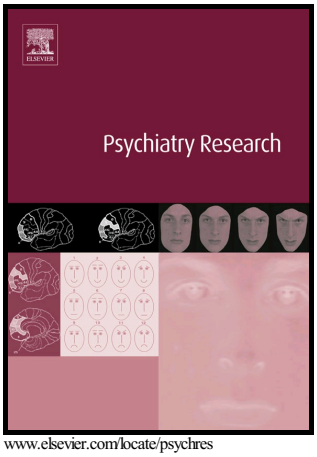
A more rational, theory-driven approach to analysing the factor structure of the Edinburgh Postnatal Depression Scale

Zoltan Kozinszky, Annamária Töreki, Emőke A. Hompoth, Robert B. Dudas, Gábor Németh



www.elsevier.com/locate/psychres

PII: S0165-1781(16)30439-5
DOI: http://dx.doi.org/10.1016/j.psychres.2017.01.059
Reference: PSY10256

To appear in: *Psychiatry Research*

Received date: 15 March 2016
Revised date: 15 January 2017
Accepted date: 21 January 2017

Cite this article as: Zoltan Kozinszky, Annamária Töreki, Emőke A. Hompoth Robert B. Dudas and Gábor Németh, A more rational, theory-driven approach t analysing the factor structure of the Edinburgh Postnatal Depression Scale *Psychiatry Research,* http://dx.doi.org/10.1016/j.psychres.2017.01.059

A more rational, theory-driven approach to analysing the factor structure of the Edinburgh Postnatal Depression Scale

Zoltan Kozinszky M.D., Ph.D.[a,b,*], Annamária Töreki Ph.D. [c], Emőke A. Hompoth[c], Robert B. Dudas M.D., Ph.D., MRCPsych[d,e,f,1], Gábor Németh M.D., Ph.D. [c,1]

[a]Department of Obstetrics and Gynaecology, Blekinge Hospital, Karlskrona, Sweden

[b]Health Research and Health Promotion Research Group, Teacher Training Faculty, College of Kecskemét, Kecskemét, Hungary

[c]Department of Obstetrics and Gynaecology, University of Szeged, Szeged, Hungary

[d]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[e]OPMHS, Cambridgeshire and Peterborough Foundation Trust, Cambridge, United Kingdom

[f]University of East Anglia, Norwich, United Kingdom

[*]Corresponding author: Zoltan Kozinszky, Utövägen 10/A, S-371 37, Karlskrona, Sweden. Telephone: +46730720835. kozinszky@gmail.com

Abstract

We endeavoured to analyze the factor structure of the Edinburgh Postnatal Depression Scale (EPDS) during a screening programme in Hungary, using exploratory (EFA) and confirmatory factor analysis (CFA), testing both previously published models and newly developed theory-driven ones, after a critical analysis of the literature. Between April 2011 and January 2015, a sample of 2,967 pregnant women (between 12th and 30th weeks of gestation) and 714 women 6 weeks after delivery completed the Hungarian version of the EPDS in South-East Hungary. EFAs suggested unidimensionality in both samples. 33 out of 42 previously published models showed good and 6 acceptable fit with our antepartum data in CFAs, whilst 10 of them showed

[1] The last two authors contributed equally

good and 28 acceptable fit in our postpartum sample. Using multiple fit indices, our theory-driven anhedonia (items 1,2) – anxiety (items 4,5) – low mood (items 8,9) model provided the best fit in the antepartum sample. In the postpartum sample, our theory-driven models were again among the best performing models, including an anhedonia and an anxiety factor together with either a low mood or a suicidal risk factor (items 3,6,10). The EPDS showed moderate within- and between-culture invariability, although this would also need to be re-examined with a theory-driven approach.

Keywords*: Edinburgh Postnatal Depression Scale; factor structure; antepartum depression; postpartum depression; theory-driven models*

1. Introduction

The Edinburgh Postnatal Depression Scale (EPDS), a 10-item self-report questionnaire, was developed for the purpose of screening for peripartum depression. Identifying latent factors may help in screening for different types of depression or to develop shorter screening instruments (Gaynes et al., 2005; Gibson et al., 2009; Kozinszky and Dudas, 2015). However, previous factor analytic studies reported factors that did not appear intuitive and mixed together low mood, anhedonia, anxiety, cognitive, and suicidal symptoms. Therefore, it is important to examine the coherence between these solutions and the usefulness of the various approaches used.

Similar to some previous authors (Lee King, 2012; Hartley et al., 2014; Cunningham et al., 2015), we posit that if a certain factor structure cannot be reproduced reliably within the same culture, it is unlikely to be useful for these purposes.

Unfortunately, the literature shows considerable variation between the published EPDS factor structures (Table 1), not just between but also within cultures.

Although, the EPDS was meant to be unidimensional (Cox et al., 1987), the authors included questions tapping into several aspects of depression. It is not unreasonable to expect that the EPDS is able to measure anxiety as well, as it was developed from the Irritability, depression and anxiety scale (Snaith et al., 1978), Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983), and The Anxiety and Depression Scale (Bedford and McIver, 1978; Cox et al., 1987). Indeed, unidimensionality could rarely be demonstrated (Berle et al., 2003). The majority of the authors seemed to detect multidimensionality with two or three factors, such as depression and anxiety (Pop et al., 1992; Astbury et al., 1994; Guedeney and Fermanian, 1998; Des Rivières-Pigeon et al., 2000; Adouard et al., 2005; Matthey, 2008; Phillips et al., 2009; Vivilaki et al., 2009; Töreki et al., 2014), anxiety and anhedonia (Chabrol and Teissedre, 2004; Montazeri et al., 2007; Tuohy and McVey, 2008) or anxiety and suicide (Brouwers et al., 2001; Ross et al., 2003; Jomeen and Martin, 2005).

The "anxiety" factor, usually including items 3, 4, and 5 (Pop et al., 1992; Astbury et al., 1994; Des Rivières-Pigeon et al., 2000; Brouwers et al., 2001; Ross et al., 2003; Jomeen and Martin, 2007; Matthey, 2008; Tuohy and McVey, 2008; Phillips et al., 2009; Lau et al., 2010; Kubota et al., 2014; Hartley et al., 2014; Cunningham et al., 2015;), appears to be the most consistent. In some studies, a factor comprising items 1 and 2 has been labelled "anhedonia" (Brouwers et al., 2001; Chabrol and Teissedre, 2004; Montazeri et al., 2007; Small et al., 2007; Tuohy and McVey, 2008; Lau et al., 2010; Kubota et al., 2014; Zhong et al., 2014; Cunningham et al., 2015;). In other studies, a

third factor including item 10 named "suicide" has emerged (Ross et al., 2003; Jomeen and Martin 2005; Small et al., 2007). Unfortunately, split-loading items have not always been reported consistently in the literature.

In terms of the usefulness of the factors, Chabrol and Teissedre (2004) identified an anxiety factor which strongly correlated with the EPDS total score and was the single significant predictor of the diagnosis of postnatal depression (PND). Brouwers et al. (2001) found that an "anxiety" factor (items 3, 4, and 5) correlated significantly with the State-trait anxiety inventory. Tuohy and McVey (2008) found good correlation of this factor with the HADS-A and the Positive and Negative Affect Scales (PANAS) - Negative subscale. This anxiety factor seems to work well when screening for anxiety disorders identified with the structured clinical interview for DSM-IV (Matthey, 2008). Touhy and McVey (2008) found significant direct correlation between items 1 and 2, a possible "anhedonia" factor, and the Negative emotions subscale of the PANAS and the HADS-A and inverse correlation with the Positive subscale of the PANAS.

There have been few attempts at examining the stability of the factor structure over the peripartum period in the same sample. Chabrol and Teissedre (2004) screened women at 2-3 days and 4-6 weeks post partum respectively, but only examined the factor structure on the first occasion, whilst Cunningham et al. (2015) reported different factor structures on admission for psychiatric treatment and before discharge in postpartum patients. Swalm et al. (2010) reported the same factor structures ante- and postpartum in a large community sample. It is worth noting, however, that their results were very different from the other Australian studies (Astbury et al., 1994; Cunningham et al.,

2015; Small et al., 2007; Matthey, 2008; Phillips et al., 2009), and appeared less than optimal.

It is of interest from an intercultural perspective that Small et al. (2007) identified different factor structures for English, Vietnamese, Turkish and Filippino speaking Australians, whilst Hartley et al. (2014) found evidence for considerable invariance between the English and Spanish versions of the EPDS across a group of Hispanic postpartum women.

In summary, the factor structure of the EPDS is not entirely consistent within or across populations or cultures, but like Lee King (Lee King, 2012), we do not think this necessarily calls into question the usefulness of a factor analytic approach. However, more work needs to be done to clarify the situation.

Although exploratory factor analysis (EFA) often produces distinct factors comprised of positively and negatively worded components (Harrington, 2009), this approach has little utility in determining the nature of these outcomes. A presumed structure cannot be imposed on the test beyond the specification of the number of latent factors. Restrictions used in the EFA model identification preclude an analysis of error covariances. There is inconsistency between studies as regards the choice of analytic approach, many used a purely exploratory approach, some a confirmatory one, whilst several authors (Phillips et al., 2009; Vivilaki et al., 2009; Töreki et al., 2014) confirmed the factors that they had identified with EPA/PCA on postpartum samples with CFA. In a non-peripartum context, item response theory-based approaches have also been used (de Cock et al., 2011).

A number of studies demonstrated and 16 studies also confirmed the multidimensionality of the EPDS, reporting the same factor structure as others previously. Phillips et al. (2009) conducted a CFA of a subset of published structural models as well as a model from their own EFA and found the best fit with the bifactorial structure of Brouwers, omitting the third factor (item 10). Jomeen and Martin (2005) have demonstrated the replicability of two models with CFA in an antepartum sample. Their optimal model was Brouwers' 3-factor one with depression, anxiety, and suicidality (item 10) factors (Brouwers et al., 2001). Lee King (2012), however, demonstrated best fit with Tuohy and McVey's (2008) three factors: depression, anxiety, and anhedonia. Cunningham et al. (2015), in postpartum patients who were admitted to psychiatric ward, found the same factor structure as Lau et al. (2010) in China.

Table 1 presents all published PCA/EFA and CFA analyses based on theoretical considerations or on the factors identified with an exploratory approach. In our own validation studies (Töreki et al., 2013; Töreki et al., 2014), we could only identify a clinically meaningful factor structure in the postpartum (but not the antepartum) sample.

In the study reported in this paper, we expected to replicate our previous findings in a much larger sample collected using the same inclusion criteria. When a measurement instrument is used across multiple populations, it is assumed that the instrument is assessing the same construct(s) on the same metric (scale). Similarly, when an instrument is used in a single population on more than one occasion, it is assumed that the scale is assessing the same construct(s) on the same metric at each time, known as measurement invariance (Widaman et al., 2010). There can be various reasons for not finding

measurement invariance, including a true lack of measurement invariance as well as methodological shortcomings.

We also wanted to examine how the above described factor structures previously reported in the literature would fit with our own data.

Accruing evidence suggests that anhedonia defines a dimension in depressive disorder that seems to be different from a dimension encompassing mood and somatic symptoms. The first appears to be associated with the underfunctioning of dopaminergic neurons, whilst the other seems to be related to a similar under-functioning in the serotonin system (Argyropoulos et al., 2013). These dimensions may occur simultaneously, but may also be present separately. Studies on the neurobiology of other emotions, e.g. anxiety (Goodwin et al., 2015) and guilt (McIatchie et al., 2016), are also beginning to identify brain areas and networks implicated, also suggesting that certain items of rating scales tapping into these symptoms may form a separate factor. Hopelessness and suicidality seemed to be linked to impulsivity (Wang et al., 2015), which appears to have its separate neurobiological underpinnings. If a factor structure really carves nature at its joints (i.e. identifies factors within the EPDS that can be related to neural mechanisms presumably linked to the illness), it should be reproducible in other samples, assuming that the compared samples have the same illness. Therefore, in this paper, incorporating recent insight from neurobiology as well as the behaviour of items of the EPDS in previous studies, we also propose a new, theory-driven method of identifying factor structures, demonstrated through our own data, in order to achieve better model fit that would enable further studies to examine measurement invariance with more precision.

2. Methods

2.1. Study design

This study used a cross-sectional design with a model comparison approach. Between 1 April 2011 and 28 January 2015, a sample of 2,967 women during pregnancy (between 12th and 30th weeks of gestation) and 714 women at 6 weeks postpartum completed the Hungarian version of the EPDS at the Department of Obstetrics and Gynaecology, University of Szeged, Hungary and 31 pregnancy care units in South-East Hungary. Only an acceptable proportion refused to participate (15% antepartum and 25% postpartum). Inclusion criteria were fluency in spoken and written Hungarian and signed informed consent. There was no exclusion due to psychiatric conditions other than peripartum depression in the context of organic causes and epilepsy, respectively, or illiteracy.

We established the factor structure of the EPDS in our sample and examined with CFA the fit onto our data of all previously published models (including those from our own validation studies) in a peripartum context available to us, as well as the models derived from our current screening sample. We also tested theory-driven models based on the literature and on neurobiological insight as to which questions should be expected to belong to the same underlying factor. We created our theory-driven models (TDM) with "anhedonia" (F1: 1, 2) the previously described "anxiety" (F2: 3, 4, 5) factor, combined with either "low mood" (F4: 8, 9; TDM 1) or "hopelessness" (F3: 6, 10; TDM 2), or both

(TDM 3). From a phenomenological perspective, we found the inclusion of item 3 (guilt) in the anxiety factor counter-intuitive and, in a theory-driven fashion, we postulated a model with "anhedonia" (items 1 and 2), "anxiety" (items 4 and 5), "low mood" (items 8 and 9), and a fourth, "suicidal risk" factor (guilt, helplessness, and self-harm thoughts - items 3, 6, and 10, respectively; TDM 4). Next, we removed the fourth (and phenomenologically least homogenous) factor and created a three-factor model with "anhedonia" (items 1 and 2), "anxiety" (items 4 and 5), and "low mood" (items 8 and 9; TDM 5). Finally, we also created a model by replacing the low mood factor with the "suicidal risk" factor (items 3, 6 and 10) (TDM 6).

The study protocol was approved by the Clinical Research Ethics Committee of the University of Szeged (Protocol 89/2011) and the study was carried out according to the principles of the Declaration of Helsinki.

2.2. Statistical analysis

The underlying dimensions of the scale were examined with exploratory factor analysis (EFA) (Tabachnik and Fidell, 2007). Cronbach's alphas (McKennell, 1970) were separately calculated for the scales in the ante- and postpartum period, with an oblique rotation of the latent factors. CFA was performed with the robust maximum likelihood estimation method (MLR) to check the fit of these factors in our dataset (Hakstian et al., 1982; Jöreskog and Sorbom, 1986; Tabachnik and Fidell, 2007), given the robustness of this estimator with non-normal, continuous data and providing Akaike information criterion (AIC). The other specified CFA models were based on models derived from the

literature via a systematic search (Table 1); items (including cross-loading ones) that loaded ≥0.40 on a factor were included in the CFA.

The metric of the factors in all models was defined by fixing the factor variable variances to 1.0. Factors were allowed to intercorrelate as would be expected of the relationship between them (anhedonia, anxiety, and depression), given models of depression (e.g., the tripartite model of depression in Mineka et al. (Mineka et al., 1998) and Watson (Watson, 2005) and the literature on the presentation of postpartum depression (e.g., Pitt (Pitt, 1968)). Error variances of the items/indicators were assumed to be uncorrelated. Factor loadings and error variances were freely estimated.

Two goodness-of-fit indicators, including including chi-square statistic (χ^2), the root mean square error of approximation (RMSEA), comparative fit index (CFI), were selected for reporting the analysis outcomes. The χ^2 statistic is a traditional measure of overall model fit, with a non-significant chi-square suggesting good fit. As a guideline, RMSEA values <0.06 denote a close fit and values below 0.11 an acceptable fit (Harrington, 2009). CFI values ranged from 0 to 1 with a value of 0.90 and greater being acceptable fit to the data (Cudeck and Browne, 1983; Jöreskog and Sorbom, 1986; Hu and Bentler, 1999; Kline, 2010).

While RMSEA and CFI are absolute indices, i.e. appropriate for evaluating an individual model alone, they are not appropriate for comparison of different models. Hence, we used the AIC (Akaike, 1974), a comparative measure of fit to compare the hypothesized models in this study. The model with the lowest AIC would be the best fitting model. It is important to note that models with complicated structures (e.g., lack of parsimony) may be penalized by AIC.

Statistical analyses were performed with the Statistical Package for Social Sciences (SPSS 17.0, Inc. Chicago, IL) software, except for CFA, for which we used MPlus 7.11 (Muthén and Muthén, 2012). EPDS total score and factor means were compared with two-sample t-test. Differences were considered significant if the two-tailed p value was less than 0.05.

3. Results

3.1. Sample demographics

The ante- and postpartum samples were similar in terms of age (Table 2). The overwhelming majority were married or lived with a partner in both samples. Primiparity and having an unplanned pregnancy were similarly prevalent in both groups. As expected, women in the puerperium had more children than their counterparts during pregnancy.

3.2. Distribution of the EPDS scores and their reliability

The distribution of the EPDS scores in the ante- and postpartum samples did not show a significant difference. The participants' EPDS scores ranged from 0 to 23 with a mean score of 3.50 (standard deviation (S.D.): 3.19) during pregnancy. Of the 2,967 screened pregnant women, 14.36% had scores above 6, indicating combined depression, while 6.6% screened positive for likely major depression with a score of 9 or higher

(Töreki et al., 2013). In the postpartum sample, the EPDS total scores varied between 0 and 25 (mean: 3.49, S.D.:3.38) and 16.1% screened positive (over 7 points) for combined and 4.2% (over 12 points) for major depression (Töreki et al., 2014).

In the antepartum sample, the EPDS showed good internal consistency (Cronbach α=0.758). The α coefficient for each item was at least 0.718, indicating acceptable homogeneity. In the postpartum sample, the EPDS showed good internal consistency (Cronbach α=0.817). The α coefficient for each item was at least 0.785, indicating good homogeneity (McKennell, 1970).

3.3. EFAs of the EPDS in our ante- and postpartum sample

An EFA with oblique rotation revealed one latent factor in the antepartum period (Table 3). Factor 1 included items 1-9, which explained 35.7% of the variance and the Cronbach alpha for Factor 1 was 0.787.

In the postpartum period, the EFA with oblique rotation on the 10 items of the EPDS revealed one latent factor. Factor 1 included questions 2-9, which explained 39.9% of the variance and Cronbach alpha for Factor 1 was 0.809.

3.4. CFAs in our own screening samples, using the factor structures identified with EFA

CFAs of the EFA models of our samples did not produce convincing results in neither the ante-, nor the postpartum sample. The RMSEA values were less favourable

than for other studied model structures both in the antepartum and in the postpartum samples (Table 4).

3.5. CFAs in our screening samples, using other models from the literature as well as our own theory-driven models

We then ran CFAs using all the models published in previous antepartum and postpartum studies as well as our theory-driven models in both our ante- and postpartum samples (Table 4).

Fit indices were significant for almost all model variants both ante- and postpartum. In our antepartum sample, 39 models - including all of our theory-driven models - showed good, and 6 acceptable fit on the basis of RMSEA values. 4 models did not fit the data. The model with the lowest AIC value was that of Matthey (2008), which only included the "anxiety" factor (items 3, 4, 5). The second lowest AIC value was produced by our Theory-driven model 5 (TDM 5), with an RMSEA value suggesting good fit. Correlations between the factors in this model ranged 0.331-0.392. The Cronbach alfa for the anhedonia factor was 0.628, for anxiety 0.612 and for low mood 0.67. Moving towards models with increasingly higher AIC values, the next one that had a good RMSEA value was our TDM 2, followed by our TDM 6.

As regards testing the models in our postpartum sample, 13 of the models (including 3 of our own theory-driven models) showed good fit, 31 models (including 3 of our theory-driven models) showed acceptable fit on the basis of RMSEA values. 5 models did not fit the data. Again, the model with the lowest AIC value was that of

Brouwers et al (2001), followed by that Matthey (2008). Our best performing theory-driven model (TDM 6) had a higher AIC value than the before mentioned, but its RMSEA was slightly better than that of the Brouwers et al. (2001) model. Correlations between the factors in TDM 6 ranged 0.320-0.498. The Cronbach alfa for the anhedonia factor was 0.594, for anxiety 0.667, and for suicidal risk 0.409. The third best model was our TDM 5, which in fact had better Cronbach alfa values than TDM 6 (anhedonia: 0.594, anxiety: 0.667, low mood: 0.691).

4. Discussion

In this study, we first endeavoured to examine if in larger, unselected antepartum and postpartum samples from the same population we would find the same factor structures with the EPDS as in our earlier validation studies (Töreki et al., 2013; Töreki et al., 2014). We found that this was not the case in either case. Second, using CFA, we looked at how well previously published factor models fitted our data and found that, with the exception of but a handful of studies, they did not fit the data particularly well or much better (especially in the postpartum sample) than our own EFA factor structures did. Third, we created a number of models in a theory-driven fashion, building on our own experience and that of previous authors with exploratory approaches as well as taking into consideration the phenomenology and neurobiology of depression, and with our best model we found extremely good fit with our antepartum data and acceptable (but better than any previously published) fit with our postpartum data.

4.1. The stability of the factor structure between our validation and screening samples

To our knowledge, we were the first to conduct a factor analysis in a validation sample, followed by another factor analysis on a larger sample from the same population. During the antepartum validation of the EPDS (n=219), a PCA produced three factors that were hard to interpret (Töreki et al., 2013). Using our larger sample (n=2,967), an EFA suggested unidimensionality, but the factor still did not explain much of the variance, nor did it appear to be psychiatrically particularly meaningful. In our postpartum validation study (n=266) (Töreki et al., 2014), a PCA provided factors with better interpretability, however, this was not replicated in our larger sample (N=714).

In our larger ante- and postpartum samples, we expected to find the same or at least similar factor structures as in our validation studies, and the rather different results were surprising. One possible explanation is that our validation studies used relatively small samples and the factor structures in our larger samples might be more reliable, although clinically still not particularly meaningful. Alternatively, peripartum depression may not be a homogenous illness (Gibson et al., 2009; Kozinszky and Dudas, 2015), and our larger samples may have included different proportions of its various forms. A similarly large-scale postpartum study from the same culture (Hungary) found two factors (Nagy et al., 2011) that were, again, different from our factors. It is of note, however, that that study used a different version of the EPDS.

Factor analytic studies so far have used PCA/EFA or CFA, or they also used CFA to check the validity of their factors identified with PCA/EFA, and reported more or less acceptable results. The finding that, using CFA, we could not confirm our factors identified with EFA in our ante- and postpartum samples was unexpected.

Other studies conducted in the same culture also found discordant findings (Australia: Astbury et al. (1994); Small et al. (2007); Matthey (Matthey, 2008); Phillips et al. (2009); Swalm et al. (2010); Cunningham et al. (2015); UK: Cox et al. (1987); Jomeen and Marteen (2005); Jomeen and Marteen (2007); Tuohy and McVey (2008); France: Guedeney and Fermanian (1998); Chabrol and Teissedre (2004); Adouard et al., 2005; Canada: Ross et al. (2003); Bowen et al. (2008) and USA: Logsdon et al. (2010); Lee King (2012); Hartley et al. (2014)), which warrants the examination of the validity of these models. Confounding factors may include not only sample characteristics, such as antepartum vs. postpartum, clinical vs. community, mixed-risk or unselected vs. high-risk pregnant women, sample size, sampling procedure (population vs. clinics and investigation centre) and time of testing, but also socioeconomic and cultural factors, race, ethnicity, and exposition to risk factors for depression.

4.2. The stability of the factor structure within and between different cultures: cross-cultural invariance?

Apart from items 3-5 and 1 and 2 tending to be on the same factor, respectively, alone or together with other items, the reported factor structures seem to be diverse. It is of note, however, that the studies tested women at different stages in the peripartum

period and used different factor analytic techniques. Although there were several studies from the UK, France, or the Netherlands, respectively, the factor structures could not be meaningfully compared due to the previously mentioned differences (see subsection 4.1). We were able to find two cultures where the studies had been done at comparable stages of the peripartum period and used similar statistical methodology; in Australia two such studies reported different factor structures, suggesting little invariance even within the same culture, whereas in Hungary the two validation studies reported fairly similar factor structures, indicating at least some degree of invariance within the same culture.

As regards invariance between cultures, out of the 48 factor structures reported in the literature, only seven were identified by more than one studies (1. Lau et al. (2010) and Cunningham et al. (2015), Coates et al. (2017) 2. Astbury et al. (1994), Des Rivieres-Pigeon et al. (2000) and Phillips et al. (2009), 3. Ross et al. (2003) and Hartley et al. (2014), Coates et al. (2017), Bina and Harrington (2016), 4. Cunningham et al. (2015), Lee King (2012), 5. Reichenheim et al. (2011), Bina and Harrington (2016), 6. Cox et al. (1987), Coates et al. (2017), 7. Zhong et al. (2014), Coates et al. (2017)), whilst the rest were all different from each other. The fact that items 3-5 seemed to form an "anxiety" factor in many of the reported studies points towards some degree of cross-cultural invariance.

To our knowledge, ours is the first study ever to have evaluated the underlying structure of the EPDS using a CFA model comparison approach examining all factor models from the literature in purposive community samples.

The literature-derived models with the best AIC values generally had poorer RMSEA values than our theory-driven models did. AIC and RMSEA values often but not

always indicate the same model as the best one. Some of our findings in both the ante- and the postpartum sample may be explained by the fact that AIC penalizes models with less parsimony. Also, the EPDS is a brief instrument covering symptoms with one or a maximum of two items and not covering certain symptoms of depression at all, which may also explain some unexpected correlations between items.

It is of note that our theory-driven models produced the best fit in the antepartum sample in terms of their AIC and RMSEA values.

4.3. Theory-driven models

We found that in our antepartum screening sample, our best theory-driven model showed better overall fit with the data than any other model previously reported in the literature, including that derived from our own antepartum validation study. This anhedonia – anxiety – low mood factor structure is phenomenologically more intuitive and provided better fit with the data than any other theory-driven combination.

In our postpartum screening sample, AIC values indicated the literature-derived Brouwers et al. (2001) model (anhedonia & low mood and anxiety) slightly better than our theory-driven anhedonia – anxiety – suicidal risk model, although our theory-driven model's RMSEA and SMRS values indicated slightly better model fit.

4.4. Methodological considerations

Although the original intention (Cox et al., 1987) was to measure one construct (depression), not detecting any factors would be somewhat unexpected, as some questions, at least on the face of it, appear to be more related (e.g. those about anxiety). Exploratory techniques, of course, have their own weaknesses; they examine correlations between all the items included in the analysis, along a latent factor present in the studied sample, without the ability to critically handle cross-loading, i.e. interpret correlations. However, often items can be grouped together in an a priori fashion according to theoretical considerations and these factors can be examined with CFA. Coates et al. (2017) created 4 models, also in a theory-driven fashion, but these models performed less well in our sample than our own models did.

We propose a new methodological approach whereby in factor analytic studies of the EPDS (or indeed other psychometric instruments) empirical findings from exploratory approaches are triangulated with our understanding of possible underlying neurobiological systems and taken into consideration when building theoretically-driven models that can then be tested with a confirmatory approach, as opposed to the relentless replication of exploratory solutions with a poor fit. We also suggest for further studies doing multi-group CFAs, e.g. according to (depression) diagnosis status. Incidentally, theory-driven models could also be used in cross-cultural replicability studies.

4.5. Strengths

We report here on one of the largest sample so far. Our study is also unique in that we used CFA to confirm in a larger sample collected from the same population the factor

structure identified by ourselves in our validation study. The previous replicability study only checked previously published models but not their own previous models (Lee King, 2012). We had separate ante- and postpartum samples and analysed them separately.

In summary, the available evidence shows significant heterogeneity in the factor structure of the EPDS within and across cultures. We found different factor structures ante- and postpartum in our larger screening samples compared to the ones identified in our validation samples, and CFAs showed poor fit indices with these new models. Some of the models previously published in the literature did well in CFAs in our screening samples, but we got the best fit indices overall with newly generated, theory-driven models based on previous experience with exploratory approaches and phenomenological considerations, namely an anhedonia – anxiety – low mood model.

Conflicts of Interests:

none.

Acknowledgement:

We would like to thank Dr Jan Zirk-Sadowski for his advice on statistical matters.

References

1. Adouard, F., Glangeaud-Freudenthal, N.M., Golse, B., 2005. Validation of the Edinburgh postnatal depression scale (EPDS) in a sample of women with high-risk pregnancies in France. Arch. Wom. Ment. Health. 8,89-95.

2. Agampodi, S.B., Agampodi, T.C., 2013. Antenatal Depression in Anuradhapura, Sri Lanka and the Factor Structure of the Sinhalese Version of Edinburgh Post Partum Depression Scale among Pregnant Women. PLoS. One. 8,e69708.
3. Akaike, H., 1974. A new look at the statistical model identification. IEEE. Transactions. Automatic. Control. 19,716-723.
4. Argyropoulos, S.V., Nutt, D.J., 2013. Anhedonia revisited: is there a role for dopamine-targeting drugs for depression? J. Psychopharmacol. 27,869-877.
5. Astbury, J., Brown, S., Lumley, J., Small, R., 1994. Birth events, birth experiences and social differences in postnatal depression. Aust. J. Public Health. 18,176-178.
6. Bedford, A., McIver, D., 1978. Foulds' "general instability" and "psychopathy" 16PF scales and their relationship to psychiatric mood state. J. Clin. Psychol. 34,417-418.
7. Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. Psychol. Bullet. 88,588.
8. Berle, J.Ø., Aarre, T.F., Mykletun, A., Dahl, A.A., Holsten, F., 2003. Screening for postnatal depression Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J. Affect. Disord. 76,151-156.
9. Bina, R., Harrington, D.,2016. The Edinburgh Postnatal Depression Scale: Screening Tool for Postpartum Anxiety as Well? Findings from a Confirmatory Factor Analysis of the Hebrew Version. Matern Child Health J. 20,904-914.
10. Bowen, A., Bowen, R., Maslany, G., Muhajarine, N., 2008. Anxiety in a socially high-risk sample of pregnant women in Canada. Canad. J. Psych., 53, 435-440.

11. Brouwers, E.P.M., van Baar, A.L., Pop, V.J.M., 2001. Does the Edinburgh Postnatal Depression Scale measure anxiety? J. Psychosom. Res. 51,659-663.
12. Chabrol, H., Teissedre, F., 2004. Relation between Edinburgh Postnatal Depression Scale scores at 2-3 days and 4-6 weeks postpartum. J. Reprod. Infant. Psychol. 22,33-39.
13. Coates, R., Ayers, S., de Visser, R., 2017.Factor Structure of the Edinburgh Postnatal Depression Scale in a Population-Based Sample. Psychol. Assess. Article in press.
14. Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br. J. Psychiatry. 150,782-786.
15. Cudeck, R., Browne, M.W., 1983. Cross-validation of covariance structures. Multivar. Behav. Research. 18,147-167.
16. Cunningham, N.K., Brown, P.M., Page, A.C., 2015. Does the Edinburgh Postnatal Depression Scale measure the same constructs across time? Arch. Wom. Ment. Health. 18,793-804.
17. de Cock, E.S., Emons, W.H., Nefs, G., Pop, V.J., Pouwer F., 2011. Dimensionality and scale properties of the Edinburgh Depression Scale (EDS) in patients with type 2 diabetes mellitus: the DiaDDzoB study. BMC. Psychiatry. 11,141.
18. Des Rivières-Pigeon, C., Séguin, L., Brodeur, J.M., Perreault, M., Boyer, G., Colin, C., Goulet, L., 2000. L'Échelle De Dépression Postnatale D'Edimbourg: Validité Au Québec Auprès De Femmes De Statut Socio-éConomique Faible [The Edinburgh Postnatal Depression Scale: the validity of its Quebec version for a population low socioeconomic status mothers]. Can. J. Commun. Ment. Health. 19,201-214.

19. Gaynes, B., Gavin, N., Meltzer-Brody, S., Lohr, K.N., Swinson, T., Gartlehner, G., Brody S, Miller W.C., 2005. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evidence Report/Technology Assessment No. 119. (Prepared by the RTI-University of North Carolina Evidence-based Practice Center, under Contract No. 290-02-0016.) AHRQ Publication No. 05-E006-2. Rockville, MD: Agency for Healthcare Research and Quality.
20. Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., Gray, R., 2009. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. Acta Psychiatr. Scand. 119,350-364.
21. Goodwin, G.M., 2015. The overlap between anxiety, depression, and obsessive-compulsive disorder. Dialogues. Clin. Neurosci. 17,249-60.
22. Guedeney, N., Fermanian, J., 1998. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results about use and psychometric properties. Eur. Psychiatry. 13,83-89.
23. Hakstian, A.R., Rogers, W.D., Cattell, R.B., 1982. The behavior of number of factor rules with simulated data. Multivar. Behav. Res. 17,193-219.
24. Harrington, D., 2009. Confirmatory factor analysis. New York: Oxford University Press.
25. Hartley, C.M., Barroso, N., Rey, Y., Pettit, J.W., Bagner, D.M., 2014. Factor structure and psychometric properties of English and Spanish versions of the Edinburgh Postnatal Depression Scale among Hispanic women in a primary care setting. J. Clin. Psychol. 70,1240-1250.

26. Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equat. Model. 6,1-55.

27. Jomeen, J., Martin, C.R., 2005. Confirmation of an occluded anxiety component within the Edinburgh Postnatal Depression Scale (EPDS) during early pregnancy. J. Reprod. Infant. Psychol. 23,143-154.

28. Jomeen, J., Martin, C.R., 2007. Replicability and stability of the multidimensional model of the Edinburgh Postnatal Depression Scale in late pregnancy. J. Psychiatr. Ment. Health. Nurs. 14,319-324.

29. Jöreskog, K.G., Sorbom, D., 1986. LISREL VI: analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. Department of Statistics, University of Uppsala, Sweden.

30. Lee King, P.A., 2012. Replicability of structural models of the Edinburgh Postnatal Depression Scale (EPDS) in a community sample of postpartum African American women with low socioeconomic status. Arch. Wom. Ment. Health. 15,77-86.

31. Kline, R.B., 2010. Principles and practice of structural equation modeling (3rd ed.). New York, New York: Guilford Press.

32. Kozinszky, Z., Dudas R.B., 2015. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J. Affect. Disord. 176,95-105.

33. Kubota, C., Okada, T., Aleksic, B., Nakamuram Y., Kunimoto, S., Morikawa, M., Shiino, T., Tamaji, A., Ohoka, H., Banno, N., Morita, T., Murase, S., Goto, S., Kanai, A., Masuda, T., Ando, M., Ozaki, N., 2014. Factor Structure of the Japanese Version of the Edinburgh Postnatal Depression Scale in the Postpartum Period. PLoS. One. 9:e103941.

34. Kwan, R., Bautista, D., Choo, R., Shirong, C., Chee, C., Mei Saw, S., Chong,Y.S., Kwek, K., Meaney, M.J., Rush, A.J., Chen, H., 2015. The Edinburgh Postnatal Depression Scale as a measure for antenatal dysphoria. J. Reprod. Inf. Psychol., 33,28-41.

35. Lau, Y., Wang, Y., Yin, L., Chan, K.S., Guo, X., 2010. Validation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale in Chengdu mothers. Int. J. Nurs. Stud. 47,1139-1151.

36. Logsdon, M.C., Usui, W.M., Nering, M., 2009. Validation of Edinburgh postnatal depression scale for adolescent mothers. Arch. Wom. Ment. Health. 12,433-440.

37. Maroto Navarro, G., García Calvente, M.d.M., Fernández Parra, A., 2005. Evaluación del estado de ánimo en el puerperio con la Escala de Depresión Postnatal de Edimburgo. [Evaluation of mood in the postpartum period with the Edinburgh Postnatal Depression Scale]. Int. J. Clin. Health Psychol. 5,305-318.

38. Massoudi, P., Hwang, C.P., Wickberg, B., 2013. How well does the Edinburgh Postnatal Depression Scale identify depression and anxiety in fathers? A validation study in a population based Swedish sample. J. Affect. Disord. 149,67-74.

39. Matthey, S., 2008. Using the Edinburgh Postnatal Depression Scale to screen for anxiety disorders. Depress. Anxiety. 25,926-931.

40. Mazhari, S., Nakhaee, N., 2007. Validation of the Edinburgh Postnatal Depression Scale in an Iranian sample. Arch. Wom. Ment. Health. 10,293–297.

41. Mclatchie, N., Giner-Sorolla R., Derbyshire, S.W.G., 2016. 'Imagined guilt' vs 'recollected guilt': implications for fMRI. Soc. Cognit. Affect. Neuroscience. 11,703-711.

42. McKennell, A., 1970. Attitude measurement: use of coefficient alpha with cluster or factor analysis. Sociology. 4,227-245.

43. Mineka, S., Watson, D., Clark, L.A., 1998. Comorbidity of anxiety and unipolar mood disorders. Annu. Rev. Psychol. 49:377.

44. Montazeri, A., Torkan, B., Omidvari, S., 2007. The Edinburgh Postnatal Depression scale (EPDS): translation and validation study of the Iranian version. BMC Psychiatry. 7,11.

45. Muthén, L. K., Muthén, B. O. (2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

46. Nagy, E., Molnár, P., Pál, A., Orvos, H., 2011. Prevalence rates and socioeconomic characteristics of post-partum depression in Hungary. Psychiatry Res. 185,113-120.

47. Odalovic, M., Tadic, I., Lakic, D., Nordeng, H., Lupattelli, A., Tasic, L. (2015). Translation and factor analysis of structural models of Edinburgh Postnatal Depression Scale in Serbian pregnant and postpartum women - Web-based study. Wom. Birth, article in press.

48. Phillips, J., Charles, M., Sharpe, L., Matthey, S., 2009. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. J. Affect. Disord. 118,101-112.

49. Petrozzi, A., Gagliardi, L., 2013. Anxious and depressive components of Edinburgh Postnatal Depression Scale in maternal postpartum psychological problems. J. Perinat. Med. 41,343-348.

50. Pitt, B., 1968. 'Atypical' depression following childbirth. Br. J. Psychiatr. 114,1325-1335.

51. Pop, V.J., Komproe, I.H., van Son, M.J.H., 1992. Characteristics of the Edinburgh Post Natal Depression Scale in the Netherlands. J. Affect. Disord. 26,105-110.

52. Reichenheim, M.E., Moraes, C.L., Oliveira, A.S., Lobato, G., 2011. Revisiting the dimensional structure of the Edinburgh Postnatal Depression Scale (EPDS): empirical evidence for a general factor. BMC Med. Res. Methodol. 11,93.

53. Ross, L.E., Gilbert, E.S.E., Sellers, E.M., Romach, M.K., 2003. Measurement issues in postpartum depression part 1: anxiety as a feature of postpartum depression. Arch. Wom. Ment. Health. 6,51-57.

54. Snaith, R.P., Constantopoulos, A.A., Jardine, M.Y., McGuffin, P., 1978. A clinical scale for the self-assessment of irritability. Br. J. Psychiatry. 132,164-171.

55. Small, R., Lumley, J., Yelland, J., Brown, S., 2007. The performance of the Edinburgh Postnatal Depression Scale in English speaking and non-English speaking populations in Australia. Soc. Psychiatr. Psychiatr. Epidemiol. 42,70-78.

56. Swalm, D., Brooks, J., Doherty, D., Nathan, E., Jacques, A., 2010. Using the Edinburgh postnatal depression scale to screen for perinatal anxiety. Arch. Wom. Ment. Health. 13,515-522.

57. Tabachnick, B.G., Fidell, L., 2007. Using Multivariate Statistics. 3rd edition. New York: Addison-Wesley-Longman.

58. Töreki, A., Andó, B., Dudas, R.B., Dweik, D., Janka, Z., Kozinszky, Z., Keresztúri, A., 2014. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. Midwifery. 30,911-918.

59. Töreki, A., Andó, B., Keresztúri, A., Sikovanyecz, J., Dudas, R.B., Janka, Z., Kozinszky, Z., Pál, A., 2013. The Edinburgh Postnatal Depression Scale: Translation and antepartum validation for a Hungarian sample. Midwifery. 29,308-315.

60. Tuohy, A., McVey, C., 2008. Subscales measuring symptoms of nonspecific depression, anhedonia, and anxiety in the Edinburgh Postnatal Depression Scale. Br. J. Clin. Psychol. 47,153-169.

61. Vivilaki, V.G., Dafermos, V., Kogevinas, M., Bitsios, P., Lionis, C., 2009. The Edinburgh Postnatal Depression Scale: translation and validation for a Greek sample. BMC Public Health. 9,329.

62. Wang, Y.Y., Jiang, N.Z., Cheung, E.F., Sun, H.W., Chan, R.C., 2015. Role of depression severity and impulsivity in the relationship between hopelessness and suicidal ideation in patients with major depressive disorder. J. Affect. Disord. 183,83-9.

63. Watson, D., 2005. Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. J. Abnorm. Psychol. 114,522-536.

64. Widaman, K.F., Ferrer, E., Conger, R.D., 2010. Factorial invariance within longitudinal structural equation models: measuring the same construct across time. Child Develop. Perspectives. 4,10-18.

65. Zhong, Q., Gelaye, B., Rondon, M., Sánchez, S.E., García, P.J., Sánchez, E., Barrios, Y.V., Simon, G.E., Henderson, D.C., Cripe, S.M., Williams, M.A., 2014. Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. J. Affect. Disord. 162,1-7.

66. Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. Acta Psychiatr. Scand. 67,361-70.

Table 1. Factor structures of the Edinburgh Postnatal Depression Scale in various languages

Country/language	*Validation against*	*Antepartum/ postpartum*	*sample size (n)*	*Method (extraction)*	*Rotation*	*% of variance*	*Factor 1/ Component 1*	*Factor 2/ Component 2*	*Factor 3/ Component 3*
The Netherlands (Pop et al., 1992)	RDC	Postpartum 4weeks	293	EFA	ORT (VAR)	56.4	7,*8*,9,10	3,4,5,*6*	1,2,*6,8*
				CFA*	LISREL (unrestricted rotation)	n/a	1,2,3,5,6,7,8,9	n/a	n/a
						n/a	1,2,7,8,9	3,4,5,6	n/a
The Netherlands (Brouwers et al., 2001)	STAI/SCL-90	Antepartum 24 weeks	197	PCA	ORT (VAR)	52.1	3,4,5,*6*	1,2,*6*,*7*,*8*,*9*	*7*,*8*,*9*,10
UK (Cox et al., 1987)	RDC	Postpartum 8 weeks	84	n/a	n/a	46	1-10	n/a	n/a
UK (Jomeen and Martin, 2005)	n/a	Antepartum 14 weeks	101	EFA/ CFA*	OBLQ (OBLM)	55.2	3,4,5	1,2,6,7,8,9	10

UK (Jomeen and Martin, 2007)	n/a	Antepartum 27-40 weeks	148	CFA	Not reported	Not reported	1,2,*8*	3,4,5,*8*	10
UK (Tuohy and McVey, 2008)	HADS, PANAS	Postpartum (mean: 6.5 months)	440	EFA	OBLQ (Direct quartimin)	61.0	3,4,5	7,8,9,10	1,2
UK (Coates et al., 2017)	n/a	Antepartum 18 and 32 weeks	12,166 - 12,100	EFA/ CFA*	OBLQ (Direct OBLM) and FIML	64	1,2	3,4,5,6	7,8,9,10
						66.1	1,2	3,4,5,6,7,8,9,10	n/a
		Postpartum 8 weeks and 8 months	11,710 - 11,195			65.5	3,4,5	1,2,6,7,8,9,10	n/a
						66	1,2,3,4,5,6,7,8,9,10	n/a	n/a
Australia (Astbury et al., 1994)	n/a	Postpartum 8-9 months	790	PCA	OBLQ (VAR)	52	3,4,5	1,2,6,7,8,9,10	n/a
Australia (Small et al., 2007)	n/a	Postpartum 6-7 months	1168	PCA	OBLQ (VAR)	>60.0	3,4,5,6,7	1,2,8	9,10
Australia (Matthey, 2008)	SCID-I	Postpartum 6 weeks	238	PCA	Unrotated	57.8	3,4,5	n/a	n/a

Australia (Phillips et al., 2009)	SCID-I, BDI-II, BAI	Postpartum 0-12 months	309	EFA/ CFA*	OBLQ (Direct OBLM)	56.44	3,4,5	1,2,6,7,8,9,10	n/a
Australia (Swalm et al., 2010)	Demographic and psychosocial risk questionnaire	Antepartum and Postpartum 6-12 weeks	4,706 3,853	PCA	ORT (VAR)	5458	Antenatal:3,4,5,*6,7,8,9* Postnatal: 3,4,5,*6,7,8,9*	Antenatal:1,2,*6,7,8,9*,10 Postnatal:1,2,*6,7,8,9*,10	n/a
Australia (Cunningham et al., 2015)	n/a	Postpartum 0-13 months	636	EFA/ CFA	OBLQ (Geomin)	Not reported	1,2,3,6,7,8,9,10	4,5	n/a
							1,2	3,4,5	7,8,9,10
							1,2	3,4,5	6,7,8,9,10
Canada (Ross et al., 2003)	BSI, HAMD	Antepartum 36 weeks – postpartum 16 weeks	150	PCA	ORT (VAR)	76.1	3,4,5,*6,7*	1,2,*6,7*,8,9	10

Canada (Bowen et al., 2008)	SCID-I	Antepartum 15 weeks	400	EFA	ORT (VAR)	Not reported	3,4,5	1,2,8	10
USA (Logsdon et al., 2010)	CES-D	Adolescent mothers	149	PCA	ORT (VAR)	60	3,4,5,6,7	1,2,8,9,10	n/a
USA (Lee King 2012)		Postpartum 1 week to 9 months	169	CFA	DWLS	Not reported	7,8,9,10	1,2	3,4,5
USA (Hartley et al., 2014)	n/a	Postpartum women, 0-10 months	220	CFA	Not reported	Not reported	1,2,8,9	3,4,5	n/a
France (Guedeney and Fermanian, 1998)	RDC	Postpartum 0-4 months (mean: 7 weeks)	87	PCA	ORT (VAR)	53.30	3,4,5,6,7,*8,9*	1,2,*8,9*	n/a
France (Chabrol and Teissedre, 2004)	n/a	Postpartum 2-3 days	299	EFA/ CFA*	ORT (VAR)	51.2	3,4,5,6,7	8,9,10	1,2
France (Adouard et al., 2005)	SCID-I	Antepartum 28-34 weeks (High-risk pregnancies)	60	PCA	ORT (VAR)	62.0	3,4,5,6,*7,9*,10	1,2,7,8,*9*	n/a
Canada	SCID-I	Postpartum	224	PCA	OBLQ	Not	3,4,5	1,2,6,7,8,9,1	n/a

(Des Rivières-Pigeon et al., 2000)		3-5 weeks			(VAR)	reported	0	
Hungary (Nagy et al., 2011)	BDI-I	Postpartum 3-26 weeks	1030	PCA	OBLQ (VAR)	53.56	3,4,5,6	1,2,7,8,9,10	n/a
Hungary (Töreki et al., 2013)	SCID-I	Antepartum 12 weeks	219	PCA*	ORT (VAR)	43.2	2,4,5,6,10	3,8,9	1,7
Hungary (Töreki et al., 2014)*	SCID-I	Postpartum 6 weeks	266	PCA/ CFA*	OBLQ (VAR)	54.6	3,4,5,6	1,2,9,10	n/a
Spain (Maroto Navarro et al., 2005)	BDI-I, SCID-I	Postpartum at discharge	75	EFA	ORT (VAR)	56	2,3,4,5,6,9	1,7,8,10	n/a
Peru (Zhong et al., 2014)	n/a	Antepartum 0-16 weeks	1517	PCA CFA	ORT (VAR) Not reported	53.0 Not reported	3,4,5,6,7,8,9,10 *8,9*,10	EFA: 1,2 1,2	n/a 3,4,5,6,7,*8,9*
Norway/Norwegian (Berle et al., 2003)	MINI	Postpartum 6-12 weeks	411	PCA	OBLQ (VAR)	46.6	1,2,3,4,5,6,7,8,9	n/a	n/a
Australia/Tagalog	n/a	Postpartum 6-7 months	106	PCA	ORT (VAR)	60.8	3,4,5,8	6,7,9,10	1,2

(Small et al., 2007)									
Australia/Turkish (Small et al., 2007)	n/a	Postpartum 6-7 months	104	PCA	OBLQ (VAR)	60.3	3,4,5,6,7,8	9,10	1,2
Australia/Vietnamese (Small et al., 2007)	n/a	Postpartum 6-7 months	103	PCA	OBLQ (VAR)	67.6	1,2,*3*,6,8,9	*3*,4,5,7	10
Iran/Persian (Montazeri et al., 2007)	Short Form Health Survey	Postpartum 8-10 weeks	100	PCA	ORT (VAR)	58.0	3,4,5,*8*	6,7,*8*,9,10	1,2
Iran/Persian (Mazhari et al., 2007)	GHQ-12, SCID-I	Postpartum (not reported in which week)	600	PCA	ORT (VAR)	Not reported	1,2,*8*	3,4,5,6,7,*8*	n/a
Greece/Greek (Vivilaki et al., 2009)*	BDI-I	Postpartum 12 weeks	120	PCA/ CFA*	ORT (VAR)	27.02	4,5,6	7,8,9	n/a
China/Mainland Chinese (Lau et al., 2010)	BDI-I, DAS, SF-12, SCID	Postpartum 3-5 days	300	PCA/ CFA*	ORT (VAR)	57.55	1,2	3,4,5	6,7,8,9,10
Brazilia/Portugese	n/a	Postpartum 0-5 months	811	PCA/ CFA*	OBLQ (Geom	73.1	1,2,6	3,4,5	7,8,9,10

(Reichenheim et al., 2011)					in)				
Italy/Italian (Petrozzi and Gagliardi, 2013)		Postpartum 0-3 days and 3 months	594	EFA	OBLQ (Promax)	Not reported	7, 8, 9, 10	3, 4, 5, 6	1,2
Japan/Japanese (Kubota et al., 2013)	RDC	Postpartum 1 month	345	EFA/ CFA*	OBLQ (Promax)	64.4	1,2	3,4,5	7,8,9
Sri Lanka/Sri Lankan (Agampodi and Agampodi, 2013)	SCID-I	Antepartum 24–36 weeks	376	PCA	ORT (VAR)	42.4	1,2,*8*	3,4,5,7,*8*, 9,10	n/a
Sweden/Swedish (Massoudi et al., 2013)	HAD-A, SCID-I	Postpartum 3 months	926	EFA	OBLQ (Direct OBLM)	50.4	1,2,3,6,7,8,9	4,5	n/a
Singapore/ English, Chinese, Malay or Tamil	n/a	Antepartum 26 and 28 weeks	920	EFA	OBLQ (Direct OBLM)	Not reported	1,2,*7,8,9,10*	3,4,5,6,*7,8,9, 10*	n/a
						Not reported	1,2	4,5	6,7,8,9 ,10

(Kwan et al., 2015						rted			
USA/Hebrew (Bina and Harrington, 2016)	n/a	Postpartum 6 weeks	715	CFA	DWSL	Not reported	1,2,6,7,8,9,10 1,2,7,8,9,10 7,8,9,10	3,4,5 3,4,5 3,4,5	n/a n/a 1,2,6
Serbia/Serbian (Odalovic et al., 2017)	n/a	Antepartum Postpartum (within one year)	76 125	PCA/ CFA*	ORT (Var)	Not reported	3,4,5,*6* 3,4,5,*6,8*	*6*,7,8,9,10 7,*8*,9,10	1,2 1,2,*6*

EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; PCA: Principal Component Analysis; RDC: Research Diagnostic Criteria; CIDI: Composite International Diagnostic Interview; CES-D: Center for Epidemiologic Studies of Depression Instrument; SCID-I: Structured Clinical Interview on DSM-IV for Axis I Disorders; MINI: Mini International Neuropsychiatric Interview for DSM-IV; BDI: Beck Depression Inventory; HADS: Hospital Anxiety and Depression Scale; PANAS: Positive and Negative Affect Scales, BAI: Beck Anxiety Inventory; BSI: Brief Symptom Inventory, HAMD: Hamilton Rating Scale for Depression, DAS: Dyadic Adjustment Scale; SF-12: standard SF-12 Health Survey; *: exploratory factor analysis was checked back with confirmatory factor analysis; n/a: not applicable; ORT: orthogonal, VAR: varimax, OBLM: oblimin, OBLQ: oblique, DWLS: diagonally weighted least squares, FIML: Full Information Maximum Likelihood

Table 2. Sociodemographic and obstetric anamnestic data in the antepartum (n=2,967) and in the postpartum sample (n=714)

	Antepartum sample (n=2,967)		Postpartum sample (n=714)	
	n	%	n	%
Age (mean±S.D.)*(year)	30.17±4.78		30.91±4.82	
≤ 24	332	11.2	76	10.64
25-30	1,010	37.1	230	32.2
≥ 31	1,536	51.7	408	57.1

Marital status				
Living with a partner (married or cohabitant)	2,528	85.2	607	85.0
Single	439	14.8	107	15.0
Number of children (mean±S.D.)*	0.63±0.86		1.68±0.89**	
Primiparity	1,478	49.8	309	43.3
Planned pregnancy	2,602	87.7	630	88.2

*: S.D. = standard deviation,

**: the postnatal sample has one child more

Table 3. Factor loadings in an EFA of the Edinburgh Postnatal Depression Scale items in the antepartum (n=2,967) and in the postpartum sample (n=714)

	Antepartum sample (n=2,967)	Postpartum sample (n=714)
Item	Factor 1	Factor 1
Item 1 (anhedonia)	0.408	0.011
Item 2 (anhedonia)	0.407	0.434
Item 3 (guilt)	0.467	0.519
Item 4 (anxiety)	0.628	0.667
Item 5 (panic attacks)	0.583	0.733
Item 6 (overwhelmed)	0.564	0.434
Item 7 (sleep disorders)	0.534	0.539

Item 8 (sadness)	0.667	0.630
Item 9 (tearfulness)	0.633	0.639
Item 10 (suicidal ideas)	0.270	0.256

EFA: exploratory confirmatory factor analysis

Table 4. Edinburgh Postnatal Depression Scale confirmatory factor analysis of models and corresponding indices based on the literature available at the time of publication evaluated in our own antepartum and postpartum samples

Reference	Factor structure	p value	df	RMSEA	CFI	TFI	AIC
Antepartum sample (n=2,967)							
Matthey (2008)†	F1:3,4,5	<0.001	0	0.00	1.00	1.00	30874.175
Theory-driven model 5†	**F1:1,2; F2:4,5; F3:8,9**	**<0.001**	**6**	**0.009**	**0.999**	**0.998**	**39631.699**
Brouwers et al. (2001)‡	F1:1,2,8; F2:3,4,5	<0.001	8	0.101	0.863	0.743	46461.865
Theory-driven model 2†	**F1:1,2; F2:3,4,5; F3:6,10**	**<0.001**	**11**	**0.026**	**0.985**	**0.972**	**48011.952**
Theory-driven model 6†	**F1:1,2; F2:4,5; F3:3,6,10**	**<0.001**	11	**0.036**	**0.973**	**0.948**	**48087.836**
Vivilaki et al. (2009)†	F1:4,5,6; F2:7,8,9	<0.001	8	0.053	0.975	0.953	48947.031
Theory-driven model 1†	**F1:1,2; F2:3,4,5; F3:8,9**	**<0.001**	**11**	**0.020**	**0.995**	**0.990**	**51774.088**
Ross et al. (2003), Hartley et al. (2014)‡	F1:1,2,8,9; F2:3,4,5	<0.001	13	0.079	0.903	0.843	52413.849
Töreki et al. (2014)‡	F1:1,2,9,10; F2:3,4,5,6	<0.001	19	0.060	0.899	0.851	54865.457
Kwan et al. (2015)†	F1:1,2; F2:4,5; F3:6,7,8,9,10	<0.001	24	0.028	0.980	0.969	55549.227
Tuohy and McVey (2008), Cunningham et al. (2015), Lee	F1:1,2; F2:3,4,5; F3:7,8,9,10	<0.001	24	0.024	0.984	0.977	58092.575

King (2012)†							
Agampodi and Agampodi (2013)†	F1:1,2,*8*; F2:3,4,5,7,*8*,9,10	<0.001	25	0.042	0.951	0.929	58388.943
Bina and Harrington (2016)†	F1:1,2,7,8,9,10; F2:3,4,5	<0.001	26	0.054	0.915	0.882	58716.796
Kubota et al. (2013)†	F1:1,2; F2:3,4,5; F3:7,8,9	<0.001	17	0.027	0.988	0.980	58768.541
Pop et al. (1992)‡	F1:1,2,3,5,6,7,8,9	<0.001	20	0.071	0.896	0.854	60007.250
Small et al. (2007)†	F1:3,4,5,8; F2:6,7,9,10	<0.001	19	0.050	0.946	0.924	60529.674
Theory-driven model 3†	**F1:1,2; F2:3,4,5; F3:6,10; F4:8,9**	**<0.001**	21	**0.022**	**0.989**	**0.982**	**60565.610**
Theory-driven model 4†	**F1:1,2; F2:4,5; F3:3,6,10; F4:8,9**	**<0.001**	24	**0.045**	**0.948**	**0.922**	**60941.904**
Mazhari et al.(2007)†	F:1,2,*8*; F2:3,4,5,6,7,*8*	<0.001	18	0.043	0.966	0.947	62822.388
Odalovic et al. (2017)†	F1:1,2,*6*; F2:3,4,5,*6*,*8*; F3:7,*8*,9,10	<0.001	30	0.027	0.980	0.969	67564.132
Odalovic et al. (2017)†	F1:1,2; F2:3,4,5,*6*; F3:*6*,7,8,9,10	<0.001	31	0.028	0.978	0.968	67585.958
Lau et al. (2010), Cunningham et al. (2015), Coates et al. (2017)†	F1:1,2; F2:3,4,5; F3:6,7,8,9,10	<0.001	32	0.030	0.975	0.964	67619.273
Montazeri et al. (2007)†	F1:1,2; F2:3,4,5,*8*; F3:6,7,*8*,9,10	<0.001	31	0.030	0.975	0.964	67614.622
Petrozzi and Gagliardi (2013)†	F1:1,2; F2:3,4,5,6; F3:7,8,9,10	<0.001	32	0.032	0.970	0.957	67669.427
Zhong et al. (2014)†	F1:1,2; F2:3,4,5,6,7,*8,9*; F3:*8,9*,10	<0.001	30	0.038	0.961	0.941	67732.430
Chabrol and Teissedre (2004)†	F1:1,2; F2:3,4,5,6,7; F3:8,9,10	<0.001	32	0.036	0.963	0.948	67739.743
Kwan et al. (2015)†	F1:1,2,*7,8,9,10*;	<0.001	30	0.045	0.946	0.920	67871.665

	F2:3,4,5,6,*7,8,9,10*						
Swam et al. (2010)†	F1:3,4,5,*6,7,8,9*; F2:1,2,*6,7,8,9*,10	<0.001	30	0.044	0.944	0.921	67873.563
Small et al. (2007)†	F1:1,2; F2:3,4,5,6,7,8; F3:9,10	<0.001	32	0.042	0.949	0.928	67883.049
Zhong et al. (2014), Coates et al. (2017)†	F1:1,2; F2:3,4,5,6,7,8,9,10	<0.001	34	0.041	0.949	0.933	67883.763
Reichenheim et al. (2011), Bina and Harrington (2016)†	F1: 1,2,6; F2:3,4,5; F3:7,8,9,10	<0.001	32	0.045	0.943	0.920	67944.197
Astbury et al. (1994), Des Rivieres-Pigeon et al. (2000), Phillips et al. (2009) , Coates et al. (2017), Bina and Harrington (2016)†	F1:1,2,6,7,8,9,10; F2:3,4,5	<0.001	34	0.051	0.920	0.894	68212.545
Pop et al. (1992)†	F1:1,2,6,8; F2:3,4,5,6; F3:7,8,9	<0.001	22	0.032	0.982	0.970	68238.820
Adouard et al. (2005)†	F1:1,2,7,8,*9*; F2:3,4,5,6,7,*9*,10	<0.001	31	0.057	0.911	0.870	68276.530
Nagy et al. (2011)†	F1:1,2,7,8,9,10; F2:3,4,5,6	<0.001	34	0.053	0.913	0.885	68282.996
Massoudi et al. (2013)†	F1:1,2,3,6,7,8,9,10; F2:4,5	<0.001	34	0.054	0.911	0.882	68310.375
Cunningham et al. (2015)†	F1:1,2,3,*4*,6,7,8,9,10; F2:*4*,5	<0.001	33	0.054	0.914	0.882	68312.375
Small et al. (2007)†	F1:1,2,8; F2:3,4,5,6,7; F3:9,10	<0.001	32	0.057	0.907	0.869	68350.043
Logsdon et al. (2010)†	F1:3,4,5,6,7; F2:1,2,8,9,10	<0.001	34	0.055	0.907	0.876	68354.512

Töreki et al. (2013)†	F1:2,4,5,6,10; F2:3,8,9; F3:1,7	<0.001	32	0.059	0.900	0.860	68409.784
Maroto Navarro et al. (2005)†	F1:1,7,8,10; F2:2,3,4,5,6,9	<0.001	34	0.059	0.893	0.858	68502.094
Cox et al. (1987), Coates et al. (2017)†	F1:1,2,3,4,5,6,7,8,9,10	<0.001	35	0.059	0.892	0.861	68514.013
Guedeney and Fermanian (1998)†	F1:1,2,*8*; F2:3,4,5,6,7,*8*,9	<0.001	24	0.052	0.948	0.922	68555.038
Pop et al. (1992)‡	F1: 1,2,7,8,9; F2:3,4,5,6	<0.001	26	0.065	0.911	0.877	68969.399
Our own antepartum screening sample EFA model, Berle et al. (2003)‡	**F1:1,2,3,4,5,6,7,8,9**	**<0.001**	**27**	**0.072**	**0.889**	**0.852**	**69192.458**
Jomeen and Marteen (2005)	F1:1,2,*8*; F2:3,4,5,*8*; F3:10	no model					
Small et al. (2007)	F1:1,2,3,6,8,9; F2:3,4,5,7; F3:10	no model					
Jomeen and Marteen (2007)	F1:1,2,6,7,8,9; F2:3,4,5; F3:10	no model					
Bowen et al. (2008)	F1:1,2,8; F2:3,4,5; F3:10	no model					

Postpartum sample (n=714)

Reference	Factor structure	p value	df	RMSEA	CFI	TFI	AIC
Brouwers et al. (2001)‡	F1:1,2,8; F2:3,4,5	<0.001	8	0.067	0.972	0.948	2897.133
Matthey (2008)†	F1:3,4,5	<0.001	0	0.000	1.000	1.000	3821.858
Theory-driven model 6‡	**F1:1,2; F2:4,5; F3:3,6,10**	**<0.001**	**6**	**0.060**	**0.965**	**0.912**	**4632.181**
Theory-driven model 5‡	**F1:1,2; F2:4,5; F3:8,9**	**<0.001**	**6**	**0.069**	**0.974**	**0.935**	**5115.409**
Vivilaki et al. (2009)‡	F1:4,5,6; F2:7,8,9	<0.001	8	0.077	0.961	0.926	6099.199

Theory-driven model 2†	**F1:1,2; F2:3,4,5; F3:6,10**	**<0.001**	**11**	**0.052**	**0.964**	**0.932**	**6143.867**
Ross et al. (2003), Hartley et al. (2014)‡	F1:1,2,8,9; F2:3,4,5	<0.001	13	0.070	0.952	0.923	6655.278
Kwan et al. (2015)†	F1:1,2; F2:4,5; F3:6,7,8,9,10	<0.001	24	0.058	0.945	0.917	6918.348
Tuohy and McVey (2008), Cunningham et al. (2015), Lee King (2012)†	F1:1,2; F2:3,4,5; F3:7,8,9,10	<0.001	24	0.054	0.952	0.927	7056.140
Agampodi and Agampodi (2013)†	F1:1,2,*8*; F2:3,4,5,7,*8*,9,10	<0.001	25	0.057	0.944	0.919	7067.069
Töreki et al. (2014)‡	F1:3,4,5,6; F2:1,2,9,10	<0.001	19	0.081	0.890	0.838	7068.577
Bina and Harrington (2016)‡	F1:1,2,7,8,9,10; F2:3,4,5	<0.001	26	0.063	0.928	0.900	7094.915
Kubota et al. (2013)‡	F1:1,2; F2:3,4,5; F3:7,8,9	<0.001	17	0.064	0.956	0.927	7259.083
Small et al. (2007)‡	F1:3,4,5,8; F2:6,7,9,10	<0.001	19	0.067	0.934	0.903	7409.432
Pop et al. (1992)‡	F1:1,2,3,5,6,7,8,9	<0.001	20	0.080	0.913	0.879	7577.122
Theory-driven model 3†	**F1:1,2; F2:3,4,5; F3:6,10; F4:8,9**	**<0.001**	**21**	**0.050**	**0.967**	**0.944**	**7768.929**
Theory-driven model 4†	**F1:1,2; F2:4,5; F3:8,9; F4:3,6,10**	**<0.001**	**24**	**0.039**	**0.929**	**0.893**	**7819.120**
Our own postpartum screening sample EFA model‡	**F1:2,3,4,5,6,7,8,9**	**<0.001**	**20**	**0.075**	**0.932**	**0.905**	**7853.912**
Mazhari et al.(2007)‡	F:1,2,*8*; F2:3,4,5,6,7,*8*	<0.001	18	0.066	0.946	0.916	7870.556
Odalovic et al. (2017)†	F1:1,2,*6*; F2:3,4,5,*6*,*8*; F3:7,*8*,9,10	<0.001	30	0.054	0.950	0.925	8395.928
Reichenheim et al. (2011), Bina and Harrington (2016)†	F1: 1,2,6; F2:3,4,5; F3:7,8,9,10	<0.001	32	0.053	0.949	0.928	8403.675

Montazeri et al. (2007)†	F1:1,2; F2:3,4,5,*8*; F3:6,7,*8*,9,10	<0.001	31	0.056	0.944	0.918	8412.724
Lau et al. (2010), Cunningham et al. (2015), Coates et al. (2017)†	F1:1,2; F2:3,4,5; F3:6,7,8,9,10	<0.001	32	0.056	0.943	0.920	8413.945
Odalovic et al. (2017)†	F1:1,2; F2:3,4,5,*6*; F3:*6*,7,8,9,10	<0.001	31	0.059	0.937	0.909	8415.236
Swam et al. (2010)‡	F1:3,4,5,*6*,*7*,8,9; F2:1,2,*6*,*7*,8,9,10	<0.001	30	0.061	0.936	0.904	8421.462
Petrozzi and Gagliardi (2013)‡	F1:1,2; F2:3,4,5,6; F3:1,7,8,9,10	<0.001	32	0.061	0.932	0.905	8432.530
Chabrol and Teissedre (2004)‡	F1:1,2; F2:3,4,5,6,7; F3:8,9,10	<0.001	32	0.064	0.925	0.894	8444.335
Kwan et al. (2015)‡	F1:1,2,7,8,9,10; F2:3,4,5,6,7,8,9,10	<0.001	30	0.070	0.916	0.874	8444.663
Astbury et al (1994); Des Rivieres-Pigeon et al. (2000); Phillips et al. (2009), Coates et al. (2017), Bina and Harrington (2016)‡	F1: 1,2,6,7,8,9,10; F2:3,4,5	<0.001	34	0.062	0.925	0.901	8447.161
Small et al. (2007)‡	F1:1,2; F2:3,4,5,6,7,8; F3:9,10	<0.001	32	0.065	0.921	0.890	8449.649
Zhong et al. (2014), Coates et al. (2017)‡	F1:1,2; F2:3,4,5,6,7,8,9,10	<0.001	34	0.064	0.920	0.894	8451.983
Massoudi et al. (2013)‡	F1:1,2,3,6,7,8,9,10; F2:4,5	<0.001	34	0.066	0.915	0.887	8462.267
Small et al. (2007)‡	F1:1,2,8; F2:3,4,5,6,7; F3:9,10	<0.001	32	0.068	0.915	0.881	8462.540
Cunningham et al. (2015)‡	F1:1,2,3,*4*,6,7,8,9,10;	<0.001	33	0.066	0.917	0.887	8464.267

	F2:*4*,5						
Adouard et al. (2005)‡	F1:1,2,7,8,*9*; F2:3,4,5,6,7,*9*,10	<0.001	31	0.072	0.906	0.864	8465.110
Nagy et al. (2011)‡	F1:1,2,7,8,9,10; F2:3,4,5,6	<0.001	34	0.066	0.914	0.886	8467.295
Logsdon et al. (2010)‡	F1:3,4,5,6,7; F2:1,2,8,9,10	<0.001	34	0.069	0.908	0.878	8476.425
Töreki et al. (2013)‡	F1:2,4,5,6,10; F2:3,8,9; F3:1,7	<0.001	32	0.074	0.899	0.858	8490.003
Cox et al. (1987), Coates et al. (2017)‡	F1: 1,2,3,4,5,6,7,8,9,10	<0.001	35	0.071	0.898	0.869	8492.023
Maroto Navaro et al. (2005)‡	F1:1,7,8,10; F2:2,3,4,5,6,9	<0.001	34	0.073	0.896	0.862	8493.150
Pop et al. (1992)†	F1:1,2,6,8; F2:3,4,5,6; F3:7,8,9	<0.001	22	0.051	0.970	0.950	8577.035
Guedeney and Fermanian (1998)‡	F1:1,*2*,*8*,*9*; F2:3,4,5,6,7,*8*,*9*	<0.001	24	0.076	0.928	0.891	8644.312
Pop et al. (1992)‡	F1: 1,2,7,8,9; F2:3,4,5,6	<0.001	26	0.078	0.917	0.886	8668.552
Berle et al. (2003)	F1:1,2,3,4,5,6,7,8,9	<0.001	36	0.229	0.000	0.000	10344.877
Jomeen and Marteen (2005)	F1:1,2,*8*; F2:3,4,5,*8*; F3:10	no model					
Small et al. (2007)	F1:1,2,*3*,6,8,9; F2:*3*,4,5,7; F3:10	no model					
Jomeen and Marteen (2007)	F1:1,2,6,7,8,9; F2:3,4,5; F3:10	no model					
Bowen et al. (2008)	F1:1,2,8; F2:3,4,5; F3:10	no model					
Zhong et al. (2014)	F1:1,2; F2:3,4,5,6,7,*8*,*9*; F3:*8*,*9*,10	no model					

95% CI: 95% confidence interval; RMSEA: root mean square error of approximation; TLI: Tucker-Lewis index; CFI: comparative fit index; AIC: Akaike information criterion

†RMSEA <0.06 (good model fit), ‡RMSEA <0.11 (acceptable model fit)

bold: our own antepartum EFA model and our theory-driven models

italics: cross-loading items

Highlights

- We found different factor structures when collecting more cases from the same population.
- The EPDS showed moderate within- and between-culture invariability, although this would also need to be re-examined with a theory-driven approach.
- Using multiple fit indices, our theory-driven anhedonia (items 1 & 2) – anxiety (items 4 & 5) – low mood (items 8 & 9) model in both the ante and the postpartum sample, and the anhedonia – anxiety – suicidal risk factors (items 3, 6, and 10) model in the postpartum sample were again among the best performing models.
- We propose a new methodological approach whereby in factor analytic studies empirical findings from exploratory approaches are triangulated with neurobiological insight when generating theoretically-driven models for testing with a confirmatory approach.