

Evaluating the SF-36 & SF-12 in Diverse Populations

Barbara C. Tilley, Rickey E. Carter,
Chengwu Yang
Medical University of South Carolina



Acknowledgments

- Analytical Team
 - Becky Boehm, Xiaobu Ye
- Collaborators
 - Neil Armstrong, J Beals, S Gregorich, Kathryn Magruder, K Markides, E Mutran, Koko Markides, Kenneth Ottenbacher, Kristen Peek, Eiseo Perez-Stable, Laura Ray, P Reed, D Rhoades, Anbesaw Selassie, A Springer
- PIs of studies contributing data
 - Anbesaw Selassie (TBI Registry);
 - Kathryn Magruder (VA PTSD Prevalence)
 - Koko Markides (Texas Hispanic Established Populations for Epidemiologic Study of the Elderly)
 - Elizabeth Mutran (North Carolina, RCMAR)
 - Eiseo Perez-Stable (California Pathfinders, RCMAR)

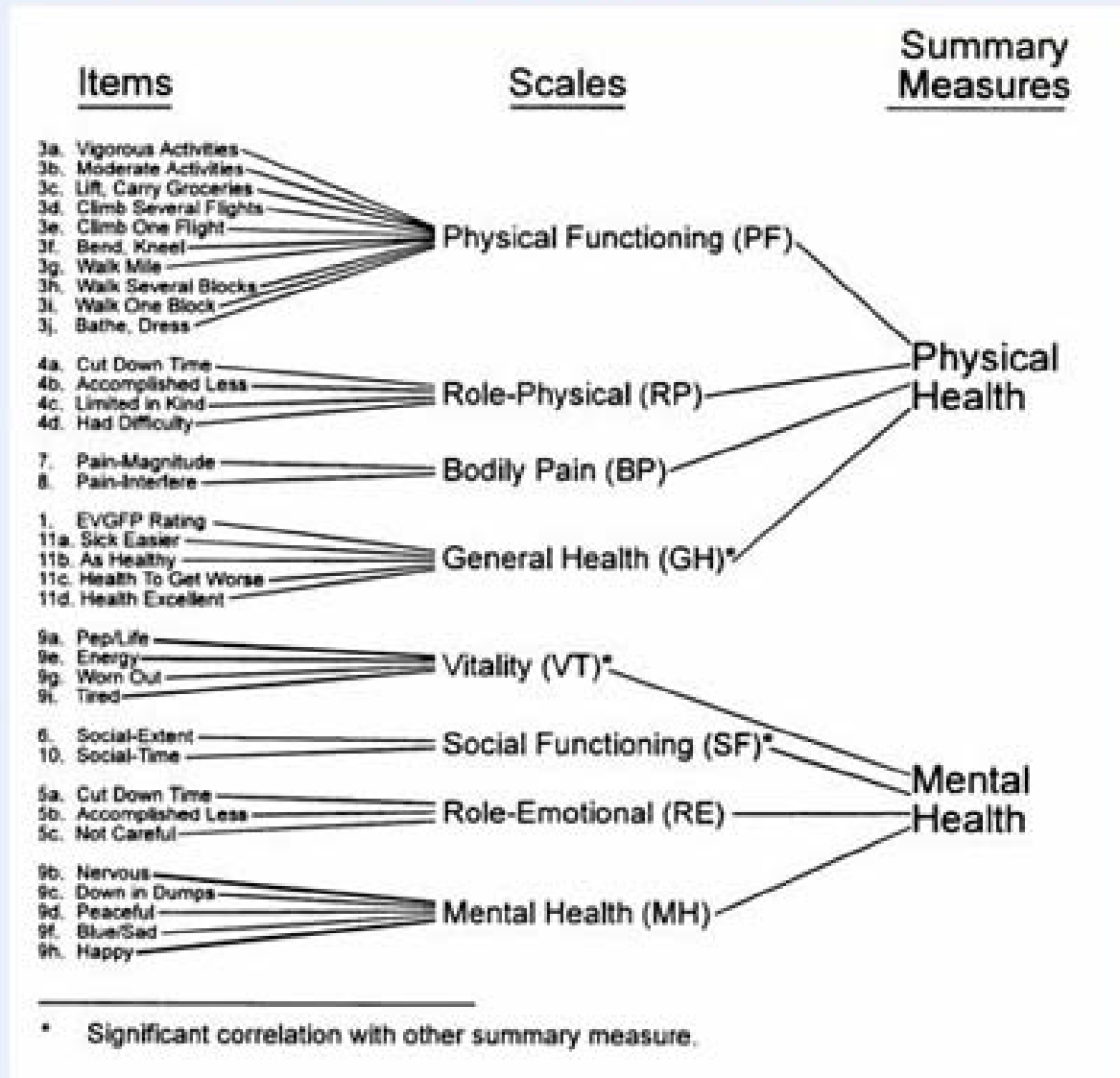
NIH-NIA: P30 AG15286, P30 AG021677

NIH-AHRQ: P01 HS10871

Primary Aims

- Aim 1: Evaluate the advantages and barriers in using existing SF-36 and SF-12 data from racially/ethnically diverse populations
- Aim 2: Validate the SF-36 and SF-12 measurement models using existing data collected in diverse populations aged 55 years or older.

SF-36[®] Measurement Model



Background

Studies in Diverse Elderly Populations

- McHornery, et al. (Med Care 1994)
 - MOS study (3 cities; HMOs and med. practices)
 - Race, age, income, education separately
 - Various properties for older, black
 - Completeness (less in older)
 - Floor/ceiling effects (little relationship to age/race)
 - Item-internal consistency (high)
 - Item-discriminant validity (high)
 - Reliability coefficient (high)
- Woolinsky & Stump (Med Care 1996)
 - Disadvantaged older black and white men/women
 - 8 factor solution acceptable
 - Possible 9th factor (health optimism)

Background Continued

- Walters et al (Age and Ageing 2001)
 - Aged 65 or more, Britain
 - Questionnaire reordered, “e.g.’s” modified to fit population
 - No CFA analysis, no presentation of racial subgroups
 - Cronbach’s alpha ≥ 0.79
 - >93% complete questionnaire
- Peek et al (Geron. 2004)
 - Older Mexican Americans in Southwest in large epi study (EPESE)
 - 8-factor solution
 - Valid, reliable
 - Factor scores higher than reported national scores

Data Sources

- South Carolina
 - MUSC TBI Registry (SF-36)
 - N=654: 87% C, 13% AA; 53% Female

- South East US
 - VA PTSD Prevalence Study (SF-36)
 - N=1005: 61% C, 39% AA; 24% Female (oversampled)

- Texas (Peek et al, 2004)
 - Community Dwelling Hispanics (SF-36)
 - N=621, 100% H; 59% Female

Data Sources Continued

- North Carolina
 - Community (26 items of SF-36)
 - N=221: 100% AA
- California Bay Area
 - Women with abnormal mammogram (SF-12)
 - N=486: 67% C, 14% AA, 12% H, 7% Multi/Other; 100% Female

Methodology

- Prepare and merge data (“Pool”)
- Analysis Overview
 - Internal consistency and summary statistics
 - Confirmatory factor analysis (CFA) – Global measurement model fit
 - CFA with covariates – Measurement invariance & population heterogeneity
 - Differential Item Function (DIF) – Response bias

Pooling issues

- SF-36 Standards
 - Theory: Font, pagination, response scales, scoring standardized
 - Practice: “Editorial” adaptations implemented
- 4 SF-36 Datasets & 4 SF-36 versions
 - 1 correct
 - 1 with only 26 questions (excluded from analysis)
 - 1 reordered questions
 - 1 with incorrect response scales for two items

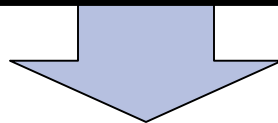
Case 1: “Has your health limited your social activities (like visiting with close friends or close relatives?” [As asked, not correct]

- Correct response scale (5 point):

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

- Questionnaire response scale (6 point):

All of the time	Most of the time	A good bit of the time	Some of the time	A little <u><i>bit</i></u> of the time	None of the time
-----------------	-------------------------	-------------------------------	------------------	--	------------------

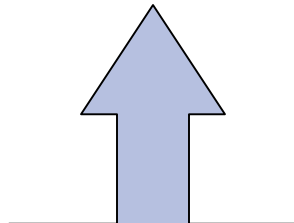


Solution: Combine into “Most of the time”

Case 2: “How much bodily pain have you had during the past four weeks?”

- Correct response scale (6 points):

None (6.0)	Very mild (5.4)	Mild (4.2)	Moderate (3.1)	Severe (2.2)	Very severe (1.0)
---------------	--------------------	---------------	-------------------	-----------------	----------------------



Questionnaire response scale (5 points):

Not at all (6.0)	Slightly 48% (4.8)	Moderately (3.1)	Quite a bit (2.2)	Extremely (1.0)
---------------------	--------------------------	---------------------	----------------------	--------------------

DIF Barriers

- Need original item response profile to assess DIF
- One dataset (Texas Hispanics) scored before arrival
- Scoring involves reverse scoring, recalibration of items, and mean imputation
- Solution: “Unscore” when possible
 - Mean imputed values often can be distinguished from rescaled items
 - Delete imputed values before DIF analysis
 - Set some items to missing

Missing Data

- For Pooled SF-36 Analysis (N=2280)
 - After Mean Imputation for difference in scale scoring:
 - 95% completed
- Analysis Plan
 - For CFA: Imputed values set to missing
 - Sensitivity Analysis: Complete Case vs. Full Information Maximum Likelihood (FIML)

Selected FIML Results

CFA Results for Pooled SF-36 Data				
Model	Chi-Square	CFI	RMSEA	
	/ df			
Full Data Set (N=2280)				
CFA: FIML	9.18	0.92	0.06	A
African American and Caucasians (N=1659)				
CFA: FIML	7.63	0.92	0.06	B
MIMIC (Race Covariate): FIML	7.46	0.91	0.06	B'
Hispanics and Caucasians (N=1805)				
CFA: FIML	7.74	0.92	0.06	C
MIMIC (Race Covariate): FIML	7.90	0.92	0.06	C'

- LRT: B' to B & C' to C $p < 0.05$

CFA Interpretation

- Data is consistent with several potential models
 - Chi-square / df test above threshold for good fit
 - MIMIC: Effect of race on measurement model appears negligible on global fit indices
 - LRT: At least one effect of race on factors may be statistically significant
- Complete case (not presented) and FIML results comparable

DIF

- DIF was not identified in sample
- Peek et al (2005): Hispanics higher factor scores
- Our results of measurement invariance and lack of DIF suggest this finding is not a result of bias or measurement error

Validation of SF-12

- SF-12 Items extracted from SF-36
- Measurement Model Fit:
 - Poor model fit indices in CFA
- CFA may not be appropriate
 - SF-12 derived from a PCA analysis of SF-36
 - Not developed as a theoretical measurement model

Take Away

- “Meta” validation logistically difficult
- All SF-36’s are not the “same”
- Creative methods to methodological challenges may lead to questionable findings
- Findings consistent with other studies of the SF-36
- SF-36 Measurement model robust



- Questions

Extra Slides



Example Unscoring

During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

As received			After		
(Scored)	N	Percent	"Unscoring"	N	Percent
1	21	3.38	1	271	43.92
2	59	9.5	2	211	34.2
3	55	8.86	3	55	8.91
3.1	2	0.32	4	59	9.56
4	211	33.98	5	21	3.4
4.2	1	0.16	Frequency Missing = 4		
5	95	15.3			
5.4	1	0.16			
6	176	28.34			

Example Analysis Plan for MUSC TBI data

