Journal of Affective Disorders 234 (2018) 174–179

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

# A comparison of the PHQ-2 and MGMQ for screening for emotional health difficulties during pregnancy

Stephen Matthey[a,*], Frances Bilbao[b]

[a] South Western Sydney Local Health District; University of Sydney and UNSW, Sydney, Australia
[b] Mums Matter Psychology, Melbourne, Australia

ARTICLE INFO

*Keywords:*
Perinatal mental health
Screening
Assessment
Depression
Distress

ABSTRACT

*Background:* Routine screening for emotional health difficulties in women during pregnancy is now advocated in several countries. There is a need therefore to compare the performance of different self-report measures to accomplish this. This study reports on the comparative performance of two such measures– the well-established PHQ-2, which aims to detect depression, and the more recent MGMQ, which aims to detect a wide array of negative emotions.

*Method:* Women (N = 2292) attending a public hospital antenatal clinic over a 14-month period completed the two measures, either on their own (72%), verbally administered by the midwife (25%), or with an interpreter (3%).

*Results:* Similar rates of women screened positive on each instrument (PHQ-2: 11.6%; MGMQ: 12.3%), but the overlap between the two measures was low. The PHQ-2 only detected 58% of the MGMQ screen positive women, while the MGMQ detected 89.5% of the PHQ-2 screen positive women. No clinically meaningful difference in screen positive rates on either measure was evident for the administration method.

*Limitations:* No demographic data were available apart from gestational age, and only about half the women presenting to the clinic during the time period were screened with the measures.

*Conclusion:* The MGMQ detected a greater proportion of women screening positive on the PHQ-2 than vice-versa. This is part due to the MGMQ's focus on a wider range of negative emotions than just depression. Accumulating evidence for this MGMQ indicates that clinical services can consider using this simple measure if they wish to screen for a broad range of negative emotions during pregnancy and postnatally.

## 1. Introduction

The argument for screening women for possible emotional difficulties during the perinatal period is now well supported (eg., Austin and Lumley, 2003; Campagne, 2004; Chaudron et al., 2004; NICE, 2014). The need for such screening has largely come from the evidence that poor emotional health in women during the perinatal period is common, with around 10% of women meeting diagnostic criteria for depression and 15% meeting diagnostic criteria for an anxiety disorder (Misri et al., 2016; Dennis et al., 2017). These moods have been shown to have significant negative impacts not only on the woman and her partner, but also on the foetus, their baby and later child development (eg., Buss et al., 2010; Glasheen et al., 2010). Research has also shown that many women with these difficulties go undetected when services simply rely on ad-hoc clinical judgment (Chaudron et al., 2004; Murray et al., 2004). Clinical practice guidelines or recommendations have therefore been introduced in several countries to improve this situation (eg, Australia: beyondblue, 2011: UK: NICE, 2014; America: Committee on Obstetric Practice, 2015; Canada: British Colombia Best Practice Guidelines for Mental Health Disorders in the Perinatal Period, 2014).

Screening, importantly, is not seen as a 'diagnostic' or 'in-depth' assessment of the woman's mood (Milgrom et al., 2011). It is a first-step, from which a screen-positive response indicates further exploration to determine the nature and extent of any emotional difficulties. This first-stage screening is usually conducted within a health setting such as an antenatal clinic or early child health centre, where women routinely come into contact with the service for reasons not primarily associated with their emotional health. Given this context, emotional health screening is not ordinarily a main focus, and thus such services are usually stretched for both resources and time to include this aspect of health care. This is an important consideration when clinical services elect to implement screening.

The choice of which screening instrument to use is therefore dependent on the time it takes to administer, as well as other factors,

* Corresponding author.
*E-mail address:* stephen.matthey@sswahs.nsw.gov.au (S. Matthey).

https://doi.org/10.1016/j.jad.2018.02.069
Received 23 August 2017; Received in revised form 12 February 2018; Accepted 25 February 2018
Available online 27 February 2018
0165-0327/ 

including its efficiency at detecting women with significant emotional health difficulties; the ease and accuracy of administration, both for the clinical staff and the client; and which emotional health problems a service wishes to screen for.

Within perinatal screening contexts, the Australian, American and Canadian guidelines have recommended the Edinburgh Postnatal Depression Scale (EPDS, or EDS outside of the postnatal period, such as in pregnancy; Cox et al., 1987, 1996) for depression screening, while the UK Guidelines recommend the 'Whooley Questions' (Whooley et al., 1997), which originated from the PRIME-MD (Spitzer et al., 1994). Another brief depression screening measure that has been used is the Primary Health Questionnaire (PHQ: Kroenke et al., 2003), which is also a derivative of the PRIME-MD, in either its 2-item version (PHQ-2: eg., Bennett et al., 2008; Chae et al., 2012), its 8-item version (PHQ-8: eg., Ashley et al., 2016), or its 9-item version (PHQ-9: eg., Davis et al., 2013; Gjerdingen et al., 2009). Due to the increasing recognition that mood screening should now encompass not just depression, but also anxiety (Grant et al., 2008; Weisberg and Paquette, 2002; Wilson et al., 2004), the UK Guidelines also recommend the use of the Generalised Anxiety Disorder questions (GAD-2: Spitzer et al., 2006).

The inclusion however of two measures within a universal screening programme – one for depression, and one for anxiety, is likely to be impractical for many services due to time restraints (Chae et al., 2012; Kabir et al., 2008). Reliable administration and scoring may also be an issue, particularly if different measures have different scoring procedures or 'screen positive' cut-off scores. Indeed, Fontein-Kuipers (2015) and Ali et al. (2016) have recently stated that perinatal research should explore ways of assessing for a variety of moods within the one measure. While some multi-mood self-report measures exist, these tend to be quite long (eg., the 21 item short version of the Depression, Anxiety and Stress Scale: DASS-21, Lovibond and Lovibond, 1995), with different cut-off scores for the different mood sub-scales (e.g., the three subscales of the DASS-21; the EPDS depression total score and the anxiety subscale score), and thus may be considered impractical within clinical settings

As a result of the above considerations, a brief screening measure, the Matthey Generic Mood Questionnaire (MGMQ: Matthey et al., 2013a) which comprises between two and four questions, has recently been developed. It is designed to detect a broad range of emotional health difficulties, and to allow for a quick assessment of the probable impact of any such negative emotion on the individual.

The study to be reported compares the performance of two of the above brief screening measures within a trial of antenatal screening in a public hospital in Victoria, Australia. These measures were the PHQ-2 (Kroenke et al., 2003), and the MGMQ (Matthey et al., 2013a). As part of its Perinatal Emotional Health Program (PEHP), the hospital chose to compare these two measures to understand their shared and unique outcomes and benefits. They were selected by the clinical service based upon their brevity, the documented supporting evidence for the PHQ-2, and the wider array of emotional difficulties that the MGMQ screens for. Both were considered to have a simple format that made it easy for clinicians to score, and also to determine what was a 'screen positive' response or score on each measure. In addition, both measures ask respondents to report on their mood over the same time period (the past two weeks), making their comparison easier to interpret.

## 2. Method

### 2.1. Participants

Participants were pregnant women attending an Australian public hospital in Victoria for antenatal care (mean gestational age: 18 weeks; sd: 5.3 weeks; range 6–39 weeks). The majority spoke English, with some requiring an interpreter. Participants were screened over a 14-month period (between March 2014 and April 2015) as part of routine antenatal care and came from the general population. Although trained midwives were instructed to screen all women in their first appointment, screening was administered at the midwives' discretion, depending on the time available for the appointment and other medical tasks to be completed.

### 2.2. Measures

#### 2.2.1. Antenatal emotional health screening form (AEHS)

This one-page self-report form incorporated the two measures described below, with the following instruction: "As you are pregnant, we would like to know how you are feeling. Please check the answer that comes closest to how you have felt in the past 2 weeks, not just how you feel today". The AEHS then consisted of the two PHQ-2 questions, followed by the two MGMQ questions.

#### 2.2.2. PHQ-2 (Kroenke et al., 2003)

This is a derivative of the PHQ-9 and PRIME-MD (Kroenke et al., 2001; Spitzer et al., 1999, 1994), and enquires about the respondent's mood over the past two weeks, It consists of the following two questions: i) I have been bothered by having little interest or pleasure in doing things, and ii) I have been bothered by feeling down, depressed or hopeless. The four possible frequency response options for each question are: 'Not at all', 'Several days', 'More than half the days', and 'Nearly everyday'. These are scored 0–3, with higher scores reflecting increasing frequency. The PHQ-2 thus has a possible range of 0–6, with a positive screen for probable major depression being a score of 3 or more in English-speaking adults (Kroenke et al., 2003).

The PHQ-2 has been shown to perform well in comparison to the EPDS (Bennett et al., 2008; Chae et al., 2012), and against other instruments and diagnostic status (Lowe et al., 2005). Kroenke et al. (2003) report its receiver operating characteristics, using a score of 3 or more against a diagnosis of major depression, to be: 83% sensitivity, 95% specificity, and, calculating from the figures provided in their paper, a positive predictive value (ppv) of 61%. The measure has also been used to report on probable rates of depressive disorders in postpartum women (Howell et al., 2010; Jewell et al., 2010; Pooler et al., 2013) and in adults outside of the perinatal period (eg., Stafford et al., 2007).

#### 2.2.3. MGMQ (Matthey et al., 2013a)

The version used in this study consists of two questions: A 'Distress' question: 'Have you felt very stressed, anxious, or unhappy, or found it difficult to cope, for some of the time?' (response options: ' Yes', 'Possibly', 'No'); and a 'Bother Impact' question (for those who answered 'Yes' or 'Possibly' to the Distress question): 'How bothered have you been by these feelings?' (response options: 'Not at all'; 'A little bit'; 'Moderately'; 'A lot'). An expanded version has two additional questions, these being a 'Reason for distress question', asking the individual why she is feeling this way (if appropriate); and a 'Wish for referral' question, asking the individual if she would like to talk to a health professional about how she is feeling (if appropriate).

Matthey et al. (2013a) demonstrated that the Distress question correctly identified 80% of women who met DSM criteria for an anxiety disorder, and 84% who scored above the cut-off threshold for possible anxiety on the anxiety subscale of the EPDS. In addition Matthey and Della Vedova (in press) found that this Distress question identified 63–70% of women screening positive for either depression or anxiety on the EPDS.

Kirkby et al. (2011) recommends the consultation of experts for determining clinically meaningful scores on new interventions, and thus a convenience sample of 26 experienced psychologists were surveyed as to the 'screen positive' threshold on the MGMQ Bother Impact question. Over three quarters of them considered that a response of 'Moderately' or 'A lot', but not that of 'A little bit', would warrant further assessment, within a screening context, and thus it is these responses ('Moderately' or 'A lot') that clinically would constitute a