
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/18197/

Link to published version: https://doi.org/10.1016/j.jad.2017.08.063

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
A systematic review of measures of mental health and emotional wellbeing in parents of children aged 0-5

Rebecca Webb, Susan Ayers and Camilla Rosan

Author Note

Rebecca Webb & Susan Ayers, Centre for Maternal and Child Health Research, City University London, EC1V 0HB

Camilla Rosan, Development and Delivery Department, Mental Health Foundation, Colechurch House, 1 London Bridge Walk, London, SE1 2SX

Correspondence concerning this article should be addressed to: Susan Ayers, Centre for Maternal and Child Health Research, City University London, EC1V 0HB. Email: Susan.Ayers@city.ac.uk. Telephone: 0044 (0)207 0405834

Funded by: National Society for Prevention of Cruelty to Children
Abstract

Background: A significant proportion of women with young children experience mental health problems and recent research suggests fathers may also be affected. This may have a long term negative impact on the child’s development with significant costs to society. Appropriate measures are therefore needed to identify parents and children at risk.

Method: This literature review aimed to identify the most reliable, evidence based global measures of mental health for parents of infants from pregnancy to 5 years postpartum (0-5 years). Literature searches were conducted on online databases and hand searches of reference lists were also carried out. Studies were included in the review if they reported information on measures of global psychological distress or wellbeing from 0-5 years postpartum.

Results: A total of 183 studies were included in the review, 19 of which directly examined the psychometric validity of an outcome measure. These studies reported information on 23 outcome measures, 4 of which had been validated in parents of children from 1-5. These were: the General Health Questionnaire (GHQ), the Symptom Checklist (SCL), the Self-Reporting Questionnaire (SRQ) and the Kessler scale (K10/6). Reliability and validity varied across studies.

Limitations: Only a small number of studies included fathers and examined psychometric validity across the entire period of early childhood.

Conclusions: The GHQ was the most frequently validated but results suggest poor reliability and validity. The SRQ and K10/6 were the most promising measures in terms of psychometric properties and clinical utility.

Keywords: Early years; parental wellbeing; measures; psychometrics; reliability; validity
A systematic review of measures of mental health and emotional wellbeing in parents of children aged 0-5

Pregnancy and the first year after birth are associated with increased risk of affective disorders with research indicating between 10 to 20% of women experience affective disorders during this time (Mann, Gilbody, & Adamson, 2010). Recent research also suggests a significant proportion of men may be affected (Cameron, Sedov, & Tomfohr-Madsen, 2016; Parfitt & Ayers, 2014). Parental mental health problems in the early years can be pre-existing, where existing mental health problems are exacerbated or retriggered during pregnancy or after birth. Alternatively, mental health problems may start during pregnancy or after birth. The most common mental health problems are anxiety, depression, post-traumatic stress disorder (PTSD), and stress-related conditions such as adjustment disorder (Cameron et al., 2016). Severe mental illness, such as puerperal psychosis, is less common but is one of the leading indirect causes of maternal death (Manktelow & Draper, 2015).

Parental mental health problems in the early years are a significant public health concern due the association between such problems and a variety of adverse outcomes for women and their partners. These include parenting distress, poor physical health, financial strain, stressful life events, low social support, and low quality partner relationship (Horwitz, Briggs-Gowan, Storfer-Isser, & Carter, 2007). Children may also be affected. For example, maternal postpartum depression is associated with less sensitive and responsive interactions between the mother and baby (Barry et al., 2015), higher rates of negative emotional expression (Murray, Kempton, Woolgar, & Hooper, 1993), and having a higher likelihood of implementing unpredictable and inconsistent parenting techniques (Beck, Shattuck, Haynie, Crump, & Simons-Morton, 1999). These factors are in turn associated with poor child development (Barry et al., 2015; Fihrer, McMahon, & Taylor, 2009; Grace, Evindar, & Stewart, 2003). Mothers with postpartum depression are also less likely to breastfeed (Dennis &
McQueen, 2009) which in some cases may be related to concerns about the effects of any psychotropic medication they may be taking on their infant (Bonari et al., 2005).

Parental mental health problems are also associated with the infant being more likely to experience mental health problems (Lang & Gartstein, 2017; Wang, 2017). This intergenerational transmission of vulnerability may be due to a number of factors, including epigenetic, environmental and parenting factors. For example, women’s mental health during pregnancy may exert specific effects on the developing foetus through neuro-biological foetal programming which may have a long-term effect on the child’s development and health. There is increasing evidence for foetal programming effects from animal research and epidemiological research showing associations between anxiety and stress in pregnancy and greater risk of a range of adverse outcomes for the child. For example, longitudinal research suggests infants of women who are anxious during pregnancy are more likely to show fearful or anxious behaviour themselves and be at greater risk of poor development and adverse outcomes such as attention-deficit hyperactivity disorder (Talge, Neal, & Glover, 2007). Poorer emotional and behavioural development may even persist into adolescence, with a study of 7,944 families in England finding that maternal anxiety and depression in pregnancy was associated with the child being twice as likely to have a mental disorder (O’Donnell et al, 2014). However, most children are not affected and the mechanisms underlying which children are affected and in which way are not well understood (Glover, 2016).

Paternal mental illness has been less studied than maternal mental illness. However, there is increasing evidence it may also have a negative impact on the couple’s relationship, father-child relationship and child developmental outcomes. For example, paternal stress has been associated with men having more negative perceptions of their marriage and baby (Zelkowitz & Milet, 1997). A meta-analysis of 28 studies of paternal depression found it was significantly associated with a decrease in positive parenting behaviors (such as warmth) and an increase in negative parenting behaviors (such as criticism) (Wilson & Durbin, 2010). Additionally, the early father-child
relationship appeared to influence later childhood outcomes: for example, father’s psychological
distress has been associated with child behavioral difficulties, emotional difficulties, social
functioning and development (Fletcher, Feeman, Garfield, & Vimpani, 2011; Kvalevaag et al., 2013).

In summary, there is substantial evidence suggesting that mental illness for men and women
during the early years may have an adverse impact on infant development and child outcomes across
a range of domains e.g. cognitive, emotional, motor and social (Milgrom, Ericksen, & Sved-
Williams, 2016; Sweeney & MacBeth, 2016). There has been some debate over the importance of
timing and the possibility that pregnancy and the first year may represent a sensitive period in infant
development, in which the infant may be more susceptible to the effects of mental health difficulties
due (in part) to the rapid and substantial neural, cognitive and socio-emotional developments that
occur during this time (Sroufe, 2005; Talge et al., 2007). However, there is some evidence that
sensitivity may continue up to age 5 with, for example, a study of a cohort of 937 adolescents in
Canada finding that those who were initially exposed to maternal depression between the ages of 2
and 5 had a two-fold increased risk of affective disorders in adolescence (Naicker, Wickham &
Colman, 2012).

The economic argument for screening and treating parental mental health problems is
compelling with evidence that the cost to society of parental mental illness is substantial. A recent
economic analysis estimated the cost to UK society of not treating women’s perinatal mental health
problems is £8.1 billion for every annual cohort of women giving birth. The majority of this cost
(72%) was attributable to long term adverse consequences for the child (Bauer, Parsonage, Knapp,
approximately 2,593 women discontinued their antidepressants and had a depressive relapse. This
resulted in maternal healthcare costs of approximately CA$1 million. Similar results have been found
in Australia, where the financial costs associated with maternal postpartum depression were
estimated to be AUS$61 million in a cohort of 70,997 women (Post and Antenatal Depression
Association, 2012). Furthermore, in the USA, Dagher, McGovern, and Dowd (2014) examined the association between depression in the postpartum period and healthcare expenditure 11 weeks after childbirth and found the mean cost for women with postpartum depression was US$1,046 compared to US$365 for women without depression (2001 prices).

From a public health perspective, a strategy for preventing perinatal mental health problems and treating them effectively therefore has the potential to prevent significant long-term burden of ill-health and problems in parents and children. A critical aspect of this is identifying efficient and acceptable measures of mental health for use with men and women during the early years. Guidelines for screening vary in their recommendations (Milgrom & Gemmill, 2014) but most focus on specific disorders, such as depression. In the UK, national guidelines recommend asking the Whooley questions (Whooley, Avins, Miranda, & Browner, 1997) to identify depression and the General Anxiety Disorder-2 question screen (Spitzer, Kroenke, Williams, & Lowe, 2006) to identify anxiety (NICE, 2014). Other countries use the Edinburgh Postnatal Depression Scale to screen for postpartum depression (Milgrom & Gemmill, 2014). However, this method of focusing on a specific disorder (in this case depression or anxiety) is unlikely to identify women or men experiencing different affective disorders such as PTSD or OCD. This means only a proportion of parents are offered treatment to reduce the possible adverse impact of mental health problems on them and their child.

Another problem is that many of the clinical guidelines for screening do not continue beyond the first year and do not include screening for fathers (NICE, 2014). As we have seen above, there is evidence that children may be affected up to age 5 so screening beyond one year is important on this front as well as the possibility of mental health problems worsening long-term or being recurrent (Brennan et al., 2000; Goodman et al., 2011). Screening fathers is similarly important because of the evidence it may have an adverse impact on fathers, their families and the child (Sweeney & MacBeth, 2016).
Therefore, to improve identification of parents with mental health problems in the early years and reduce the negative impact on child outcomes, a global measure is needed that can be used with mothers and fathers across the early years from 0 to age 5. This literature review aimed to identify reliable, valid and acceptable measures of parental mental health to use with parents during pregnancy and up to 5 years postpartum to identify those with mental health problems. The results of this review will contribute to the wider literature on how best to screen and measure parental wellbeing and outcomes during pregnancy and up to 5 years postpartum.

**Method**

**Identification of Relevant Papers**

The literature searches and study selection were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). The principal method of identifying studies suitable for the review involved searching the following databases: MEDLINE, PsycINFO, PsychARTICLES, EMBASE, Web of Science, and Scopus. Searches were conducted up to and including February 2016. Search terms formed three categories (the 0-5 early years period, psychological wellbeing or distress and measures) and included, but were not limited to: *natal OR *partum OR parent* AND psychological OR mental OR problem OR disorder OR wellbeing AND measure* OR questionnaire OR scale. Combinations of all search terms and specific syntax used is available in the online supplementary material. In addition, the reference lists of identified articles and existing reviews were hand searched to identify additional relevant papers.

**Inclusion and Exclusion Criteria**

Studies were included if they were quantitative studies reporting empirical research that examined or included global measures of mental or emotional wellbeing in men and/or women who were expectant or new parents during pregnancy and up to 5 years postpartum. Global measures
were defined as those that examined overall mental health or emotional wellbeing. The 0-5 early years period was defined as from conception to 5 years postpartum. Studies were included if they used and reported psychometric or normative data on global measures of mental health or emotional wellbeing in men and/or women during this period (see Table 1). Studies were excluded if they: (1) Reported information for measures of specific affective disorders/symptoms such as anxiety or depression rather than global wellbeing or mental health; (2) If all or some of the sample measurements were outside the 0-5 early years period; (3) The paper was not available in English; (4) Non-empirical papers; (5) Dissertation or conference abstracts; (6) Qualitative studies; (7) Reviews of intervention studies where no information regarding the measure was reported before the intervention (i.e. at baseline).

Studies using measures originally developed in English that were translated for use in other countries have been included in the final tables (see online supplementary materials) to provide information on which languages different measures have been translated into and their validity in different countries.

**Study Selection and Data Extraction**

The literature searches provided a total of 24,327 citations, with a further 73 papers identified through hand searching the reference lists of key papers. After removing duplicates and screening through abstracts, titles and full texts according to the inclusion and exclusion criteria, 183 studies remained for inclusion in the review. Most of these studies used a measure of psychological wellbeing as part of a broader research study but reported some psychometric information. Only 19 of the studies included directly examined the psychometric validity of a measure. Detailed results of the literature search and screening are shown in Figure 1.

A data extraction sheet (see online supplementary materials) was designed and used to extract all relevant information from the full text of eligible studies. This included: (1) the measure used; (2) the language of the measure; (3) the country the study took place in; (4) the amount of participants
in the study; (5) the participant group; (6) participant demographics (including parental gender and age of children); (7) the design of the study; (8) norming data for the measure; (9); cut off scale of the measure; (10) how the measure was administered; and (11) any information provided on the reliability and validity of the measure.

Risk of Bias

Risk of bias was assessed in the 19 studies identified that explored psychometric properties of measures of parental mental health using the criteria set out by Brink and Louw (2012). This criteria was developed to help researchers critically appraise reliability and validity studies. It includes 13 items, that can be scored as ‘yes’ ‘no’ or ‘not applicable’, with items that were scored as ‘yes’ being assigned a score of 1, and items that were scored as ‘no’ being assigned a score of 0. These were then averaged (excluding the answers scored n/a) and multiplied by 100. Studies that scored between 0-33 were labelled as having a high risk of bias, those scoring between 34 – 66 were labelled as having a medium risk of bias, studies that scored between 67-100 were labelled as having a low risk of bias. The majority of studies (n = 16) had a low risk of bias, and 3 studies were scored as having a medium risk of bias. Details of this process are given in the online supplementary material.

Results

Study Characteristics

Details of all 183 studies included in the review are given in online supplementary materials. Nineteen of these studies were validation studies which directly examined the psychometric properties of measures of global parental mental health and wellbeing. All the other studies included in the review (n = 164) did not directly examine psychometric properties of a measure of parental mental health and wellbeing but used one such measure as part of a research study and reported relevant information such as reliability or means which can be used as norms.
Studies were conducted in 46 different countries, the most common being the USA (n = 27), Norway (n = 18), Australia (n = 18), and the UK (n = 17). All measures evaluated were available in English but as a result of the number of countries that the research was conducted in, norms, reliability and validity from a range of translated versions can also be extracted.

Samples included mothers only, fathers only, and co-parenting pairs. Data were collected in pregnancy, after birth, in the early years, and across the full 0-5 time course. Sample sizes ranged from 14 to 106,935. The majority of studies reported information on the reliability or validity of measures during the perinatal period in mothers. Only a small number of studies were found that included fathers (n = 7 with fathers only; n = 29 with mothers and fathers). Few studies explored validity across the later period of early childhood from age one to five (n = 24).

**Measures Identified by the Review**

The studies in the review used a total of 24 different measures of general psychological health or wellbeing, details of these can be found in Table 2. These are categorised into measures of: (1) current mood; (2) negative symptoms; (3) positive wellbeing; (4) positive and negative symptoms; (5) somatic symptoms; and (6) other measures. Measures that include somatic items (i.e. physical symptoms) are listed separately because these may be confounded by normal symptoms of pregnancy and postpartum. Of the measures identified, only four had been directly evaluated for psychometric properties in parents of children from 0 to 5. These measures and the number of studies that reported their psychometric properties are:

1. General Health Questionnaire (Goldberg & Williams, 1988) 12 item version (GHQ-12, n = 11); 28 item version (GHQ-28, n = 7); 30 item version (GHQ-30, n = 4) and 36 item version (GHQ-36, n = 1).

2. Self-Reporting Questionnaire (SRQ-20, World Health Organisation, 1994, n = 6)
3. (Hopkins) Symptom Check List (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974) 25 item version (HSCL-25), (n = 1); and 5 item version (HSCL-5), (n = 5)

4. Kessler Psychological Distress Scale (Kessler et al., 2003) 6 item version (K6, n = 3); and 10 item version (K10, n = 2)\(^1\)

Non-validation studies showed that the Brief Symptom Inventory (BSI, Derogatis, 1993) was a widely used measure. The psychometric properties of these five measures are given in Table 3 and outlined in the next section.

A few studies created their own measure with no adequate psychometric evaluation. The absence of psychometric information does not necessarily mean they are invalid measures as many were created for specific populations or for use as very brief assessment tools. Information on these is given in online supplementary materials for researchers who might be interested in using and evaluating these.

Psychometric properties of maternal global mental health outcome measures

General Health Questionnaire

The most extensively validated measure was the GHQ but the results suggest it has poor validity in early years parenting samples. There were discrepancies reported in the structure of the GHQ, with 2 studies finding a 2-factor model to offer a superior fit to the data (Ip & Martin, 2006; Spiteri, Jomeen, & Martin, 2013) suggesting it may not be a unitary measure of psychological wellbeing. Furthermore, Aderibigbe and Gureje (1992) found that only 3 of the 4 subscales of the GHQ made a significant contribution to the discriminatory power of the GHQ. These findings suggest the structural validity of the GHQ is questionable. Despite this, the GHQ appears to have good discriminatory powers (Navarro et al., 2007; Tran et al., 2011), although the EPDS (one of the most cited measures of perinatal depression used in the literature and clinical practice) did

\(^1\) One study examined both the K10 and the K6.
outperform the GHQ in two studies (Navarro et al., 2007; Tran et al., 2012). With regards to reliability of the measure, internal consistency was found to be acceptable to high, however test-retest reliability was poor. For example, Spiteri found a low correlation ($r = .38$) across all four time points in which they tested the measure. Due to the poor validity and test-retest reliability of the GHQ in the early years, this measure cannot be recommended for use with parents during early years. This is backed up by the conclusions of Prady et al. (2013) who asserted that there is very little evidence for the GHQ’s use in primary care, due to the fact that women from different ethnic groups respond in very different ways to the items on the questionnaire making the results difficult to analyse.

**Kessler Psychological Distress Scale**

The Kessler Psychological Distress Scale was promising. Internal consistency was good to high ($\alpha = .73$ to .90). The K6 and K10 performed well at discriminating diagnosis of major depression (Tesfaye et al., 2010), panic disorder and social phobia (Spies et al., 2009). The K10 and K6 yielded high sensitivity and specificity at the optimal cut off point, again indicating good discriminant validity. However, test-retest reliability was not measured by any of the validity studies, so no conclusions can be made about its ability to discriminate psychological distress across time points. Despite this, a non-validation study reported that father’s postpartum psychological distress at one time point correlated with their psychological distress at time three ($r = .48$; Giallo et al., 2014), suggesting the measure may be able to be successfully used across time points. Furthermore, the K10 is currently being used within the UK NHS children and young people’s tier one psychological services with parents, and is due to be rolled out in tier 2 children and adolescent mental health services. It will therefore be widely used in NHS settings in the UK. However, only one study validated this measure and there was no information available for UK samples so it is difficult to make firm recommendations on this basis.

**Self Report Questionnaire**
The SRQ was also promising with high discriminant validity with AUCs ranging from 0.74 to 0.826. Furthermore, this measure had high sensitivity and specificity when using the optimal cut off point. The measure had acceptable criterion validity, with a medium correlation between the SRQ and the Comprehensive Psychological Rating Scale ($r = 0.5$; Hanlon et al., 2008). Additionally, one study also found a medium correlation with the EPDS ($r = 0.53$; Abou-Saleh, 1997). No studies measured test-retest reliability so no conclusions can be made about its ability to discriminate psychological distress across time points. However, again only 3 studies validated this measure so it is difficult to make firm recommendations on this basis.

**The Hopkins Symptom Checklist**

The (Hopkins) Symptom Checklist (HSCL) was only validated by one study. This showed that the HSCL had poorer AUCs than the SRQ and K10, however the specificity and sensitivity of the measure was higher, as was the internal consistency. Non-validation studies showed acceptable to good internal consistency ($\alpha = 0.71$ to 0.87). However, no other studies provided validation information and no studies assessed test-retest reliability. With the lack of data on this measure, it is difficult to provide firm conclusions on the usefulness of this measure in the perinatal period.

**The Brief Symptom Inventory**

The Brief Symptom Inventory was the fourth most frequently used measure after the GHQ, HSCL and SRQ, with eighteen studies using this measure in early years parenting populations. However, no studies have validated the BSI during this time. Internal consistency was acceptable to high ($\alpha = 0.57$ to 0.99) and the BSI had the highest test-retest reliability of all the measures discussed ($r = .68$ to .91) (see Table 4 for comparisons of the measures). Only one study provided validity data and found that women’s scores on the BSI predicted scores on the EPDS and Spielberg Trait Anxiety Inventory showing it is correlated with other measures of the same construct (Cak et al., 2015). The BSI is copyrighted so needs to be purchased for use.

**Psychometric properties of paternal global mental health outcome measures**
The number of studies that recruited just only fathers was low \( (n = 7) \). One of these was a validation study (Tran et al., 2012), and 6 were non-validation studies. Studies were carried out in Asia \( (n = 1) \), Europe \( (n = 1) \), and Australasia \( (n = 5) \). More studies recruited both mothers and fathers \( (n = 29) \) and all of these were non-validation studies. The studies that recruited only fathers used the GHQ \( (n = 4) \), MHI-5 \( (n = 2) \), K6/K10 \( (n = 1) \), and HSCL \( (n = 3) \).

**General Health Questionnaire**

Tran et al. (2012) validated the GHQ in Vietnamese men whose partners were pregnant, or had recently given birth. Internal consistency was good for the GHQ (Cronbach’s a = 0.70). Discriminant validity was also good, with Tran et al. (2012) finding that it outperformed the EPDS and the Zung SAS in terms of discriminant validity (GHQ: AUC = 79.2; EPDS: AUC = 76.7; Zung SAS: AUC = 77.5). When the GHQ was used with fathers it performed better than it did with mothers (Tran et al., 2011). Using a recommended cut off of 0/1 its sensitivity was found to be 75.6% and its specificity was 74.7%. Condon, Boyce and Corkindale (2004) reported that there were no significant differences in GHQ scores from pregnancy to 12 months postpartum suggesting strong test re-test reliability. However, the same pattern was found for the MHI-5 which could suggest test-retest reliability is inflated by the stability of fathers’ mood rather than the reliability of the measure. The other two studies that used this measure did not report psychometric properties (Boyce, Condon, Barton, & Corkindale, 2007; Roberts, Bushnell, Collings & Purdie, 2006).

**Kessler Psychological Distress Scale.**

Only one study reported psychometric information for the K6. Giallo, Cooklin, Wade, D’Esposito, and Nicholson (2014) reported that fathers’ scores on the K6 at baseline predicted scores at repeated follow up points; suggesting this measure produces consistent results over time.

**Hopkins Symptom Checklist**
For the HSCL, Kvalevaag et al. (2014) reported that fathers’ scores on the SCL-5 were strongly correlated with their scores on the SCL-25 \((r = 0.92)\); suggesting the shortened version is as valid as the longer version.

**Mental Health Inventory 5 (Mental Health scale of the SF-36)**

As reported above, Condon, Boyce and Corkindale (2004) found that there were no significant differences in MHI-5 scores from pregnancy to 12 months postpartum suggesting good test re-test reliability or strong stability in fathers’ mood across time.

**Discussion**

This review aimed to identify reliable, valid and acceptable measures of parental mental health to use with parents during pregnancy and up to 5 years postpartum. The review found a wide range of measures that are available for measuring global parental mental health and wellbeing. However, very little research directly validated these measures for use with parents in the 0-5 early years. Four measures (General Health Questionnaire, Kessler Psychological Distress Scale, Self Report Questionnaire, Hopkins Symptom Checklist) were validated for use during this time but overall there was fairly limited psychometric data. The majority of studies explored the validity of measures during the perinatal period (0-1 year) in mothers, and only a small number of studies were found that included fathers or explored validity across the entire period of early childhood (0-5 years).

It is important to ensure that there are valid and reliable measures available to enable effective screening of parental mental health and wellbeing in the early years. There is evidence that untreated parental mental health issues can adversely impact the infant and the family (Barry et al., 2015; Kinsella & Monk, 2009; Talge et al., 2007) therefore the ability to screen for these difficulties is important in health care practice. Valid screening tools will enable parental mental health problems to be identified early through universal services, such as maternity and primary care services (Hogg, 2013).
A number of characteristics are important in a screening tool. Firstly, validity and effectiveness are critical if the screening tool is to accurately identify parents who require help. Other characteristics are also important in clinical practice, such as brevity, cost and availability in different languages. In clinical settings it is important that screening measures are brief and easy to complete due to time and funding constraints for both the individual and the healthcare system (Papanicolas & Smith, 2013; Squires, 2011). Similarly, lack of resources may prohibit the use of screening tools that need to be paid for. The availability of questionnaires in different languages to ensure that the measure is suitable for everyone.

The questionnaires identified in this review have advantages and disadvantages. The GHQ is available in 19 languages and has shortened versions (e.g. GHQ-12), making it appealing in the health care context. However due to its poor validity and reliability in this population it cannot be recommended for use in early years parents. The BSI has been widely used (with acceptable internal consistency) but has not been validated for use in this population, is very long (53 items), and has to be purchased. The BSI also includes questions about somatic symptoms, which may be confounded by normal physical symptoms of pregnancy and postpartum. The lack of validity information, length and cost therefore mean it also cannot be recommended for use – particularly in publically funded health services. The HSCL had good internal consistency and discriminant validity with mothers and has been used with fathers. However, it has many versions ranging from 5 to 90 items and many of these need to be purchased. Although it is available in 10 languages for perinatal populations, the multiple versions mean there is limited validation and languages for some of the versions. This, coupled with the length and cost therefore mean it is also unlikely to be suited to use in clinical settings.

The remaining two measures had promising psychometric properties. The SRQ had good discriminant validity and acceptable criterion validity with mothers. It is fairly brief (20 items), free to use and available in five languages. The K10/6 had good internal consistency and good
discriminant validity with mothers and has also been used with fathers. It is brief (6 items), free to use and available in four languages. At the current time these measures therefore seem most appropriate for use with parents of children aged 0-5. However, both these measures have not been widely validated so more research is needed to evaluate them in this population in different cultural contexts.

It is worth also considering newer measures that are available that have not yet been used in early years parenting populations but which may be promising in the future. For example, the CORE-10 is a broad, 10-item measure that includes symptoms (anxiety, depression, insomnia, re-experiencing) and functioning (coping, support) with emerging evidence of good psychometric properties with pregnant women (Coates, Ayers, de Visser, unpublished). This review therefore highlights the need for further validation of the promising measures, or alternatively examination of newer tools to identify psychological distress in the early years.

Limitations

There are a number of limitations to the current review. The first is the sparsity of studies that directly validated measures in the early years. This makes it difficult to draw conclusions or make firm recommendations about the ideal measure to screen for emotional wellbeing in this population. Second, some of the measures that were most validated were developed many years ago. The amount of time since development means older measures are more likely to have been validated in different populations so may be over-represented in reviews such as this one. For example, the GHQ (developed in 1988) has 5 validation studies and 23 non-validation studies reporting psychometric information. In contrast, the Kessler (developed in 2003) has 1 validation study and 4 non-validation studies. As mentioned, there may also be other new measures that are promising but have not yet been evaluated. The amount of psychometric information available is therefore not necessarily a measure of quality and is influenced by time since the measure was developed. This needs to be
considered when drawing conclusions, particularly in relation to newer measures that have not been widely validated.

There are also methodological issues that need considering. A key issue is the potential confounding of measures of maternal psychological wellbeing by normal symptoms of pregnancy and postpartum. The length of time between measures of test-retest reliability also varied between studies and it is possible that the length of time influenced estimates of reliability. In addition, two of the studies were considered to have a medium risk of bias in how the validation was carried out, therefore this should be taken into account before drawing any conclusions.

Finally, although a wide range of screening tools were identified for screening general psychological distress only four of these have been validated in the early years period. More research needs to be done to validate these measures in this population.

**Conclusions**

The Kessler Psychological Distress Scale and Self Report Questionnaire were both identified as promising global measures of parental mental health from 0 to 5 years in terms of psychometric properties and clinical utility. However, this review also shows that more research is needed to validate available measures of general psychological distress in early years parenting populations. It is possible that other measures of psychological distress, such as the CORE-10, which are not included in this review may also be promising. Therefore, future research should also examine the validity of scales not included here to see if they are valid for use in the early years with both mothers and fathers. Identifying valid and reliable measures will enable effective screening for perinatal distress and potentially prevent the long term negative impact of parental mental health problems.
References


cognitive development and behavior: a review and critical analysis of the literature. *Arch

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating

Detecting perinatal common mental disorders in Ethiopia: Validation of the self-reporting
doi:http://dx.doi.org/10.1016/j.jad.2007.10.023

NSPCC.

correlates, and persistence of maternal depression. *J Womens Health (Larchmt), 16*(5), 678-
691. doi:10.1089/jwh.2006.0185

Hoagwood, K. E. (2007). Barriers to the identification and management of psychosocial

Education Limited.

Questionnaire (GHQ-12) in Chinese women during pregnancy and in the postnatal period.
*Psychol Health Med, 11*(1), 60-69. doi:10.1080/1354850050155750


doi:10.1097/GRF.0b013e3181b52df1


discontinuation during pregnancy. *Can J Psychiatry, 54*(6), 399-408.
doi:10.1177/070674370905400607

maternal mood in pregnancy on childhood psychopathology. *Dev Psychopathol, 26*(2), 393-
403. doi: 10.1017/s0954579414000029

agenda for policy, information and research. *European Observatory on Health Systems and
Policies Series* Retrieved from
http://www.euro.who.int/__data/assets/pdf_file/0009/244836/Health-System-Performance-
Comparison.pdf


Post and Antenatal Depression Association. (2012, 18/04/2017). The cost of perinatal depression in
Australia FINAL REPORT. Retrieved from

The psychometric properties of the subscales of the GHQ-28 in a multi-ethnic maternal
doi:10.1186/1471-244x-13-55

period and child development: a prospective population study. *Lancet, 365*(9478), 2201-
2205. doi:10.1016/s0140-6736(05)66778-5


Table 1. Psychometric measures of validity and reliability and their definitions

<table>
<thead>
<tr>
<th>Psychometric measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Reliability measure. Measures the extent to which the items within a scale are consistent. This is usually calculated by correlating the score of one half of the measure with the other half or using Cronbach’s alpha $\alpha$ (which is the average of all possible split-half tests) (Howitt &amp; Cramer, 2008)</td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>Reliability measure. This calculates a measure consistency over time. The larger the correlation, the better the consistency.</td>
</tr>
<tr>
<td>Discriminant validity</td>
<td>Refers to the degree to which the scale measures what it intends to. The two main ways of measuring this is through the sensitivity (%) of a measure to pick up known ‘cases’ and the specificity (%) of a measure to excluded known ‘non-cases’. A receiver operating characteristic curve (ROC) analysis can also been performed. The area under this curve (AUC) provides an indication of a scales ability to discriminate between those with and without a diagnosis (Hanley &amp; McNeil, 1982). It has been suggested that an AUC of above 0.5 suggests the measure is performing above chance (Gill,</td>
</tr>
</tbody>
</table>
Butterworth, Rodgers, & Mackinnon, 2007; Spies et al., 2009).

Criterion validity

Refers to the extent to which a measure is related to an outcome that it is intended to be related to (i.e. an anxiety measure should correlate strongly with a diagnosis of GAD).
<table>
<thead>
<tr>
<th>Measure</th>
<th>Abbreviation</th>
<th>No. of Items</th>
<th>What the scale measures</th>
<th>Time frame of response</th>
<th>No. used measure</th>
<th>No. Psychometric</th>
<th>Countries measure used in</th>
<th>Language of measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures of current mood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Affect Adjective Checklist ©</td>
<td>MAACL</td>
<td>132</td>
<td>State and trait positive and negative affect</td>
<td>Not reported</td>
<td>1</td>
<td>1</td>
<td>South America</td>
<td>Spanish</td>
</tr>
<tr>
<td>Positive and Negative Affect Scale</td>
<td>PANAS</td>
<td>20</td>
<td>Positive affect, negative affect</td>
<td>Right now OR past week</td>
<td>3</td>
<td>2</td>
<td>Australia</td>
<td>English</td>
</tr>
<tr>
<td>Profile of Mood States ©</td>
<td>POMS</td>
<td>65</td>
<td>Anger, confusion, depression, fatigue, tension, vigour</td>
<td>Right now OR past week</td>
<td>5</td>
<td>2</td>
<td>Australia</td>
<td>English</td>
</tr>
<tr>
<td>University of Wales Institute of Science &amp; Technology Mood Adjective Checklist</td>
<td>UMACL</td>
<td>24</td>
<td>UMACL: Energetic arousal, tense arousal, hedonic tone, UWIST</td>
<td>Right now</td>
<td>1</td>
<td>1</td>
<td>Poland</td>
<td>Polish</td>
</tr>
<tr>
<td></td>
<td>UWIST</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
OUTCOME MEASURES OF PARENTAL WELLBEING

includes anger

<table>
<thead>
<tr>
<th>Measures of negative symptoms</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Wellbeing Scale</td>
<td>18</td>
<td>Depression, anxiety, irritability (outward and inward directed)</td>
<td>Last few days</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Crown-Crisp Experiential Index</td>
<td>CCEI</td>
<td>Depression, anxiety, phobia, obsessional, somatic anxiety, hysterical personality</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>General Health Questionnaire ©</td>
<td>GHQ-6</td>
<td>GHQ12: General psychopathology;</td>
<td>Last 7 days</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>Questionnaire ©</td>
<td>GHQ-12</td>
<td>GHQ28: Depression, anxiety/insomnia, somatic, social dysfunction; GHQ30: Excludes items related to physical illness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHQ-28</td>
<td>GHQ-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Angola, Australia, Arabic, Chilean, Belgium, Brazil, Chinese, Danish, Chile, China, Dutch, English, Denmark, Finland, French, Iranian, France, Hong Kong, Italian, Japanese, India, Iran, Italy, Maltese, Mirpuri, Japan, Malta, Nepal, Nepalese, Nigerian, Netherlands, New Norwegian, Polish,
<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Scale</th>
<th>Items</th>
<th>Description</th>
<th>Last Measured</th>
<th>Last Measured</th>
<th>Last Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopkins Symptom Check List</td>
<td>HSCL</td>
<td>58</td>
<td>SCL58: Depression, anxiety, somatisation,</td>
<td>Last week</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Or SCL</td>
<td>25</td>
<td>anxiety, somatisation, obsessive-compulsive,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interpersonal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>obsessive-compulsive, interpersonal sensitivity;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SCL25: Depression, anxiety.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>SCL15: subscale of the; SCL25 which measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>depression; SCL5: five items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kessler Psychological Distress</td>
<td>K6</td>
<td>6</td>
<td>General distress as an indication of serious</td>
<td>30 days</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Scale</td>
<td>K10</td>
<td>10</td>
<td>mental illness</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Countries where the measures are used:
- Zealand, Norway, Spanish, Urdu, Nigeria, Poland, Saudi Arabia, Spain, UK, USA, Vietnam,
- Australia, Bosnia & Herzegovina, Canada, English, Dutch, Egypt, France, French, German,
- Germany, Italy, Israel, Italian, Norwegian, Netherlands, Norway, Spanish, Swiss,
- Tanzania, USA,
- Australia, Japan, New Zealand, South Africa, Japanese, Zulu,
### OUTCOME MEASURES OF PARENTAL WELLBEING

<table>
<thead>
<tr>
<th>Measure</th>
<th>Scale</th>
<th>Language(s)</th>
<th>Measure Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marcé Clinical Checklist</td>
<td>53</td>
<td>English</td>
<td>Clinician-administered scale to identify psychological disorders</td>
</tr>
<tr>
<td>Mental Health Inventory</td>
<td>MHI-5</td>
<td>Australian, USA</td>
<td>MHI-5: General distress Last month 5 2</td>
</tr>
<tr>
<td></td>
<td>MHI-18*</td>
<td>English, Spanish</td>
<td>MHI-18: Depression, anxiety, behavioural control, positive affect</td>
</tr>
<tr>
<td>Self-Reporting Questionnaire</td>
<td>SRQ-20</td>
<td>Bangladesh, Brazil, Ethiopia, Ghana, Pakistan, UK, United Arab Emirates</td>
<td>General distress. Last 30 days 21 6</td>
</tr>
<tr>
<td>WHO Well-being Index</td>
<td>WHO-5</td>
<td>English</td>
<td>Positive wellbeing, physical energy, interest in life Two weeks 1 0 USA</td>
</tr>
<tr>
<td>Warwick-Edinburgh</td>
<td>WEMW</td>
<td>USA</td>
<td>SWEMWBS (7 items) Two weeks 1 1</td>
</tr>
</tbody>
</table>
### Measures of positive and negative symptoms

<table>
<thead>
<tr>
<th>Measure</th>
<th>Code</th>
<th>Items</th>
<th>Description</th>
<th>Timeframe</th>
<th>Country</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological General Wellbeing Inventory</td>
<td>PGWBI</td>
<td>22</td>
<td>Positive wellbeing, depression, mood, self-control, general health and vitality</td>
<td>Past month</td>
<td>Sweden, USA</td>
<td>English, Swedish</td>
</tr>
<tr>
<td>Wellbeing Questionnaire</td>
<td>W-BQ</td>
<td>12</td>
<td>Positive wellbeing, negative wellbeing, energy</td>
<td>1</td>
<td>UK</td>
<td>English</td>
</tr>
</tbody>
</table>

### Measures that include somatic symptoms

<table>
<thead>
<tr>
<th>Measure</th>
<th>Code</th>
<th>Items</th>
<th>Timeframe</th>
<th>Country</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief Symptom Inventory</td>
<td>BSI</td>
<td>Depression, anxiety, somatisation, hostility</td>
<td>Last 7 days</td>
<td>Finland, Netherlands, Portugal, Turkey</td>
<td>Dutch, Finnish, English, Portuguese</td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Description</td>
<td>Language</td>
<td>Country</td>
<td>Frequency</td>
<td>Language</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Outcomes of Parental Wellbeing</strong></td>
<td>Obsessive-compulsive, interpersonal sensitivity, phobia, paranoia, psychoticism</td>
<td>USA</td>
<td>Turkish</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kellner Symptom Questionnaire (KSQ)</strong></td>
<td>Depression, anxiety, anger-hostility, somatic symptoms</td>
<td>USA</td>
<td>English</td>
<td>Past week</td>
<td>USA</td>
</tr>
<tr>
<td><strong>Primary Care Mental Disorders Screening Questionnaire (PRIME-MD)</strong></td>
<td>Depression, anxiety, somatic symptoms, substance use, eating/weight change</td>
<td>Sweden, USA</td>
<td>NR</td>
<td>Two weeks</td>
<td>USA</td>
</tr>
<tr>
<td><strong>Somatic &amp; Psychological Health Report (SPHERE)</strong></td>
<td>SPHERE-12: Psychological health; physical symptoms and fatigue</td>
<td>New Zealand</td>
<td>English</td>
<td>Past few weeks</td>
<td>USA</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Stress, support from</td>
<td>Switzerland</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile</td>
<td>Short Form Quality of Life Scale ©</td>
<td>SF-12</td>
<td>12</td>
<td>Mental health-related quality of life (MCS), physical health-related quality of life (PCS)</td>
<td>16</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
<td>----</td>
<td>------------------------------------------------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>Chinese Health Questionnaire</td>
<td>CHQ</td>
<td>12</td>
<td></td>
<td>Non-UK measure</td>
<td>3</td>
</tr>
<tr>
<td>Shona Symptom Questionnaire</td>
<td>SSQ</td>
<td>2</td>
<td></td>
<td>Non-UK measure</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3. Summary of psychometric properties and norms of widely used and evaluated measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>No. of studies</th>
<th>No. of validation studies</th>
<th>AUC</th>
<th>Sensitivity (r)</th>
<th>Specificity (r)</th>
<th>Alpha</th>
<th>Test-retest (r)</th>
<th>No. of studies in late perinatal period</th>
<th>No. of studies with fathers</th>
<th>Population based means</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHQ</td>
<td>23</td>
<td>5</td>
<td>0.72-</td>
<td>74.7-</td>
<td>56.5-</td>
<td>.64-</td>
<td>.24-</td>
<td>6</td>
<td>4</td>
<td>UK: T1 mean 12.09 (SD 5.73); T2 mean 11.16 (SD 5.46) (Van Bussell et al. 2006); India: Pregnant women 3rd trimester mean 3.30 (SD = .15) (Bhat et al., 2015); Norway: Mothers: T1 mean 22.0; T2 mean 17.2; T3 mean 16.7; Fathers: T1 mean 16.4; T2 mean 17.9; T3 mean 15.9. Percentage defined to have psychological distress, mothers: T1: 37%, T2: 21%, T3: 19%, fathers: T1: 13%, T2: 11%, T3: 11% (Skari</td>
</tr>
</tbody>
</table>
## OUTCOME MEASURES OF PARENTAL WELLBEING

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>T1 Mean</th>
<th>T2 Mean</th>
<th>T1 SD</th>
<th>T2 SD</th>
<th>T1 25th</th>
<th>T1 75th</th>
<th>T2 25th</th>
<th>T2 75th</th>
<th>et al. 2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kessler</strong></td>
<td>4</td>
<td>0.86-50-54-82.7</td>
<td>0.73- .48</td>
<td>2</td>
<td>1</td>
<td>0.91</td>
<td>84.6</td>
<td>.90</td>
<td></td>
<td><strong>Australia:</strong> Mothers T1 mean 3.20 (SD=3.01); T2 mean 2.95 (SD = 3.10) (Giallo et al., 2014); <strong>New Zealand:</strong> First time parents mean 13.4 (SD=5). Subsequent parents mean 13.2 (SD=3.8) (McKenzie &amp; Carter, 2013)</td>
</tr>
<tr>
<td><strong>SRQ</strong></td>
<td>6</td>
<td>0.74-59.2-80-</td>
<td>.78- NR</td>
<td>1</td>
<td>0</td>
<td>0.826</td>
<td>83</td>
<td>85.4</td>
<td>.85</td>
<td><strong>Ethiopia:</strong> Median score (25th centile, 75th centile): in pregnancy 2 (0, 4); two months postpartum 0 (0, 2); prevalence of CMD: 12% in pregnancy, 4.6% in postpartum period (Senturk et al., 2012)</td>
</tr>
<tr>
<td><strong>HSCL</strong></td>
<td>6</td>
<td>0.495-88 89</td>
<td>.80- NR</td>
<td>3</td>
<td>3</td>
<td>0.797</td>
<td></td>
<td>.93</td>
<td></td>
<td><strong>Norway:</strong> Mothers mean 1.42 (Kvalevaag et al. 2014)</td>
</tr>
<tr>
<td><strong>BSI</strong></td>
<td>10</td>
<td>NR NR NR</td>
<td>.57 -.68-</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td>.99</td>
<td>.91</td>
<td><strong>Netherlands:</strong> Mean 0.26 during pregnancy (N = 4848)</td>
</tr>
</tbody>
</table>

**NOTES:** * non-validation studies only. NR = not reported
OUTCOME MEASURES OF PARENTAL WELLBEING