Published in final edited form as:

Appl Res Qual Life. 2012 March 1; 7(1): 63–81. doi:10.1007/s11482-011-9148-4.

Measuring Quality of Life with SF-36 in Older Americans with Traumatic Brain Injury

Chengwu Yang, MD, MS, PhD,

Pennsylvania State University College of Medicine, 600 Centerview Drive, Suite 3400H, Hershey, PA 17036

Anbesaw W. Selassie, MPH, DrPH,

Medical University of South Carolina, 135 Cannon Street, Suite 302M, Charleston, SC 29425

Rickey E. Carter, PhD, and

Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905

Barbara C. Tilley, PhD

University of Texas Health Science Center at Houston, 1200 Herman Pressler Drive, RAS E833, Houston, TX 77030

Introduction

Traumatic brain injury (TBI) is an acute injury to the brain resulting from rapid acceleration and deceleration of the cranial cavity. It is estimated that each year approximately 1.4 million Americans sustain a TBI [Langlois et al. 2006]. Such events are likely to occur in severe motor vehicle crashes, falls from elevated heights, assaults, etc. Typical manifestations of the disease may range from simple fatigue to loss of consciousness and are dependent on the severity of the brain injury [Styrke et al. 2007]. As with most cerebrovascular events, early treatment is vital to ensuring a successful recovery; however, there is a need to monitor broad-based functioning during the recovery process.

The Medical Outcomes Study (MOS) Short-Form Health Survey (SF-36) is a generic measure of eight domains of general physical and mental health [Ware and Sherbourne 1992]. Uses of the instrument include: assessment of patients in a clinical practice setting; differentiating treatment effects in a research study; and describing the general health of populations [Ware 2000]. As such, SF-36 is currently the most widely used general health assessment survey in the world, and has been translated to many languages for international use [Ware 2000]. Furthermore, SF-36 appears well suited to measuring the multiple systemic manifestations of TBI [Findler et al 2001; Bullinger and the TBI Consensus Group 2002].

Acceptable reliability and factor structure of an instrument are critical to the interpretation of instrument scores in clinical practice [Cook and Beckman 2006]. Reliability is typically

measured by the Cronbach's coefficient alpha [DeVellis 2003], and factor structure is best evaluated by CFA [Zumbo 2007].

Although SF-36 has been shown by numerous researchers to be robust in terms of reliability and validity in many populations [de Vet et al 2005], we are unaware of other research validating SF-36 in an older, racially-diverse population with traumatic brain injury. The objective of this study was to assess SF-36 regarding, a) reliability and factor structure; b) the impact of injury severity and race on the items and latent domains, through a latent variable modeling approach (MIMIC), predicated on demonstration of satisfactory psychometric properties in a representative sample of older, racially-diverse TBI patients. Without such an assessment, different racial groups and different severity groups among TBI patients would need different interpretations of the SF-scores, and their scores may not be comparable.

Methods

SF-36 measurement model

SF-36 was introduced by Ware [Ware and Sherbourne 1992] and assesses eight health domains or latent traits (Figure 2) using 35 questions. The second question on the instrument ("Compared to one year ago, how would you rate your health in general now?") measures health transition and does not contribute to any of the scale scores [Ware and Sherbourne 1992]. For each of these eight health domains, a scale score can be calculated with the higher scores representing better health status. In addition, two summary measures representing overall physical and mental health have been proposed [Ware et al 1994], but for the purpose of this research, we focus our attention on the eight traditional health domains. A complete description of the scoring and measurement model is described elsewhere [Ware et al 1993].

South Carolina traumatic brain injury follow-up registry

The South Carolina Traumatic Brain Injury Follow-up Registry (SCTBIFR) is a population-based data collection activity on individuals discharged from non-federal hospitals in South Carolina. The operations and technical details of the registry are described elsewhere [Pickelsimer et al 2006].

To obtain a representative sample of TBI cases in South Carolina, the registry utilized a two-stage, random sampling design. The population for this registry is individuals that were at least 15 years old at time of the injury and discharged alive from one of 62 non-federal hospitals with an ICD-9-CM code of 800–801.9, 803–804.9, 850–854.1 or 959.01 [Selassie et al 2005].

Severity of injury was measured by the Abbreviated Injury Scale (AIS) and the Injury Severity Score (ISS) through the translation of ICD-9-CM codes with ICDMAP-90 [Center for Injury Research Policy of the Johns Hopkins University School of Public Health 1997]. The AIS has 6 grades: Minor, Moderate, Serious, Server, Critical, and Maximum (untreatable, unsurvivable, vegetative stage), respectively for AIS 1–6. While AIS is a measure of the severity of injury to one body region, ISS provides an overall score for

patients with multiple injuries. The ISS score equals to the sum of the squares of the three highest AIS [Copes et al 1988]. ISS score values range from 1 to 75, with higher scores indicating greater compound severity [MacKenzie et al 1989]. Since this study is focusing on the effect of brain injury, the AIS score of brain is chosen in the MIMIC modeling.

Cohort included in this analysis

All whites (non-Latino Whites) and blacks (African Americans), age 55 years and older, who provided follow-up information through June 2005 were included in this analysis. Older was defined as age 55 and older rather than the traditional age 65 and over for White, non-Latinos. Due to the high burden of chronic disease morbidity and early mortality, the black population ages 55 and over has a health status similar to white, non-Latinos aged 65 and over [Tilley et al 2003]. Thus blacks ages 55–64 need to be considered in analyses of older adults.

Statistical analysis

Data were organized and scored according to SF-36 manual [Ware et al 1993]. Preliminary descriptive analysis of basic demographic characteristics was implemented. The psychometric properties of SF-36 in this sample were evaluated using Cronbach's alpha, followed by three other assessments. First, a confirmatory factor analysis (CFA) of the eight-factor structure was conducted to confirm the proposed factor structure of SF-36. Second, to investigate the effects of race and AIS on the eight latent traits measured by SF-36, multiple-indicator multiple-causes (MIMIC) analysis was conducted [Muthén and Muthén 2003]. Third, DIF analysis using the MIMIC models was conducted to determine if the effects of race and AIS on the latent traits were contaminated by DIF effects [Muthén and Muthén 2003, Jones 2006]. If SF-36 factor structure is confirmed and there is no DIF, then the scale scores calculated according to SF-36 manual would be compared with weighted norms.

Descriptive analyses were conducted using the SAS System version 9.1 (Cary, NC) and CFA and MIMIC models were estimated using MPlus version 5 (Los Angeles, CA). The remainder of this section details the analytical methods implemented and an overview of the analysis. The cases' disposition and analysis procedure are presented in Figure 1.

Data preparation and missing data accommodation

Items were scored according to SF-36 Manual [Ware et al 1993]. In the event of missing data (partially completed assessments), the data imputation algorithm substitutes a person-specific mean of the completed items for the missing items provided the respondent answered at least 50 percent of the items in the corresponding health domain [Ware et al 1993]. This approach has been implemented in the calculation of scale means for descriptive analyses; however, it is not amenable to CFA and MIMIC, because imputing values for exogenous variables based on the model to be validated would inflate model fit indices. For CFA and MIMIC (described below), the WLSMV (weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic that uses a full weight matrix [Muthén and Muthén 1998–2007]) estimator in Mplus was used, and the WLSMV extraction combined with Full

Information Maximum Likelihood (FIML) were used to augment the covariance matrix for missingness [Muthén and Muthén 1998–2007]. To assess the sensitivity to missing data of this analysis approach, a “complete case” analysis was also conducted using only data from subjects with complete assessments for both of CFA and MIMIC.

Reliability analysis

Cronbach’s alpha was calculated for each of the eight subscales of SF-36, using the 551 cases with complete information on all of the 35 items. To compare the reliability between blacks and whites, this analysis was implemented on each of the two sub-datasets: 74 blacks and 477 whites. Cronbach’s alphas of 0.70–0.79, 0.80–0.89, and 0.90–1.00 were ‘adequate’, ‘very good’, and ‘excellent’, respectively [Kline 2005].

Confirmatory factor analysis

The adherence of this population’s data to the published factor structure of SF-36 was examined using CFA. Two of the 35 items, items 1 and 21 were considered continuous after recalibration according to SF-36 manual. The remaining 33 items were treated as ordinal. The WLSMV estimator was used for CFA estimation. The delta parameterization was chosen for its better performance over the theta parameterization [Muthén and Asparouhov 2002].

Three types of model fit indices were used to assess the fit of the theoretical SF-36 measurement model in this population [Brown 2006]. Weighted Root Mean Square Residual (WRMR) assessed absolute fit, where less than or equal to 1.00 implies a good fit [Muthén and Muthén 2003]. The root mean square error of approximation (RMSEA) was used to assess the parsimony of the model, with values of less than 0.08 suggesting adequate fit [Brown 2006]. The comparative fit index (CFI) and Tucker-Lewis index (TLI) assessed comparative fit, with cutoff values for good fit set to 0.95 [Muthén and Muthén 2003; Brown 2006].

MIMIC modeling

MIMIC models were used to study the effects of race and AIS on the latent traits represented by the eight domains in SF-36, i.e., the population heterogeneity of the latent traits among different populations defined by race and AIS. The models are fit using regressions with latent variables [Zumbo 2007]. The population heterogeneity of the latent traits is assessed by testing whether the regression coefficients are non-zero for the covariates in a model with the eight SF-36 latent traits as outcomes (Figure 2).

MIMIC models were also utilized to assess whether there is differential item functioning (DIF) of race and AIS on individual SF-36 items that are not mediated by the eight latent traits [Muthén and Muthén 2003]. This was done by checking the modification indices (MI) [Muthén and Muthén 1998–2007] of DIF in an iterative fashion [Jones 2006]. The MI of a DIF from a covariate to a single SF-36 item was calculated by comparing the differences in chi-squares of the model before and after adding the DIF of the covariate. The MI of a DIF has an approximate chi-square distribution with one degree of freedom [Muthén and Muthén 1998–2007]. If the MI for a possible DIF is greater than 3.84, then it indicates there is a DIF

effect. This procedure was done iteratively by adding the DIF effect with the biggest MI greater than 3.84, and re-checking the remaining MI. This procedure was repeated until no more remaining MI was greater than 3.84. If there is no MI greater than 3.84, no DIF was detected. If some DIF are identified, then the relative change of regression coefficients from race or AIS to the eight latent traits before and after adding the DIF effects to the model will be used to assess the degree to which the population differences in the latent traits are contaminated by DIF [Jones 2006].

Sensitivity analysis

In order to evaluate the effect of missing data on SF-36 items, CFA and MIMIC analysis were implemented on the overall dataset and the dataset with complete information (Figure 1). Model fit indices and regression coefficients were presented.

Selection of norms

If SF-36 is shown to be valid (adequate reliability, valid factor structure, measurement invariant over race and AIS) in this population, the eight scale scores calculated according to the SF-36 manual can be compared to other populations to better understand how TBI may affect general health status. For this comparative analysis, we calculated a weighted mean reflecting the age and gender composition of our sample using the age-and-gender-specific norms for the eight scale scores provided with the instrument [Ware et al 1993]. These weighted norms are the basis for comparisons of scale scores to scores from other populations.

Results

Of the 2,118 enrolled into the Traumatic Brain Registry, 668 (32%) were identified as older individuals with a TBI. Of these, a total of 14 (2.1%) were eliminated due to missing injury severity status (n=2) or a racial classification with too few observations for model estimation (n=1 Asian descent; n=11 Hispanic). Therefore, 654 subjects comprised the full analysis sample. Of the 654, 551 (84.3%) answered all of SF-36 items and an additional 72 individuals answered enough questions in each scale to have all missing data imputed by the mean scale response. Thus, summary scale scores was calculated for 623 (95%) of the cases (Figure 1).

Selected demographics, including severity of brain injury, are summarized in Table 1. More than half (53%) have severe to critical injuries (AIS 4–5), and no case with either minor (AIS=1) or maximum (AIS=6) injury. The average ISS score is 13.5, but this is an index of overall body injury severity, not specific for TBI. Thus, this sample represents cases with significant brain injury.

The means and standard deviations of the eight scale scores obtained through scoring of the items according to SF-36 manual are listed in table 2, overall and by race. These scores are composite scores, not the latent traits measured by the SF-36 items and used in the MIMIC modeling. Among the eight scale scores, the differences of the means between blacks and whites are within 4 units, except for vitality and role-emotional.

Reliability

The Cronbach's alphas for the 8 scale scores for SF-36 using only individuals with complete assessments ($n=551$) are presented in table 2. Internal consistency for the overall sample is good, with all of the Cronbach's alphas greater than 0.80. Table 2 also presents the internal consistencies stratified by race. With 87 percent of the sample being whites, the race-specific estimates of whites are comparable to the sample as a whole. Estimates for internal consistency of blacks are lower but most of these estimates are greater than 0.70, suggesting acceptable internal consistency.

CFA

The model fit indices for the overall sample and for the two subset samples - whites and blacks - are also summarized in table 2. For each of the three datasets (overall, white, black), results are reported for observations with complete information on all SF-36 items and for overall sample.

For the overall sample, although the absolute fit indices (WRMR) are above cutoff value (1.0), they are close to it. And the comparative fits (CFI, TLI) are greater than 0.95, and parsimony corrections (RMSEA) are less than 0.08, therefore, the proposed eight-domain factor structure of SF-36 is supported. Results for white cases are comparable to the overall sample, which is reasonable given that over 80% of the overall cases are white. For black cases, although the RMSEA is greater than 0.08, the RMSEA is less than the rejection criteria (0.10 [Brown 2006]) and the absolute fit and the comparative fits are fall in the acceptable range. These results suggest that SF-36 can be used to measure the eight domains in the study sample, either as a racial diverse sample, or within each of the white and black races.

MIMIC

MIMIC models can be used to study the effects of the covariates on the eight latent traits (domains) measured by SF-36. In particular, the effects of the two covariates (race, AIS) on the eight latent traits measured by SF-36, assessed through MIMIC modeling are summarized in table 3. The MIMIC models are run separately on all of the study population and on those with complete information. An effect of AIS on the eight latent traits was not detected. However, race has some effect on the latent traits. Controlling for AIS, blacks have higher vitality and lower role-emotional compared with whites ($p\text{-value} < 0.05$).

The MIMIC modeling did not detect any DIF effect of for AIS or race on any of the 35 items used in the SF-36. This lack of detectable DIF implies measurement model invariance over AIS and race, and implies that differences in the latent traits of VT and RE between blacks and were not contaminated by DIF.

Sensitivity analysis

The results of the CFA and the MIMIC modeling using all cases and the complete set of cases were comparable (table 2, 3). These results suggest missingness has little effect on the results.

Application of SF-36 summary scores

On average, all of the scale scores of this sample are lower than the age and gender weighted norms of the general U.S. population [Ware et al 1993], with a higher difference in the four physical health scale scores (Figure 3). Our analysis of the measurement properties of the SF-36 in the study population suggests these observed differences are a result of lower function as opposed to the lack of measurement invariance in this population.

Discussion

SF-36 has been widely used in a number of epidemiologic and clinical settings. Indeed, it is likely the most commonly used instrument for the assessment of health status in the United States. Our objective was to assess the reliability and validity of SF-36 in an older sample of individuals, all of whom have traumatic brain injury (TBI). Our study includes 668 older TBI cases, broadly representative of the racial distribution in South Carolina, and is a subset of the population-based sample used to construct the TBI registry for South Carolina. The proportion of older whites and blacks in the final analysis (87.3% vs. 12.7%) is relatively comparable with the 2000 SC census data (78.3% vs. 20.9%) for this age group, the statewide TBI hospital discharge data (80.5% vs. 17.4%), and the TBI follow-up registry data (83.3% vs. 15.0%). Although there is evidence of attenuation of the proportion of older blacks in the subsequent datasets, this is partially explained by the higher refusal rates among blacks (46.9%) to participate in the study as compared to whites (39.6%). However, this difference is not statistically significant ($p=0.08$).

Our sample on average had lower health scores compared to population norms. These lower scores appear to be directly attributable to the lower health functioning of older persons with TBI [Flanagan et al 2005] as opposed to any fundamental measurement non-invariance or lack of validity of SF-36 in this population. The measurement invariance allows use of SF-36 for prediction of health status, with no additional statistical adjustment for injury severity or race. It should be noted that the sample from which these cases were drawn is restricted to individuals who were 55 years of age or older at the time of injury; hence, our findings are broadly interpretable to only an older population of individuals with TBI.

Our analyses also show remarkable robustness of the models to the various test statistics used to assess model fit. Different indices were used for assessment of model fit with consistent results irrespective of the model statistic. Analyses conducted on those with complete information ($n=551$) were in agreement with total cohort ($n=654$) adjusted for missing data. Hence, missing items did not alter our findings.

The discordant finding between injury severity measures (AIS) and SF-36 scale scores is quite perplexing. Not only did the eight scale scores of SF-36 failed to be associated with injury severity (data not shown), but also the latent variable analysis (MIMIC modeling) results indicate that injury severity did not predict SF-36 domain scores (table 3). This is primarily explained by an attribute of the injury severity measure (AIS). AIS is mainly intended to predict mortality after trauma, not long-term residual functional deficits [Association for the Advancement of Automotive Medicine 1990]. This has been noted in previous studies that failed to show the predictive abilities of injury severity measures on

cognitive functioning [Horner et al 2008] as well as psychosocial functioning and general health [McCarthy et al 2006]; while the predictive power of severity on risk of death is affirmed [Selassie et al 2005].

The significantly higher vitality and lower role-emotional scores noted among blacks suggest sociocultural differences in perception of dealing with physical and emotional stressors. Higher vitality score that express hardiness, strength, energy, not feeling worn out or tired even when confronting external stressors has been described as a coping style in African Americans reflecting personality traits often referred as 'John Henryism' [James 1994]. Persons with this trait are highly goal-oriented and succumbing to illness is perceived as weakness despite the realization that they may not have the emotional support and the resources needed to attain their goals [James 1984]. This latter phenomenon perhaps explains the discordance noted between high vitality and lower role emotional score, i.e., while acknowledging higher strength and energy there is self awareness of achieving less than what is desired.

A multivariate regression analysis of the eight scale scores obtained according to SF-36 manual on race and AIS also shows that only the effects of race on validity and role-emotional scale scores are significant (data not shown), which seems to be consistent with the results of MIMIC modeling. However, there are distinct differences between MIMIC and multivariate regression. First, multivariable regression cannot assess validity of the factor structure of SF-36 in the study sample. Second, multivariable regression cannot assess whether the effects of race and/or AIS on the scale scores are contaminated by DIF effects. MIMIC modeling is a powerful statistical analysis tool in a validity study and in assessing the effects of covariates on latent traits [Zumbo 2007].

This study has noteworthy limitations. First, there is differential participation rate between the races contributing to response bias. Fewer blacks than whites participated in the study relative to the original sample. However, the influence of this bias appears to be minimal since the rates are not statistically different. Second, while much of the data were gathered from medical records, a portion is self-report, which can introduce recall bias. Recall bias can be introduced when a person forgets or underestimates an exposure or when the exposure is exaggerated. Nonetheless, valid results have been shown from self-report from persons with TBI [Findler et al 2001]. Third, we used unconventionally lower cutoff age of 55 years to offset the early mortality experience of blacks, likely contributing to higher functional heterogeneity in functional limitations between the groups (mean age blacks=69.3 vs. whites 73.2; $P<0.05$). However, this is likely to have little influence on the analyses of reliability and factor structure as well as latent variable modeling since our finding demonstrated the measurement model to be invariant to injury severity and racial characteristics.

Despite the aforementioned limitations, our study has several noteworthy strengths. To our knowledge, this is the first study that investigated the factor structure of SF-36 in TBI. For instance, a widely cited study on this subject [Findler et al 2001] affirmed the reliability and validity of SF-36 on TBI patients without assessing its factor structure on the subjects analyzed. Second, this study is a subset of a large population-based cohort and a

representative sample of persons with TBI lending external validity to our findings. Third, while other studies have been published on smaller samples with inconsistent definitions of TBI, this study relied on a standard CDC case definition with chart reviews. Thus there is minimal case ascertainment bias in our findings. Fourth, our study employed a latent variable modeling (MIMIC) technique that is capable of assessing the effects covariates on latent traits while adjusting for DIF.

Taken together, this research supports the notion that injury severity and race do not confound the reliability and validity of SF-36 among older persons with TBI. One can take confidence in using SF-36 as a broad spectrum measure of health status in this population. Additional research on SF-36 among those with TBI could include an assessment of additional racial and ethnic categories, to examine test/retest reliability, and to evaluate this approach in other geographic locations. Additionally, new studies might include individuals with less severe forms of traumatic brain injury.

Acknowledgments

This study was supported in part by grant U17/CCU421926 from the National Center of Injury Prevention and Control (NCIPC), Centers of Disease Control and Prevention (CDC) (PI: Anbesaw Selassie). The opinions and conclusions expressed are solely those of the authors and should not be construed as representing the opinions and policy of the CDC. Other support was provided by EXCEED P01 1HS10871 from the Agency for Healthcare Research and Quality, and South Carolina Cooperative for Healthy Aging in Minority

References

- Association for the Advancement of Automotive Medicine. Des Plaines, IL: 1990. p. 1-4.
- Brown, TA. Confirmatory factor analysis for applied research. New York: The Guilford Press; 2006.
- Bullinger M. the TBI Consensus Group. Quality of Life in patients with TBI: basic issues, assessment, and recommendations - Results of Consensus Meeting. *Restor Neurol Neurosci*. 2002; 20(3-4): 111-124. [PubMed: 12454360]
- Center for Injury Research Policy of the Johns Hopkins University School of Public Health. ICDMAP-90 software. 1997
- Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med*. 2006; 119(2):166.e7-166.e16. [PubMed: 16443422]
- Copes WS, Champion HR, Sacco WJ, et al. The Injury Severity Score revisited. *J Trauma*. 1988; 28(1):69-77. [PubMed: 3123707]
- DeVellis, RF. Scale Development: Theory and Applications. 2nd Edition. Sage Publications; 2003.
- de Vet HC, Ader HJ, Terwee CB, et al. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of SF-36. *Qual Life Res*. 2005; 14(5):1203-1218. discussion 1219-1221, 1223-1224. [PubMed: 16047498]
- Findler M, Cantor J, Haddad L, et al. The reliability and validity of the SF-36 health survey questionnaire for use with individuals with traumatic brain injury. *Brain Inj*. 2001; 15(8):715-723. [PubMed: 11485611]
- Flanagan S, Hibbard M, Gordon W. The impact of age on TBI. *Phys Med Rehabil Clin N Am*. 2005; 16(1):163-177.
- Horner DM, Selassie AW, Lineberry LA, et al. Predictors of Psychological Symptoms one year after TBI: A Population-based Epidemiological Study. *J Head Trauma Rehabil*. 2008; 23(2):74-83. [PubMed: 18362761]
- James SA. John Henryism and the health of African Americans. *Culture, Medicine, and Psychiatry*. 1994; 18(2):163-182.

- James SA, LaCroix AZ, Kleinbaum DG, et al. John Henryism and blood pressure differences among black men: The role of occupational stressors. *J Behav Med.* 1984; 7(3):259–275. [PubMed: 6481796]
- Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: detecting differential item functioning using MIMIC modeling. *Med Care.* 2006; 44(Suppl 3):S124–S133. [PubMed: 17060819]
- Kline, RB. *Principles and Practice of Structural Equation Modeling.* New York, NY: The Guilford Press; 2005.
- Langlois, JA.; Rutland-Brown, W.; Thomas, KE. *Traumatic brain injury in the United States: Emergency department visits, hospitalizations, and deaths.* Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2006.
- MacKenzie EJ, Steinwachs DM, Shankar B. Classifying trauma severity based on hospital discharge diagnosis. *Med Care.* 1989; 27(4):412–422. [PubMed: 2649755]
- McCarthy ML, Dikmen SS, Langlois JA, et al. The Psychosocial Health Outcomes Reported At One Year Post Injury Among A Population-Based Sample Of Persons With Traumatic Brain Injury. *Arch Phys Med Rehabil.* 2006; 87:953–961.
- Muthén BO, Asparouhov T. Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. 2002 Mplus Web Notes: No. 4. www.statmodel.com.
- Muthén, LK.; Muthén, BO. *Mplus User's Guide.* Fifth Edition. Los Angeles, CA: Muthen & Muthen; 1998–2007.
- Muthén LK, Muthén BO. *Mplus Short Courses: Traditional Latent Variable Modeling Using Mplus.* 2003 www.statmodel.com.
- Pickelsimer EE, Selassie AW, Gu JK, et al. A population-based outcomes study of persons hospitalized with traumatic brain injury: operations of the South Carolina Traumatic Brain Injury Follow-up Registry. *J Head Trauma Rehabil.* 2006; 21(6):491–504. [PubMed: 17122680]
- Selassie AW, McCarthy ML, Ferguson PL, et al. Risk of post-hospitalization mortality among persons with Traumatic Brain Injury, South Carolina 1999–2001. *J Head Trauma Rehabil.* 2005; 20(3): 248–260.
- Styrke J, Stalnacke BM, Sojka P, et al. Traumatic brain injuries in a well-defined population: epidemiological aspects and severity. *J Neurotrauma.* 2007; 24(9):1425–1436. [PubMed: 17892405]
- Tilley, BC.; Wisdom, K.; Sadler, ZE., et al. Overcoming barriers to the inclusion of older African Americans in health outcome research. In: Curry, L.; Jackson, J., editors. *The science of Inclusion: Recruiting and Retaining Racial and Ethnic, Elders in the Health Research.* Vol. Chapter 8. Washington, DC: The Gerontological Society of America; 2003. p. 63–71.
- Ware JE Jr. SF-36 health survey update. *Spine.* 2000; 25(24):3130–3139. [PubMed: 11124729]
- Ware, JE.; Kosinski, M., et al. *SF-36 Physical and Mental Health Summary Scales: A User's Manual.* Booston, MA: The Health Institute; 1994.
- 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(6):473–483. [PubMed: 1593914]
- Ware, JE.; Snow, KK.; Kosinski, M., et al. *SF-36 Health Survey Manual and Interpretation Guide.* Boston, MA: New England Medical Center, The Health Institute; 1993.
- Zumbo, BD. Validity: foundational issues and statistical methodology. In: Rao, CR.; Sinharay, S., editors. *Handbook of Statistics.* Vol 26. Psychometrics. Elsevier; 2007. p. 45–79.

Fig. 1.
Flow chart for patient disposition and statistical analysis

Figure 2. The 8-factor MIMIC model with race and AIS as covariates for the SF-36.

The MIMIC model has two parts. The structural part is

$$\eta = \alpha + B\eta + \Gamma X + \zeta,$$

and the measurement part is

$$Y^* = v + \Lambda\eta + KX + \epsilon.$$

Here, η is the latent variables (SF-36 domains), B is the regression coefficients among latent variables (not shown), Γ is the regression coefficients of the latent variable on covariates, Y^* (not shown) is the continuous latent response variables, which represent Y through thresholds, Λ is factor loadings, K is the possible direct effect (DIF) from covariates to SF-36 items (dash lines).

Fig. 2.
Diagram of the MIMIC model

Fig. 3.
Comparison of SF-36 domain scores between weighted norms and TBI sample

Table 1

Demographic characteristics of the study sample (n=654)

Variable	Level	N	%
Age		72.4 ± 10.5	
Gender	Male	306	46.8
	Female	348	53.2
Race	Caucasian, non-Hispanic	571	87.3
	African American	83	12.7
Annual Income	<10K	199	33.7
	10K<= <20K	185	31.4
	20K<= <50K	159	26.9
	>=50K	47	8.0
	Missing	64	
Education	Less than high school	113	17.5
	Some high school	95	14.7
	High graduate	208	32.3
	Some college	111	17.2
	College graduate	118	18.3
	Missing	9	
Brain Injury Severity			
Injury Severity Score, ISS		13.5 ± 6.9	
Abbreviated Injury Score, AIS	1. Minor	0	0
	2. Moderate	192	29.4
	3. Serious	113	17.3
	4. Severe	329	50.3
	5. Critical	20	3.0
	6. Maximum	0	0

Table 2

Scale scores and reliabilities of SF-36 domains, and CFA results

		Overall	White	Black
Scale Scores				
All Cases (SF-36 Domains)	N	654	571	83
	Physical Functioning (PF)	41.3 ± 34.2	41.8 ± 34.5	38.2 ± 31.9
	Role - Physical (RP)	44.3 ± 41.2	44.6 ± 41.6	42.2 ± 38.0
	Bodily Pain(BP)	57.3 ± 28.7	57.8 ± 29.0	54.1 ± 26.7
	General Health (GH)	50.8 ± 26.3	51.3 ± 26.8	47.9 ± 22.8
	Vitality(VT) ^a	44.0 ± 26.3	43.2 ± 26.8	49.6 ± 21.4
	Social Functioning(SF)	67.6 ± 34.2	67.1 ± 34.6	70.9 ± 31.2
	Role – Emotional(RE)	74.0 ± 39.0	75.3 ± 38.7	65.5 ± 39.8
	Mental Health(MH) ^b	69.8 ± 24.2	69.6 ± 24.7	70.8 ± 19.9
Cronbach's Alpha				
Complete Cases (SF-36 Domains)	N	551	477	74
	PF	0.95	0.95	0.94
	RP	0.86	0.88	0.77
	BP	0.89	0.90	0.81
	GH	0.81	0.83	0.69
	VT	0.87	0.88	0.69
	SF	0.89	0.90	0.81
	RE	0.87	0.89	0.76
	MH	0.89	0.90	0.79
CFA Results				
Complete Cases	N	551	477	74
	CFI	0.963	0.972	0.956
	TLI	0.993	0.994	0.971
	RMSEA	0.068	0.066	0.095
	WRMR	1.064	0.984	0.843
All Cases	N	654	571	83
	CFI	0.960	0.966	0.960
	TLI	0.992	0.993	0.975
	RMSEA	0.070	0.069	0.089
	WRMR	1.176	1.095	0.836

^a. N = 639, 557, and 82 for overall, white, and black.

^b. N = 635, 554, and 81 for overall, white, and black.

Table 3

Summary of MIMIC modeling results ^a

Model Fit Indices	All Cases (N = 654)					Complete Cases(N = 551)				
	CFI	TLI	RMSEA	WRMR		CFI	TLI	RMSEA	WRMR	
	0.962	0.992	0.060	1.129		0.968	0.993	0.063	1.021	
Covariate Effects	race ^b		AIS			race ^b		AIS		
	gamma	p-value	gamma	p-value		gamma	p-value	gamma	p-value	
PF	-0.094	0.458	-0.063	0.170		-0.080	0.546	-0.060	0.230	
RP	-0.061	0.640	-0.037	0.467		-0.047	0.732	-0.032	0.557	
BP	-0.149	0.234	0.039	0.396		-0.088	0.517	0.040	0.424	
GH	-0.133	0.263	-0.008	0.871		-0.092	0.477	-0.007	0.898	
VT	0.279	0.015	-0.014	0.760		0.275	0.023	-0.026	0.594	
SF	0.104	0.441	0.004	0.934		0.119	0.402	-0.003	0.953	
RE	-0.304	0.026	-0.061	0.264		-0.300	0.036	-0.083	0.168	
MH	0.006	0.962	-0.052	0.259		-0.045	0.880	-0.061	0.228	

^a . See table 2 for explanation of abbreviation of SF-36 domains.

^b , race is coded as 0-White, 1-Black.