Reimagining the General Health Questionnaire as a measure of emotional wellbeing: A study of postpartum women in Malta

M.C. Spiteri, Faculty of Health Sciences, University of Malta
J. Jomeen, Faculty of Health and Social Care, University of Hull
C.R. Martin, Faculty of Society and Health, Buckinghamshire New University; West London Mental Health NHS Trust

Abstract

Background

Postpartum health has been subject to a focus on psychological morbidity, despite positive associations between postpartum recovery and maternal emotional wellbeing. There are currently many validated tools to measure wellbeing and related concepts, including non-psychiatric morbidity. The General Health Questionnaire, 12 items (GHQ-12) is one such instrument, widely used and validated in several languages. Its use in postpartum settings has been documented with disagreement about the instrument's utility in this population, particularly in relation to scoring method and threshold. The GHQ-12 has never been translated into Maltese. This study explored the psychometric properties of the GHQ-12 in a Maltese postpartum population to consider if the use of a different scoring method (visual analogue scale) in the GHQ-12 can determine postpartum wellbeing.

Methods

One hundred and twenty-four postpartum women recruited from one hospital in Malta completed the translated and adapted GHQ-12 as a wellbeing measure (GHQ-12(WB)) at four postpartum time points. The psychometric properties of the GHQ-12(WB) were explored using confirmatory factor analysis, discriminant and divergent validity and reliability analysis.

Results

The GHQ-12(WB) demonstrated good divergent and known-groups validity and internal consistency. No models offered a good fit to the data. The overall consistent best-fit to the data was an eight item, two factor model (GHQ-8). Model fit improved across all models in terms of CFI at 13 weeks.

Conclusion

Findings generally support the reliability and validity of the Maltese version of the GHQ-12(WB). Model fit changes over time reflect the dynamic nature of postpartum recovery. Further evaluation of the GHQ-8(WB) is recommended.

Keywords

Postpartum; GHQ-12; Wellbeing; Psychometric; General Health

1. Introduction

The optimum state of health in postpartum women is articulated in the World Health Organisation's (WHO) definition of general health as the physical, mental and social wellbeing of a person.[1] The term wellbeing is further defined by several other authors as 'being equivalent to the state of set conditions which fulfil or enable a person to work or fulfil her realistic chosen and biological potentials'[2] and '…creatively build strong and positive relationships with others and contribute the their chosen community'.[3]

Wellbeing is currently a prevailing theme in the health care arena and suggests a concentration of capabilities and positive emotions rather than negative emotions, illness or disability.[4 and 5] Wellbeing aims to encompass multiple domains which can be experienced spatially, temporally, inter-personally, bodily, in mood and personal identity.[6] It is undoubtedly a complex construct and there is recognition of the overlap with other related concepts such as quality of life and general health.[4] In a perinatal context wellbeing is a well-used idiom but as in other areas it is used interchangeably with other terms, perhaps reflecting common characteristics between these constructs.

Postpartum health has been well considered and acknowledged to combine many factors. As well as a return to physical health, becoming a parent is a major life transition and a time of crucial psychological adjustment, as both lifestyles and relationships alter.[7] The individual potential of postpartum women to health may be placed within the context of their family, support network[8] and the nursing of their newborn babies, which is further affected by the woman's views, beliefs and attitudes, culture, race and religion. The weeks and months after giving birth have been identified as a time of considerable pleasure for the mother, but also as a time of considerable stress.[9 and 10] The recognition of this multitude of factors has led to a focus in the international literature[11] and policy on postpartum psychological morbidity,[12] despite an acknowledgement of a positive association between postpartum recovery and maternal emotional wellbeing.[13] Inevitable physical recovery correlates with improving psychological profiles[14] and during that recovery, new mothers undergo the process of attaining their maternal identity that consists of developing an attachment with their baby, competence in mothering behaviours and experiencing pleasure when interacting with their baby.[15] This process of personal growth in becoming a mother is described as a process of appreciation, discovery, learning and acceptance of the women's new role which results in a positive and worthwhile experience.[16] This fits with the notion of wellbeing as encapsulating the presence of positive capabilities and emotions.[4] Pregnancy and the postpartum period are inherently dynamic for women and will inevitably represent both a physical and psychological challenge.[14] However, it is and should be viewed a normative rather than pathologic process. The wellbeing perspective offers a potent framework to encompass and exemplify the holistic nature of new motherhood, one which can involve a multi-faceted and evolving continuum ranging from a positive to a negative sense of wellbeing and where the domains of the experience are inherently implicated in each other but one which is intrinsically normal.

This raises important questions about how current assessment often seeks to identify 'deviation from the norm' in particular poor psychological health, rather than focus on identifying positive adjustment, which raises a further issue about how postpartum 'health' can be assessed within a positive normative frame rather than from negative focus. This could facilitate a normalising context and help practitioners to promote a positively oriented model of postpartum recovery.

2. Assessing wellbeing

There are currently many validated tools developed to measure wellbeing – taking different conceptualisations of the term as their starting point – and many to measure concepts that relate to wellbeing. Existing instruments focus on quality of life, happiness, satisfaction or aim to assess depression, anxiety, and stress. One such instrument designed to identify depression/general non-psychiatric morbidity is the GHQ-12.[17] This is a shortened version of the original 60 item screening tool, which has been extensively translated and validated.[18, 19, 20 and 21] Its use in postpartum settings has been documented; however this same literature is contradictory in its assessment of the instrument's utility in this population. Whilst some studies suggest the GHQ-12 is a useful instrument,[22 and 23] questions have been raised about scoring methods; appropriate threshold and factorial stability.[24, 25 and 26] These authors have urged differing degrees of caution in its use to identify psychological distress in postpartum women.

Despite the apparent limitations of the GHQ-12 in postpartum women, in the absence of a measure of wellbeing designed for or validated in a postpartum population and the potential congruence of the concepts of general health and wellbeing – acknowledged by other authors[4 and 5] – the GHQ-12 was considered as a potentially useful instrument to assess wellbeing in a Maltese group of postpartum women. The instrument has not previously been translated into Maltese and hence there is currently no evidence of the key psychometric properties of the instrument in this group. This was also considered in light of the critique and concerns about the GHQ-12 in terms of factorial structure and stability, as useful to detect postpartum psychological distress as well as a consideration of the value of assessing postpartum wellbeing rather than morbidity. The aim of this study, therefore, was to consider the psychometric properties of the Maltese GHQ and explore its potential use of as a measure of general wellbeing rather than psychological distress, specifically to address the following questions:

1. What is the underlying factor structure of the Maltese GHQ?
2. Is the underlying factor structure stable over time?
3. Is the Maltese version of the GHQ-12 internally consistent?
4. Does the Maltese version of the GHQ-12 demonstrate good test-retest reliability?
5. Can the use of a different scoring method in the GHQ-12 determine postpartum wellbeing?

3. Methods

3.1. Measure: the GHQ-12

3.1.1. Translation

The GHQ-12 had not previously been used in a Maltese population so was subject to a translation process. In the first phase the GHQ-12 was translated into Maltese by an experienced and professional translator of English. During the second phase, a preliminary validation of the translation was conducted by back translation, a standard method used for translating a research instrument from English to non-English languages.[27] The translation of the GHQ-12 back into English was performed by a Maltese teacher fluent in English. In the third phase, the English translation derived from the Maltese translation was compared with

the original English version of the GHQ-12 and a consensus was reached on the final Maltese version. This last phase was completed by a group of three bilingual Maltese nurse educators. Once finalised, the GHQ-12 was administered to a sample of five postpartum mothers to test its readability.

3.1.2. Scoring

To overcome the criticisms of scoring method and thresholds and to facilitate the use of the measure as one of wellbeing rather than a case detector for psychological morbidity, an alternative approach to scoring was employed. One approach which facilitates a scaling of the intensity of the mother's feelings associated with their emotional health is the use of the visual analogue scales (VAS). This facilitates an element of self-evaluation which is particularly pertinent to the dynamic postpartum context. The advantage of VAS, therefore, over a categorical scale is that they do not limit subjects to a number of possible responses but offer a continuum, which offers both greater sensitivity and enables potentially finer distinctions to be made.[28] In line with the conceptualisation of wellbeing, these scales were on a positive trend and were scored by measuring the distance from the low end (zero point) to the specified place on the mother's mark. Permission to adapt the scale into VAS was given by the publishers of the GHQ-12, nferNelson (personal communication 5th June 2003).

3.1.3. Reliability testing

The instrument was then subject to reliability testing with four different groups of nursing students. The GHQ-12 in English (Likert scales and VAS), Maltese (VAS) and back translation (VAS) versions were administered separately to groups. The test–retest was carried out with a three week interval. Reliability testing of the GHQ-12 was performed on the English version (Likert scale and VAS), the Maltese version and back translation. The stability of the scales was performed by test–retest reliability for each question of the four versions, using a paired samples test (paired t-test), all of which gave a p-value of <0.05. These results demonstrate test–retest agreement and reliability of the Maltese version of the GHQ-12. The internal consistency of the GHQ-12 in the four versions was also examined.

3.2. Piloting of the GHQ-12 (wellbeing)

A pre-testing of Maltese GHQ-12 (WB) was also carried out with a convenience sample of 10 postpartum mothers with a mean age of 24. Feedback revealed some minor modifications in re-phrasing. Re-piloting of the tool then followed on a larger convenience sample of another 39 postpartum mothers who satisfied the inclusion criteria for the main study.

3.3. Design

The study was approved by the Local Research Ethics Committee and institutional permissions were gained. The study used a within-subjects design with participants recruited to the study within the first 36 h of admission to the postpartum ward. Data was collected at four time points; within 48 h postpartum, 10 days postpartum, 6 weeks postpartum and 13 weeks postpartum.

3.4. Participants

One hundred and forty four postpartum women were recruited from one hospital in Malta. Inclusion criteria specified mothers with Maltese citizenship, delivering their 1st, 2nd or 3rd babies, who delivered spontaneously, at term (>38 week gestation), with a vaginal delivery. The exclusion criteria consisted of mothers with caesarean section, medical complications, psychological disorders, with pre-term or sick babies. One hundred and twenty four (86%) women completed the measure at all four time points and thus represent the data reported in this study. The age of mothers ranged from 20 and 40 years (mean age 27.8 years). All participants were volunteers and gave informed consent to be involved in the study. Only 12 women (8.3%) of the sample were unmarried. 96 (66.7%) were primigravid, 44 (30.6%) were having their second baby and 4(2.8%) were having their third baby. 72.9% had a 2 day hospital stay with the remaining staying between 3 and 5 days.

3.5. Procedure

All participants were recruited from the postpartum ward of the hospital in Malta, where the majority of births occur, across a 6 month time period. Women who fulfilled the inclusion criteria were approached by the author (CS) to explain the aim of the study. Consented women were then asked to complete the GHQ-12(WB) and a demographic sheet. Subsequent questionnaires were mailed to the woman prior to the identified time points and then collected in person by the author (CS) a week later.

3.6. Statistical analysis

Statistical analysis was conducted using PASW version 18,[29 and 30] Analysis of Moment Structures (AMOS) version 18[31] and Mplus version 3.[32]

3.7. Confirmatory factor analysis

Evaluation of a psychometric measure can be conducted, in part, using confirmatory factor analysis (CFA). The GHQ-12 has been established to be, despite being conceived and scored as a uni-dimensional measure, a scale that is intrinsically multi-dimensional.[26] Eleven CFA models were evaluated, including ten multi-dimensional models and one uni-dimensional at each of the observation points, thus forty-four CFA models of the GHQ-12 were evaluated in total. For the purposes of brevity, the models evaluated will not be described in detail here but are usefully summarised in Ip and Martin[26] and Hankins.[33] The focus of the CFA within the context of this study, is not only to determine the best model fit, but also to evaluate consistency of model fit across discrete time periods, given that childbirth and perinatal period represents a period of dynamic physiological, psychological and social change. Observations over multiple time periods postpartum allow the scrutiny of the possible impact of such interactional factors on the tool. A key focus of model evaluation through CFA is the determination of model fit and parsimony. A maximum-likelihoods (ML) approach to model estimation was used, this being consistent with the assumption of multivariate normality.[34, 35 and 36]

The robustness of parametric tests against violations of the fundamental parametric assumptions[37] have resulted in the contemporary use of ordinal or ordered categorical data – the common reality of questionnaire data – with these statistical techniques.[38, 39, 40 and 41] However, data exhibiting significant deviation from the normal distribution assumption can lead to an erroneous outcome based on assumed parametric acceptable data distributional characteristics and consequently, an incorrect and potentially misleading interpretation of

statistical findings.[37, 42, 43 and 44] Therefore, each of the GHQ-12 items distributional characteristics were examined in detail to determine deviation from assumed normality, which could have a deleterious impact on the CFA and SEM. Skew and kurtosis characteristics of each item were examined and those exhibiting any significant deviation from normality were rejected from the GHQ-12 item pool prior to further statistical analysis based on normality assumptions. The criteria for item rejection based on univariate skew and kurtosis characteristics, was based on absolute skew values equal to, or greater than 3 and absolute kurtosis values of equal to, or greater than 10, based on the non-normality cut-off recommendations of Kline.[45]

Multiple goodness of fit tests[46] were used to evaluate the models, these being the comparative fit index (CFI),[47] and the root mean squared error of approximation (RMSEA). A CFI greater than 0.90 indicates an acceptable fit to the data[46, 48, 49 and 50] while a CFI equal to or greater than 0.95 indicates a good fit to the data.[49] A RMSEA with values of less than 0.08 indicate an acceptable fit to the data[51] while values of less than 0.05 indicate a good fit to the data.[52] A statistically significant χ^2 indicates a significant proportion of variance within the data is unexplained by the model,[46] though a significant χ^2 statistic is often observed as an artefact of trivial variations in data,[49] hence model evaluation is almost universally determined by model fits statistics such as CFI and RMSEA.[34] Finally, given the focus on model comparison, the expected cross-validation index (ECVI)[51] is used to compare baseline models. The ECVI was originally developed within the context of assessing the possibility that a model cross-validates with other samples of similar size and from a similar population. [34 and 51] The ECVI statistic can be computed for each model and models can consequently be compared on the basis of the absolute value of this statistic. The model with the lowest ECVI value is considered to have the largest replication potential.[34]

3.8. Divergent validity

Divergent validity was determined by correlating GHQ-12 scale scores at each observation point with the participant's age. It was predicted that there would be no significant relationship between GHQ-12 scores and participant's age.

3.9. Known-groups discriminant validity

Known-groups discriminant validity was evaluated by testing for differences in GHQ-12 scores in response to perineal birth trauma. A perinatal birth trauma score was calculated from each participant and a medium-split conducted to categorise equal numbers of participants into either a low or high perineal birth trauma groups of 72 participants. It was hypothesised that those in the high perineal birth trauma group would have significantly lower GHQ-12 scores compared to those in the low perineal trauma group at the 48 h and 10 ten days observation points.

3.10. Internal consistency

An internal consistency analysis of the GHQ-12 at each observation point was conducted to ensure that the measures satisfied the criteria for clinical and research purposes using the Cronbach coefficient alpha statistical procedure. A Cronbach's alpha reliability statistic of 0.70 is considered as the minimum acceptable criterion of instrument internal reliability.[35 and 36]

3.11. Test–retest reliability

Test retest reliability was evaluated using Pearson's correlation coefficients between 48 h and the 13 weeks observation point. A test–retest reliability coefficient of 0.80 has been suggested as the threshold for acceptability for a psychometric test.[35]

4. Results

The mean GHQ scores of participants at each observation point were 760.28 (165.12), 833.10 (200.64), 906.23 (185.71) and 934.88 (204.05) at 48 h, 10 days, 6 weeks and 13 weeks observation points respectively. One-way within-subject analysis of variance revealed a statistically significant difference in GHQ-12 scores as a function of observation point, $F_{(3,369)} = 32.15$, $p < 0.001$, with GHQ-12 scores increasing over time.

4.1. Multivariate normality

The distribution of the GHQ-12 items at the 48 h observation point revealed generally no evidence of significant skew or kurtosis (skew range 0.48–1.21, kurtosis range 0.04–1.64). At the 10 days observation point a similar pattern emerges (skew range 0.39–2.77, kurtosis range 0.08–2.09) with the exception of GHQ-12 item 12 which revealed evidence of kurtosis (11.83). Similarly, the 6 weeks observation point a generally distributionally satisfactory pattern emerges (skew range 0.66–3.14, kurtosis range 0.50–3.56), again, with the exception of GHQ-12 item 12 which revealed evidence of kurtosis (12.31). Finally, at 13 weeks, a consistent pattern emerges with the second and third observations (skew range of 0.99–3.59, kurtosis range 0.01–5.82) with the exception of GHQ-12 item 12 which revealed evidence of kurtosis (21.53) (see Table 1).

Table 1.

Summary of the Cronbach's alpha for the GHQ-12 in the four versions.

Variables	Parameter	Original Likert	Original VAS	Maltese VAS	Back translation VAS
GHQ-12	Cronbachs Alpha	0.8678	0.8598	0.9228	0.9278

4.2. Evaluation of models

The results of the CFA for each model at each observation point are summarised in Table 2, Table 3, Table 4 and Table 5. During the first observation point the two-factor model of Graetz[53] offered the best fit to the data. However, during each of the following observations, the overall best-fit to the data in terms of consistency was that of Kalliath et al.,[54] however, none of the models evaluated offered a good fit to the data based on both CFI and RMSEA. It was also noteworthy that model fit improved across all models in terms of CFI at the 13 weeks observation point.

Table 2.

Factor structure of the GHQ-12 at 48 h postpartum determined by testing the fit of models derived from factor analysis and SEM models related to negativity/positivity item bias.

Model (90%)	χ^2 (df)	p	CFI	RMSEA	(90%)[a] ECVI
Doi and Minowa (2 factor)	115.59 (43)	0.001	0.86	0.11 (0.08–0.13)	1.13 (0.93–1.38)
Doi and Minowa (3 factor)	123.81 (51)	0.001	0.87	0.10 (0.08–0.12)	1.24 (1.04–1.50)
Graetz (2 factor)	74.80 (34)	0.001	**0.90**	**0.09** (0.06–0.12)	**0.82** (0.67–1.02)
Graetz (3 factor)	152.82 (51)	0.001	0.83	0.12 (0.10–0.14)	1.45 (1.21–1.73)
Kilic et al.	171.53 (53)	0.001	0.80	0.12 (0.10–0.15)	1.55 (1.30–1.85)
Martin	133.85 (41)	0.001	0.81	0.13 (0.10–0.15)	1.29 (1.07–1.56)
Politi et al.	147.30 (52)	0.001	0.84	0.11 (0.09–0.13)	1.39 (1.17–1.67)
Unitary	198.06 (54)	0.001	0.75	0.14 (0.12–0.16)	1.72 (1.45–2.05)
Valenced (+/−)	165.94 (53)	0.001	0.81	0.12 (0.10–0.14)	1.51 (1.26–1.81)
Worsley and Gribbin	172.52 (51)	0.001	0.79	0.13 (0.11–0.15)	1.58 (1.33–1.89)
Kalliath et al.	90.98 (19)	0.001	0.81	0.16 (0.13–0.20)	0.87 (0.69–1.11)
Hankins positive errors	101.48 (39)	0.001	0.89	0.11 (0.08–0.13)	1.25 (1.07–1.49)
Hankins negative errors	130.72 (39)	0.001	0.84	0.13 (0.10–0.15)	1.46 (1.23–1.73)

Note: Bold indicates best model fit as a function of model fit index criterion and scoring method. *Abbreviations*: comparative fit index (CFI), root mean squared error of approximation (RMSEA), expected cross-validation index (ECVI).

a   Denotes 90% confidence interval (CI).

Table 3.

Factor structure of the GHQ-12 at 10 days postpartum determined by testing the fit of models derived from factor analysis and SEM models related to negativity/positivity item bias.

Model (90%)	χ^2 (df)	p	CFI	RMSEA	(90%)[a] ECVI
Doi and Minowa (2 factor)	173.50 (43)	0.001	0.86	0.15 (0.12–0.17)	1.69 (1.43–2.00)
Doi and Minowa (3 factor)	205.09 (51)	0.001	0.84	0.14 (0.12–0.17)	1.98 (1.70–2.31)
Graetz (2 factor)	135.60 (34)	0.001	0.85	0.14 (0.12–0.17)	1.38 (1.16–1.66)
Graetz (3 factor)	197.04 (51)	0.001	0.85	0.14 (0.12–0.16)	1.92 (1.65–2.25)
Kilic et al.	252.69 (53)	0.001	0.79	0.16 (0.14–0.18)	2.28 (1.96–2.66)
Martin	167.97 (41)	0.001	0.83	0.15 (0.12–0.17)	1.68 (1.42–1.98)
Politi et al.	212.92 (52)	0.001	0.83	0.15 (0.13–0.17)	2.02 (1.73–2.36)
Unitary	277.09 (54)	0.001	0.77	0.17 (0.15–0.19)	2.44 (2.10–2.83)
Valenced (+/−)	222.75 (53)	0.001	0.82	0.15 (0.13–0.17)	2.07 (1.77–2.42)
Worsley and Gribbin	213.71 (51)	0.001	0.83	0.15 (0.13–0.17)	2.04 (1.75–2.38)
Kalliath et al.	76.20 (19)	0.001	**0.90**	**0.14** (0.11–0.18)	**0.88** (0.72–1.10)
Hankins positive errors	171.16 (39)	0.001	0.86	0.15 (0.13–0.18)	1.91 (1.65–2.22)
Hankins negative errors	154.56 (39)	0.001	0.88	0.15 (0.12–0.17)	1.81 (1.57–2.11)

Note: Bold indicates best model fit as a function of model fit index criterion and scoring method. *Abbreviations*: Comparative fit index (CFI), Root mean squared error of approximation (RMSEA), Expected cross-validation index (ECVI).

a   Denotes 90% confidence interval (CI).

Table 4.

Factor structure of the GHQ-12 at 6 weeks postpartum determined by testing the fit of models derived from factor analysis and SEM models related to negativity/positivity item bias.

Model (90%)	χ^2 (df)	p	CFI	RMSEA	(90%)[a] ECVI
Doi and Minowa (2 factor)	182.40 (43)	0.001	0.85	0.15 (0.13–0.17)	1.75 (1.48–2.07)
Doi and Minowa (3 factor)	211.30 (51)	0.001	0.83	0.15 (0.13–0.17)	2.02 (1.74–2.36)
Graetz (2 factor)	131.15 (34)	0.001	0.86	0.14 (0.12–0.17)	1.35 (1.13–1.62)
Graetz (3 factor)	181.11 (51)	0.001	0.86	0.13 (0.11–0.15)	1.81 (1.55–2.13)
Kilic et al.	221.80 (53)	0.001	0.82	0.15 (0.13–0.17)	2.07 (1.77–2.42)
Martin	150.85 (41)	0.001	0.86	0.14 (0.11–0.16)	1.56 (1.32–1.85)
Politi et al.	214.38 (52)	0.001	0.86	0.15 (0.13–0.17)	2.03 (1.74–2.37)
Unitary	222.77 (54)	0.001	0.82	0.15 (0.13–0.17)	2.06 (1.77–2.41)
Valenced (+/−)	214.37 (53)	0.001	0.83	0.15 (0.13–0.17)	2.02 (1.73–2.36)
Worsley and Gribbin	184.55 (51)	0.001	0.86	0.13 (0.11–0.16)	1.84 (1.57–2.15)
Kalliath et al.	82.14 (19)	0.001	**0.89**	0.15 (0.12–0.19)	**0.92** (0.75–1.15)
Hankins positive errors	148.15 (39)	0.001	**0.89**	**0.14** (0.12–0.16)	1.75 (1.51–2.04)
Hankins negative errors	169.43 (39)	0.001	0.86	0.15 (0.13–0.18)	1.90 (1.64–2.21)

Note: Bold indicates best model fit as a function of model fit index criterion and scoring method. *Abbreviations*: comparative fit index (CFI), root mean squared error of approximation (RMSEA), expected cross-validation index (ECVI).

a   Denotes 90% confidence interval (CI).

Table 5.

Factor structure of the GHQ-12 at 13 weeks postpartum determined by testing the fit of models derived from factor analysis and SEM models related to negativity/positivity item bias.

Model (90%)	χ^2 (df)	p	CFI	RMSEA	(90%)[a] ECVI
Doi and Minowa (2 factor)	138.59 (43)	0.001	0.92	0.12 (0.10–0.15)	1.44 (1.22–1.72)
Doi and Minowa (3 factor)	156.05 (51)	0.001	0.91	0.12 (0.10–0.14)	1.64 (1.40–1.93)
Graetz (2 factor)	90.27 (34)	0.001	0.94	0.11 (0.08–0.13)	1.06 (0.89–1.29)
Graetz (3 factor)	149.27 (51)	0.001	0.92	0.12 (0.09–0.14)	1.60 (1.36–1.87)
Kilic et al.	179.25 (53)	0.001	0.90	0.12 (0.10–0.15)	1.71 (1.46–2.01)
Martin	113.00 (41)	0.001	0.93	**0.11** (0.09–0.13)	1.29 (1.10–1.54)
Politi et al.	163.54 (52)	0.001	0.91	0.12 (0.10–0.14)	1.67 (1.43–1.97)
Unitary	173.44 (54)	0.001	0.90	0.12 (0.10–0.14)	1.72 (1.46–2.02)
Valenced (+/−)	164.25 (53)	0.001	0.91	0.12 (0.10–0.14)	1.67 (1.42–1.96)
Worsley and Gribbin	148.37 (51)	0.001	0.92	0.12 (0.09–0.14)	1.58 (1.35–1.86)
Kalliath et al.	55.49 (19)	0.001	**0.95**	0.12 (0.08–0.15)	**0.74** (0.61–0.92)
Hankins positive errors	145.44 (39)	0.001	0.91	0.14 (0.11–0.16)	1.73 (1.50–2.02)
Hankins negative errors	114.51 (39)	0.001	0.94	0.12 (0.09–0.14)	1.51 (1.32–1.77)

Note: Bold indicates best model fit as a function of model fit index criterion and scoring method. *Abbreviations*: Comparative fit index (CFI), root mean squared error of approximation (RMSEA), expected cross-validation index (ECVI).

a   Denotes 90% confidence interval (CI).

4.3. Divergent validity

No significant correlation was observed between the GHQ-12 total score and participants age at 48 h ($r = 0.10$, $p = 0.22$), 10 days ($r = 0.08$, $p = 0.33$), 6 weeks ($r = 0.03$, $p = 0.76$) and 13 weeks postpartum ($r = 0.01$, $p = 0.91$). The common variance explained between GHQ-12 scores and participants age at each observation point was <1%.

4.4. Known-groups discriminant validity

The mean GHQ-12 scores for the low and high perineal trauma groups are shown in Table 6 for each observation point (48 h and 10 days perinatal). A significant difference between groups differentiated by the perineal trauma type was observed on GHQ-12 scores at 48 h postpartum, $t_{(142)} = 4.07$, $p < 0.001$, and at 10 days postpartum, $t_{(138)} = 3.17$, $p < 0.001$ in the predicted direction.

Table 6.

Mean GHQ-12 scores at 48 h and 10 days postpartum as a function of perineal birth trauma classification. Standard deviations in parentheses.

Variable	Low perineal trauma ($N = 72$)	High perineal trauma ($N = 72$)
GHQ-12 (48 h)	813.46 (130.55)	707.10 (179.26)
GHQ-12 (10 days)	885.14 (181.73)	781.06 (206.34)

4.5. Internal consistency

Calculated Cronbach's alpha of the GHQ-12 scale at 48 h, 10 days, 6 weeks and 13 weeks were 0.85, 0.91, 0.92 and 0.95 respectively.

4.6. Test–retest reliability

Pearson's r was calculated between GHQ-12 scores for the 1st and 4th observation point and though statistically significant ($p = 0.001$) did not reach Kline's[35] threshold for test–retest reliability ($r = 0.31$).

5. Discussion

The Maltese GHQ-12 as a measure of wellbeing, utilising VAS scoring demonstrated good divergent and known-groups discriminant validity, consistent with the literature on perineal trauma which clearly suggests that perineal pain decreases progressively as mothers regain their general health.[55] This further supports the use of a VAS method of scoring as one able to determine wellbeing rather than psychological morbidity. Further research which assesses the GHQ-12(WB) against other measures of wellbeing would be useful to further establish the value of the VAS scale in this regard. In line with other studies the GHQ-12(WB) was observed to have good internal consistency, suggesting that the GHQ-12(WB) may be a useful clinical screening tool in the postpartum context to determine wellbeing.

The test–retest characteristics of the GHQ-12(WB) in this study, however, not reach Kline's criterion. This is not uncommon and consistent with previous research.[26] Klein suggests that if used in single groups for research purposes then a lower correlation level may be acceptable.[35] Alternately one suggestion offered by Ip and Martin[26] is that the event of birth and its associated emotional aspects make a test-retest procedure within a normal threshold of acceptability unfair on the test.[26] The same argument can be adopted here in that the postpartum period is normally characterised by a dynamic and fluctuating emotional state.

The findings of the CFA support the generally accepted notion that the GHQ-12 is a multi-dimensional measure, with two factor models offering the best solutions at all data points, a finding consistent with other translated versions of the instrument.[19 and 26] The factor structure is most generally consistent with Kalliath et al.[54] and the findings of Ip and Martin[26] which focus on a two factor correlated short eight-item scale. Taken with the positive findings related to validity and internal consistency, this suggests that an 8 item scale version of the GHQ -12 (WB) may have potential clinical utility. That the factor models tested do not provide a good fit on all fit indices may be directly related to the use of VAS as a scoring method and support use of the GHQ-12 to determine wellbeing. The GHQ-12(WB) was specifically used as a means to measure positive general health and emotional wellbeing. Hence the original focus of the GHQ-12 on mental health, particularly anxiety or depression and societal dysfunction, in this sense was not utilised.

The improvement of model fit across all models, irrespective of factor structure at the fourth observation point does indicate degree of variability in factorial stability over time. However, this may be more indicative of the impact of the 'change state' of the women in the postpartum period, which causes women to respond differently to the questionnaire by the later postpartum period. Whilst this seems intuitively reflective of the natural state of postpartum recovery, it significantly highlights the importance of the dynamics of the population under investigation on the measurement stability, reliability and veracity of psychometric instruments. This perhaps further supports the use of a general wellbeing approach rather than attempts to measure specific domains of psychological morbidity.

6. Conclusion

Findings from the current study seem to generally support the reliability and validity of the Maltese version of the GHQ and offer promise of its utility as a potential measure of wellbeing rather than one of psychological morbidity. The postpartum period is a time when there might be a number of threats to mood state which do may not necessarily constitute morbidity but may affect a sense of general wellbeing, making the instrument's relevance in this context particularly pertinent. The current study has some limitations including small sample size and its use in one particular context. These could be addressed by future replication studies, evaluation against other measure of wellbeing and use in the context of postnatal depression to further determine the measures discriminant validity. Investigation of the GHQ-12(WB) in other languages would seem merited. Of additional value would be establishment of its value in a pregnant population and across the continuum of the perinatal period. Further research is also required to establish the full extent of the utility of Kalliath et al's[54] two factor, eight-item version of the GHQ in postpartum women. Despite the tenuous findings re factor structure and the test–retest issues, the results here appear encouraging enough to support further investigation of this instrument.

Ethical approval

References

1. CB World Health Organization (WHO) & UNICEF. Declaration of Alma-Alta. Geneva: Switzerland: World Health Organization; 1978.

2. Lucey P. Social determinates of health. Nursing Economics 2007;25(2):103–9.

3. Beddington J, Cooper CL, Field J, Goswami U, Huppert FA, Jenkins R, et al. The mental wealth of nations. Nature 2008;455(23):1057–60.

4. Kinderman P, Schwannauer M, Tai S. The development and validation of a general measure of well-being: the BBC well-being scale. Quality of Life Research

2011;20:1035–42.

5. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick– Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health and Quality of Life Outcomes 2007;5(63). Available at http://www.hqlo.com/content/5/1/63 [accessed 3,.0. 2013].

6. Galvin KT, Les Todres L. Kinds of well-being: a conceptual framework that provides direction for caring. International Journal of Qualitative Studies on Health and Well-being 2011;6:1–13.

7. Magill-Cuerden J. Birthwrite nurturing and supporting mothers. A hidden skill in midwifery. British Journal of Midwifery 2006;14(6):374.

8. Darvell R, Skirton H, Farrand P. Psychological factors that impact on women's experiences of first-time motherhood: a qualitative study of the transition. Midwifery 2010;26:357–66.

9. Glazener CMA. Parental perceptions and adaptation to parenthood. British Journal of Midwifery 2005;13(9):578–83.

10. Glazener CMA. Women's health after childbirth. In: Henderson C, Bick D, editors. Perineal care: an international issue. London: Quay Books Division; 2005. p. 11–7.

11. Yelland J, Sutherland G, Brown SJ. Postpartum anxiety, depression and social

health: findings from a population-based survey of Australian women. BMC Public Health 2010;10:771. Available at http://www.biomedcentral.com/14712485/10/771 (accessed 25.04.13].

12. Centre for Maternal and Child Enquiries (CMACE). Saving mothers' lives: reviewing maternal deaths to make motherhood safer: 2006–2008. British Journal of Obstetrics and Gynaecology 2011;118(Supp. 1).

13. Brown S, Lumley J. Physical health problems after childbirth and maternal depression at six to seven months postpartum. British Journal of Obstetrics and Gynaecology 2000;107:1194–201.

14. Jomeen J, Martin CR. The impact of choice of care on psychological health

15. Olsson A, Lundqvist M, Faxelid E, Nissen E. Women's thoughts about sexual life after childbirth: focus group discussions with women after childbirth. Scandinavian Journal of Caring Science 2005;19:381–7.

16. Martell LK. Heading towards the new normal: a contemporary postpartum experience. Journal of Obstetric Gynecologic & Neonatal Nursing 2001;30:496–506.

17. Goldberg DP, Williams P. A user's guide to the General Health Questionnaire. NFERNelson: Basingstoke; 1998.

18. Doi Y, Minowa M. Factor structure of the 12-tiem General Health questionnaire in the Japanese adult population. Psychiatry and Clinical Neurosciences 2003;57: 379–83.

19. Montazeri A, Harirchi AM, Shariati M, Garmaroudi G, Ebadi M, Fateh A. The 12- item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. Health and Quality of Life Outcomes 2003;1(66.). Available at http://www.hqlo.com/content/1/1/66 (accessed 25.04.13).

20. Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care. Acta Psychiatrica Scandinavica 1999;100:462–8.

21. Jacobs KS, Bhugra D, Mann AH. The validation of the 12-item General Health Questionnaire in ethnic Indian women living in the United Kingdom. Psychological Medicine 1997;27:1215–7.

22. Lee DTS, Yip ASK, Chui HFK, Leung TYS, Chung TKH. Screening for postnatal depression: are specific instruments mandatory. Journal of Affective Disorders 2001;63:233–8.

23. Navarro P, Ascaso C, Garcia-Esteve L, Aguado J, Torres A, Martin-Santos R. Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. General Hospital Psychiatry 2007;29:1–7.

24. Werneke U, Goldberg DP, Yalcin I, Ustun BT. The stability of the factor structure of the General Health Questionnaire. Psychological Medicine 2000;30:823–9.

25. Martin CR, Jomeen J. Is the 12-item General Health Questionnaire (GHQ-12)

confounded by scoring method during pregnancy and following birth? Journal of Reproductive and Infant Psychology 2003;21:267–78.

26. Ip WP, Martin CR. Psychomtric properties of the 12-item General Health Questionnaire in Chinese women during pregnancy and the postnatal period. Psychology Health and Medicine 2006;11(1):60–9.

27. European Commission Directorate-General for Translations. 2005 English Style Guide—a handbook for authors and translators in the European Commission. 5th ed. 2005. Available at http://ec.europa.eu/comm/translation/writing/style guides/english/style guide en.pdf [accessed 10.11.05].

28. Waltz CF, Strickland OL, Lenz ER. Measurement in nursing and health research. 3rd ed. New York: Springer Publishing Company; 2005.

29. SPSS. PASW statistics 18 core system user's guide. Chicago, IL: SPSS Inc.; 2009.

30. SPSS. PASW advanced statistics 18. Chicago, IL: SPSS Inc.; 2009.

31. Arbuckle JL. AMOS 18 user's guide. Chicago, IL: AMOS Development Corporation; 1995–2009.

32. Muthen LK, Muthen B.O.. Mplus user's guide. 3rd ed. Los Angeles, CA: Muthen and Muthen; 1998–2004.

33. Hankins M. The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? Clinical Practice and Epidemiology in Mental Health 2008;4(10). Available at: http://www.cpementalhealth.com/content/4/1/10 [accessed 10.03.13].

34.   Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. 2nd ed. Routledge/Taylor & Francis Group: New York, NY US; 2010.

35.   Kline P. The handbook of psychological testing. London: Routledge; 1993.

36.   Kline P. A psychometrics primer. London: Free Association Books; 2000.

37.   Martin CR, Thompson DR. Design and analysis of clinical nursing research studies. London: Routledge; 2000.

38.   Colman AM, Norris CE, Preston CC. Comparing rating scales of different lengths: equivalence of scores from 5-point and 7-point scales. Psychological Reports 1997;80(2):355–62.

39.   Friedrich F, Alexandrowicz R, Benda N, Cerny G, Wancata J. The criterion validity of different versions of the General Health Questionnaire among non-psychiatric inpatients. Social Psychiatry and Psychiatric Epidemiology 2011;46(7): 635–41.

40.   Kind P, Barmby P. Defending attitude scales. In: Russell M, Kavanaugh M, editors. Assessing students in the margins: challenges, strategies, and techniques. Charlotte, NC, US: IAP Information Age Publishing; 2011. p. 117–35.

41.   Shulruf B, Hattie J, Dixon R. Factors affecting responses to Likert type questionnaires: introduction of the ImpExp, a new comprehensive model. Social Psychology of Education 2008;11(1):59–78.

42.   Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods 2004;9(4):466–91.

43.   Lubke GH, Muthe´n BO. Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. Structural Equation Modeling 2004;11(4):514–34.

44.   Muthen B, Kaplan DA. Comparison of some methodologies for the factor analysis of non-normal Likert variables: a note on the size of the model. British Journal of Mathematical and Statistical Psychology 1992;45(1): 19–30.

45.   Kline RB. Principles and practice of structural equation modeling. 2nd ed. New York: Guilford Press; 2005.

46.   Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88:588–606.

47.   Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin 1990;107:238–46.

48.   Bollen KA. Structural equations with latent variables. New York: John Wiley & Sons; 1989.

49.   Hu LT, Bentler PM. Evaluating model fit. In: Hoyle RH, editor. Structural equation modelling: concepts, issues and applications. Thousand Oaks, CA: Sage; 1995. p. 76–99.

50.   Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. Psychological Bulletin 1988;103: 391–410.

51.   Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136–62.

52.   Schumaker RE, Lomax RG. A beginner's guide to structural equation modelling. Hillsdale, NJ: Lawrence Erlbaum Associates; 1996.

53.   Graetz B. Multidimensional properties of the General Health Questionnaire.

Social Psychiatry and Psychiatric Epidemiology 1991;26:132–8.

54.   Kalliath TJ, O'Driscoll MP, Brough P. A confirmatory factor analysis of the General Health Questionnaire-12. Stress Health 2004;20:11–20.

55.   Andrews V, Thakar R, Sultan AH, Jones PW. Evaluation of postpartum perineal pain and dyspareunia—a prospective study. European Journal of Obstetrics & Gynecology and Reproductive 2008;137(2):152–6.