Psychiatry Research 264 (2018) 182–191

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Establishing a coherent and replicable measurement model of the Edinburgh Postnatal Depression Scale

Colin R. Martin[a], Maggie Redshaw[b,*]

[a] Professor of Perinatal Mental Health, Faculty of Health Sciences, University of Hull, UK
[b] Senior Research Fellow, Policy Research Unit in Maternal Health and Care, National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, UK

ARTICLE INFO

Keywords:
Edinburgh Postnatal Depression Scale
EPDS
Factor structure
Validity
Psychometrics

ABSTRACT

The 10-item Edinburgh Postnatal Depression Scale (EPDS) is an established screening tool for postnatal depression. Inconsistent findings in factor structure and replication difficulties have limited the scope of development of the measure as a multi-dimensional tool. The current investigation sought to robustly determine the underlying factor structure of the EPDS and the replicability and stability of the most plausible model identified. A between-subjects design was used. EPDS data were collected postpartum from two independent cohorts using identical data capture methods. Datasets were examined with confirmatory factor analysis, model invariance testing and systematic evaluation of relational and internal aspects of the measure. Participants were two samples of postpartum women in England assessed at three months (n = 245) and six months (n = 217). The findings showed a three-factor seven-item model of the EPDS offered an excellent fit to the data, and was observed to be replicable in both datasets and invariant as a function of time point of assessment. Some EPDS sub-scale scores were significantly higher at six months. The EPDS is multi-dimensional and a robust measurement model comprises three factors that are replicable. The potential utility of the sub-scale components identified requires further research to identify a role in contemporary screening practice.

1. Introduction

Postnatal depression (PND) represents a significant mental health concern with an average of 13% of women experiencing this distressing condition O'Hara and Swain (1996), though reported rates differ considerably, for example Banti et al. (2011). The impact of PND is pervasive, with robust evidence of deleterious impact not only on the woman herself (Pope et al., 2013; Wisner et al., 2013), but also on her baby (Dahlen et al., 2015; Fairbrother and Woody, 2008; Jennings et al., 1999; Milgrom and Holt, 2014) and partners (Cameron et al. 2017). Paradoxically, given the impact of PND, universal screening for all postnatal women is currently not policy (American College of Obstetricians and Gynecologists' Committee on Obstetric American College of Obstetricians and Gynecologists' Committee on Obstetric Practice, 2015), current practice in the UK being to consider a brief screen by health professionals using two identification questions and a follow up to a positive response to either question with a validated screening measure or a referral (National Institute for Health and Care Excellence, 2015). The most widely used screening measure for PND is the Edinburgh Postnatal Depression Scale (EPDS) developed by J. L. Cox et al. (1987). A driver in the development of the EPDS was the avoidance of items which could be influenced by physical symptoms (J. L. Cox et al., 1987), a critical aspect in approaching screening given the significant physiological changes that accompany pregnancy and childbirth. The EPDS has endured as the most widely used PND screening measure (Moraes et al., 2017; Smith et al., 2016).

Despite, the extensive use of the EPDS as a screening instrument, the measure has also been noted for some contradictory observations in terms of its measurement structure. The measure itself was originally developed to be a unitary measure of (postnatal) depression, however, a multitude of studies have demonstrated the EPDS to have an underlying multi-dimensional factor structure (Brouwers et al., 2001; Gollan et al., 2017; Jomeen and Martin, 2007; Matthey, 2008; Phillips et al., 2009; Reichenheim et al., 2011; Ross et al., 2003; Tuohy and McVey, 2008). The findings of such studies constructively suggest that the EPDS may comprise sub-scale domains of potential and added clinical value (Matthey, 2008). At the same time they indicate that the tool itself does not appear to measure what it was designed to measure (depression) and consequently may be limited in terms of both screening

* Corresponding author.
E-mail addresses: c.r.martin@hull.ac.uk (C.R. Martin), maggie.redshaw@npeu.ox.ac.uk (M. Redshaw).

https://doi.org/10.1016/j.psychres.2018.03.062
Received 20 November 2017; Received in revised form 9 March 2018; Accepted 22 March 2018
Available online 23 March 2018

effectiveness (Matthey and Agostini, 2017) and links to a coherent clinical and unidimensional model of postnatal depression (Gollan et al., 2017). Nevertheless, the notion of a multi-dimensional underlying structure to the EPDS need not necessarily detract from its clinical utility, with identification of robust independent sub-scales embedded within the tool not anticipated by the instrument developers (Matthey, 2008). However, there must be consideration of structural stability, and the multidimensional structure of the EPDS and the embedded sub-scales, should be replicable across groups, for example, depressed/non-depressed, white/black minority ethnic, high social economic status (SES)/low SES (Matthey and Agostini, 2017). This has not been found to be the case, with evidence of wide variation in the items assigned to factors across a range of studies, even within the context of two-factor, or three-factor models which have been the most pervasive factorial determinations of measurement studies of the EPDS (Chabrol and Teissedre, 2004; Jomeen and Martin, 2007; Pallant et al., 2006; Ross et al., 2003; Tuohy and McVey, 2008). Interpretation of the content of underlying factor domains within the EPDS has thus been problematic, due to inconsistent factor structure, with most two factor model solutions reporting domains of anxiety and depression, though the domains themselves have been indicated by different individual items across studies (Reichenheim et al., 2011). Clearly, such unreconciled differences across studies are unsatisfactory in terms of theoretical coherence and practical clinical interpretation. The possibility that the underlying structure of the EPDS may indeed map onto a theoretically robust multi-dimensional model of depression could be inferred by the study of Tuohy and McVey (2008) who described the third factor in their tri-dimensional analysis as representing 'anhedonia'. This observation is not only consistent with an important component of the tri-dimensional model of depression suggested by Clark and Watson (1991) but also resonates with the finding of a tri-dimensional structure which includes an anhedonia domain to the Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983), another screening measure that has been frequently used within the perinatal field (Jomeen and Martin, 2008a; Meades and Ayers, 2011; Tohotoa et al., 2012). Reichenheim et al., (2011) conducted an elegant study examining the underlying factor structure of the EPDS, finding evidence for three factors but ultimately recommending the use of a unitary total score to best represent the measurement model of tool. This was premised on the basis of a superior fit of a bi-factor model comprising a general factor and three specific factors, however it has been suggested that superior fit of bifactor models could be due to a 'method effect' in contrast to the empirical superiority of the underlying model which should be specified on conceptual and theoretical grounds (Morgan et al., 2015).

It is noteworthy also that the majority of studies examining the measurement properties of the EPDS have been cross-sectional in design. This is important as the recommendations of not only when to screen for PND but indeed, when PND may be diagnosed as a disorder distinct from major depressive disorder vary dramatically from birth to twelve months depending on the timing of assessment. A number of these cross-sectional studies have recruited across a broad sample range post-partum, for example, women from birth to ten months post-partum (Hartley et al., 2014), between birth and one year (Phillips et al., (2009) and much closer to the birth at 2–3 days postpartum (Teissedre and Chabrol, 2004). For such studies to be compared, a fundamental assumption must be that the underlying structure should be consistent across time. In a large sample (N ~ 1200) study strong evidence was found for a tri-dimensional structure to the EPDS that was consistent in both antenatal and postnatal samples (Coates et al., 2016).

A relatively small number of studies have examined the longitudinal structure of the EPDS and findings from these studies are potentially helpful given the clinical reality of variations in screening times and screening opportunities for PND. A study was conducted on the measurement of women's mental health at admission and at discharge to psychiatric mother and baby units (Cunningham et al., 2015). Uniquely, this study focused on a clinical group with a confirmed psychiatric diagnosis and incorporating implicitly the effect of intervention on outcome. It was observed that the EPDS comprised two distinct factors on admission and three distinct factors on discharge and concluded that women may interpret EPDS items in characteristically different ways as a function of their degree of psychological/psychiatric distress (Cunningham et al., 2015). The finding from this study that the EPDS measures different constructs at different time points is far-reaching in terms of screening practice and research. However, an important caveat, recognised by the investigators themselves, was that the sample represented a distinct population with diagnosed and significant mental illness requiring in-patient admission and that of course, consequentially, therapeutic intervention represented an inevitable component of the journey between admission and discharge. It is therefore difficult to conclude whether factorial instability would generalise to populations without severe mental illness (Cunningham et al., 2015). This is particularly salient given that the majority of women following birth do not develop PND and the time that they may be screened for PND may vary. A critical issue, specifically, is whether the most robust empirically-derived factorial structure of the EPDS is replicable and consistent in normal population samples drawn at different postpartum intervals.

The objectives of the current study are to:

1. Evaluate comparative model fit of empirically-derived multi-dimensional models of the EPDS against the single factor model.
2. Evaluate comparative model fit of the equivalent tri-dimensional model of the EPDS against a bifactor model of the EPDS as proposed by Reichenheim et al. (2011).
3. Demonstrate the replicability and stability of the best-fit model of the EPDS across time.
4. Determine the measurement coherence of mean EPDS scores across time points.
5. Evaluate the equivalence of EPDS total and sub-scale internal consistency across time points.
6. Determine the equivalence of EPDS total and sub-scale correlational relationships between time points.
7. Evaluate case classification rate concordance between the conventional 10-item EPDS and the recent 7-item EPDS suggested by Gollan et al. (2017).

2. Methods

2.1. Participants

Data were collected from a randomly selected sample of women in England at either three months (time point 1) or six months (time point 2) postpartum, these being two separate samples thus the use of a between-subjects design. The sample was drawn by the Office for National Statistics who managed the mailing. A questionnaire was sent to each woman selected, with an invitation letter and an information leaflet, followed by a further questionnaire and reminder as appropriate. Women aged less than 16 years were excluded as were those whose babies had died in the months after birth. Completion of the questionnaire was taken as implicit consent to participate. No incentives were offered for questionnaire return.

2.2. Design

A between-subjects design was used to investigate the study objectives in this secondary analysis study. To address objectives 1 (evaluate EPDS model fit) and 2 (tri-dimensional/bifactor model comparison) data were collapsed between time points and single-factor, tri-dimensional and bifactor models compared. To address objective 3 (replicability and stability of best-fit model) the most convincing model found evaluating objectives 1 and 2 would be evaluated using data stratified