Journal of Clinical Epidemiology 59 (2006) 208–215

Journal of Clinical Epidemiology

Different factor loadings for SF36: The Strong Heart Study and the National Survey of Functional Health Status

Janette Beals[a,*], Thomas K. Welty[a], Christina M. Mitchell[a], Dorothy A. Rhoades[a], Jeun-Liang Yeh[b], Jeffrey A. Henderson[a,c], Spero M. Manson[a], Dedra S. Buchwald[d]

[a]*American Indian and Alaska Native Programs, University of Colorado at Denver and Health Sciences Center, F800, P.O. Box 6508, Aurora, CO 80045, USA*

[b]*Center for Epidemiologic Research, College of Public Health, University of Oklahoma Health Sciences Center, 801 N.E. 13th Street, CHB #100, Oklahoma City, OK 73104, USA*

[c]*Black Hills Center for American Indian Health, 701 St. Joseph Street, Suite 204, Rapid City, SD 57701, USA*

[d]*Department of Medicine, University of Washington, 325 Ninth Avenue, Box 359780, Seattle, WA 98104, USA*

Accepted 11 July 2005

Abstract

Background: To increase our understanding of the psychometric characteristics and factor structure of the SF36® in older American Indian populations.

Methods: Between 1993 and 1995, SF36® data were collected from 3,488 Phase II participants of the Strong Heart Study (SHS) between the ages of 48 and 81. Comparison data were provided by an age- and gender-matched sample ($n = 695$) from the National Survey of Functional Health Status (NSFHS) conducted in 1989 and 1990.

Results: Generally, the basic psychometric analyses showed that the SF36® performed adequately in these older American Indians. Exploratory factor analyses indicated that a one-factor model best fit the data for both older groups. On the other hand, confirmatory factor analyses showed that a two-factor model with correlated factors provided a superior fit to the data than a one-factor model. An assumption of equivalent factor loadings for the SHS and NSFHS groups was untenable.

Conclusion: These analyses demonstrate that use of summary scores assuming a differentiated physical/mental functioning structure is likely improper in at least some populations. The SF36® provides an important opportunity to understand cultural differences in the conceptualization and measurement of health-related quality of life.

Keywords: Health status indicators; SF36®; Crosscultural; American Indian; Reliability; Factor analysis

1. Introduction

The past century saw tremendous progress in the health status of American Indians—especially in the prevention and treatment of infectious diseases [1]. In recent decades, however, infectious diseases have been replaced by chronic conditions such as heart disease, cancer, and diabetes as the leading causes of death in this population [2]. Accordingly, the goal of health care shifted from focused, time-limited interventions, such as acute care, to long-term management of chronic diseases and their physical and psychologic concomitants. As a consequence, the need increased for reliable and valid assessments of patient perceptions of the impact of chronic diseases, such as that provided by measures of Health-Related Quality of Life (HRQoL).

As one of the most widely used measures, the Medical Outcomes Study 36-item Short Form Health Survey (SF36®) [3,4] provides an excellent opportunity for clinical researchers to investigate the crosscultural equivalence of this measure and, indeed, of HRQoL itself [5–10]. Measurement equivalence includes the common assessments of reliability and validity; factor analytic methods provide additional information—specifically, whether the same dimensions underlie such measures. Such methods have been especially important with the SF36®, where exploratory factor analyses (EFA) of the SF36® data from a national probability sample of the United States confirmed that two dimensions—physical and mental components of HRQoL—efficiently explain the pattern of covariances among eight SF36® subscales. Using the factor scores from these analyses, the Physical and Mental Component Summaries (PCS and MCS, respectively) have become common SF36® scale scores [11,12]. However, in using these

* Corresponding author. Tel.: 303-724-1453; fax: 303-724-1474.
E-mail address: jan.beals@uchsc.edu (J. Beals).

0895-4356/06/$ – see front matter
doi: 10.1016/j.jclinepi.2005.07.010

component scores, we assume that the two-factor structure and factor scores are equally valid for all samples—an assumption that has not been tested for many populations, particularly ethnic minorities.

To date, only one small clinic-based study has reported on the functioning of the SF36® in an American Indian sample [13]. The current study takes an important next step in understanding the utility of the SF36® among American Indians. Using data from the Strong Heart Study (SHS), an epidemiological study of cardiovascular disease [14–16], we present results from a large, older American Indian group. These data were analyzed in tandem with age- and gender-matched data from the National Survey of Functional Health Status (NSFHS)—the sample of the U.S. general population upon which the SF36® Version 1 norms are based [17].

Using the standard SF36® psychometric protocol, we hypothesized that the SF36® would function similarly among the American Indians and the age- and gender-matched NSFHS sample. Further, to assess the crosscultural measurement equivalence of the PCS and MCS, we completed both EFAs and confirmatory factor analyses (CFAs) of both groups. As we embarked on this work, we hypothesized that the conceptual distinction between physical and mental health that forms the basis of the two-factor structure would be less clear in these American Indian samples. In particular, the mind–body dualism underlying this conceptualization is largely Euro-American in origin [18,19] and may not be shared by cultures with a more holistic view of health, including many American Indian populations [19]. Thus, we expected that an assumption of measurement equivalence would be untenable.

2. Methods

2.1. Samples

The study design of the SHS has been discussed in detail elsewhere (http://caihr.ouhsc.edu/) [14–16]. Briefly, it is a large, ongoing, longitudinal population-based study of risk factors for cardiovascular disease. Three Centers (Dakotas, Oklahoma, and Arizona) coordinate the data collection from 13 tribes. The study cohort consisted of 4,549 persons over 45 who were seen in the first (Phase I) examination conducted between July 1989 and January 1992. Participation rates were 72% for the Arizona Center, 62% for the Oklahoma Center, and 55% for the Dakota Center, respectively. The data presented here are from the Phase II reexamination conducted between July 1993 and December 1995. Overall, 88% of the surviving Phase I participants returned for Phase II [16]. The SF36® was included in a personal interview portion of the reexamination; 87% (n = 3,488) of the Phase II participants completed at least one item of the SF36® and were included here.

The NSFHS used a 1990 cross-sectional sample from the 1989 and 1990 General Social Survey sample frames. Conducted by the National Opinion Research Center [20], the population of inference was the noninstitutionalized U.S. population [21]. Altogether, a single member from each of the 1,537 households sampled in 1989 and 1,372 in 1990 was approached to complete the survey. The overall response rate for the NSFHS was 77% [21]. Respondents were randomly assigned to a mail survey or telephone-administered interview format; method of administration did not have a substantial effect [22]. The SHS groups ranged in age from 48 to 81 years; the NSFHS data were assigned random numbers within gender and age blocks (typically spanning 5 years); NSFHS participants were chosen for inclusion in proportion to the SHS distributions of these blocks, providing an age- and gender-match of NSFHS to SHS with a final NSFHS sample of 695. The NSFHS data were obtained from the Inter-University Consortium for Political and Social Research [17].

2.2. Measures

The SF36® contains eight subscales: Physical Functioning, Role—Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role—Emotional, and Mental Health. A single-item measure of change over the past year in self-perceived health status (Health Transition) is also included. All subscales were recoded to range from 0 to 100, and most assessed the absence of limitations such that 0 implied extreme limitations and 100 signified no limitations on that particular dimension. Three subscales deviated from this pattern. For the General Health, Vitality, and Mental Health subscales, those with no limitations scored 50 while those with higher than average general health, vitality, or mental health scored as high as 100.

Also included in some analyses were the demographic variables that we found in common between the datasets: age, gender, educational status (less than high school, high school, some college), and marital status (married, unmarried).

2.3. Analyses

The SF Health Outcomes Scoring Software [23] was used to assess item and scale performance as well as to create the eight scales that were used in the factor analyses.

Both exploratory and confirmatory factor analyses (EFA and CFA, respectively) were conducted. EFA is the classic method for determining the number and composition of dimensions that can efficiently explain the covariances between measures—in this case, the SF36® subscales. As mentioned previously, EFA methods have been central to the development of the PCS and MCS summary scores. Here, we used the EFA module available in SPSS 13.0 [24]; both the one- and two-component results were reported for the SHS group and for both the age- and gender-matched NSFHS subsample and the full NSFHS sample. As in earlier work, principal components analyses with varimax rotation were used to derive the factor loadings [21].

In contrast to EFA, CFA methods allow greater flexibility for testing alternative models against one another (for instance, one- vs. two-factor models) and to test whether the parameters (e.g., factor loadings) can be assumed to be equivalent across groups. Here, we used M*plus* [25]. For each CFA model, the χ^2 statistic, degrees of freedom, comparative fit index (CFI), and root mean square error of approximation (RMSEA) are reported. A CFI over .9 [26] and RMSEA under .05 [27] indicate adequate fit of the model to the data.

3. Results

A demographic comparison verified that age and gender did not differ between SHS and NSFHS. The SHS group was less likely to be currently married (53% vs. 63%, $P < .001$) and more likely to have not finished high school (47% vs. 25%, $P < .001$).

The psychometric summary may be found in Table 1. The Social Functioning subscale was less likely to be considered a scaling success and was more likely to have an unacceptable level of internal consistency. Floor effects (a substantial percentage of group receiving the 0 or lowest possible score) were most often observed for the Role—Physical and Vitality subscales; for both of these, the SHS samples was more likely to score 0 even controlling for demographic differences. Ceiling effects (a substantial percentage scoring 100) were most common on the Role—Physical, Bodily Pain, Social Functioning, and Role—Emotional subscales. Controlling for demographic differences, the SHS group was more likely to have ceiling effects for Bodily Pain and Health Transition but less likely to report the maximum score for Role—Physical and General Health than was the NSFHS sample. The pattern of these statistics and the percentages of participants with complete data and computable scores were similar across groups.

We used the same EFA methods as Ware and colleagues [21,28] to begin our exploration of the dimensionality of the subscales. As seen in Table 2, when EFAs were completed on both the older SHS and NSFHS groups, the Kaiser criterion indicated one-factor solutions, where the full NSFHS sample yielded the commonly reported two-factor solution. Review of the eigenvalues suggested that, for all groups, the first component was strong, with the second one either just over or just under the Kaiser criterion of 1.0. The total variance explained by the two-component solution for SHS was 60% compared to 72% for the older NSFHS subsample and 68% for the full NSFHS sample.

We then moved to CFA methods, which allow for the testing of alternative models, including those across groups. As shown in Table 3, the first set of alternative models compared the one-factor structure with two sets of two-factor structures: one in which no subscales crossloaded (Model 2-1) [12,29] and the more common model in which General Health, Vitality, and Social Functioning were

Table 1
Scaling success, internal consistency, floor and ceiling effects, and computable scores: SHS and NSFHS samples, aged 48–81

	Scaling success[a]		Internal consistency		Floor effects			Ceiling effects		
Scale	SHS %	NSFHS %	SHS	NSFHS	SHS	NSFHS	ph[d]	SHS	NSFHS	ph[d]
Physical functioning	100	100	0.92	0.92	1.5	1.2		17.4	17.8	
Role—physical	100	100	0.92	0.89	***25.8***[c]	15.1	***	***51.2***	***59.1***	**
Bodily pain	85.7	85.7	0.80	0.81	0.4	1.2		***29.4***	23.0	***
General health	97.1	97.1	0.73	0.83	0.1	0.4		2.4	5.8	***
Vitality	100	100	0.74	0.88	0.2	1.2	*	1.5	1.7	
Social functioning	42.8	92.9	***0.60***[b]	***0.66***	0.3	1.0		***51.5***	***53.7***	
Role–emotional	100	100	0.90	0.85	15.5	10.8		***70.6***	***70.6***	
Mental health	100	100	0.75	0.84	9.1	0.1		1.0	6.2	
Health transition					5.4	5.9		9.1	4.5	***

	Complete data and computable scores	
	SHS	NSFHS
	n = 3,493	n = 695
Complete data (%)[e]	98.7	98.5
Consistent responses	86.9	88.9
Estimable scale scores with half-scale MDE[f]	98.9	98.3

[a] Percent of item-hypothesized-scale correlations significantly greater than item-competing-scale correlations (definite scaling success).
[b] Those with internal consistency of less than .7 highlighted in ***bold italics***.
[c] Those with floor or ceiling effects in excess of 25% highlighted in ***bold italics***.
[d] Post-hoc tests of statistically significant differences between groups in floor/ceiling effects controlling for age, gender, education and marital status, * P < .05, ** P < .01, *** P < .001.
[e] Percent of respondents who had complete data for a scale (range across eight scales).
[f] Percent of respondents for whom a scale score could be calculated (range across eight scales after missing item replacement when less than half the items were missing).

Table 2
Results of exploratory principal component analyses: SHS and NSFHS samples (aged 48–81 and full sample)

Rotated components						
			NSFHS			
	Strong Heart Study		Aged 48–81		Full sample (18–94)	
	One-component solution					
	HRQoL	*h^2*	*HRQoL*	*h^2*	HRQoL	h^2
Physical functioning	***0.70***	*49.0%*	***0.77***	*58.9%*	**0.71**	50.3%
Role—physical	***0.71***	*51.0%*	***0.82***	*67.1%*	**0.79**	61.6%
Bodily pain	***0.66***	*44.4%*	***0.80***	*63.9%*	**0.75**	56.4%
General health	***0.73***	*53.4%*	***0.78***	*60.5%*	**0.77**	59.1%
Vitality	***0.75***	*56.0%*	***0.85***	*71.5%*	**0.79**	61.6%
Social functioning	***0.71***	*49.9%*	***0.78***	*61.5%*	**0.76**	58.4%
Role—emotional	***0.64***	*41.4%*	***0.66***	*42.9%*	**0.65**	41.9%
Mental health	***0.65***	*42.5%*	***0.76***	*58.1%*	**0.71**	49.8%

	Strong Heart Study			Aged 48–81			Full sample (18–94)		
	Two-component solution								
	Mental	Physical	h^2	Physical	Mental	h^2	*Physical*	*Mental*	*h^2*
Physical functioning	0.17	**0.82**	70.5%	**0.85**	0.16	73.7%	***0.85***	*0.12*	*73.0%*
Role—physical	0.21	**0.80**	68.8%	**0.77**	0.34	71.0%	***0.81***	*0.27*	*73.3%*
Bodily pain	0.29	**0.64**	49.7%	**0.76**	0.32	67.7%	***0.76***	*0.28*	*65.3%*
General health	0.51	0.53	53.4%	**0.79**	0.25	68.0%	***0.70***	*0.38*	*62.6%*
Vitality	**0.66**	0.40	59.3%	**0.68**	0.50	71.5%	*0.48*	***0.65***	*64.2%*
Social functioning	**0.66**	0.33	55.1%	0.53	**0.60**	63.9%	*0.43*	***0.67***	*63.0%*
Role–emotional	0.59	0.32	44.9%	0.16	**0.88**	79.8%	*0.17*	***0.78***	*64.2%*
Mental health	**0.88**	0.03	78.2%	0.35	**0.80**	76.1%	*0.17*	***0.87***	*78.2%*

	Strong Heart Study			Aged 48–81			Full sample (18–94)		
	Eigenvalues and variance explained								
	1st Comp.	2nd Comp.	Total	1st Comp.	2nd Comp.	Total	1st Comp.	2nd Comp.	Total
Eigenvalue	3.86	0.94		4.84	0.88		4.39	1.05	
Variance explained	48.3%	11.7%	60.0%	60.6%	10.9%	71.5%	54.9%	13.1%	68.0%
Rotated eigenvalue	2.42	2.38		3.40	2.32		2.90	2.54	
Rotated variance explained	30.3%	29.7%	60.0%	42.5%	29.0%	71.5%	36.2%	31.8%	68.0%

h^2 indicates the communality or item variance explained.
Italics indicates the preferred solution according to the Kaiser criterion.
Bold indicates a component loading of .6 or above.

indicators for both the *Physical* and *Mental* factors (Model 2-2) [21,30]. For both SHS and NSFHS older sample, the two-factor model with crossloadings provided the best fit. The allowance of correlated residuals (Models 2-1r and 2-2r) accommodates uneven patterns of covariances between subscales (e.g., Role—Emotional and Role—Physical were reasonably hypothesized to be correlated with one another because of their similar format). Table 3 shows that the addition of correlated residuals improved the fit of all two-factor models, yet the common two-factor structure continued to provide a better fit.

The common two-factor model with crossloading (Model 2-2r) was then tested across groups to determine whether the loadings could be assumed equivalent either across both the SHS and older NSFHS groups. The difference χ^2 (107.4, 9 *df*) between these models indicated that significant differences in factor loadings existed across samples.

CFA models allow inclusion of other variables, in this case the common demographics, in the models. Model 2-2r-demo, where the demographic variables were not constrained equal across groups, fit better than did the constrained Model (2-2r-demo =). However, examination of the parameters suggested that age had an equivalent relationship with the *Physical* and *Mental* latent factors across SHS and older NSFHS sample (Model 2-2r-age =) and, indeed, the model with age constrained equal across groups did not differ from that allowing demographics to be estimated independently across groups (Model 2-2r-demo); thus, Model 2-2r-age = served as the final model.

Table 4 presents parameter estimates for the final model. As indicated in the multigroup analyses, many of the unstandardized loading estimates differed; in particular, Bodily Pain, General Health, Vitality, and Social Functioning were significantly higher for the older NSFHS sample on the *Physical* factor while General Health, Social

Table 3
Comparison of confirmatory factor analytic models: SHS and older NSFHS sample aged 48–81

Within samples tests of alternative models

		SHS				NSFHS			
Model	Description	χ^{2a}	*df*[b]	CFI[c]	RMSEA[d]	χ^2	*df*	CFI	RMSEA
1	One-factor model	1064.2	20	0.885	0.124	312.2	20	0.900	0.148
1r	**Correlated residuals:** RE with RP; MH with VT; RP with PF; RE with SF; MH with SF; MH with RE;	185.2	14	0.981	0.06	73.8	14	0.980	0.080
2-1	Two-factor model with no crossloadings: Physical by PF, RP, BP, and GH; Mental by MH, VT, SF, and RE	836.3	19	0.910	0.113	250.6	19	0.923	0.135
2-1r	**Correlated residuals:** MH with VT; VT with GH; MH with GH	442.4	16	0.953	0.089	202.3	16	0.938	0.132
2-2	Two-factor model with no crossloadings: Physical by PF, RP, BP, GH, VT, and SF; Mental by MH, VT, SF, RE and GH.	617.4	16	0.934	0.105	130.7	16	0.962	0.100
2-2r	**Correlated residuals:** RE with RP; MH with VT; VT with GH; MH with GH	96.1	12	0.991	0.045	54.0	12	0.986	0.073

Multisample tests of two-factor model

						Nested comparisons			
	Model	χ^2	df	CFI	RMSEA	Alt. Model	Diff χ^{2e}	*df*	*P*[f]
2-2r	Loadings freely estimated	150.1	24	0.990	0.051				
2-2r=	All loadings constrained equal across 2 groups	257.5	33	0.981	0.058	2-2r	107.4	9	0.0001
2-2r-demo	Demographic variables added	490.5	72	0.966	0.055				
2-2r-demo=	Demographics constrained =	516.7	80	0.965	0.053	2-2r-demo	26.2	8	0.005
2-2r-age=	Only age constrained =	495.9	74	0.966	0.054	2-2r-demo	5.4	2	0.072

Abbreviations: PF, Physical functioning; RP, Role—physical; BP, Bodily pain; GH, General health; VT, Vitality; SF, Social functioning; RE, Role—emotional; MH, Mental health.

[a] Chi-squared test for model fit.
[b] Degrees of freedom.
[c] Comparative fit index.
[d] Root mean square error of approximation.
[e] Difference in chi-squared between nested models.
[f] Probability that nested models are different from one another.

Functioning, and Role—Emotional were higher for SHS on the *Mental* factor. When we examined the relationships of the demographic variables to the latent *Physical* and *Mental* factors, we found that age was negatively related to *Physical* but not significantly related to *Mental*. Male gender was positively correlated to both *Physical* and *Mental* for SHS but not for the older NSFHS sample. Education was positively correlated to both these latent factors for both groups, but the correlation was significantly greater for *Physical* in the older NSFHS sample than in SHS. Finally, being married was positively related only to *Mental*, although only for the NSFHS older sample. Further, after controlling for these demographics, the correlations between the *Physical* and *Mental* factors decreased slightly, from .67 to the .61 reported here.

As was true of the EFAs, more of the variance in the subscales was explained in the older NSFHS sample than in the SHS group. Furthermore, the demographic variables explained more of the variability in the *Physical* latent factor for NSFHS (14.5%) than SHS (9.9%); these variables explained similar, small amounts of the variability for the *Mental* latent variable.

4. Discussion

Using the standard SF36® psychometric protocol, the subscale-level statistics reported here were generally comparable across SHS and the age- and gender-matched NSFHS group. Perhaps most notable here was the poor performance of the Social Functioning subscale—a pattern also reported by others [31–33]. However, when we moved to the factor analyses, the most intriguing findings came to light. Using the same EFA methods as Ware et al. [21] in constructing the PCS and MCS summary scores, we found that a one-factor rather than two-factor solution was generated—not only for the American Indian group, as hypothesized, but also for the age- and gender-matched NSFHS sample. The CFAs shed further light: We found that the two-factor solution provided a superior fit to the one-factor but that the resulting *Physical* and *Mental* latent variables

Table 4
Final parameter estimates of two-factor models (Model 2-2r-age=) for the SHS and older NSFHS samples

	SHS		NSFHS		
	Unstand. estimate[b]	Stand. estimate[c]	Unstand. estimate[b]	Stand. estimate[c]	ph[a]
Factor loadings					
Physical health					
Physical functioning	1.00[d]	0.74	1.00[d]	0.80	
Role—physical	1.54	0.74	1.60	0.82	
Bodily pain	0.73	0.59	0.96	0.80	***
General health	0.47	0.49	0.83	0.75	***
Vitality	0.43	0.46	0.68	0.61	***
Social functioning	0.13	0.12	0.46	0.39	***
Mental health					
General health	0.37	0.19	0.01	0.01	***
Vitality	0.39	0.21	0.36	0.24	
Social functioning	1.43	0.65	0.72	0.45	***
Role—emotional	2.56	0.67	1.66	0.73	***
Mental health	1.00[d]	0.61	1.00[d]	0.82	
Correlation between physical and mental health					
	0.61		0.61		
Relationships of demographics to physical and mental latent variables					
Physical					
Age	−0.41	−0.20	−0.42	−0.17	
Male	5.23	0.12	0.27	0.01	*
Education	4.69	0.17	7.78	0.30	*
Married	1.41	0.03	3.08	0.07	
Mental					
Age	0.07	0.03	0.07	0.04	
Male	2.76	0.13	0.98	0.03	
Education	2.29	0.18	3.38	0.17	
Married	0.56	0.03	4.20	0.14	*
Variance explained (r^2)					
Physical functioning	55.5%		62.5%		
Role—physical	53.1%		69.8%		
Bodily pain	36.1%		62.2%		
General health	39.9%		55.1%		
Vitality	38.4%		61.8%		
Social functioning	54.2%		58.5%		
Role—emotional	44.7%		53.9%		
Mental health	37.6%		66.2%		
Physical	9.9%		14.5%		
Mental	5.6%		5.1%		

[a] Post-hoc test of differences between SHS and NSFHS: * $P < .05$, ** $P < .01$, *** $P < .001$.
[b] Unstandardized estimate.
[c] Standardized estimate.
[d] Marker variable set to 1.00. The remaining estimates for each factor were calculated in reference to the marker variable.

were highly correlated. Further, we found that we could not assume measurement equivalence across the groups and the models explained more of the variability for the NSFHS than for SHS groups, even controlling for demographic differences.

Our results are similar to some existing crosscultural factor analytic studies but differ from others. Many crosscultural efforts have found the two-factor structure [12,28,30,34–37]. However, others have reported one-factor structures—for example, among older Maori and Pacific Islanders [38] and British multiple sclerosis patients [29]. Others have also reported that the greatest differences were found for the loadings for those subscales present on both the *Physical* and *Mental* latent factors [31,38,39]. Unlike these previous efforts, however, the Role—Emotional subscale in our study remained tied to the mental health factor.

The results presented here and those cited above have implications both for the use of the PCS/MCS summary scores and for the crosscultural study of HRQoL. Because we could not assume the factor loadings were equivalent, use of the PCS and MCS would likely introduce some biases when used uncritically. We are not the first to raise this issue [29,39–42]. Indeed, the creators of the SF36® advise users that the PCS and MCS scores should be used only after careful consideration of the subscale scores [11]. Researchers in other fields have criticized use of factor scores [43–45] and suggest alternatives including unit-weighting [45], where each variable is coded as 0/1, depending on the magnitude of its factor loading. This strategy may also address other criticisms of the PCS and MCS [40,41].

More importantly, however, lines of inquiry such as those presented here provide the opportunity to improve our understanding of HRQoL in a crosscultural context and, in this case, perhaps across age groups as well. For at least some populations, physical and mental functioning are not well differentiated. White factor score methods with orthogonal rotations allow us to mathematically manage data to render the PCS and MCS scores unrelated to one another, such manipulations may obscure important group differences. While we laud the extensive efforts of the SF authors and others [46,47] to develop measures that operate similarly across groups, we are concerned that the content and implications of cultural differences have not been discussed in the literature to the degree warranted.

We join others in arguing that understanding HRQoL at a basic level could be an important link in addressing health disparities among U.S. ethnic/racial minority populations [10]. The importance of subjective medicine, or an appreciation of health from the patient's point of view, is increasingly emphasized as critical to understanding the propensity of patients to seek and accept help for medical conditions—and by extension may be critical to comprehending health disparities. Both qualitative and quantitative methods are important [10] to understand both the conceptual and measurement equivalence of measures such as the SF36®. Although recent attention has been focused on the advances provided by the cognitive aspects of survey methodology [48], we urge further inclusion of ethnographic methods as well [49,50]. Together with the expanded array of psychometric tools, including not only CFA but also Item Response Theory [51,52] and other methods, the potential for better understanding HRQoL is enhanced [10,48].

4.1. Limitations

Limitations to the current study relate to differential sampling methodologies, data collection methods, and analytic strategies between the SHS and NSFHS. We used the NSFHS data to provide a broader context for the older American Indian groups. However, the NSFHS was based on a multilevel household sampling strategy, while the SHS groups derived from the second wave of data collection where the sample frame included all tribal members. Although the SHS participants included all older American Indians residing on or near 13 tribal communities at the initiation of Phase I, its restricted age range limits our findings. Further, here the data from the 3 SHS Centers were combined; although preliminary analyses indicated few differences, some detail was inevitably lost. Similarly, our results based on the SHS are unlikely generalizable to the more than 300 federally recognized American Indian tribes and the more than 200 Alaska Native entities [53]. The response rates for the SF36® were adequate for the NSFHS and Oklahoma Center, but low for the Arizona and Dakota Centers. Last, we chose to emulate the approach of Ware et al. [21] with factor analyses at the subscale level rather than the item level; a full second-order factor model would likely yield somewhat different results.

5. Conclusions and clinical implications

In this first examination of the factor structure of the eight SF36® subscales among American Indian populations, we found that an assumption of measurement equivalence between this group of older American Indians and an age- and gender-matched sample of the U.S. population was unsupported. Furthermore, the uncorrelated two-factor structure that forms the basis of the PCS and MCS scores proved especially problematic. Because the SHS and NSFHS groups were defined by both age/gender and culture, the results do not clearly indicate which is of greater importance to understanding the underlying dimensionality of the SF36®. Among older populations, as the burden of disease increases, physical and mental functioning may be less differentiated than in younger populations. Cultural influences may also be operational; indeed, we envision the SF36® as an important instrument for further study of cultural equivalence—especially as the focus on health disparities increases.

Acknowledgments

The Strong Heart Study (SHS) was conducted by cooperative agreements U01-HL41642, U01-HL41652, and U01-HL41654 from the National Heart, Lung, and Blood Institute. The authors acknowledge the assistance and cooperation of the participating tribes and the Indian Health Service facilities that serve those tribes. The authors also thank the study participants, the Directors of SHS clinics, Betty Jarvis, Martha Stoddart, Beverly Price, Marcia O'Leary, Dr. Tauqeer Ali, Alan Crawford and their staffs, and Dr. Barbara Howard, Dr. Elisa Lee, and the other SHS investigators who contributed to this work. The opinions expressed in this report are those of the authors and not necessarily those of the Indian Health Service.

The data from the National Survey of Functional Health Status 1990 were made available by the Inter-University Consortium for Political and Social Research.

Manuscript preparation was supported by P01 MH42473 (Manson, PI), P30 AG15297 (Manson, PI), P01 HS10854 (Manson, PI), and P60 MD000507 (Manson, PI).

References

[1] Brenneman GR, Handler AO, Kaufman SF, Rhoades ER. Health status and clinical indicators. In: Rhoades ER, editor. American Indian Health: innovations in health care, promotion, and policy. Baltimore, MD: The Johns Hopkins University Press; 2000.
[2] U.S. Department of Health and Human Services. Trends in Indian health 1998–1999. Rockville, MD: Department of Health and Human Services; 2001.
[3] Ware JE Jr, Sherbourne CD. The MOS 36-item Short Form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473.
[4] Ware JE Jr. Standards for validating health measures: definition and content. J Chronic Dis 1987;40:473–80.
[5] Angel RJ, Frisco ML. Self-assessments of health and functional capacity among older adults. J Mental Health Aging 2001;7:119–35.
[6] Liang J. Assessing cross-cultural comparability in mental health among older adults. J Mental Health Aging 2001;7:21–30.
[7] Napoles-Springer AM, Stewart AL. Use of health-related quality of life measures in older and ethnically diverse populations. J Mental Health Aging 2001;7:173–9.
[8] Teresi JA. Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical, and mental health. J Mental Health Aging 2001;7:31–40.
[9] Teresi JA, Holmes D. Some methodological guidelines for cross-cultural comparisons. J Mental Health Aging 2001;7:13–9.
[10] Stewart AL, Napoles-Springer AM. Advancing health disparities research: Can we afford to ignore measurement issues? Med Care 2003;41:1207–20.
[11] Ware JE, Kosinski M. Interpreting SF-36 summary health measures: A response. Qual Life Res 2001;10:405–13; discussion 415–20.
[12] Peek MK, Ray L, Patel K, Stoebner-May D, Ottenbacher KJ. Reliability and validity of the SF-36 among older Mexican Americans. Gerontologist 2004;44:418–25.
[13] Johnson J, Nowatzki T, Coons S. Health-related quality of life of diabetic Pima Indians. Med Care 1996;34:97–102.
[14] Lee EN, Welty TK, Fabitz RR, Cowan LD, Ngoc AL, Oopik AJ, et al. The Strong Heart Study—a study of cardiovascular disease in American Indians: design and methods. Am J Epidemiol 1990;132:1141–55.
[15] Howard BV, Welty TK, Fabsitz RR, Cowan LD, Oopik AJ, Le N, et al. Risk factors for coronary heart disease in diabetic and nondiabetic Native Americans. Diabetes 1992;41:4–11.
[16] Howard BV, Lee ET, Cowan LD, Devereux RB, Galloway JM, Go OT, et al. Rising tide of cardiovascular disease in American Indians. The Strong Heart Study. Circulation 1999;99:2389–95.
[17] Ware JE Jr. National Survey of Functional Health Status. John E. Ware J, New England Medical Center [producer] 1991, editor. Ann

Arbor, MI: Inter-University Consortium for Political and Social Research; 1995.
[18] Kirmayer LJ. Cultural variations in the response to psychiatric disorders and emotional distress. Soc Sci Med 1989;29:327–39.
[19] Manson SM. Mental health services for American Indians and Alaska Natives: Need, use, and barriers to effective care. Can J Psychiatry 2000;45:617–26.
[20] Thalji L, Haggerty CC, Rubin R, Berckmans TR, Pardee BL. 1990 National Survey of Functional Status: Final report. Chicago, IL: National Opinion Research Center; 1991.
[21] Ware JE Jr. SF-36 physical and mental health summary scales: A user's manual. Boston, MA: The Health Institute; 1994.
[22] McHorney CA, Kosinski M, Ware JE Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: Results from a national survey. Med Care 1994;32:551–67.
[23] QualityMetric Incorporated. SF health outcomes scoring software. Lincoln, RI: QualityMetric Incorporated; 2004.
[24] SPSS Inc. SPSS. 13.0.1 ed. Chicago, IL: SPSS Inc.; 2004.
[25] Muthen LK, Muthen BO. M*Plus* user's guide. 3rd ed. Los Angeles: Muthen and Muthen; 1998–2004.
[26] Bollen KA. Structural equations with latent variables. New York: John Wiley and Sons; 1989.
[27] Wolinsky FD, Stump TE. A measurement model of the Medical Outcomes Study 36-Item Short-Form Health Survey in a clinical sample of disadvantaged, older, black, and white men and women. Med Care 1996;34:537–48.
[28] Ware JE Jr, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1159–65.
[29] Hobart J, Freeman J, Lamping D, Fitzpatrick R, Thompson A. The SF-36 in multiple sclerosis: why basic assumptions must be tested. J Neurol Neurosurg Psychiatry 2001;71:363–70.
[30] Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998;51:1139–47.
[31] Fuh J-L, Wang S-J, Lu S-R, Juang K-D, Lee S-J. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community. Qual Life Res 2000;9:675–83.
[32] Ren XS, Amick B, 3rd, Zhou L, Gandek B. Translation and psychometric evaluation of a Chinese version of the SF-36 Health Survey in the United States. J Clin Epidemiol 1998;51:1129–38.
[33] Stadnyk K, Calder J, Rockwood K. Testing the measurement properties of the Short Form-36 Health Survey in a frail elderly population. J Clin Epidemiol 1998;51:827–35.
[34] Lewin-Epstein N, Sagiv-Schifter T, Shabtai EL, Shmueli A. Validation of the 36-item short-form Health Survey (Hebrew version) in the adult population of Israel. Med Care 1998;36:1361–70.
[35] Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. J Epidemiol Community Health 2003;57:259–63.
[36] Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of Life in rural and urban populations in Lebanon using SF-36 Health Survey. Health Qual Life Outcomes 2003;1:30. Available at: http://www.hqlo.com/content/pdf/1477-7525-1-30.pdf.
[37] Wolinsky FD, Miller DK, Andresen EM, Malmstrom TK, Miller JP. Health-related quality of life in middle-aged African Americans. J Gerontol B Psychol Sci Soc Sci 2004;59:S118–23.
[38] Scott KM, Sarfati D, Tobias MI, Haslett SJ. A challenge to the cross-cultural validity of the SF-36 health survey: Factor structure in Maori, Pacific and New Zealand European ethnic groups. Soc Sci Med 2000;51:1655–64.
[39] Fukuhara S, Ware JE Jr, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. J Clin Epidemiol 1998;51:1045–53.
[40] Simon GE, Revicki DA, Grothaus L, Vonkorff M. SF-36 summary scores: are physical and mental health truly distinct? Med Care 1998;36:567–72.
[41] Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res 2001;10:395–404.
[42] Wilson D, Parsons J, Tucker G. The SF-36 summary scales: problems and solutions. Soz Praventivmed 2000;45:239–46.
[43] McDonald RP, Mulaik S. Determinacy of common factors: a nontechnical review. Psychol Bull 1979;86:297–306.
[44] McDonald RP, Burr EJ. A comparison of four methods of constructing factor scores. Psychometrika 1967;32:381–401.
[45] Grice JW, Harris RJ. A comparison of regression and loading weights for the computation of factor scores. Multivariate Behav Res 1998;33:221–47.
[46] Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.
[47] Skevington SM. Advancing cross-cultural research on quality of life: Observations drawn from the WHOQOL development. Qual Life Res 2002;11:135–44.
[48] Bjorner JB, Ware JE Jr, Kosinski M. The potential synergy between cognitive models and modern psychometric models. Qual Life Res 2003;12:261–74.
[49] Momenzadeh S, Posner N. Iranian migrants' discourses of health and the implications for using standardized health measures with minority groups. J Immigrant Health 2003;5:173–80.
[50] Zargooshi J. Quality of life of Iranian kidney "donors". J Urol 2001;166:1790–9.
[51] Mungas D, Reed BR. Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. Stat Med 2000;19:1631–44.
[52] Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: Application to cognitive assessment measures. Stat Med 2000;19:1651–83.
[53] Bureau of Indian Affairs. Indian entities recognized and eligible to receive services from the United States Bureau of Indian Affairs: Federal Register; 2000.