Reliability, validity and responsiveness of the short form-36 health survey: Findings from the women's health study of Accra, Ghana

FAUSTINA FREMPONG-AINGUAH[1, 2]
ALLAN HILL[1]

Résumé

Mesurer l'état de santé des individus ou des populations est un premier pas important vers l'évaluation des besoins de santé et de l'impact des interventions en matière de santé. Les statistiques fiables sur l'état de santé de la population générale sont rares, en particulier dans les pays en développement comme le Ghana. Cet article mesure la validité et la fiabilité d'un outil standard largement utilisé dans les pays développés, le questionnaire SF-36 (Short-Form 36), pour évaluer l'état de santé des femmes en milieu urbain au Ghana. L'étude est basée sur un échantillon représentatif de 2.814 enquêtées de la Women's Health Study of Accra. Nous avons utilisé le test alpha de Cronback, le test de fiabilité fractionnée, les matrices multi-questions et multi-réponses et l'analyse en composante principale pour mesurer la validité du SF-36 dans le contexte ghanéen. Les taux de réponse à l'enquête et pour chaque domaine de santé sont supérieurs à 97 %. Les tests de validation des réponses satisfont les critères psychométriques rigoureux de cohérence internes et la fiabilité était supérieure au critère suggéré de 0,40 pour tous les éléments du SF-36 excepté un élément de la sous-échelle de vitalité. Les scores normalisés du SF-36 indiquent une grande variabilité et une cohérence interne acceptable (alpha de Cronbach > 0,70). Suivant une simple rotation orthogonale, deux composantes principales ont été identifiées et interprétées comme les composantes de la santé physique et mentale, identiques à celles trouvées dans d'autres contextes d'études. Ces composantes représentent 62,6 % de la variance des scores d'échelle dans la population étudiée. L'absence de corrélations fortes entre les huit différentes sous-échelles du SF-36 démontre que chaque échelle mesure un concept de

---

1.   Division of Social Statistics, Faculty of Social and Human Sciences, University of Southampton.

2.   Regional Institute for Population Studies, University of Ghana.

8

Reliability, validity and responsiveness of the short form-36 health survey:
Findings from the women's health study of Accra, Ghana

santé distinct. L'étude prouve que le SF-36 convient à l'évaluation de l'état de santé des femmes à Accra et, par extension, à l'évaluation de l'état de santé des femmes dans d'autres zones urbaines du Ghana.

Mots-clés

Santé physique et mentale, Short-Form 36 (SF-36), validation, fiabilité, femmes adultes, Accra, Ghana.

Summary

Measuring the health status of individuals or a population is an important first step in the assessment of health care needs and the evaluation of the impact of health interventions. Reliable population-wide statistics on the health status are scarce especially in developing countries including Ghana. This paper assesses the validity and reliability of a standard instrument widely used in high income countries, the Short Form-36, as a tool to evaluate health status among women age 20 years and older in urban Ghana. The study draws on a representative sample of 2'814 respondents from the Women's Health Study of Accra. We used multi-trait, multi-item matrix methods, Cronbach's alpha estimates, half-split reliability tests and principal component analysis to evaluate the validity of SF-36 in the Ghanaian context. The response rate and the rate of completion for each health domain was over 97%. Item-level validation satisfied the rigorous psychometric criteria for internal consistency and reliability was above the suggested 0.40 criterion for all SF-36 scoring items apart from one item under Vitality sub-scale. The normed SF-36 scores showed wide variability and acceptable internal consistency (Cronbach's alpha > 0.70). Following simple orthogonal rotation two principal components were identified and interpreted as Physical and Mental health components, identical to those found in other study settings. These components accounted for 62.6% of the reliable variance in scale scores across the study population. The absence of strong correlations between the various eight SF-36 sub-scales, demonstrate that each scale measures a distinct concept of health. The study provides evidence that SF-36 is suitable for the assessment of health status among women in Accra and, by extension, suitable for assessment of the health status of women in urban areas of Ghana.

Key words

Physical and Mental health; Short-Form 36 (SF-36); validity; reliability; adult females; Accra; Ghana.

Introduction

Reliable, objective and practical measures of health outcomes are vital for the assessment of levels, changes and differentials in the health of individuals or populations. Such measures are needed to assess health needs and the impact of health services on health outcomes. Reliable

statistics about the health status of whole populations are scarce and often incomplete, especially in developing countries, including Ghana (Moore, Gould, Keary, 2003; O'Donnell *et al.*, 2008). Policy makers, health practitioners and researchers usually rely on proxy measures of morbidity, such as mortality records and estimates, as well as physician-based diagnoses in health facilities to estimate population health outcomes. Riley (2001) however, amongst others, has shown how, during the health and mortality transitions, mortality and morbidity evolved on different trajectories. The inherent challenges associated with assessing individual health status for large populations have led to the increasing use of subjective measures (Kuhn, Rahman, Menken, 2004; Lohr, 1988) to complement measures based on medical outcomes. Measures of subjective health have widely been used to assess health status, trends and differentials among different populations since the 1980s. Such approaches, however, have seldom been used to assess the health status of any African population, including Ghana (Wanyonyi, Stones, Sequeira, 2010; Wyss *et al.*, 1999). This paper evaluates the reliability, validity and responsiveness of the Short Form-36 instrument to examine intra-urban morbidity patterns among women in Accra, Ghana.

Background

Since the second half of the twentieth century, the emphasis on health status measurements has shifted from a focus on traditional measures of survival and frank disability often included in censuses (e. g. blindness, deafness and missing limbs) to include measures of subjective health (Hunt, McEwen, McKenna, 1986; Kind, Gudex, 1991; Lohr, 1988). This is because human populations throughout the world are now living longer as a result of rising living standards, technological advancements and medical science (Kilama, 2009). This has contributed to a shift in disease patterns, resulting in an increased burden of chronic diseases (Hunt, McEwen, McKenna, 1986; Lynch, 2003) and an interest in the measurement of the impact of health beyond the absence of disease (Sabbah *et al.*, 2003). It is now widely acknowledged that a thorough assessment of the benefits of good health must examine both clinical outcomes and self-perceived dimensions of morbidity (Kind, Gudex, 1991; Chen, Li, Kochen, 2005).

Subjective health measures have been used in many clinical as well as large population surveys, especially in industrialised countries. Generic instruments used as aids in clinical and non-clinical studies include the

10

Reliability, validity and responsiveness of the short form-36 health survey:
Findings from the women's health study of Accra, Ghana

Short Form-36 Health Survey, the Nottingham Health Profile, the Sickness Impact Profile, and the World Health Organisation Quality of Life Assessment (WHOQoL) among others. These measures include physical functioning, general well-being, social functioning, and psychological well-being, alone or in combination, to assess health status among different population sub-groups (Jenkinson, Wright, Coulter, 1994; McFadden *et al.*, 2009; Mishra *et al.*, 2004). These are valuable sources of information as they are simple to collect, reflect individuals' overall health status (Mishra *et al.*, 2004; Quesnel-Vallée, 2007), are sensitive to some pre-clinical conditions (McFadden *et al.*, 2009) and reflect health trajectories (Chen, Cohen, Kasen, 2007). One of the most widely used generic instruments is the Short Form-36 (Muller-Nordhorn, Roll, Willich, 2004; Macran, Clarke, Joshi, 1996).

Short Form-36 (SF-36) is a multi-item scale comprising 36 questions developed to assess the health of the general population as well as that of patients across eight health domains. It includes Physical Functioning (PF), Role limitation due to Physical health problems (RP), Bodily Pain (BP), Role limitations due to Emotional problems (RE), General Health perception (GH), Social Functioning (SF), Vitality (VT) and Mental Health (MH) (Fukuhara *et al.*, 1998; Ware, Sherbourne, 1992; Mishra *et al.*, 2004). In addition, the eight domain scales can be combined to yield two summary scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS), related to physical functioning and mental health/well-being respectively. The SF-36 also includes a single item that provides an indication of perceived change in health.

Other shorter versions such as SF-20, SF-12, SF-8 and SF-6 have also been developed, but none has gained much acceptance and widespread use as the SF-36 until now (Garratt *et al.*, 2002; Butterworth, Crosier, 2004; Ware, Gandek, 1998b). The SF-36 health study was listed as an evaluation tool by the International Quality of Life Assessment (IQOLA) Project in 1991 (Ware, Gandek, 1998a). Since then it has been translated into a number of languages (QualityMetrics, 2011) and used around the world (including Australia, UK, Japan, India, Lebanon, Sweden, Tanzania, China, Kenya and South Africa). It has been found to be a valid and reliable outcome measure across several populations (Burholt, Nash, 2011; Wyss *et al.*, 1999; Wanyonyi, Stones, Sequeira, 2010; Sullivan, Karlsson, Ware, 1995; McInerney *et al.* 2008). The SF-36 has proved useful in comparing the health of general and specific population sub-groups (Macran, Clarke, Joshi, 1996; Jenkinson, Wright, Coulter, 1994; Mishra *et al.*, 2004), screening individual patients, and estimating the relative bur-

den of different diseases (Muller-Nordhorn, Roll, Willich, 2004; Ruta *et al.*, 1994).

Notwithstanding the relevance of subjective measures to a population's health, to date there have been relatively few applications focused on sub-Saharan Africa, especially Ghana (Arku *et al.*, 2011; Debpuur *et al.*, 2010; Mwanyangala *et al.*, 2010; Wanyonyi, Stones, Sequeira, 2010; Wyss *et al.*, 1999). A few studies conducted in Ghana applied self-reported general health and mental health measures to assess their relationships with housing quality (Arku *et al.*, 2011), or among older adults in rural settings (Debpuur *et al.*, 2010). The health status of the adult population, especially older women, is of great concern, as this group has long been neglected in policies and service provision. Health care in Ghana is mainly organised around curative services and acute conditions and has focused attention on child health and the major tropical diseases (de-Graft Aikins, 2007). Although, these policies have led to a reduction in child mortality rates and a drop in some tropical diseases, large gaps remain concerning the level and extent of adult health problems unrelated to infectious diseases, tropical or otherwise.

In the past, morbidity patterns in Ghana and other developing countries were driven mainly by infectious and parasitic diseases with often fatal outcomes (Lynch, 2003). The prevailing assumption that chronic conditions do not pose significant public health problems in the region is gradually changing (Ekpenyong *et al.*, 2012; de-Graft Aikins, 2007) with rapid urbanisation and modernisation in urban centres. Degenerative diseases and their risk factors in Ghana are fast increasing especially in urban areas among the adult population making this group less productive (Agyei-Mensah, de-Graft Aikins, 2010; Duda, Hill 2005). This situation is compounded by the fact that a sizeable proportion of Ghanaians (40%) do not utilise formal health facilities (Ghana Statistical Service, 2008) due to past health policies including the twenty years of implementation of the «cash and carry» health care system following enforced structural adjustment reforms (Agyepong, Adjei, 2008).

Accra, the capital of Ghana, accounts for a quarter of the country's urban population (Ghana Statistical Service, 2012) and like other rapidly growing cities in the sub-region is moving through demographic and epidemiologic transitions with its associated dual burden of disease, when communicable diseases (such as malaria and diarrhoeal diseases) persist while non-communicable diseases emerge (mainly obesity, hypertension, diabetes and cardio-vascular conditions (Agyei-Mensah, de Graft Aikins, 2010). Since the 1960s, Accra has been a destination for migrants from less well-endowed regions and rural areas (Songsore,

12

Reliability, validity and responsiveness of the short form-36 health survey:
Findings from the women's health study of Accra, Ghana

2003). The city's population density increased over 280% within 26 years, from 441 persons per square kilometre in 1984 to 1,236 km² in 2010 (Ghana Statistical Service, 2012). Rapid urbanisation and technological advances have led to behavioural changes which are impacting morbidity and mortality patterns.

Self-reported indicators of health are increasingly used as valid indicators of morbidity within the general population. It is believed that adapting the SF-36 to study the health status of women in Accra can provide insights into the dimensions of morbidity of importance for health service providers, policy-makers and health researchers and offers guidance to others collecting and analysing similar measures of health in the region. In view of the limited use of population self-reported health status in Ghana, this study seeks to assess the practicality and value of using the SF-36 in the Ghanaian context by examining the responsiveness, reliability and validity of SF-36 items among women in Accra. Specifically, the study seeks to test the following assumptions: (1) there is widespread acceptance of the SF-36 health items in terms of concepts, data completeness and distribution of response patterns; (2) there is item convergence, divergence and acceptable scaling success rates; (3) there is internal consistency and reliability for each item and the eight domain scales, including ceiling and floor effects; and (4) the structural validity for the two summary measures can be tested and validated. The study hypothesized that at the item level, acceptability, validity and reliability would be supported if assumptions 1-3 were satisfied. Additionally, for the domains and summary scores, validity and reliability would be achieved if assumptions 3 and 4 were fulfilled. When these assumptions are satisfied it would demonstrate that these domains possess the desired properties, thereby being suitable for studying the health status among the general population in Accra and other urban areas of Ghana. This paper deals solely with these technical aspects of the instrument. Empirical findings and interpretations are presented elsewhere (Hill *et al.*, 2007; Darko *et al.*, 2012; DeMenil *et al.*, 2012).

Materials and methods

The study draws on the Women's Health Study of Accra (WHSA) conducted in 2008/9 by the Institute of Statistical, Social and Economic Research, Ghana and the Department of Global Health and Population, Harvard School of Public Health. Although the WHSA comprises two waves to date, the data used here from Wave II are analysed as a cross-

sectional survey using a multi-stage sampling design. Further details are reported elsewhere (The WHSA writing team, 2011). The updated 2000 population and housing census master list of Enumeration Areas (EAs) in the Accra Metropolitan and Ledzokuku Krowor Municipal Areas (hereafter Accra) constituted the sampling frame. At the first stage, 200 EAs were selected using probability proportional to size from 1'731 EAs, with an estimated population of 2.2 million living in roughly 364'000 households (Ghana Statistical Service, 2005). To ensure that all women aged 18 years and older usually resident in Accra were duly represented, an extensive mapping and listing exercise was carried out in the selected EAs. This listed individuals formed the frame for the next stage of selection.

A total of 3'068 women aged 20 years and older were selected for the study. Interviews were conducted using a structured questionnaire, after respondents' had given their consent. The protocol comprised a household roster and demographic information, an individual interview for socio-economic information, the SF-36 questions, self-reported illnesses, health care utilisation, medical history, lifestyle patterns, among others. The survey instrument was translated and pre-tested to make it suitable for the study population. Experienced field personnel were recruited, trained and used to gather the information between October 2008 and June 2009 (The WHSA writing team, 2011). The SF-36 was administered before respondents were interviewed about additional health and other illnesses so that any discussion of health problems did not influence respondents' answers to the SF-36 questions.

Analytical approach

The acceptability and quality of the data was analysed by examining the number of incomplete responses to SF-36 items and the distribution of the responses. The distribution of dichotomous items response scale was tested for «aggregate endorsement frequencies», using the 90/10 percentage distribution as recommended (The WHOQOL Group, 1998).

Validity was evaluated using three approaches: convergence, divergence and structural validity. A multi-trait, multi-method matrix was used to assess convergent and divergent validity. Convergence measured the degree to which items that theoretically should be related to each other are indeed observed to be interrelated. Similarly, divergence measured the extent to which SF-36 items that theoretically should not be related

to each other are also observed to discriminate between the scales. Convergence was achieved if the correlation coefficient between each item and its related scale equalled or was greater than 0.40 after accounting for overlap. To support the divergence hypothesis, the correlation between each item and its hypothesized sub-scale should be higher than with other scales. Principal components analysis with the varimax rotation method was used to test for structural validity of the scales and the two summary scores (PCS and MCS). The Kaiser-Meyer-Olkin (KMO) statistic and Bartlett's test of sphericity were carried out to check for sampling suitability and factor analysis.

Reliability tests were validated by assessing the extent to which items within a scale were correlated with each other, using Cronbach's alpha coefficient. Internal and consistent reliability was considered adequate if $\alpha \geq 0.70$ for group comparison (Ware, Gandek, 1998b). Similarly, a half-split reliability test and Cronbach's alpha were employed to measure the internal stability and reliability of the sub-scales.

Each of the SF-36 items raw scores were calibrated, summed and transformed on to a scale ranging from 0 to 100, with higher scores indicating better health status. Descriptive statistics were computed for items and sub-scales, including ceiling and floor effects. Data preparation and analyses were carried out using SPSS version 20.0 (SPSS Inc., Chicago, IL, USA) for Windows. Two-tailed tests of significance and a p-value < 0.05 was considered statistically significant.

Results

Of the 3'068 women sampled, interviews were completed for 2'814, yielding a response rate of 91.8%. The mean age of respondents was 42.6 years, with an age range from 20-95 years. Seventeen per cent had no formal education while 28.1% had secondary or higher education. The majority (81.8%) were ever married with 24.5% having had five or more children. Thirty-seven per cent of the respondents were obese, with an additional 28.2% being overweight. One-third of women interviewed had experienced multiple disease symptoms four weeks prior to the survey.

TABLE 1   Percentage distribution of SF-36 item frequency, 2008-9

Short Form-36 items	Response distribution of each category in percentages (%)						
Item content	1	2	3	4	5	6	Missing
Physical functioning (PF)							
Moderate activity	4.0	11.8	84.2				0.0
Vigorous activity	18.7	27.5	53.8				0.0
Lifting or carrying	5.2	10.1	84.7				0.0
Climb flight stairs	5.3	11.5	83.2				0.0
Climb several flights	15.4	25.4	59.2				0.0
Bending or kneeling	6.4	17.9	75.7				0.0
Walking short	2.2	5.0	92.8				0.0
Walking medium	4.3	11.4	84.4				0.0
Walk long distance	11.3	21.7	67.0				0.0
Bathing or dressing	1.3	5.2	93.5				0.0
Role Physical (RP)							
Cut down on time	22.4	77.6					0.0
Accomplished less	23.2	76.8					0.0
Limited in activity	17.4	82.5					0.1
Difficulty performing	23.5	76.5					0.0
Bodily Pain (BP)							
Body pain	0.4	6.3	28.3	22.2	15.8	26.9	0.2
Interfere	1.3	5.4	11.5	25.0	56.5		0.3
General Health (GH)							
Overall health	1.8	11.4	47.6	26.7	12.2		0.2
Get sick easier[a]	1.6	10.4	3.3	38.1	46.6		0.1
Healthy as others	4.8	14.3	7.5	43.5	29.6		0.2
Health to worsen[a]	0.6	11.1	12.4	33.3	42.4		0.2
Excellent health	3.6	19.7	1.5	53.9	21.2		0.1
Role Emotional (RE)							
Cut down work	17.2	82.5					0.4
Less done like	17.7	81.9					0.4
Not carefully	18	81.6					0.4
Social Functioning (SF)							
Interfere social	1.7	3.9	9.8	15.4	69.2		0.0
Time limit	1.8	2.4	4.1	8.1	15.5	67.4	0.7
Vitality (VT)							
Full of life	1.3	3.9	18.7	23.6	36.4	16.0	0.3
Lots of energy	1.5	4.8	14.8	24.8	38.7	15.1	0.3
Worn out[a]	0.3	2.3	4.6	18.8	26.6	46.9	0.4
Feel tired[a]	1.3	7.3	7.9	43.7	26.5	12.9	0.3
Mental Health (MH)							
Nervous[a]	0.3	5.3	3.2	17.8	14.5	58.5	0.3
Unhappy[a]	0.5	2.5	3.4	14.6	24.0	54.6	0.2
Calm	0.9	4.9	19.9	23.2	36.9	13.9	0.2
Downhearted[a]	0.4	2.2	3.2	15.0	23.8	55.0	0.4
Happy	1.0	4.3	18.0	23.0	39.1	14.3	0.4
Health change question	1.8	22.3	47.2	17.3	10.5		0.9

Abbreviations: blank cell = Not applicable; [a] All items showing negative order were reverse coded. Respondents were 2'814. Figures are weighted by sampling weights.

The distribution of each item, per cent of response categories, item and sub-scale label and per cent missing data are presented in Table 1. All 36 items, except the «health transition» question, were included in the computation processes, with each item contributing to only one scale. The distribution of items showed that all response categories were fully utilised (Table 1). Less skewed distributions were observed for items with wide response categories. Results of the sensitivity aggregate endorsement for each dichotomous item revealed acceptable sensitivity, with all seven items (RP and RE) exceeding the 90-10 percentage distribution criterion (Table 1). Missing data for the SF-36 items was generally low (less than 0.9%) and evenly spread over the items. This suggests that SF-36 items had good acceptability among the respondents. Since very few items were missing, the study did not substitute for the missing data but employed list-wise criteria during the analysis.

The tests of convergence and divergence validity were assessed to determine whether the items hypothesized to measure a particular sub-scale actually measured that construct. Table 2 presents the multi-trait, multi-method item correlation matrix showing the relationship between each item to its hypothesized SF-36 sub-scale, as well as the item's relationship with the other seven health sub-scales. Columns 5-12 represent correlations between each sub-scale and all items, with the figures on grey background depicting the convergence score of the matrix. The results showed that item internal consistency with their hypothesized scales, after correction for overlap, exceeded the 0.40 criterion apart from one item-scale correlation, *«feeling tired»* of the Vitality scale, at 0.38. In addition, the SF-36 items demonstrated linear association and contributed substantial information to the total scale score. An item's contribution was considered substantial, if each item-scale correlation varied from 0.40 to 0.90 across items in the same scale. The correlations generally confirmed the hypothesized associations of the items with their respective scales.

TABLE 2   Item descriptive statistics, and multi-trait,
multi-item correlation matrix between items and eight scales

Scale				Eight SF-36 subscales							
Item	Mean	SD	CI	PF	RP	BP	GH	SF	RE	VT	MH
Physical Functioning											
Moderate	2.80	0.49	2.78-2.82	.76	.37	.32	.38	.38	.18	.31	.17
Vigorous	2.35	0.78	2.32-2.37	.72	.33	.33	.47	.31	.15	.38	.22
Lifting/carrying	2.80	0.52	2.78-2.82	.77	.36	.33	.41	.36	.17	.39	.22
Flight of stairs	2.78	0.53	2.76-2.80	.82	.40	.32	.39	.39	.23	.35	.20
Several flights	2.44	0.74	2.41-2.47	.81	.40	.36	.48	.35	.18	.41	.23
Bending/kneeling	2.69	0.58	2.67-2.71	.82	.38	.35	.40	.39	.21	.39	.20
Walking short	2.91	0.36	2.90-2.92	.71	.35	.27	.30	.38	.19	.32	.15
Medium distance	2.80	0.49	2.78-2.82	.84	.41	.32	.40	.43	.24	.38	.19
Long distance	2.56	0.69	2.53-2.59	.82	.42	.36	.43	.37	.22	.44	.23
Bathing/dressing	2.92	0.31	2.91-2.93	.57	.27	.28	.27	.36	.14	.30	.15
Role Physical											
Cut down time	1.78	0.42	1.76-1.80	.41	.87	.40	.35	.43	.36	.36	.22
Accomplish less	1.77	0.42	1.75-1.79	.44	.90	.41	.39	.44	.36	.37	.23
Limited in work	1.83	0.38	1.81-1.84	.43	.83	.37	.33	.44	.33	.32	.20
Difficulty	1.77	0.42	1.75-1.79	.45	.85	.42	.40	.42	.37	.40	.23
Bodily Pain											
Bodily pain	4.42	1.30	4.37-4.47	.30	.32	.58	.40	.27	.20	.45	.28
Interference	4.87	1.40	4.82-4.92	.44	.46	.58	.48	.51	.30	.50	.31
General Health											
Overall health	3.66	0.92	3.63-3.69	.45	.37	.46	.53	.35	.26	.47	.33
Get sick easier	4.18	1.01	4.14-4.22	.35	.29	.34	.56	.30	.23	.42	.30
Healthy as others	3.79	1.16	3.75-3.83	.37	.29	.36	.64	.29	.18	.36	.24
Health to worsen	4.06	1.02	4.02-4.10	.31	.20	.38	.47	.19	.11	.26	.20
Excellent health	3.70	1.12	3.66-3.74	.44	.35	.43	.72	.34	.22	.48	.32
Social Functioning											
Interfere social	4.46	0.94	4.43-4.49	.41	.43	.39	.34	.62	.51	.35	.34
Time limit	5.46	1.08	5.42-5.50	.42	.41	.40	.37	.62	.38	.41	.32
Role Emotional											
Cut down work	1.83	0.38	1.82-1.84	.23	.38	.25	.24	.47	.87	.31	.39
Accomplish less	1.82	0.38	1.81-1.83	.22	.36	.27	.27	.47	.88	.32	.41
Not carefully	1.82	0.39	1.81-1.83	.22	.35	.27	.24	.46	.84	.30	.37
Vitality											
Full of life	4.38	1.15	4.33-4.42	.41	.28	.36	.39	.30	.23	.52	.45
Lots of energy	4.40	1.14	4.36-4.44	.49	.32	.38	.39	.33	.20	.54	.42
Worn out	5.11	1.04	5.07-5.15	.23	.26	.36	.36	.29	.25	.42	.40
Feel tired	4.26	1.10	4.22-4.30	.16	.24	.36	.33	.25	.22	.38	.27
Mental Health											
Nervous	5.17	1.18	5.13-5.21	.07	.12	.22	.18	.20	.21	.22	.41
Unhappy	5.23	1.04	5.19-5.27	.16	.14	.14	.19	.26	.29	.32	.57
Calm	4.32	1.14	4.28-4.36	.20	.16	.25	.25	.22	.27	.47	.52
Downhearted	5.26	1.01	5.22-5.30	.19	.18	.22	.28	.29	.34	.39	.51
Happy	4.38	1.12	4.34-4.42	.24	.22	.28	.34	.29	.29	.51	.55

Abbreviations: SD – Standard deviation; CI – Confidence interval. Note: Figures on grey background represent the correlation corrected for overlap; weighted by survey weights.

18

Reliability, validity and responsiveness of the short form-36 health survey:
Findings from the women's health study of Accra, Ghana

Divergent validity was tested by examining each row's scores in the matrix by comparing the figure on grey background with all other figures (see Table 2). Divergence was supported if the correlation between an item and its hypothesized construct was higher than the correlation between that item and all other scales. The results showed that the correlation between each item and its hypothesized scale was greater than the correlation between that item and alternative scales, confirming item discriminant validity for all items. The coefficient range of the discriminant validity scaling success and reliability within individual levels revealed that discriminant validity achieved a 100% success rate for all eight scales. Similarly, with the exception of the VT scale (75%), complete success rates (100%) were achieved in the test of convergence with their hypothesized scales.

Descriptive statistics, ceiling and floor effects of health status among the respondents as identified by eight sub-scales are presented in Table 3. The number of items included in each scale ranged between 2 and 10. The transformed mean scores, standard deviations, median and confidence intervals are also reported. The transformed mean scores and standard deviation indicate where along a scale continuum (0-100) the majority of the respondents in the study population are likely to be found. Adopting a five percent trimmed mean, the results show that on average women in Accra assessed their health states better on all eight scales. The proportion of women achieving lower scores on the eight scales was highest for VT (28.8%) and lowest for SF (10.1%). Ceiling and floor effects show the percentage of women with the worst, or best, possible health score for each scale. In general, minimal floor effects were observed for all scales except RP and RE, having the heaviest floor effects of 15.8% and 14.0% respectively (Table 3). This means that approximately 1 in 7 women interviewed had severe limitations in relation to «cutting down on» or «accomplishing less work» due to physical or emotional challenges in the four weeks prior to the study.

The degree of internal reliability, stability and correlation coefficients among eight sub-scales are presented in Table 4. The overall reliability scale for the eight scales using Cronbach's alpha was 0.82 and standardised alpha coefficient yielded 0.86. The reliability coefficient for the eight scales revealed that none of the items when removed would greatly improve their internal consistency and reliability. The correlation matrix showed that all eight scales were linearly correlated. For example, PF maintained moderate association with GH, SF, VT, RP and BP while maintaining a weak correlation with MH and RE scales. On the other hand, MH scale correlated moderately with the VT and RE sub-scales.

TABLE 3 Descriptive statistics and score distribution for eight health scales

Scale	Transformed scores (0-100) for eight scales*							
	PF	RP	BP	GH	SF	RE	VT	MH
Items	10	4	2	5	2	3	4	5
Mean	85.2	78.3	72.9	71.8	87.7	82.2	70.7	77.4
SD	21.0	37.9	23.9	19.7	20.4	35.8	15.3	14.8
95% CI	85.1	77.9	73.3	69.7	86.6	80.8	70.3	76.9
	-86.7	-80.7	-74.9	-72.2	-88.1	-83.5	-71.5	-78.0
Trimmed mean	88.7	82.5	75.5	71.2	89.9	85.7	71.5	78.2
Median	95.0	100.0	74.0	75.0	100.0	100.0	70.0	80.0
Range (0-100)								
% floor	0.9	15.8	0.2	0.0	0.6	14.0	0.0	0.0
% ceiling	46.7	71.5	25.7	5.8	58.8	77.8	1.9	5.5
Total	2'814	2'814	2'808	2'814	2'813	2'804	2'809	2'808

*Transformed score = [(raw scale score = minimum)/range]*100; CI = Confidence interval.
% floor, % ceiling – per cent at the lowest and highest possible scale scores, respectively.

Internal consistency estimates were greater than 0.70, the minimum standard for group comparisons, for all scales except the Vitality scale. This ranged from 0.69 for Vitality to 0.94 for Role- Physical. A split-half reliability test with different splits was used to compare the stability of the eight scales yielded 0.76 and 0.79 respectively, confirming the relative stability of SF-36.

TABLE 4 Reliability coefficients and inter-scale correlations (eight scales)

Index score	PF	RP	BP	GH	SF	VT	RE	(α)
Physical Functioning								0.93
Role Physical	0.51							0.94
Bodily Pain	0.46	0.46						0.73
General Health	0.57	0.43	0.53					0.81
Social Functioning	0.54	0.49	0.46	0.45				0.76
Vitality	0.52	0.42	0.56	0.57	0.46			0.69
Role Emotional	0.25	0.40	0.28	0.27	0.45	0.33		0.93
Mental Health	0.26	0.25	0.34	0.38	0.37	0.56	0.41	0.71
Split half test								
Split [A]								0.76
Split [B]								0.79

Split [A] = reliability scales = PF, RP, SF, RE; Split [B] = BP, VT, GH and MH; P < 0.001 for all correlations.

TABLE 5   Correlations between eight scales and rotated principal
               components

Index score	Hypothesised association		Rotated Principal Components	
	Physical	Mental	PCS	MCS
Physical functioning	*a*	*c*	0.84	0.06
Role Physical	*b*	*c*	0.68	0.24
Bodily Pain	*a*	*c*	0.73	0.23
General Health	*a*	*c*	0.77	0.20
Social Functioning	*b*	*b*	0.62	0.45
Vitality	*b*	*b*	0.65	0.46
Role Emotional	*c*	*a*	0.16	0.81
Mental Health	*c*	*a*	0.22	0.79
Variance explained			62.6%	
Kaiser – Meyer-Olkin (KMO) sampling adequacy			0.86	
Bartlett's Test of Sphericity χ^2			8'708.9***	
Degrees of freedom			28	
Number of cases			2'797	

Notes: a = strong association (r ≥ 0.70); b = moderate association (0.30 < r < 0.70); c = weak association (r ≤ 0.30).

Structural validity was evaluated using principal component analysis for the eight scales by testing whether the observed data collected during the study correlated with the hypothetical structure of the two summary scores (PCS and MCS). Table 5 shows the correlation between the eight scales and the two rotated components. The analysis included 99.4% of valid cases and the analysis supports the two-factor solution, accounting for 62.6% of the total variance explained. The correlations confirmed the hypothesized association between the eight scales with PCS and MCS scores obtained from the first and second components. Items PF, GH, and BP were strongly correlated with PCS (a = r ≥ .70), but had a weak association with MCS. The MH and RE scores, however, showed a strong correlation with MCS. Vitality and SF had moderate associations with both PCS and MCS. The results further showed the KMO measure of sampling adequacy statistics to be 0.86 and the Bartlett's test of Sphericity (χ^2 = 8708.9, df = 28) at p < 0.0001, satisfying the requirement underlying the factorial analysis.

Discussion

The study aimed at evaluating the responsiveness, reliability and validity of SF-36 by testing data completeness and the distribution of respon-

ses, the convergence and divergence among items, estimating internal consistency and reliability, as well as the structural validity underlying the two summary scores in a rapidly growing city of Africa. The very low missing item information supports the notion that respondents understood the concepts articulated in the SF-36, making it acceptable and suitable for the study of the general population, as reported previously in other populations (Sabbah *et al.,* 2003; Wanyonyi, Stones, Sequeira, 2010; Wyss *et al.*, 1999; Sullivan, Karlsson, Ware, 1995). The various scoring processes and groupings within the eight scales showed considerable consistencies. For example, the mean distributions for the ten items (PF scale) ranged from 2.35-2.92 and this was comparable to findings from the study conducted in 11 countries from the IQOLA project (Gandek *et al*. 1998). Gandek and colleagues (1998) found the mean range for the 10 items (PF) for the United Kingdom, Denmark and Sweden were 2.25-2.92; 2.36-2.91; and 2.34-2.91 respectively. The results suggest that the assumption of checking for acceptability, completeness of data and response distribution were satisfactory and relevant to the study of women's health in Accra.

Convergence and divergence validity were assessed by testing the scaling assumption, scoring and ordering of item means, within scales, and the relationship to the hypothesised scales. The scoring items were linearly related and contributed sufficient information to their respective scales. Item divergence showed that all items successfully discriminated between their hypothesized scale and other measuring constructs. These findings compare favourably to previous studies conducted in, Lebanon (Sabbah *et al*., 2003), China (Qu *et al*., 2009; Zhou *et al*., 2013; Zhang *et al*., 2012) and other developed countries (Ware, Gandek, 1998a) using the SF-36. Item convergence was achieved for all with the exception of «feeling tired». This item may not be sensitive among women in Accra and would require further investigations. A plausible reason for this is that the socially assigned roles of women in this society are very stressful. For example, practical duties of household chores, caring for children and making a living exerts much pressure, making them feel constantly tired. The validity checks showed that, none of the items or scales when deleted would significantly improve the score reliability.

Testing the hypothesized element of the two summary scores showed that the factorial analysis of the scales yielded results identical to those conducted in other study settings (Butterworth, Crosier, 2004; Sullivan, Karlsson, Ware, 1995; Qu *et al*., 2009). The total variance of 62.7% accounted for by the two-factor solution is similar to that found in other studies. For example, in a study of the factor structure of Lebanon, Sab-

bah *et al.* (2003) reported that the two factor solution accounted for 63% of the total variance. This is slightly lower than was found in 10 industrialised countries 68-74% (Gandek *et al.*, 1998), although it satisfied the threshold criterion of ≥ 60%. Evaluation of inter-scale correlations helped to demonstrate that each scale measures a distinct concept of health. This reinforces the multi-dimensional nature of health (Macran, Clarke, Joshi, 1996; Wynia *et al.*, 2009).

The study also revealed that PF, BP, GH and RP correlated highly with PCS while MH and RE loaded heavily on the MCS.

It is worth pointing out that the correlation between social functioning scale and MCS in this study was lower than hypothesized by the SF-36 developers. Secondly, although both MH and RE loads strongly onto the MCS, this study found the correlation between RE and MCS to be slightly higher. This finding is consistent with the developers' hypothesis that PF leans strongest towards PCS. On the contrary, they specified that MH tends strongly towards MCS, but in the case of Ghana, RE loaded most strongly.

The study is not without some limitations. It is restricted to women and is urban-centred whereas almost half the population lives in rural areas. Accra has become a multi-ethnic city, however, with different language groups and a wide range of socio-economic classes of people and different age cohorts, all of which were well represented in the survey. It should be noted that other studies suggest that women assess their health status less well than men do. Focusing on women in urban districts gives a fair idea of the suitability of SF-36 items. The strong effect of age on both the physical and mental health norm-based scores (Figure 1) underlines the importance of correctly recording age at survey and using standardized age groups for national and international comparisons. Age effects differ between the two sets of health measures both in Accra and elsewhere.

Note from Figure 1 that the norm-based numerical scores both for the PCS and the MCS domains are above the 50% thresholds for younger women, indicating good health amongst this age group but with steep declines in the PCS measures beyond age 60. Generally, the scales are normed to produce a 50% score on the eight health domains, representing the attained values in a «normal» or reference population. We recommend further testing of this tool in the entire country, and for both sexes, to provide a more comprehensive assessment of morbidity among the Ghanaian population.

FIGURE 1   Norm-based scores on eight health domains
by age for Accra women 2008/9
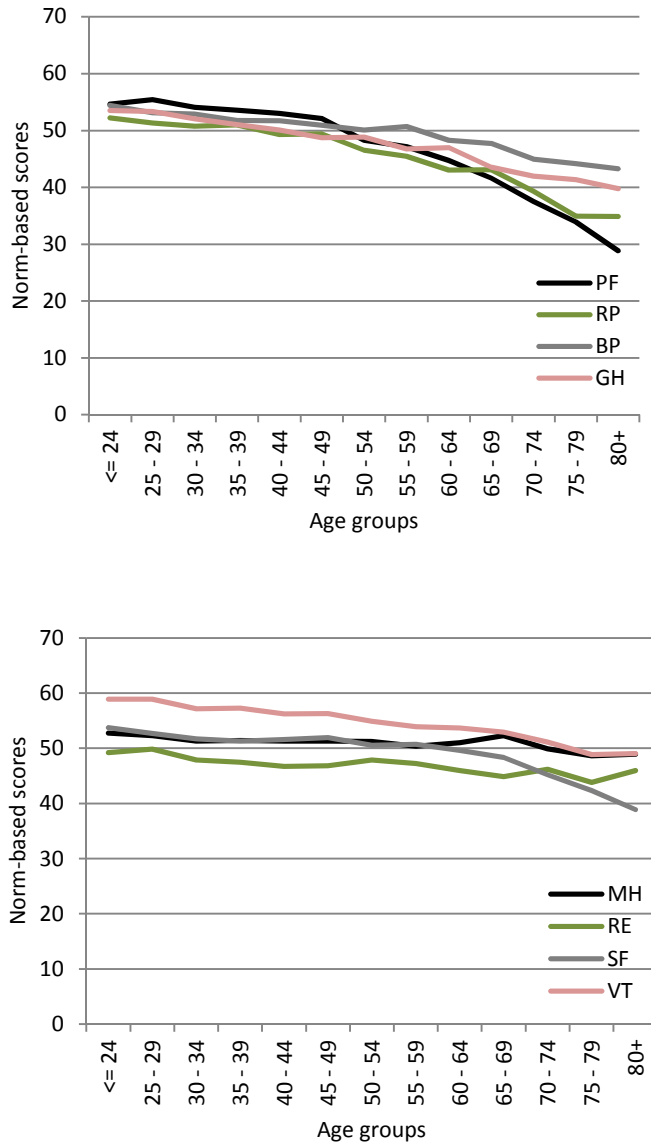
Key: PF = Physical Functioning; RP = Role limitation Physical; BP = Bodily Pain; GH = General Health; MH = Mental Health; RE = Role limitation Emotion; SF = Social Functioning; VT = Vitality.

Conclusions

The study has demonstrated the validity of SF-36 data collected in the second wave of the WHSA study. The items and scales were shown to be psychometrically sound, with good internal consistency and stability, achieving good divergence validity and very high reliability. The results further support the underlying two-factor structure with factors related to the Physical and Mental Component Scores. The pattern of factor loadings, explained variance and the profile of the scale means were consistent with previous studies conducted in other areas. It can be concluded that the assumptions underlying the SF-36 evaluation are satisfied and should be therefore suitable for assessment of the health status among women in urban areas of Ghana.

Acknowledgements

References

AGYEI-MENSAH S., **DE-GRAFT AIKINS A.** (2010), «Epidemiological Transition and the Double Burden of Disease in Accra, Ghana», *Journal of Urban Health: Bulletin of the New York Academy of Medicine,* 87 (5), pp. 879-897, http://dx.doi.org/10.1007/s11524-010-9492-y.

AGYEPONG I. A., **ADJEI S.** (2008), «Public Social Policy Development and Implementation: A Case Study of the Ghana National Health Insurance Scheme», *Health Policy and Planning*, 23, pp. 150-160, http://dx.doi.org/10.1093/heapol/czn002.

ARKU G., **LUGINAAH I.**, **MKANDAWIRE P.**, **BAIDEN P.**, **ASIEDU A. B.** (2011), «Housing and Health in Three Contrasting Neighbourhoods in Accra, Ghana», *Social Science & Medicine,* 72, pp. 1864-1872, http://dx.doi.org/10.1016/j.socscimed.2011.03.023.

BURHOLT V., **NASH P.** (2011), «Short Form-36 (SF-36) Health Survey Questionnaire: Normative Data for Wales», *Journal of Public Health*, 17.

BUTTERWORTH P., **CROSIER T.** (2004), «The Validity of the SF-36 in an Australian National Household Survey: Demonstrating the Applicability of the Household Income and Labour Dynamics in Australia (HILDA) Survey to Examination of Health Inequalities», *BMC Public Health*, http://dx.doi.org/10.1186/1471-2458-4-44.

CHEN H., **COHEN P.**, **KASEN S.** (2007), «Cohort Differences in Self-Rated Health: Evidence from a Three-Decade, Community-Based, Longitudinal Study of Wom», *American Journal of Epidemiology,* 166 (4), pp. 439-446, http://dx.doi.org/10.1093/aje/kwm100.

CHEN T.-H., **LI L.**, **KOCHEN M. M.** (2005), «A Systematic Review: How to Choose Appropriate Health-Related Quality of Life (HRQOL) Measures in Routine General Practice?», *Journal of Zhejiang University SCIENCE,* 6B (9), pp. 936-940, http://dx.doi.org/10.1631/jzus.2005.B0936.

DARKO R., **ADANU R.**, **DUDA R.**, **DOUPTCHEVA N.**, **HILL A.** (2012), «The Health of Adult Women in Accra, Ghana: Self-Reporting and Objective Assessments 2008-2009», *Ghana Medical Journal,* 46 (2), pp. 50-57.

DE-GRAFT AIKINS A. (2007), «Ghana's Neglected Chronic Disease Epidemic: A Developmental Challenge», *Ghana Medical Journal,* 41 (4), pp. 154-159.

DEBPUUR C., **WELAGA P.**, **WAK G.**, **HODGSON A.** (2010), «Self-Reported Health and Functional Limitations Among Older People in the Kassena-Nankana District, Ghana», *Global Health Action Supplement,* 2, pp. 54-63, http://dx.doi.ord/10.3402/gha.v3i0.2151.

DEMENIL V., **OSEI A.**, **DOUPTCHEVA N.**, **HILL A. G.**, **YARO P.**, **DE-GRAFT AIKINS A.** (2012), «Symptoms of Common Mental Disorders and their Correlates Among Women in Accra, Ghana: A Population-Based Survey», *Ghana Medical Journal,* 46 (2), pp. 95-103.

DUDA R. B., HILL A. G. (2005), «The Burden of Co-Morbid Illness in Women Residing in Accra, Ghana», *American Journal of Tropical Medicine and Hygiene,* 73 (6), pp. 60-60, 180 suppl.

EKPENYONG C. E., UDOKANG N. E., AKPAN E. E., SAMSON T. K. (2012), «Double Burden, Non-Communicable Diseases and Risk Factors Evaluation in sub-Saharan Africa:The Nigerian Experience», *European Journal of Sustainable Development,* 1 (2), pp. 249-270.

FUKUHARA S., WARE J. E. JR., KOSINSKI M., WADA S., GANDEK B. (1998), «Psychometric and Clinical Tests of Validity of the Japanese SF-36 Health Survey», *Clinical Epidemiol,* 51 (11), pp. 1045-1053.

GANDEK B., WARE J. E. JR., AARONSON N. K., ALONSO J., FUKUHARA S., KAASA S., LEPLEGE A., SULLIVAN M. (1998), «Tests of Data Quality, Scaling Assumptions, and Reliability of the SF-36 in Eleven Countries: Results from the IQOLA Project», *Clinical Epidemiol* 51 (11), pp. 1149-1158, http://dx.doi.org/10.1016/S0895-4356(98)00106-1.

GARRATT A., SCHMEDIT L., MACKINTOSH A., FITZPATRICK R. (2002), «Quality of Life Measurement: Bibliographic Study of Patients Assessed Health Outcome Measures», *British Medical Journal,* 324, pp. 1417-1419, http://dx.doi.org/10.1136/bmj.324.7351.1417.

Ghana Statistical Service (2005), *Population Data Analysis Report: Socioeconomic and Demographic Trends*, Accra, Ghana Statistical Service.

Ghana Statistical Service (2008), *Ghana Living Standards Survey: Report of the Fifth Round (GLSS 5), 145*, Accra, Ghana Statistical Service.

Ghana Statistical Service (2012), *2010 Population and Housing Census: Summary Report of Final Results, 103*, Accra, Ghana Statistical Service.

HILL A. G., DARKO R., SEFFAH J., ADANU R. M. K., ANARFI J. K., DUDA R. B. (2007), «Health of Urban Ghanaian Women as Identified by the Women's Health Study of Accra», *International Journal of Gynecology & Obstetrics,* 99 (2), pp. 150-156, http://dx.doi.org/10.1016/j.ijgo.2007.05.024.

HUNT S. M., MCEWEN J., MCKENNA S. P. (1986), *Measuring health status*. Great Britain, Croom Helm Ltd.

JENKINSON C., WRIGHT L., COULTER A. (1994), «Criterion Validity and Reliability of the SF-36 in a Population Sample», *Quality of Life Research,* 3 (1), pp. 7-12, http://dx.doi.org/10.1007/BF00647843.

KILAMA W. L. (2009), «The 10/90 Gap in Sub-Saharan Africa: Resolving Inequities in Health Research», *Acta Tropica,* pp. 112S: S8-S15.

KIND P., GUDEX C. (1991), «The HQM: Measuring Health Status in the Community», WOODCOCK, PEARSON (eds), U. o., York.

KUHN R., RAHMAN O., MENKEN J. (2004), *Relating Self-Reported and Objective Health Indicators to Adult Mortality in Bangladesh*, Institute of Behavioral Science, Population and Aging Center.

LOHR K. N. (1988), «Outcome Measurement: Concepts and Questions», *Inquiry,* 25 (1), pp. 37-50.

LYNCH S. M. (2003), «Cohort and Life-Course Patterns in the Relationship Between Education and Health: A Hierarchical Approach», *Demography,* 40 (2), pp. 309-331, http://dx.doi.org/10.1353/dem.2003.0016.

MACRAN S., CLARKE L., JOSHI H. (1996), «Women's Health: Dimensions and Differentials», *Social Science & Medicine,* 42 (9), pp. 1203-1216, http://dx.doi.org/10.1016/0277-9536(95)00432-7.

MCFADDEN E., LUBEN R., BINGHAM S., WAREHAM N., KINMONTH A.-L., KHAW K.-T. (2009), «Does the Association Between Self-Rated Health and Mortality Vary by Social Class?», *Social Science & Medicine,* 68 (2), pp. 275-280, http://dx.doi.org/10.1016/j.socscimed.2008.10.012.

MCINERNEY P. A., NCAMA B. P., WANTLAND D., BHENGU B. R., MCGIBBON C., DAVIS S. M., CORLESS I. B., NICHOLAS P. K. (2008), «Quality of Life and Physical Functioning in HIV-Infected Individuals Receiving Antiretroviral Therapy in KwaZulu-Natal, South Africa», *Nursing and Health Sciences,* 10, pp. 266-272, http://dx.doi.org/10.1111/j.1442-2018.2008.00410.x.

MISHRA G. D., BALL K., DOBSON A. J., BYLES J. E. (2004), «Do Socioeconomic Gradient in Women's Health Widen Over Time and With Age?», *Social Science & Medicine,* 58, pp. 1585-1595, http://dx.doi.org/10.1016/S0277-9536(03)00368-X.

MOORE M., GOULD P., KEARY B. S. (2003), «Global Urbanization and Impact on Health», *International Journal of Hygiene and Environmental Health,* 206, pp. 269-278, http://dx.doi.org/10.1078/1438-4639-00223.

MULLER-NORDHORN J., ROLL S., WILLICH S. N. (2004), «Comparison of the Short Form (SF) – 12 Health Status Instrument with the SF-36 in Patients with Coronary Heart Disease», *Heart,* 90 (5), pp. 523-527, http://dx.doi.org/10.1136/hrt.2003.013995.

MWANYANGALA M. A., MAYOMBANA C., URASSA H., CHARLES J., MAHUTANGA C., ABDULLAH S., NATHAN R. (2010), «Health Status and Quality of Life Among Older Adults in Rural Tanzania», *Global Health Action Supplement 2*, pp. 36-44, http://dx.doi.org/10.3402/gha.v3i0. 2142.

O'DONNELL O., VAN DOORSLAER E., WAGSTAFF A., LINDELOW M. (2008), Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and their Implementation.

QU B., GUO H. Q., LIU J., ZHANG Y., SUN G. (2009), «Reliability and Validity Testing of the SF-36 Questionnaire for the Evaluation of the Quality of Life of Chinese Urban Construction Workers», *The Journal of International Medical Research,* 37, pp. 1184-1190, http://dx.doi.org/10.1177/147323000903700425.

QualityMetrics (2011), *2011 Products & services catalog*, Q. M. INCORPORATED (ed), United States of America, 74.

28

Reliability, validity and responsiveness of the short form-36 health survey: Findings from the women's health study of Accra, Ghana

QUESNEL-VALLÉE A. (2007), «Self-Rated Health: Caught in the Crossfire of the Quest for 'True' Health», *International Journal of Epidemiology,* 36, pp. 1161-1164, http://dx.doi.org/10.1093/ije/dym236.

RILEY J. C. (2001), *Rising Life Expectancy a Global History*, United States of America, Cambridge University Press.

RUTA D. A., ABDALLA M. I., GARRATT A. M., COUTTS A., RUSSELL I. T. (1994), «SF 36 Health Survey Questionnaire: I. Reliability in two Patient Studies», *Quality in Health Care,* 3 (4), pp. 180-185, http://dx.doi.org/10.1136/qshc.3.4.180.

SABBAH I., DROUBY N., SABBAH S., RETEL-RUDE N., MERCIER M. (2003), «Quality of Life in Rural and Urban Populations in Lebanon Using SF-36 Health Survey», *BioMed Central,* 1 (30), http://dx.doi.org/10.1186/1477-7525-1-30.

SONGSORE J. (2003), *Regional Development in Ghana: The Theory and the Reality*, Accra, Woeli Publishing Services.

SULLIVAN M., KARLSSON J., WARE J. E. JR. (1995), «The Swedish SF-36 Health Survey-I Evaluation of Data Quality, Scaling Assumptions, Reliability and Construct Validity Across General Populations», *Social Science & Medicine,* 41 (10), pp. 1349-1358, http://dx.doi.org/10.1016/0277-9536(95)00125-Q.

The WHOQOL Group (1998), «The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties», *Social Science & Medicine,* 46 (12), pp. 1569-1585, http://dx.doi.org/10.1016/S0277-9536(98)00009-4.

The WHSA writing team (2011), «Women's Health Study of Accra Wave II (WHSA-II)», *ISSER Technical Publication Series*, Accra, ISSER, 112.

WANYONYI S. Z., STONES W., SEQUEIRA E. (2010), «Utility of the Short Form (SF)-36 Health Related Quality of Life Questionnaire as a Measure of Outcome Among Contraceptive Users», *Journal of Obstetrics and Gynaecology of Eastern and Central Africa,* 23 (1), pp. 14-18.

WARE J. E. JR., GANDEK B. (1998a), «Methods for Testing Data Quality, Scaling Assumptions and Reliability: The IQOLA Project Approach», *Clinical Epidemiol,* 51 (11), pp. 945-952, http://dx.doi.org/10.1016/S0895-4356(98)00085-7.

WARE J. E. JR., GANDEK B. (1998b), «Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project», *Clinical Epidemiol,* 51 (11), pp. 903-912, http://dx.doi.org/10.1016/S0895-4356(98)00081-X.

WARE J. E., SHERBOURNE C. D. (1992), «The MOS 36-Item Short-Form Health Survey: I. Concept and Framework and Item Selection», *Medical Care,* 30 (6), pp. 473-483, http://dx.doi.org/10.1097/00005650-199206000-00002.

WYNIA K., **MIDDEL B.**, **DE RUITER H.**, **VAN DIJK J. P.**, **LOK W. S.**, **DE KEYSER J. H. A.**, **REIJNEVELD S. A.** (2009), «Adding a Subjective Dimension to an ICF-Based Disability Measure for People with Multiple Sclerosis: Development and Use of a Measure for Perception of Disabilities», *Disability and Rehabilitation,* 31 (12), pp. 1008-1017, http://dx.doi.org/10.1080/09638280802509488.

WYSS K., **WAGNER A. K.**, **WHITING D.**, **MTASIWA D. M.**, **TANNER M.**, **GANDEK B.**, **KILIMA P. M.** (1999), «Validation of the Kiswahili Version of the SF-36 Health Survey in a Representative Sample of an Urban Population in Tanzania», *Quality of Life Research,* 8, pp. 111-120.

ZHANG Y., **QU B.**, **LUN S.**, **GUO Y.**, **LIU J.** (2012), «The 36-Item Short Form Health Survey: Reliability and Validity in Chinese Medical Students», *International Journal of Medical Sciences,* 9 (7), pp. 521-526, http://dx.doi.org/10.7150/ijms.4503.

ZHOU K. N., **ZHANG M.**, **WU Q.**, **JI Z. H.**, **ZHANG X. M.**, **ZHUANG G. H.** (2013), «Reliability, Validity and Sensitivity of the Chinese (Simple) Short Form-36 Health Survey Version 2 (SF-36v2) in Patients with Chronic Hepatitis B», *Journal of Viral Hepatitis,* 20, pp. e47-e55, http://dx.doi.org/10.1111/jvh.12030.