

**This item is the archived peer-reviewed author-version of:**

The general health questionnaire as a measure of emotional wellbeing in pregnant women

**Reference:**

Fontein-Kuipers Yvonne, Jomeen Julie, Dilles Tinne, van Rompaey Bart.- The general health questionnaire as a measure of emotional wellbeing in pregnant women  
Journal of Mental Health, Training & Practice - ISSN 1755-6228 - 14:6(2019), p. 447-456  
Full text (Publisher's DOI): <https://doi.org/10.1108/JMHTEP-05-2019-0028>  
To cite this reference: <https://hdl.handle.net/10067/1625070151162165141>

# The General Health Questionnaire as a measure of emotional wellbeing in pregnant women.

Yvonne Kuipers, Julie Jomeen, Tinne Dilles, Bart Van Rompaey

The Journal of Mental Health Training, Education and Practice

Publication date: 3 September 2019

## Abstract

### Purpose

To measure reliability, validity and accuracy of the 12-item General Health Questionnaire (GHQ-12) as a measure of emotional wellbeing in pregnant women; utility and threshold in particular.

### Design/ methodology/ approach

We measured self-reported emotional wellbeing responses of 164 low-risk pregnant Dutch women with the GHQ-12 and a dichotomous case-finding item ('Gold standard'). We established internal consistency of the 12 GHQ-items (Cronbach's coefficient  $\alpha$ ); construct validity: factor analysis using oblimin rotation; convergent validity (Pearson's correlation) and discriminatory ability (Area Under the Receiver Operating Characteristics Curve and Index of Union); and external validity of the dichotomous criterion standard against the GHQ-12 responses (sensitivity, specificity, likelihood ratios, predictive values), applying a cut-off value of  $\geq 12$  and  $\geq 17$  respectively.

### Findings

A  $\alpha$  coefficient of .85 showed construct reliability of the 12-items GHQ. The GHQ-12 items in the pattern matrix showed a three-dimensional factorial model: factor 1. *Anxiety & Depression*; factor 2. *Coping*; and factor 3. *Significance/ Effect on life*, with a total variance of 59%. The GHQ-12 showed good accuracy (0.84;  $p = <.001$ ) and external validity ( $r = .57$ ;  $p = <.001$ ) when the cut-off value was set at the  $\geq 17$  value. Using a cut-off value of  $\geq 17$  demonstrated higher sensitivity (72.32% versus 41.07%) but lower specificity (32.69% versus 55.77%) compared to the commonly used cut-off value of  $\geq 12$ .

### Research limitations/ implications

Findings generally support the reliability, validity and accuracy of the Dutch version of the GHQ-12. Further evaluation of the measure, at more than one timepoint during pregnancy, is recommended.

### Practical implications

The findings offer promise of the GHQ-12 utility as a potential measure of antenatal emotional wellbeing and women's emotional responses and coping mechanisms with reduced antenatal emotional wellbeing.

**Social implications**

Adapting the GHQ-12 cut-off value enables effective identification of reduced emotional wellbeing to provide adequate care and allows potential reduction of anxiety among healthy pregnant women who are incorrectly screened as positive.

**Originality/ value**

A novel aspect is adapting the threshold of the GHQ-12 to  $\geq 17$  in antenatal care.

**Keywords**

GHQ-12; Emotional wellbeing; Pregnancy; Psychometric; Factor-analysis; Validation

**Citation**

[Kuipers, Y.](#), [Jomeen, J.](#), [Dilles, T.](#) and [Van Rompaey, B.](#) (2019), "The general health questionnaire as a measure of emotional wellbeing in pregnant women", *The Journal of Mental Health Training, Education and Practice*, <https://doi.org/10.1108/JMHTEP-05-2019-0028>

## **Introduction**

Although pregnancy is primarily thought of as a physiological process – its life-altering aspects together with the social, physical and emotional changes and challenges associated with being pregnant and becoming a mother – mark it as a dynamic psychological process (Jomeen and Martin, 2008). It is therefore no surprise that women's emotional wellbeing can be affected during pregnancy (Martin and Jomeen, 2003; Magill-Cuerden, 2006). Maternal emotional wellbeing is described as a good or satisfactory sense of welfare, health and happiness (Ridner, 2004). Antenatal emotional wellbeing has been described as a complex construct of psychological mood states such as depression, anxiety, stress, fear and worry – being either pregnant or non-pregnant originated - often co-occurring and being inter-related during the antenatal period (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). Antenatal emotional health manifests itself in a varying continuum ranging from a positive to a negative sense of wellbeing, *i.e.* from the presence, reducing to the absence of emotional wellbeing. A positive sense of antenatal emotional wellbeing enables a woman to function in society, to meet the demands of everyday life and to have the ability to respond effectively to change or adverse or unexpected events (Ryan and Deci, 2001; Ridner, 2004; Cruwys-Ververda, 2005; Lucey, 2007; Beddington *et al.*, 2008). A negative sense of antenatal emotional wellbeing can be a predictor for preterm labour and low infant birth weight and can result in adverse short- and long-term postpartum mental health effects for both mother and child, including postpartum depression and post-traumatic stress (Mulder, Robles de Medina, Huizink, Van den Bergh, Buitelaar and Visser, 2002; Heron, O'Connor, Evans, Golding and Glover, 2004; van Son, Verkerk, van der Hart, Komproe and Pop, 2005; Leigh and Milgrom, 2008; Loomans *et al.*, 2013; Biaggi, Conroy, Pawlby and Pariante, 2016), behaviour and emotional problems in children, and problematic family relationships (Mennes, Stiers, Lagae and Van den Bergh, 2006; Robinson *et al.*, 2008; Goodyer and Cooper, 2011; Biaggi, Conroy, Pawlby and Pariante, 2016).

Emotional wellbeing among Dutch pregnant women has been studied – in women with healthy pregnancies – showing reduced emotional wellbeing in every 4 to 5 pregnant women with varying prevalences for depression, anxiety and pregnancy-related anxiety (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). Emotional wellbeing among women with physiological *i.e.* low-risk pregnancies, has been measured with psychometric instruments with a rather pathologic character focusing on symptomatology of mental wellbeing such as depression and anxiety (Nast, Bolten, Meinschmidt and Helhammer, 2013), for example the Edinburgh Postpartum Depression Scale and the State-Trait Anxiety Inventory (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). Within the perspective that pregnancy is primarily thought of as a physiological process, a healthy event and intrinsically normal, it would be of benefit to assess antenatal emotional wellbeing within a similar orientated normalizing context, utilising a corresponding instrument.

### ***General Health Questionnaire (GHQ-12)***

The 12-item General Health Questionnaire (GHQ-12) is a well-known self-administered instrument designed for measuring and identifying public mental wellbeing, allowing to predict emotional wellbeing in a general population in the primary care setting (Martin and Jomeen, 2003; Böhnke and Croudace, 2016). The GHQ-12 has been translated into Dutch (Koeter and Ormel,

1991) and validated in Dutch primary care populations (Hoeymans, Garssen, Westert and Verhaak, 2004; Verhaak, 2005). The GHQ-12 has been used in western pregnant (Martin and Jomeen, 2003; Yp and Martin, 2006) and postnatal populations (Yp and Martin, 2006; Aguado, Campbell, Ascaso, Navarro, Garcia-Esteve and Luciano, 2012; Spiteri, Jomeen and Martin, 2013), showing to measure various constructs of maternal antenatal emotional wellbeing such as depression, anxiety, stress and functioning (Ip and Martin, 2006; Doi and Minowa, 2003). The discussion about the optimal cut-off point for the GHQ-12 for pregnant women to determine caseness in pregnancy, suggests a higher threshold level for women with uncomplicated pregnancies - allowing the manifestation of antenatal physical symptoms or pregnancy discomfort in addition to reduced emotional wellbeing to inflate the GHQ score (Martin and Jomeen, 2003) - albeit that adapting the cut-off value for a healthy pregnant population has never been further investigated.

The GHQ-12 has never been used to measure emotional wellbeing of Dutch pregnant women. Bearing in mind that the Dutch childbearing population has a specific culture and a more physiological approach towards pregnancy and childbirth, compared to women in other high-income countries, as well as that healthy pregnant women represent a general population, receiving primary care (de Vries *et al.*, 2013). It would therefore be of interest to learn how the GHQ-12 measures emotional wellbeing in this low-risk population. Additionally, these women seem the opportune target population to determine the optimal GHQ-12 threshold for use in healthy pregnancies – women belonging to a primary care population for which the GHQ-12 has been designed (Martin and Jomeen, 2003; Böhnke and Croudace, 2016). For the purpose of this study we wanted to examine the psychometric properties of the GHQ-12 in a healthy Dutch pregnant population to consider if this measure can determine maternal antenatal emotional wellbeing. In order to fulfil this purpose, we sought answers to the following questions:

1. Is the GHQ-12 reliable, valid and accurate to evaluate the level of antenatal emotional wellbeing in women in the Netherlands with a healthy pregnancy?
2. What is a suitable cut-off value of the GHQ-12 when compared to self-reported emotional wellbeing?

## **Methods**

We performed a diagnostic accuracy study among a Dutch pregnant population. The data was collected using self-completed questionnaires. We included the Dutch version of the GHQ-12 and an additional self-reported emotional wellbeing case-finding question in our questionnaire. The data were collected between 1 September and 30 November 2016. We aimed to include women in the Netherlands who received midwife-led primary care. Women receiving this type of care are women with healthy uncomplicated *i.e.* low-risk pregnancies that do not require obstetric care/ have obstetric complications. We included women with good comprehension of the Dutch language, 18 years of age or older and during any trimester of pregnancy. Women receiving secondary or tertiary care and/ or women with (acute) psychiatric conditions were excluded. We recruited pregnant women through the websites and Facebook accounts of midwifery practices in the south-west region of the Netherlands. The study was conducted in accordance with the Helsinki declaration (WMA, 2018). Because of the noninvasive character of the study, ethical approval was not required according to Dutch ethical research standards (reference WC2016–055; <https://english.cmo.nl/investigators/legal-framework-for-medical-scientific-research/your->

research-is-it-subject-to-the-wmo-or-not). The questionnaire included a privacy notice explaining confidentiality and data handling. Participation was voluntary and informed consent was obtained (via box ticking) before the questionnaire could be completed.

### **Measures**

The GHQ-12 consists of 12 items, describing mood states - each item assessing the level of emotional wellbeing over the past few weeks. We used the 4-point scale (less than usual, no more than usual, rather more than usual, or much more than usual), recognising to produce an acceptable distribution of scores per item (Pilar Sánchez-López and Dresch, 2008). The categories generate a total score ranging from 0 to 36, with higher scores indicating reduced emotional wellbeing (Goldberg and Blackwell, 1970). Total scores of  $\geq 12$  are commonly used as a cut-off value to identify positive testing (Lundin, Hallgren, Theobald, Hellgren and Torgé, 2016). We used a secondary dichotomous criterion standard/ measure to assess self-reported/ subjected sense of emotional wellbeing, formulated as a case-finding question: *Over the last couple of weeks, have you been feeling well, mood-wise; not bothered by being nervous, anxious, depressed, worrying and/ or not bothered by having little interest in doing things?* either to be answered positively (no) or negatively (yes). In this study a 'no' was considered as a positive test. The formulation of this criterion standard was based on the NICE (2014) antenatal mental health identifying questions – being regarded as an acceptable and recommended method of assessment of antenatal emotional wellbeing (NICE, 2014; Hanley, 2015). Our criterion standard represents the specific wording of the case-finding Whooley-items and the 2-item Generalized Anxiety Disorder Scale (GAD-2). The general formulation 'have you been feeling well' followed by more formal assessment (*i.e.* Whooley-items and GAD-2) shows a combined holistic and clinical approach, it enhances the external validity of the probe question as well as this form of questioning aligns with addressing emotional wellbeing in antenatal care (NICE, 2014; Hanley, 2015). A relationship between the heightened GHQ-12 scores and positive tests is expected (NICE, 2014; Fontein-Kuipers, 2015; Fontein-Kuipers and Jomeen, 2019). Furthermore, brief demographic and personal information was collected.

### **Data Analysis**

Internal consistency was assessed by calculating Cronbach's coefficient  $\alpha$  value, to verify if the items of the GHQ-12 measured the same concept, *i.e.* if the GHQ-12 measures a single underlying concept by using multiple items (Terwee *et al.*, 2007). A value of 0.7 was considered as the lower limit of internal consistency (Field, 2013). To assess construct validity, the oblimin rotation was performed, based on the assumption that factors that form the questionnaire have correlations between them (Costello and Osborne, 2005; Ip and Martin, 2006; Spiteri, Jomeen and Martin, 2013). Before conducting factor analysis, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test were conducted to evaluate whether the sample was large enough to perform a satisfactory factor analysis. A KMO of 0.8 was considered as adequate and Bartlett's test was used with a significance level  $p < 0.05$  (Field, 2013). Factor loading for each item was taken as the minimum of 0.40 (Costello and Osborne, 2005; Field, 2013). Finally, a scree plot analysis was performed to verify the number of factors that could be extracted. We used the Receiver Operating Characteristics (ROC) analysis to determine the accuracy of the GHQ-12, which are reported as Area Under the Curve (AUC). AUC with a value approaching 1.0 indicated a high sensitivity and specificity (Lalkhen and McCluskey, 2008). Subsequently, we ascertained the discriminatory ability *i.e.* the optimal cut-off value of the GHQ-12 with the Index of Union

(IU) method, utilizing the sensitivity and specificity 1- values nearest to the UAC value (Unal, 2017).

We identified the high scores of the GHQ-12 using  $\geq 12$  as a commonly used cut-off value to identify positive testing (Lundin *et al.*, 2016) and the cut-off value identified by the IU. The sets were recoded in either 'yes' when above cut-off level and 'no' when below cut-off level. We ascertained the rates of true and false positives and true and false negatives for the GHQ-12 in both sets, based on the respective cut-off values. External validity was assessed using 2×2 contingency tables of weighted prevalences. Agreement between the GHQ-12 and our dichotomous criterion standard were analysed in both sets, using standard diagnostic performance measures: sensitivity (the proportion of true positives correctly identified by the test), specificity (the proportion of true negatives correctly identified by the test), likelihood ratio (shows how much more likely a woman is to get a positive test if she has reduced emotional wellbeing, compared with a woman with good emotional wellbeing), positive predictive value (the proportion of women with positive test results who are correctly identified) and negative predictive value (the proportion of women with negative test results who are correctly identified). The positive response to the dichotomous criterion standard was regarded as the standard against which the test was compared; using the GHQ-12 total score as the criterion for possible caseness. Caseness is a probabilistic term—whereby, respondents would be more likely to experience reduced emotional wellbeing (Jackson, 2007). Statistical analyses were performed using SPSS 24.0 and the MedCalc Diagnostic test evaluation calculator (2018). The validity outcomes based on both cut-off values were compared.

### ***Sample size***

There are various ways to determine the sample size for factor analysis. We considered the four most often used rules of thumb: (i) including at least 100 participants (Gorsuch, 1983); (ii) applying a subjects-to-variable ratio of at least 5:1 (Bryant and Yarnold, 1995; Costello and Osborne, 2005), implying 5 times 12 items for the GHQ-12; (iii) including ten to 15 participants per variable (Field, 2013); (iv) accepting either communalities in the 0.5 range, requiring a sample between 100 and 200 participants, or accepting all communalities above .06 in a sample of less than 100 participants (MacCallum, Widaman, Zhang and Hong, 1999; Costello and Osborne, 2005; Field, 2013). *A priori* we aimed to include between a minimum of 60 and a maximum of 180 participants.

### **Results**

We received 164 completed questionnaires. The participants in our study were aged between 24 and 45 years (Mean 30.9;  $\pm 3.85$ ), were predominantly of Dutch origin (96%) and had fairly high levels of education (87%). The participants were predominantly employed (90%) and in a relationship (99%). The participants had a mean gestational age of 27 weeks ( $\pm 8.61$ ; range 9 to 41 weeks). The sample contained 38% nulliparous and 62% parous women (Table 1). The mean total GHQ-12 score was 13.28 ( $\pm 5.88$ ; range 3 to 29) (Table 1).

Table 1. Demographic and personal characteristics participants (n= 164)

	Mean (SD $\pm$ ) range	N (%)
<b>Age in years</b>	30.09 ( $\pm$ 3.85) 24-45	
<b>Gestational age in weeks</b>	27 ( $\pm$ 8.61) 9-41	
<b>Nulliparous*</b>		62 (37.8)
<b>Parous*</b>		102 (62.2)
<b>Partnered</b>		163 (99.4)
<b>Working (paid) job</b>		144 (90)
<b>Ethnicity</b>		
<b>Respondent born in the Netherlands</b>		157 (95.6)
<b>Respondent born in other Western country</b>		4 (2.5)
<b>Respondent born in non-Western country</b>		3 (1.9)
<b>Education</b>		
<b>Low level of education</b>		21 (13.1)
<b>Medium level of education</b>		40 (25)
<b>High level of education</b>		103 (61.9)
<b>Emotional wellbeing</b>		
<b>GHQ-12</b>	13.28 ( $\pm$ 5.88) 3-29	
<b>GHQ-12 cut-off <math>\geq</math> 12</b>		86 (55.5)
<b>GHQ-12 cut-off <math>\geq</math> 17</b>		48 (29.3)
<b>Over the last couple of weeks, have you been feeling well, mood-wise...?<sup>1</sup> Positive ('no')</b>		50 (31.1)

\*Nulliparous: women who never have given birth; Parous: women who have given birth once or more

<sup>1</sup> Over the last couple of weeks, have you been feeling well, mood-wise; *not* bothered by being nervous, anxious, depressed, worrying and/ or *not* bothered by having little interest in doing things?

### **Reliability analysis**

The total instrument was found to have an  $\alpha$  coefficient of .85, showing good internal consistency.

### **Construct validity**

First exploratory factor analysis was performed. The KMO was .842 and Bartlett's test of sphericity showed a significant result ( $X^2$  640.1,  $p < .001$ ); both tests indicating the appropriateness of performing factor-analysis in the presented sample, showing an adequate sample size. The principal component analysis with oblimin rotation was performed and a three-factor structure was loaded that jointly accounted for 59% of the variance. The scree plot showed inflexions that justified to retain components 1 to 3. The GHQ-12 items in the pattern matrix showed a three-dimensional factorial model including values of 0.66 or higher, suggesting factor 1 to represent *Anxiety & Depression* (item 2, 5, 6, 9, 10), factor 2 *Coping* (item 4, 8) and factor 3 representing *Implications/ Effect on life* (item 1, 3, 7, 11, 12). The first factor is a major factor, accounting for more than half of the total variance of the GHQ-12, whereas factors 2 and 3 are minor factors. The structure matrix showed item 5 (Have you recently felt constantly under strain? *i.e.* founding everything getting on top of you) to be a common theme between factor 1 and 3 and 10 between factor 1 and 2 (Table 2). The factors show negligible and low inter-correlation (1 and 2  $r = .22$ ; 1 and 3  $r = .46$ ; 2 and 3  $r = .17$ ).



Table 2. Factor structure of the GHQ-12 using oblimin rotation

<b>GHQ-12 items</b>	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>
% of variance per factor	39.4%	10.2%	9.4%
<b>2. Have you recently lost much sleep over worry?</b>	.72		
<b>5. Have you recently felt constantly under strain?</b>	.80		.41
<b>6. Have you recently felt you couldn't overcome your difficulties?</b>	.72		
<b>9. Have you recently been feeling unhappy or depressed?</b>	.74		
<b>10. Have you recently been losing confidence in yourself?</b>	.76	.55	
<b>4. Have you recently felt capable of making decisions about things?</b>		.81	
<b>8. Have you recently been able to face up to problems?</b>		.76	
<b>3. Have you recently felt that you were playing a useful part in things?</b>			.81
<b>7. Have you recently been able to enjoy your normal day-to-day activities?</b>			.75
<b>11. Have you recently been thinking of yourself as a worthless person?</b>			.66
<b>12. Have you recently been feeling reasonably happy, all things considered?</b>			.70

Factor 1: Anxiety & Depression

Factor 2: Coping

Factor 3: Implications/ Effect on life

The total GHQ-12 score and the criterion standard showed a significant correlation ( $r = .57$ ;  $p = <.001$ ). The ROC curve with an AUC of .84 (95% CI .79 to .90;  $p = <.001$ ) showed good accuracy. Using sensitivity and specificity 1- values nearest to the UAC value as coordinates of the accurate cut-off point of the GHQ-12, indicated the  $\geq 17$  value.

### ***External validity***

Using a cut-off value of  $\geq 12$  (conventional method), 55.5% of the participants showed caseness. The Index of Union (IU) method showed  $\geq 17$  as the optimal cut-off value. When applying a cut-off value of  $\geq 17$ , 29.3% of the participants showed caseness. Thirty-one percent of the participants answered positive on the constructed item, representing the participants who had not been feeling well, mood-wise over the last couple of weeks, *i.e.* participants with a subjected sense of reduced antenatal emotional wellbeing (see Table 1).

With a cut-off value of  $\geq 12$ , the GHQ-12 showed higher specificity than sensitivity. Compared to a cut-off value of  $\geq 17$ , the GHQ-12 had an increased sensitivity but specificity decreased. With a cut-off value of  $\geq 12$ , 66% would go undetected; with a cut-off value of  $\geq 17$ , 31% would go undetected (false negatives); 23% (cut-off  $\geq 12$ ) and 35% (cut-off  $\geq 17$ ) were incorrectly identified as test positive (false positive). The likelihood ratios indicated that when a cut-off value of  $\geq 12$  is used, a positive test is less likely to occur in women with reduced emotional wellbeing than when a cut-off value of  $\geq 17$  is applied. When a cut-off value of  $\geq 12$  is utilised, a negative test is more likely to occur in women with reduced emotional wellbeing than when a cut-off value of  $\geq 17$  is applied. The negative predictive value (NPV) at both cut-off values showed a rather inconclusive ability for negative results (Table 3).

Table 3. GHQ-12 as indicator for ‘not feeling well, mood-wise’

	<b>GHQ-12 cut-off <math>\geq 12</math></b>	<b>GHQ-12 cut-off <math>\geq 17</math></b>
<b>True positive (n)</b>	46	81
<b>False negative (n)</b>	66	31
<b>False positive (n)</b>	23	35
<b>True negative (n)</b>	29	17
<b>Sensitivity (95% CI)</b>	41.07% (31.86 to 50.76)	72.32% (63.07 to 80.36)
<b>Specificity (95% CI)</b>	55.77% (41.33 to 69.53)	32.69% (20.33 to 47.11)
<b>Positive Likelihood Ratio</b>	.93 [.64 to 1.35]	1.07 [.86 to 1.34]
<b>Negative Likelihood Ratio</b>	1.06 [.79 to 1.41]	.85 [.52 to 1.38]
<b>PPV (95% CI)</b>	66.67% (57.83 to 74.47)	69.83% (64.97 to 74.28)
<b>NPV (95% CI)</b>	30.53% (24.79 to 36.93)	35.42% (25.12 to 47.27)

PPV = Positive predictive value; NPV = Negative predictive value

## Discussion

The Dutch GHQ-12 as a measure of antenatal emotional wellbeing demonstrated good divergent validity as well as multi-dimensionality of the scale. The  $\alpha$  coefficient in our study was fairly similar to those in other studies with samples of pregnant women (Martin and Jomeen, 2003; Ip and Martin, 2006), indicating that the GHQ-12 is a reliable scale to measure emotional wellbeing in pregnancy. The reported mean scores of a pregnant population assessed using the GHQ-12 Likert method (Martin and Jomeen, 2003) were similar to those of our sample, enhancing the reliability of our findings. The prevalence of reduced emotional wellbeing with a cut-off value of  $\geq 17$  corresponded with earlier Dutch epidemiological studies that used other psychometric measures in similar pregnant populations (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). Adapting the cut-off value for low-risk pregnant women seems justified. The demographic characteristics in our study were analogous to women in similar studies performed in the Netherlands, allowing generalisability of the findings to comparable populations (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016).

Our principal component analysis showed a three-factor model. The ‘Anxiety & Depression’ and ‘Coping’ and ‘Implications/ Effect on life’ factors have also been identified in previous studies among pregnant and postpartum women (Ip and Martin, 2006; Spiteri, Jomeen and Martin, 2013). Hence, the similarities between this study and previous studies strengthen the validity of our findings. The ‘Depression & Anxiety’ factor accounted for more than half of the total variance of the GHQ-12, acknowledging the psychometric value of the measure in a general population of pregnant women (Martin and Jomeen, 2003; Böhnke and Croudace, 2016) but also strongly emphasising the GHQ-12’s multi-dimensional approach of emotional wellbeing. For research purposes, the GHQ-12 seems a good alternative to summing scores of different depression and anxiety measures into one construct (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). The factors ‘Coping’ and ‘Implications/ Effect on life’ enhanced the importance of discussing women’s responses to and uncertainties with reduced emotional wellbeing as well as their (self-) management abilities, including social and professional networks. Emotional wellbeing is integral to coping behaviour (Fontein-Kuipers, Ausems, Budé, van Limbeek, de Vries and Nieuwenhuijze,

2015). Ergo, the individual potential of pregnant women to their emotional wellbeing may be viewed within the context of their circumstances, their networks and coping abilities (Fontein-Kuipers, van Limbeek, Ausems, de Vries and Nieuwenhuijze, 2015; Fontein-Kuipers, van Limbeek, Ausems, de Vries and Nieuwenhuijze, 2017).

When predictive values were considered in addition to sensitivity and specificity levels, this emphasised the preference for applying the cut-off value of  $\geq 17$  for healthy Dutch pregnant women. The higher rate of true positives, showed a better discriminative power and a somewhat better predictive ability (van Stralen, Stel, Reitsma, Dekker, Zoccali and Jager, 2009; Lalkhen and McCluskey, 2008). As predictive values are greatly influenced by prevalence rates, our results cannot be transferred to other healthcare settings with different prevalence of reduced antenatal emotional wellbeing or to childbearing populations with dissimilar characteristics than those of our sample of women. When using the GHQ-12 as a screening tool for caseness, the fact that there is no (near) 100% sensitivity when the cut-off value of  $\geq 17$  is applied seems to be acceptable, considering that it is highly unlikely that most reference tests, and in particular purposively designed probe questions like our question, will have a sensitivity and specificity of 100% (van Stralen, Stel, Reitsma, Dekker, Zoccali and Jager, 2009). However, pregnant healthy women might benefit when adapting the cut-off value, as this is likely to reduce anxiety among healthy pregnant women who are incorrectly screened as positive. Health care practitioners would be better able to identify women with reduced antenatal emotional wellbeing that is not affected by physical discomfort during pregnancy (Martin and Jomeen, 2003) to coordinate adequate (follow-up) care. Our findings suggest a more relaxed cut-off value than commonly used cut-offs, supporting the importance of establishing specific cut-off values in a pregnant population. It would be of merit to test this cut-off value in a larger population of healthy pregnant women before generalising these findings to a wider population of low-risk women.

Although the GHQ-12 total score demonstrated a significant correlation with our probe question, which we regarded as our criterion standard, external validity might be strengthened by comparing the scores of the probe question against a structured clinical interview to confirm this finding. Nevertheless, we believe the correlation to be reliable because the formulation of our probe question was based on valid screening items. Further, we have to bear in mind that our data has been collected at one point during pregnancy which can be regarded as a limitation of our study as emotional wellbeing is known to have a fluctuating nature through the course of pregnancy (Biaggi, Conroy, Pawlby and Pariante, 2016; Fontein-Kuipers, Ausems, de Vries and Nieuwenhuijze, 2016). Cross-sectional comparison of emotional wellbeing at different times of pregnancy in addition to repeated measures of the GHQ-12, could provide more information about reliability, validity and accuracy and offer a future direction of further research.

## **Conclusion**

Findings from the current study generally support the reliability, validity and accuracy of the Dutch GHQ-12 and offer promise of its utility as a potential measure of antenatal emotional wellbeing and women's emotional responses and coping mechanisms with reduced emotional wellbeing. In the present study, with the Likert scoring method, the best trade-off between sensitivity and specificity is given by a threshold of  $\geq 17$ , potentially offering benefits for both pregnant women and maternity healthcare professionals. The findings from our study, when

considered in context of our work on this measure, suggests promising opportunities for the antenatal use of the GHQ-12 and are sufficiently robust to indicate the value of further investigation.

### **Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### **References**

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., Luciano, JV. (2012) Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment*, Vol. 19 No. 4, pp. 517-25.

Beddington, J., Cooper, C.L., Field, J, et al. (2008). The mental wealth of nations. *Nature*, Vol. 455 No. 23, pp. 1057–60.

Biaggi, A., Conroy, S, Pawlby, S., Pariante, M. (2016). Identifying the women at risk of antenatal anxiety and depression: A systematic review. *Journal of Affective Disorders*, Vol. 191, pp. 62-77. doi: [10.1016/j.jad.2015.11.014](https://doi.org/10.1016/j.jad.2015.11.014)

Böhnke, J.R., and Croudace, T. J. (2016). Calibrating well-being, quality of life and common mental disorder items: psychometric epidemiology in public mental health research. *The British Journal of Psychiatry*, Vol. 209 No. 2, pp. 162-8

Bryant, F.B., Yarnold, P.R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR. (Eds.). *Reading and Understanding Multivariate Statistics*. Washington: American Psychological Association, pp. 99-136.

Costello, A.B., Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, Vol. 10 No. 7, pp. 1-9

Cruwys-Ververda, C. (2005). *Jongleren met salutogenese* [Juggling with salutogenesis]. Westerbork: Academie Gradatim.

De Vries, R., Nieuwenhuijze, M., Buitendijk, S., the members of Midwifery Science Work Group. (2013). What does it take to have a strong and independent profession of midwifery? Lessons learned from the Netherlands. *Midwifery*, Vol 29 No. 10, pp. 1122-1128. <http://dx.doi.org/10.1016/j.midw.2013.07.007>

Doi, Y., Minowa, M. (2003). Factor structure of the 12-item General Health Questionnaire in the Japanese general adult population. *Psychiatry and Clinical Neuroscience*, Vol. 57 No. 4, pp. 379-83

Field, A. (2013). *Discovering statistics using SPSS (3<sup>rd</sup> Ed.)* London: SAGE Publications Ltd.

Fontein-Kuipers, Y. (2015) Reducing maternal distress in pregnancy: what is the best approach? *Current Opinion in Obstetrics & Gynecology*, Vol. 27 No. 2, pp. 128-32. DOI:10.1097/GCO.0000000000000155

Fontein-Kuipers, Y., Ausems, M., Budé, L., van Limbeek, E., de Vries, R., Nieuwenhuijze, M. (2015). Factors influencing maternal distress among Dutch women with a healthy pregnancy. *Woman and Birth*, Vol. 28 No. 3, pp. e36-e43

Fontein-Kuipers, Y., Ausems, M., de Vries, R., Nieuwenhuijze, M. (2016). The effect of WazzUp Mama?! An antenatal intervention to prevent or reduce maternal distress in pregnancy. *Archives of Women's Mental Health*. Vol. 19 No. 5, pp. 779-88. DOI: 10.1007/s00737-016-0614-8

Fontein-Kuipers, Y., van Limbeek, E., Ausems, M., de Vries, R., Nieuwenhuijze, M. (2017). Responding to maternal distress: from needs assessment to effective intervention. *International Journal of Public Health*, Vol. 62 No. 4, pp. 441-52.

Fontein-Kuipers, Y., Jomeen, J. (2019). Validity and accuracy of the Whooley questions to identify maternal distress in Dutch pregnant women". *The Journal of Mental Health Training, Education and Practice*. Vol. 14 No. 1, pp.33-45 <https://doi.org/10.1108/JMHTEP-06-2018-0034>

Goldberg, D.P., Blackwell, B. (1970). Psychiatric illness in general practice. A detailed study using a new method of case identification. *British Medical Journal*, Vol. 1 No. 5707, pp. 439-43.

Goldberg, D., Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Windsor: NFER- Nelson.

Goodyer, I. and Cooper, P. (2011), "Maternal postnatal depression and the development of depression in offspring up to 16 years of age". *Journal of American Academy Child Adolescent Psychiatry*, Vol. 50 No. 5, pp. 460-70.

Gorsuch, R.L. (1983). *Factor Analysis*. Hillsdale, NJ: Erlbaum.

Hanley, J. (2015). *Listening visits in perinatal mental health: A guide for health professionals and support workers*. London: Routledge.

Heron, J., O'Connor, T., Evans, J., Golding, J. and Glover, V. (2004), "The course of anxiety and depression through pregnancy and the postpartum in a community sample". *Journal of Affective Disorders*, Vol. 80 No. 1, pp. 65-73.

Hoeymans, N., Garsen, A.A., Westert, G.P., Verhaak, P.F.M. (2004). Measuring mental health of the Dutch population: a comparison of the GHQ-12 and the MHI-5. *Health Qual Life Outcomes*, Vol. 2 No. 23, pp. 1-6. doi: [10.1186/1477-7525-2-23](https://doi.org/10.1186/1477-7525-2-23)

Ip, W.Y., Martin, C. (2006). Psychometric properties of the 12-item General Health Questionnaire

(GHQ-12) in Chinese women during pregnancy and in the postnatal period. *Psychology, Health & Medicine*, Vol. 11 No. 1, pp. 60-9.

Jackson, C. (2007). The General Health Questionnaire. *Occupational Medicine*, Vol. 57 No. 1, p. 79.

Jomeen, J., Martin, C.R. (2008). The impact of choice of care on psychological health outcomes for women during pregnancy and the postnatal period. *Journal of Evaluation in Clinical Practice*, Vol. 14 No. 3, pp. 645–8.

Koeter, M.W.J., Ormel, J. (1991). *General Health Questionnaire, Nederlandse bewerking: Handleiding* [Dutch version: Manual] Lisse: Swets Test Services.

Lalkhen, A., McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, Vol. 8 No. 6, pp. 221-3.

Leigh, B. and Milgrom, J. (2008), “Risk factors for antenatal depression, postnatal depression and parenting stress”. *BMC Psychiatry*, Vol. 16 No. 8, p. 24, pp. 1-11.

Loomans, E., van Dijk, A., Vrijkotte, T., van Eijsden, M., Stronks, K, Gemke, R. and Van den Bergh, B. (2013), “Psychosocial stress during pregnancy is related to adverse outcomes: results from a large multi-ethnic community-based birth cohort”. *Eur Journal Public Health*, Vol. 23 No. 3, pp. 485-91.

Lucey, P. (2007). Social determinates of health. *Nursing Economics*, Vol. 25 No. 2, pp. 103–9.

Lundin, A., Hallgren, M., Theobald, H., Hellgren, C., Torgé, M. (2016). Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health*, Vol. 136, pp. 66-74

Magill-Cuerden, J. (2006). Birthwrite nurturing and supporting mothers. A hidden skill in midwifery. *British Journal of Midwifery*, Vol. 14 No. 6, p. 374.

Martin, C.R., Jomeen, J. (2003). Is the 12-item General Health Questionnaire (GHQ-12) confounded by scoring method during pregnancy and following birth? *Journal of Reproductive and Infant Psychology*, Vol. 21 No. 4, pp. 267-78.

MacCallum, R.C., Widaman, K.F., Zhang, S., Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, Vol. 4 No. 1, pp. 84-99.

MedCalc easy-to-use statistical software. 2018. Ostend, Belgium: MedCalc Software. [https://www.medcalc.org/calc/diagnostic\\_test.php](https://www.medcalc.org/calc/diagnostic_test.php) Accessed 19 December 2018

Mennes, M., Stiers, P., Lagae, L. and Van den Bergh, B. (2006), “Long-term cognitive sequelae of antenatal maternal anxiety: involvement of the orbitofrontal cortex”. *Neurosci Biobehav Reviews*, Vol. 30 No. 8, pp. 078-86.

Mulder, E., Robles de Medina, P., Huizink, A., Van den Bergh, B., Buitelaar, J. and Visser, G. (2002), "Prenatal maternal stress: effects on pregnancy and the (unborn) child". *Early Human Development*, Vol. 70 No. 1-2, pp. 3-14.

Nast, I., Bolten, M., Meinlschmidt, G., Helhammer, D.H. (2013). How to measure prenatal stress? A systematic review of psychometric instruments to assess psychological stress during pregnancy. *Paediatric and Perinatal Epidemiology*, Vol. 27 No. 4, pp. 313-22.

NICE (2014), "Antenatal and postnatal mental health: clinical management and service guidance: updated edition (update 2017)", NICE Clinical Guidelines No. 192, National Institute for Health and Care Excellence, London.

Petkovska, M.S., Bojadziev, M.I., Stefanovska, V.V. (2015). Reliability, Validity and Factor Structure of the 12-Item General Health Questionnaire among General Population. *Open Access Macedonian Journal of Medical Sciences*, Vol 3 No. 3, pp. 478-83. [10.3889/oamjms.2015.075](https://doi.org/10.3889/oamjms.2015.075)

Pilar Sánchez-López del, M., Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema*, Vol. 20 No. 4, pp. 839-43.

Ridner, S. (2004). Psychological distress: a concept analysis. *Journal of Advanced Nursing*, Vol. 45 No. 5, pp. 536-45.

Robinson, M., Oddy, W., Li, J., Kendall, G., de Klerk, N., Silburn, S., Zubrick, S., Newnham, J., Stanley, F. and Mattes, E. (2008), "Pre- and postnatal influences on preschool mental health: a large-scale cohort study". *The Journal of Child Psychology and Psychiatry*, Vol. 49 No. 10, pp. 1118-28.

Ryan, R., Deci, E. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, Vol. 52, No. 1, pp. 41-66.

Spiteri, M.C., Jomeen, J., Martin, C.R. (2013). Reimagining the General Health Questionnaire as a measure of emotional wellbeing: A study of postpartum women in Malta. *Women and Birth*, Vol. 26, No. 4, pp. e105-e11.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M. and de Vet, H.C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*, Vol. 60 No. 1, pp. 34-42.

Unal, I. (2017). *Defining an optimal cut-point value in ROC analysis: An alternative approach*. Computational and Mathematical Methods in Medicine. <https://doi.org/10.1155/2017/3762651>

Van Son, M., Verkerk, G., van der Hart, O., Komproe, I. and Pop, V. (2005), "Prenatal depression, mode of delivery and perinatal dissociation as predictors of postpartum posttraumatic stress: an empirical study". *Clinical Psychology & Psychotherapy*, Vol. 12 No. 4, pp. 297-312.

Van Stralen, K., Stel, V., Reitsma, J., Dekker, F., Zoccali, C., Jager, K. (2009). Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. *Kidney International*, Vol. 75, No. 12, pp. 1257–63.

Verhaak, P.F.M. (2005). Mental health in the Dutch population and in general practice: 1987-2001. *British Journal of General Practice*, Vol. 55 No. 519, pp. 770-75.

WMA. (2018) Declaration of Helsinki – Ethical principles for medical research involving human subjects. Ferney-Voltaire, France: World Medical Association.