis. *J Adv Nurs*. 2019;75(11):2753–2765. https://doi.org/10.1111/jan.14136.

13 Sun JW, Li JH, Zhang X, et al. Perinatal depression: data-driven subtypes derived from life history and mindfulness and personality. *J Affect Disord*. 2020;276:335–344. https://doi.org/10.1016/j.jad.2020.07.040.

14 Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. *Br J Psychiatry*. 1987;150(6):782–786.

15 Cochran AL, Pingeton BC, Goodman SH, et al. A transdiagnostic approach to conceptualizing depression across the perinatal period in a high-risk sample. *J Abnorm Psychol*. 2020;129(7):689–700. https://doi.org/10.1037/abn0000612.

16 Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *J Affect Disord*. 2017;208:191–197. https://doi.org/10.1016/j.jad.2016.10.019.

17 Di Florio A, Putnam K, Altemus M, et al. The impact of education, country, race and ethnicity on the self-report of postpartum depression using the Edinburgh Postnatal Depression Scale. *Psychol Med*. 2017;47(5):787–799. https://doi.org/10.1017/S0033291716002087.

18 Chiu YHM, Sheffield PE, Hsu HHL, Goldstein J, Curtin PC, Wright RJ. Subconstructs of the Edinburgh postnatal depression scale in a multi-ethnic inner-city population in the U.S. *Arch Womens Ment Health*. 2017;20(6):803–810. https://doi.org/10.1007/s00737-017-0765-2.

19 Williamson JA, Hara MW, Stuart S, Hart KJ, Watson D. Assessment of postpartum depressive symptoms: the importance of somatic symptoms and irritability. *Assessment*. 2015;22(3):309–318. https://doi.org/10.1177/1073191114544357.

20 Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. *N Engl J Med*. 1999;341(18):1329–1335. https://doi.org/10.1056/NEJM199910283411801.

21 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatr*. 1961;4(6):561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004.

22 Ji S, Long Q, Newport DJ, et al. Validity of depression rating scales during pregnancy and the postpartum period: impact of trimester and parity. *J Psychiatr Res*. 2011;45(2):213–219. https://doi.org/10.1016/j.jpsychires.2010.05.017.

23 Lubke GH, Muthén B. Investigating population heterogeneity with factor mixture models. *Psychol Methods*. 2005;10(1):21–39. https://doi.org/10.1037/1082-989X.10.1.21.

24 Nieser KJ, Cochran AL. Addressing heterogeneous populations in latent variable settings through robust estimation. *Psychol Methods*. https://doi.org/10.1037/met0000413. Published online 2021.

25 First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders: Clinician Version: Administration Booklet*. Washington, DC: American Psychiatric Publishing; 1997.

26 Silverman ME, Reichenberg A, Savitz DA, et al. The risk factors for postpartum depression: a population-based study. *Depress Anxiety*. 2017;34(2):178–187. https://doi.org/10.1002/da.22597.

27 Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck depression inventory: twenty-five years of evaluation. *Clin Psychol Rev*. 1988;8(1):77–100. https://doi.org/10.1016/0272-7358(88)90050-5.

28 Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol*. 2006;62(1):123–146. https://doi.org/10.1002/jclp.20213.

29 Bernstein DP, Stein JA, Newcomb MD, et al. Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse Negl*. 2003;27(2):169–190. https://doi.org/10.1016/S0145-2134(02)00541-0.

30 *The Math Works*. Natick, Massachusettts: Inc. MATLAB; 2021a. Published online 2021.

31 Rubin DB, Thayer DT. EM algorithms for ML factor analysis. *Psychometrika*. 1982;47(1):69–76. https://doi.org/10.1007/BF02293851.

32 Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educ Psychol Meas*. 2013;73(6):913–934. https://doi.org/10.1177/0013164413495237.

33 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Published online 2022. https://www.R-project.org/.

34 Muthén LK, Muthén BO. *Mplus User's Guide*. 8th ed. Published online 2017. http://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

35 Kuo JR, Goldin PR, Werner K, Heimberg RG, Gross JJ. Childhood trauma and current psychological functioning in adults with social anxiety disorder. *J Anxiety Disord*. 2011;25(4):467–473. https://doi.org/10.1016/j.janxdis.2010.11.011.

36 Simon NM, Herlands NN, Marks EH, et al. Childhood maltreatment linked to greater symptom severity and poorer quality of life and function in social anxiety disorder. *Depress Anxiety*. 2009;26(11):1027–1032. https://doi.org/10.1002/da.20604.

37 Shahar B, Doron G, Szepsenwol O. Childhood maltreatment, shame-proneness and self-criticism in social anxiety disorder: a sequential mediational model: shame and self-criticism in social anxiety. *Clin Psychol Psychother*. 2015;22(6):570–579. https://doi.org/10.1002/cpp.1918.

38 Beesdo K, Bittner A, Pine DS, et al. Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. *Arch Gen Psychiatr*. 2007;64(8):903–912. https://doi.org/10.1001/archpsyc.64.8.903.

39 Coelho HF, Murray L, Royal-Lawson M, Cooper PJ. Antenatal anxiety disorder as a predictor of postnatal depression: a longitudinal study. *J Affect Disord*. 2011;129(1–3):348–353. https://doi.org/10.1016/j.jad.2010.08.002.

40 Duffy ME, Mueller NE, Cougle JR, Joiner TE. Perceived burdensomeness uniquely accounts for suicidal ideation severity in social anxiety disorder. *J Affect Disord*. 2020;266:43–48. https://doi.org/10.1016/j.jad.2020.01.116.

41 Wenzel A, Haugen EN, Jackson LC, Brendle JR. Anxiety symptoms and disorders at eight weeks postpartum. *J Anxiety Disord*. 2005;19(3):295–311. https://doi.org/10.1016/j.janxdis.2004.04.001.