



Published in final edited form as:

Qual Life Res. 2010 November ; 19(9): 1323–1331. doi:10.1007/s11136-010-9691-8.

## Confirmatory Factor Analysis of the Validity of the SF-12 for Persons with and without a History of Stroke

**Ozioma C. Okonkwo, PhD,**

Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD

**David L. Roth, PhD,**

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL

**LeaVonne Pulley, PhD, and**

Department of Health Behavior and Health Education, University of Arkansas for Medical Sciences, Little Rock, AK

**George Howard, DrPH**

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL

### Abstract

**Purpose**—To assess the validity of the Physical and Mental Component Summary scores (PCS and MCS) of the 12-item Short Form Health Survey (SF-12), a measure of health-related quality of life (HRQoL), among persons with a history of stroke.

**Methods**—Persons with ( $n = 2,581$ ) and without ( $n = 38,066$ ) a reported history of stroke were enrolled in the REasons for Geographic And Racial Differences in Stroke (REGARDS) study. Confirmatory factor analysis methods were used to evaluate the fit of a 2-factor model that underlies the PCS and MCS and to examine the equivalence of the factors across both study groups.

**Results**—The 2-factor model provided good fit to the data among individuals with and those without a self-reported history of stroke. Item factor loadings were found to be largely invariant across both groups, and correlational analyses confirmed that the two latent factors were highly related to the PCS and MCS scores, calculated by the standard scoring algorithms. The effect of stroke history on physical health was more than twice its effect on mental health.

**Conclusions**—The psychometric measurement model that underlies the PCS and MCS summary scores is comparable between persons with and without a history of stroke. This suggests that the SF-12 has adequate validity for measuring HRQoL not only in the general population, but also in cohorts following stroke.

### Keywords

Stroke; SF-12; psychometrics; confirmatory factor analysis; factor invariance

### Introduction

Health-related quality of life (HRQoL) is considered a major outcome indicator especially in chronic diseases, such as stroke, for which a complete recovery is often not observed [1]. As a result, health care providers are paying more attention to patients' perceptions of their health

status and the effectiveness of the treatments they are receiving. Self-report measures of HRQoL are increasingly being incorporated into clinical practices and research studies, but the validity of these measures for specific patient populations is often unexamined, and may be questionable.

A growing number of studies use the 12-item Short-Form Health Survey [SF-12; 2] to assess HRQoL among patient populations. This includes persons with a history of stroke or stroke symptoms [3-6]. The SF-12 is a shortened version of the SF-36 [7], a widely used self-report instrument for measuring HRQoL. Like the SF-36, scores on the SF-12 are summarized by two composite scores: a physical component summary (PCS) and a mental component summary (MCS). Previous studies have reported that the PCS and MCS from either the SF-36 or SF-12 have satisfactory reliability and validity among stroke patients [3,8-12]. However, in an extensive study by Hobart and colleagues [13], the SF-36 was administered to 177 patients with a history of ischemic stroke and the item responses from this instrument were subjected to rigorous psychometric evaluation. Using a scorecard approach that classified whether the SF-36 items and scales met 8 scaling assumptions, these investigators concluded that summary scores generated from the SF-36 and, by extension, the SF-12 have “limited validity as outcome measures after stroke (p. 1348).”

Limitations of many of the previous validity studies, including the one by Hobart and colleagues [13], include the use of relatively small convenience samples identified from hospitals or rehabilitation programs and the failure to directly compare the psychometric properties of the SF-12 (or the SF-36) among persons with a history of stroke vis-à-vis persons without such history. Furthermore, although the PCS and MCS are obtained from an algorithm based on previous factor analysis results, to our knowledge no previous studies have formally examined the underlying factor structure of the SF-12 items among persons with a history of stroke. Confirmatory factor analysis (CFA) is well-suited for this purpose for several reasons [14-18]. First, CFA can evaluate the fit of competing factor models in terms of how well the item-level variances and covariances are explained by underlying latent factors. Secondly, it can determine whether a specific factor model provides comparable fit to the item-level data from multiple samples. Thirdly, CFA can be used to provide powerful model-based tests of differential item functioning (DIF) between two or more groups [15]. With its connections to item response theory, CFA is superior for evaluating the psychometric properties of items as contributors to composite summary scores across multiple groups than subjective scorecard approaches that rely on simple counts of the number of items that pass arbitrary scaling criteria.

In the present paper, we used CFA to directly examine the validity of the PCS and MCS from the SF-12 in persons with and without a history of stroke using data from a large national epidemiologic study.

## Method

### Participants

The analyses were based on data from the REasons for Geographic And Racial Differences in Stroke (REGARDS) study, a national, epidemiological study of adults 45 years and older in the United States [19]. The purpose of REGARDS is to determine the causes for the excess stroke incidence and mortality in the Southeastern United States and among African Americans. Recruitment began in January of 2003 and ended in October of 2007. A stratified sampling method was used such that approximately 50% of the participants were from the Stroke Belt region and 50% from the other regions of the United States. Within each geographic designation, the sampling design called for the study cohort to be approximately 50% African American and 50% White, 50% male and 50% female. The analyses reported in this paper are based on the 40,647 participants who completed the SF-12 as part of their baseline telephone

interview. This sample includes 2,581 persons who reported a history of stroke and 38,066 persons who did not.

### Procedures and Measures

Participants were recruited via mail and telephone contacts from a commercially available nationwide list of households that was purchased through Genesys, Inc. The list was then stratified by geographic region, and the demographic stratifying and eligibility variables were then assessed during the initial telephone contact. For eligible and consented participants, other demographic characteristics, health behaviors, history of cardiovascular procedures, history of stroke, psychosocial factors, general health, and cognitive function were collected through a computer-assisted telephone interview. A subsequent in-home visit was scheduled, and written informed consent, blood and urine samples, electrocardiogram, and blood pressure measures were obtained from the home visit.

Additional methodological details concerning the REGARDS study are available elsewhere [19,20]. This study and all of its recruitment, informed consent, and data collection procedures were approved by the Institutional Review Board (IRB) of the University of Alabama at Birmingham (Protocol # F020925004) and by the IRB of each participating institution.

During the telephone interview, participants provided information regarding their age, gender, race, education, income, marital status, and living arrangement. The participant's self-reported history of stroke was assessed with the interview question, "Were you ever told by a physician that you had a stroke?" Those who answered "Yes" were classified as having a self-reported history of stroke.

Information on HRQoL was obtained using version 1 of the SF-12. The PCS and MCS scores were obtained using the standard scoring algorithm provided by the instrument's developers [2]. The 4-item version of the Center for Epidemiological Studies-Depression scale (CESD-4) was used to screen for depressive symptoms [21]. The CESD-4 has been shown to have sufficient reliability and validity in comparison with full 20-item CESD [21]. In addition to history of stroke, participants were also asked whether they had ever experienced a heart attack/myocardial infarction or ever been diagnosed with hypertension or diabetes. The number of disease conditions endorsed (range = 0-4) was used as an index of disease comorbidity (DIS).

### Data analysis

Cronbach's  $\alpha$  coefficients [22] were calculated to estimate the internal consistencies of the SF-12's PCS and MCS in this cohort. Instruments with Cronbach's  $\alpha$  of .70 or greater are considered to have satisfactory internal consistency [23]. Correlations among the PCS, MCS, CESD-4, and DIS were computed to ascertain the construct validity of the PCS and MCS in our sample.

Next, confirmatory factor analysis methods were applied to the raw item data to investigate the factor structure of the SF-12 and to compare factor solutions for persons with and without a history of stroke. CFA is a form of structural equation modeling that is used to evaluate the fit of proposed measurement models [24]. Two measurement models were tested: a 1-factor model that forced all SF-12 items to load on a single latent variable; and a 2-factor model that specified the five physical functioning, role-physical, and bodily pain items to load on a physical health factor only, the four role-emotional and mental health items to load on a mental health factor only, and the three items assessing general health, vitality, and social functioning to load on both the physical and mental health factors. The physical and mental health factors were allowed to be correlated in the 2-factor model. In addition, for both the 1- and 2-factor models, a correlated residual was estimated between the two physical functioning items that

ask about the extent to which health limits participation in “moderate activities” and “climbing several flights of stairs.” Previous CFA studies on the SF-12 have allowed for this correlated residual due to the similarity in the wording of these two items and their response options [25,26].

All CFA models were estimated and tested using version 5.1 of the Mplus analysis system [27]. Because the SF-12 items are measured on an ordinal scale with a limited number of response options, weighted least squares estimation (WLSMV option in Mplus) was used to model the observed polychoric correlation matrix (CATEGORICAL option in Mplus). The factors were identified by fixing the factor loading for item 4 (on the physical health factor) and item 6 (on the mental health factor) to be 1. Chi-square goodness-of-fit tests that compare the observed data to the data that can be reproduced on the basis of the model were used to assess the fit of each model. The degrees of freedom for these chi-square tests were calculated from the empirical data using the algorithm associated with the WLSMV option as described by Muthén and Muthén [27]. Because the chi-square test is highly sensitive to sample size, additional fit indices were examined including the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). These fit indices take into account overall model fit and the complexity of the model. The CFI ranges from 0 to 1, with higher values representing better fit. Values above .90 and .95 are considered indicative of reasonable and excellent fit, respectively [28]. For the RMSEA, lower values represent better fit, with a value of .08 or less considered indicative of reasonable fit and .05 or less indicating very close fit to the observed data [28,29].

Nested model comparisons were conducted to examine the incremental fit of the 2-factor model over the 1-factor model and to examine the invariance of these factor models across participants with and without a history of stroke. The DIFFTEST option was used in conjunction with the WLSMV estimation method when conducting chi-square tests of the statistical significance of any improvements in fit between nested models. The invariance models were conducted using a multiple group CFA approach in which separate models were run that first constrained the factor loadings to be equal across stroke history and non-stroke history groups and then allowed these loadings to be different between the two groups. In both types of models, the item thresholds were constrained to be equal across the two groups and the residual variances of the items (estimated using the THETA parameterization option in Mplus) were fixed at 1 in the no-stroke history group and freely estimated in the stroke history group. Latent variable means, variances, and the covariance were estimated and allowed to differ across the two groups. Under these specifications, the factor loadings were compared in the context of assuming equal item thresholds and unequal item residual variances across the two groups. If the constrained model was found to fit the data nearly as well as the more unconstrained model, then this was considered evidence that the factor loading estimates are equivalent between the two groups under these specifications. In this case, that would mean that the sensitivity of any one of the 12 items as indicators of the underlying latent constructs is similar across the stroke and non-stroke history groups. Such a finding would support the validity of the SF-12 for measuring HRQoL after stroke events, especially if the underlying latent factors are also highly correlated with the standard PCS and MCS component summary scores. Conversely, if the unconstrained model was found to fit the data much better than the constrained model, then this would suggest that at least some of the SF-12 items are not assessing the same constructs in the same manner in both groups, and might call into question the validity of the SF-12 component summary scores for use in measuring HRQoL in one or both groups.

After fitting the SF-12 item measurement model, the correlations of the actual PCS and MCS composite scores with the latent factors were examined to determine whether the latent factors estimated with REGARDS data closely resembled the SF-12 component summary scores as calculated by the standard scoring algorithms. Differences between participants with and

without a history of stroke on the latent factors and on the component summary scores were estimated in standardized effect sizes (standard deviation units). We expected participants with a history of stroke to have significantly worse physical and mental health compared to participants without a history of stroke, and these effects would be evident on both the standard summary scores and the latent factors before and after adjusting for demographic variables.

## Results

### Descriptive characteristics

Table 1 displays demographic information for the REGARDS participants by stroke history status. Participants who reported a history of stroke were older, were more likely to be men, and were more likely to be African American compared to participants who did not report a history of stroke.

### Reliability and construct validity

The SF-12's PCS and MCS demonstrated good internal consistencies within the entire sample, within subjects with a history of stroke, and within subjects without a history of stroke. Specifically, Cronbach's  $\alpha$  for the PCS was .868 in the entire sample, .845 in the stroke-history group, and .866 in the non-stroke history group. For the MCS, Cronbach's  $\alpha$  was .816 in the entire sample, .812 in the stroke-history group, and .813 in the non-stroke history group. These values clearly exceed the recommended threshold of .70 for acceptable reliability [23]. Further, they show that the PCS and MCS have very similar reliabilities in both stroke history and non-stroke history groups.

Pearson's correlations among the PCS, MCS, CESD-4, and DIS within the entire sample were: -.23 (CESD-4 and PCS), -.66 (CESD-4 and MCS), -.34 (DIS and PCS), and -.10 (DIS and MCS). Within the stroke history group, these correlations were -.20 (CESD-4 and PCS), -.68 (CESD-4 and MCS), -.22 (DIS and PCS), and -.07 (DIS and MCS). Finally, within the non-stroke history group, the correlations were -.23 (CESD-4 and PCS), -.66 (CESD-4 and MCS), -.30 (DIS and PCS), and -.08 (DIS and MCS). All correlations were significant at  $p < .0001$ . Although the CESD-4 and DIS are arguably rudimentary measures, this pattern of relationships supports the construct validity—i.e., convergent and discriminant validities—of the SF-12's component summary scores.

### Confirmatory Factor Analyses of the 1- and 2-Factor Models

The CFA results for the entire sample showed that the 1-factor model with all items loading on a single “general health” latent factor showed poor overall fit to the observed data ( $\chi^2 = 25,910.06$ ,  $df = 27$ ,  $p < .0001$ , CFI = 0.865, RMSEA = 0.154). The 2-factor model that forms the basis of the PCS and MCS was found to provide much better fit ( $\chi^2 = 4,625.76$ ,  $df = 32$ ,  $p < .0001$ , CFI = 0.976, RMSEA = 0.059), and the nested comparison revealed that the 2-factor model fit significantly better than the 1-factor model ( $\chi^2 = 10,765.47$ ,  $df = 3$ ,  $p < .0001$ ). The RMSEA for the 2-factor model suggested that some minor improvements in fit might still be possible, but overall very good to excellent fit was observed for this model.

The standardized factor loadings for the 2-factor model are displayed in Table 2. All loadings equal to 0 were fixed to be 0 in the model identification process. All non-zero loadings in the total sample analysis were significantly different from zero ( $p < .0001$ ). A moderate positive correlation of 0.54, which was significantly greater than zero ( $p < .0001$ ), was found between the physical health latent factor (PHLF) and the mental health latent factor (MHLF). An alternative 2-factor model that forced PHLF and MHLF to be completely independent and uncorrelated did not provide good fit to the observed data ( $\chi^2 = 26,132.00$ ,  $df = 15$ ,  $p < .0001$ , CFI = 0.864, RMSEA = 0.207). The 2-factor model with the moderate correlation between the

PHLF and MHLF was therefore adopted as the optimal SF-12 measurement model and used in subsequent analyses to examine the invariance of the SF-12 items as indicators of these latent factors across the stroke history and non-stroke history groups.

### Multiple Group Invariance Analyses

The multiple group CFA model with the factor loadings constrained to be equal across stroke history and non-stroke history groups provided excellent overall fit to the observed data ( $\chi^2 = 4003.74$ ,  $df = 84$ ,  $p < .0001$ ,  $CFI = 0.980$ ,  $RMSEA = 0.048$ ). When the equality constraints on the factor loadings were relaxed in the unconstrained model, a statistically significant improvement in fit was observed in the nested model comparison ( $\chi^2 = 155.94$ ,  $df = 11$ ,  $p < .0001$ ), but the overall model fit statistics were worse for this unconstrained model ( $\chi^2 = 4261.66$ ,  $df = 83$ ,  $p < .0001$ ,  $CFI = 0.977$ ,  $RMSEA = 0.052$ ). In essence, both the RMSEA and CFI (which take into account both statistical fit and the complexity of a model) were slightly better for the constrained model, suggesting that the statistical improvements in fit for the unconstrained model were rather minimal and trivial when weighed against its increased complexity.

The standardized factor loadings for the unconstrained multiple group 2-factor model are presented in Table 2. All non-zero factor loadings from the multiple group analysis were significantly different from zero at the  $p < .0001$  level except for the loading of the general health item on the MHLF for the stroke history group, which was more modest but still significantly different from zero ( $p = .04$ ). An inspection of these loadings revealed no loading differences greater than 0.11 between stroke history and non-stroke history groups. In fact, 10 of the 15 factor loadings differed by 0.05 or less between the two groups, and only one of the 15 loadings differed by  $\sim 0.10$  between the two groups. However, supplemental DIF analyses (results not shown) indicated that many of these rather small loading differences were statistically significant ( $ps < .01$ ), with the loadings for the stroke history group tending to be slightly lower. The statistical significance of these loading differences was likely driven by our large sample size, as opposed to indicating meaningful between-group differences. The standardized correlation between the PHLF and MHLF was also very similar across groups, ranging from 0.544 in those with a history of stroke to 0.535 in those without a history of stroke.

Given the similar fit statistics of the constrained and unconstrained models and the small differences in the factor loadings when unconstrained estimates were allowed, these findings support an interpretation of “approximate invariance” [30] such that the association between any one of the SF-12 items and the underlying latent factors is largely equivalent across the stroke and non-stroke history groups.

### Associations between the SF-12 Latent Variables and Component Summary Scores

Even though the 2-factor model resembled the structure for calculating the standard PCS and MCS scores from the SF-12, the factor loadings for the PHLF and the MHLF from this analysis are specific to the REGARDS sample and may not correspond well with the weights used in the standard calculation of the PCS and MCS. The CFA models were therefore extended to add additional observed variables and examine the correlations between the latent factors and the component summary scores, and to compare participants with and without a history of stroke on both the latent factors and component summary scores.

The standardized correlations between the REGARDS-specific latent factors, the calculated component summary scores, and participant age are listed in Table 3. All correlations are significantly different from zero ( $ps < .001$ ). As expected, the PHLF was highly correlated with the PCS ( $r = 0.97$ ) and the MHLF was highly correlated with the MCS ( $r = 0.96$ ), indicating that the REGARDS-specific latent factors closely resembled the standard component summary

scores, with the notable exception that the latent factors are moderately correlated with each other ( $r = .54$ ) whereas the component summary scores are not ( $r = .08$ ). Both latent factors and the component summary scores showed very similar correlations with participant age, with older participants reported slightly worse physical health but better mental health than younger adults.

Comparisons between participants with and without a history of stroke on the latent factors and the component summary scores are summarized in Table 4. In these models, the effects represent the differences between the two groups in standard deviation units. For example, in the entire sample, the standard deviation for PCS was 10.61. The product of 10.61 and -0.70 (i.e., the unadjusted estimate for PCS in Table 4) is -7.43. Thus, the mean PCS score of the stroke history group is approximately 7.43 points lower than that of the non-stroke history group, as reflected in the means in Table 1 (within rounding error). Effects in Table 4 are presented both before and after adjusting for age, race, and gender effects. As expected, participants with a history of stroke reported significantly more physical and mental health problems than participants without a history of stroke, and these effects were highly significant statistically both before and after adjusting for demographic covariates. The physical health differences (PHLF and PCS) were notably stronger than the mental health differences (MHLF and MCS), and the REGARDS-specific latent factors (PHLF and MHLF) tended to yield somewhat stronger group differences than their respective component summary scores (PCS and MCS, respectively).

## Discussion

The results of this study indicate that the measurement properties of the SF-12 items are remarkably similar for persons with and without a history of stroke in the context of a 2-factor measurement model that postulated separate but correlated physical and mental health factors. To our knowledge, this is the first study that has examined the factor structure of the SF-12 items among persons with a history of stroke using CFA. A prior study that investigated the “factor structure” of the SF-12 in stroke patients used a more exploratory principal components analysis approach [10], and other researchers who have examined the SF-12’s measurement properties among other patient populations have also utilized more exploratory factor analysis procedures [31-33].

The excellent fit of the 2-factor model, the similarity of the estimated factor loadings across persons with and without a history of stroke, and the close correspondence between the sample-specific latent factors and the PCS and MCS scores as calculated with the standard scoring algorithms combine to strongly support the validity of the SF-12 and its component summary scores for assessing general HRQoL in research and clinical work with persons who report a history of stroke. Presumably this would also include those who are still recovering from recent stroke events. These findings are particularly noteworthy in light of the concerns raised by Hobart and colleagues [13], who characterized the component summary scores provided by the SF-36 and, by extension, the SF-12, as having limited validity as measures of health outcomes following stroke. Our results demonstrate that SF-12 items have very similar psychometric scaling properties in the context of this 2-factor model for persons with and without a self-reported history of stroke, thus providing support for the validity of the SF-12 composite scores for measuring HRQoL in both groups. Other studies have also reported satisfactory performance of the SF-12 as an index of HRQoL among stroke patients [3-5,10, 34]. Relatively brief measures of HRQoL are needed in many situations, particularly in large-scale epidemiologic studies such as REGARDS where instrument length and ease of administration are key considerations [2,3]. It is encouraging that, among persons with a history of stroke, a popular and standardized instrument such as the SF-12 appears to have psychometric properties that are comparable to those observed in the general population.

A moderate correlation was observed between the physical health and mental health latent factors, indicating that overall physical health is not completely independent of mental health. This moderate correlation is inconsistent with the near zero correlation between the SF-12 PCS and MCS scores, which is the result of a scoring algorithm based on an orthogonal factor rotation procedure [2]. Results from prior studies have also questioned the forced orthogonality of the SF-12 and SF-36 summary scores [35-38]. This artificial orthogonality of the PCS and MCS scores may undermine the sensitivity of these summary scores to capture veritable improvements in physical and mental health over time. Others have also shown that correlated PCS and MCS summary scores more accurately reflect the underlying raw data compared to orthogonal summary scores [26,39,40]. Interestingly, Wilson and colleagues [39] demonstrated that, of three factor extraction methods tested, SF-36 summary scores derived from a confirmatory factor analysis with correlated physical and mental health factors provided the best fit to the data.

Comparisons between persons with and without a history of stroke revealed that those with a history of stroke reported poorer physical and mental health compared to those without this self-reported history. These effects were found on both the REGARDS-specific latent factors and the standard SF-12 component summary scores. Both latent and component summary measures estimated the impact of stroke on physical functioning to be about twice as large as its impact on mental functioning. Prior investigators have similarly found evidence of diminished physical and mental health and HRQoL among persons who have experienced a stroke. For example, Carod-Artal and colleagues [41] reported that stroke patients continued to report significant levels of depression, restriction in psychosocial functioning, and dependence in activities of daily living one year post stroke. Other studies have documented similar deficits as far out as four years post stroke [42]. Previous analyses of REGARDS data have shown that even persons who report symptoms suggestive of stroke but deny a history of stroke or any transient ischemic attacks (included as persons without a history of stroke in the present analyses) report poorer HRQoL compared to symptom-free persons [6]. Taken together, these findings suggest that even relatively mild strokes can have adverse effects on a person's physical and mental health, including emotional difficulties caused by the awareness of an increased risk for future strokes [43].

Limitations of the present study include the reliance on self-report data for a positive history of stroke. Even so, prior epidemiological studies have demonstrated the validity of self-report in establishing a history of stroke [44]. The SF-12 was only administered on a single occasion in this study, so we were unable to examine other potentially important SF-12 psychometric properties such as test-retest reliability or sensitivity to change over time. However, results from other studies suggest that the SF-12 summary scores have satisfactory test-retest reliabilities among stroke patients [4,9,10]. We also acknowledge that, because respondents were non-institutionalized and appeared competent to provide informed consent, these findings may not be generalizable to all stroke survivors. Finally, we note that because the development model fit statistics have largely occurred within the context of parametric maximum likelihood estimations, their application to non-parametric ordinal models should be made cautiously. Aside from these limitations, we note that this paper is unique in several ways. It was conducted with a large national sample of persons who were not selected because of their history of stroke or stroke-related impairments and includes large numbers of participants with and without a self-reported history of stroke. This contrasts with prior studies that have largely relied on convenience samples from medical clinics or rehabilitation programs to examine questions concerning HRQoL following stroke.

In summary, this is the first study to our knowledge that has used CFA methods to examine the hypothesized 2-factor structure of the SF-12 items for persons who report a history of stroke. We demonstrated that this 2-factor model provides very good fit to the observed data and shows



remarkable similarity across persons with and without a history of stroke. These findings support the validity of the SF-12 summary scores for assessing general HRQoL in persons with a history of stroke, and the use of this relatively brief instrument could yield useful data, especially in large sample observational studies where questionnaire administration time and reduction of research burden are important considerations. Finally, our findings argue that, to be valid, an evaluation of an instrument's psychometric properties must be driven by adequate statistical methodology, include an appropriate comparison group, and use sufficiently large samples. This is of considerable clinical, research, and public policy interest due to the potential for disservice when otherwise bona fide instruments are rejected on the basis of conclusions reached using less than adequate methodologies.

## Acknowledgments

This research project is supported by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NINDS or the National Institutes of Health. Representatives of the funding agency have been involved in the review of the manuscript but not directly involved in the collection, management, analysis or interpretation of the data. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at <http://www.regardsstudy.org> Additional funding was provided by an investigator-initiated grant from NINDS (R01 NS045789, David L. Roth, PI). Representatives from the NINDS did not have any role in the design and conduct of the study, the collection, management, analysis, and interpretation of the data, or the preparation or approval of the manuscript.

## References

1. Stewart, AL.; Ware, JE. Measuring functioning and well-being: The Medical Outcomes Study approach. London: Duke University Press; 1992.
2. Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care* 1996;34:220–233. [PubMed: 8628042]
3. Lim LLY, Fisher JD. Use of the 12-item short-form (SF-12) health survey in an Australian heart and stroke population. *Quality of Life Research* 1999;8:1–8. [PubMed: 10457733]
4. Muller-Nordhorn J, Nolte CH, Rossnagel K, Jungehulsing GJ, Reich A, Roll S, et al. The use of the 12-item short-form health status instrument in a longitudinal study of patients with stroke and transient ischaemic attack. *Neuroepidemiology* 2005;24:196–202. [PubMed: 15802924]
5. Xie J, Wu EQ, Zheng ZJ, Croft JB, Greenlund KJ, Mensah GA, et al. Impact of stroke on health-related quality of life in the noninstitutionalized population in the United States. *Stroke* 2006;37:2567–2572. [PubMed: 16946158]
6. Howard G, Safford MM, Meschia JF, Moy CS, Howard VJ, Pulley L, et al. Stroke symptoms in individuals reporting no prior stroke or transient ischemic attack are associated with a decrease in indices of mental and physical functioning. *Stroke* 2007;38:2446–2452. [PubMed: 17673720]
7. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;30:473–481. [PubMed: 1593914]
8. Anderson C, Laubscher S, Burns R. Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke* 1996;27:1812–1816. [PubMed: 8841336]
9. Bohannon RW, Maljanian R, Landes M. Test-retest reliability of short form (SF)-12 component scores of patients with stroke. *International Journal of Rehabilitation Research* 2004;27:149–150. [PubMed: 15167113]
10. Bohannon RW, Maljanian R, Lee N, Ahlquist M. Measurement properties of the short form (SF)-12 applied to patients with stroke. *International Journal of Rehabilitation Research* 2004;27:151–154. [PubMed: 15167114]
11. O'Mahony PG, Rodgers H, Thomson RG, Dobson R, James OF. Is the SF-36 suitable for assessing health status of older stroke patients? *Age and Ageing* 1998;27:19–22. [PubMed: 9504362]
12. Dorman P, Dennis M, Sandercock P. How do the scores on the EuroQol relate to scores on the SF-36 after stroke? *Stroke* 1999;30:2146–2151. [PubMed: 10512920]

13. Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: Uses and abuses of the SF-36. *Stroke* 2002;33:1348–1356. [PubMed: 11988614]
14. Hatcher, L. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. Cary, NC: SAS Institute, Inc; 1994.
15. Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *Journal of Applied Psychology* 2006;91:1292–1306. [PubMed: 17100485]
16. Keller SD, Ware JE, Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology* 1998;51:1179–1188. [PubMed: 9817136]
17. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin* 1993;114:552–566. [PubMed: 8272470]
18. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 1995;7:286–299.
19. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and design. *Neuroepidemiology* 2005;25:135–143. [PubMed: 15990444]
20. Roth DL, Perkins M, Wadley VG, Temple EM, Haley WE. Family caregiving and emotional strain: Associations with quality of life in a large national sample of middle-aged and older adults. *Quality of Life Research* 2009;18:679–688. [PubMed: 19421895]
21. Melchior LA, Huba GJ, Brown VB, Reback CJ. A short depression index for women. *Educational and Psychological Measurement* 1993;53:1117–1125.
22. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
23. Nunnally, JC. *Psychometric theory*. New York: McGraw-Hill; 1978.
24. Bollen, KA. *Structural equations with latent variables*. New York: Wiley; 1989.
25. Maurischat C, Ehlebracht-König I, Kühn A, Bullinger M. Factorial validity and norm data comparison of the Short Form 12 in patients with inflammatory-rheumatic disease. *Rheumatology International* 2006;26:614–621. [PubMed: 16179999]
26. Wilson D, Tucker G, Chittleborough C. Rethinking and rescoring the SF-12. *Sozial- und Präventivmedizin* 2002;47:172–177. [PubMed: 12238299]
27. Muthén, LK.; Muthén, BO. *Mplus User's Guide*. 4th. Los Angeles, CA: Muthen & Muthen; 1998-2007.
28. Kline, RB. *Principles and practice of structural equation modeling*. 2nd. New York: Guilford; 2005.
29. Hu LT, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1–55.
30. Millsap, RE. Four unresolved problems in studies of factorial invariance. In: Maydeu-Olivares, A.; McArdle, JJ., editors. *Psychometrics: A festschrift to Roderick P McDonald*. Mahwah, NJ: Erlbaum; 2005. p. 153-171.
31. Salyers MP, Bosworth HB, Swanson JW, Lamb-Pagone J, Osher FC. Reliability and validity of the SF-12 health survey among people with severe mental illness. *Medical Care* 2000;11:1141–1150. [PubMed: 11078054]
32. Gandhi SK, Salmon JW, Zhao SZ, Lambert BL, Gore PR, C K. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clinical Therapeutics* 2001;23:1080–1098. [PubMed: 11519772]
33. Amir M, Lewin-Epstein N, Becker G, Buskila D. Psychometric properties of the SF-12 (Hebrew version) in a primary care population in Israel. *Medical Care* 2002;40:918–928. [PubMed: 12395025]
34. Pickard AS, Johnson JA, Penn A, Lau F, Noseworthy T. Replicability of SF-36 summary scores by the SF-12 in stroke patients. *Stroke* 1999;30:1213–1217. [PubMed: 10356102]
35. Johnson JA, Maddigan SL. Performance of the RAND-12 and SF-12 summary scores in type 2 diabetes. *Quality of Life Research* 2004;13:449–456. [PubMed: 15085917]
36. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? *Quality of Life Research* 2001;10:395–404. [PubMed: 11763202]

37. Simon GE, Revicki DA, Grothaus L, Vonkorff M. SF-36 summary scores: are physical and mental health truly distinct? *Medical Care* 1998;36:567–572. [PubMed: 9544596]
38. Maurischat C, Herschbach P, Peters A, Bullinger M. Factorial validity of the Short Form 12 (SF-12) in patients with diabetes mellitus. *Psychology Science Quarterly* 2008;50:7–20.
39. Wilson D, Parsons J, Tucker G. The SF-36 summary scales: Problems and solutions. *Sozial- und Praventivmedizin* 2000;45:239–246. [PubMed: 11210594]
40. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health and Quality of Life Outcomes* 2007;5:54–61. [PubMed: 17825096]
41. Carod-Artal J, Egido JA, González JL, de Seijas V. Quality of life among stroke survivors evaluated 1 year after stroke: Experience of a stroke unit. *Stroke* 2000;31:2995–3000. [PubMed: 11108762]
42. Niemi ML, Laaksonen R, Kotila M, Waltimo O. Quality of life 4 years after stroke. *Stroke* 1988;19:1101–1107. [PubMed: 3413807]
43. Duncan PW, Samsa GP, Weinberger M, Goldstein LB, Bonito A, Witter DM, et al. Health status of individuals with mild stroke. *Stroke* 1997;28:740–745. [PubMed: 9099189]
44. Engstad T, Bonna KH, Viitanen M. Validity of self-reported stroke: The Tromso Study. *Stroke* 2000;31:1602–1607. [PubMed: 10884460]

**Table 1**

## Background Characteristics of Study Participants

Variable	History of stroke (n = 2,581)	No history of stroke (N = 38,066)	p value
Age, mean (SD)	68.26 (9.32)	64.46 (9.62)	<.0001
Race, n (%)			
African American	1,435 (56)	17,380 (46)	<.0001
White	1,146 (44)	20,686 (54)	
Gender, n (%)			
Female	1,340 (52)	21,636 (57)	<.0001
Male	1,241 (48)	16,430 (43)	
PCS, mean (SD)	39.27 (11.38)	46.68 (10.39)	<.0001
MCS, mean (SD)	51.45 (10.70)	53.88 (8.56)	<.0001

PCS = Physical Component Summary; MCS = Mental Component Summary.

**Table 2**

Standardized Factor Loadings for the 2-Factor Model of the SF-12

SF-12 item:	Unconstrained Multiple Group Analysis					
	Total Sample (N = 40,647)		History of Stroke (N = 2,581)		No History of stroke (N = 38,066)	
	PHLF	MHLF	PHLF	MHLF	PHLF	MHLF
1. General health	-0.58	-0.12	-0.62	-0.05	-0.57	-0.12
2. Limited in moderate activities	0.75	0	0.71	0	0.75	0
3. Limited in climbing several stairs	0.71	0	0.64	0	0.70	0
4. Accomplished less due to physical health	0.87	0	0.83	0	0.87	0
5. Limited in kind of work	0.92	0	0.89	0	0.92	0
6. Accomplished less due to emotional health	0	0.93	0	0.94	0	0.93
7. Not work as carefully	0	0.94	0	0.94	0	0.93
8. Pain interfere with work	-0.78	0	-0.67	0	-0.78	0
9. Physical/emotional interfere with social	0.56	0.38	0.57	0.33	0.55	0.39
10. Felt calm and peaceful	0	-0.67	0	-0.69	0	-0.67
11. Have a lot of energy	-0.45	-0.35	-0.48	-0.28	-0.44	-0.35
12. Felt downhearted and blue	0	0.74	0	0.76	0	0.74

Note: PHLF = Physical Health Latent Factor; MHLF = Mental Health Latent Factor.

All zero factor loadings were fixed to be zero in the model identification process. All non-zero factor loadings were significantly different from zero ( $p < .0001$ ), except the loading of the general health item on the MHLF for the stroke history group.

The correlation between PHLF and MHLF was .54 ( $p < .0001$ ) in the total sample, .59 ( $p < .0001$ ) in the stroke history group, and .53 ( $p < .0001$ ) in the non-stroke history group.

Items with negative factor loadings represent those SF-12 questions on which higher scores indicated poorer health status whereas items with positive factor loadings represent the SF-12 questions on which higher scores reflected better health.

**Table 3**

Correlations between the Latent Factors, Standard Component Summary Scores, and Age

	1. PHLF	2. MHLF	3. PCS	4. MCS	5. Age
1. PHLF	1.00				
2. MHLF	0.54	1.00			
3. PCS	0.97	0.26	1.00		
4. MCS	0.27	0.96	0.08	1.00	
5. Age	-0.09	0.14	-0.09	0.13	1.00

Note: PHLF = physical health latent factor; MHLF = mental health latent factor; PCS = Physical Component Summary; MCS = Mental Component Summary.

**Table 4**  
Effects of Self-Reported History of Stroke on Latent Factor Scores and Standard Component Summary Scores

Outcome variable	No Covariate Adjustments		Effects After Adjusting for Race, Gender, and Age	
	Estimate*	SE	Estimate*	SE
PHLF	-0.77	0.02	-0.75	0.02
MHLF	-0.36	0.02	-0.43	0.02
PCS	-0.70	0.02	-0.66	0.02
MCS	-0.28	0.02	-0.33	0.02

\*The estimate is the difference, in standard deviation units, on the outcome variable between participants with a history of stroke and those without a history of stroke.

Note: PHLF = physical health latent factor; MHLF = mental health latent factor; PCS = Physical Component Summary; MCS = Mental Component Summary.