THEORY AND METHODS

Retest reliability of surveillance questions on health related quality of life

E M Andresen, T K Catlin, K W Wyrwich, J Jackson-Thompson

J Epidemiol Community Health 2003;**57**:339–343

See end of article for authors' affiliations

Correspondence to:
Dr E M Andresen, Saint Louis University School of Public Health, Department of Community Health, Salus Center, 3545 Lafayette Avenue Suite 300, St Louis, MO 63104-1399, USA; andresen@slu.edu

Accepted for publication
8 January 2003

Study objectives: Health related quality of life (HRQoL) is an important surveillance measure for monitoring the health of populations, as proposed in the American public health plan, *Healthy People 2010*. The authors investigated the retest reliability of four HRQoL questions from the US Behavioral Risk Factor Surveillance System (BRFSS).

Design: Randomly sampled BRFSS respondents from the state of Missouri were re-contacted for a retest of the HRQoL questions. Reliability was estimated by κ statistics for categorical questions and intraclass correlation coefficients for continuous questions.

Setting: Missouri, United States.

Participants: 868 respondents were re-interviewed by telephone about two weeks after the initial interview (mean 13.5 days). Participants represented the adult, non-institutionalised population of Missouri: 59.1% women; mean age 49.5 years; 93.2% white race.

Main results: Retest reliability was excellent (0.75 or higher) for Self-Reported Health and Healthy Days measures, and moderate (0.58 to 0.71) for other measures. Reliability was lower for older adults. Other demographic subgroups (for example, gender) showed no regular pattern of differing reliability and there was very little change in reliability by the time interval between the first and second interview.

Conclusions: Retest reliability of the HRQoL Core is moderate to excellent. Scaling options will require future attention, as will research into appropriate metrics for what constitutes important population group differences and change in HRQoL.

Within the public health arena, emphasis has been placed on measuring outcomes such as "quality of life" and, more specifically, "health related quality of life" (HRQoL). Health related quality of life refers to the "physical, psychological, and social domains of health, seen as distinct areas that are influenced by a person's experiences, beliefs, expectations, and perceptions".¹ HRQoL can itself be defined as a health issue of importance, and variations by person, place and time can be important to understanding population differences.^{2–6} In addition, HRQoL measures can be outcomes of other health events such as access to care,⁷ impact of chronic conditions such as arthritis,⁸ or the effects of aging.⁹ Therefore, the surveillance of HRQoL is important to monitoring changes in the health of people and populations.

One of the two overarching goals stated in the American public health plan, *Healthy People 2010*, is to increase the quality and years of healthy life.¹⁰ This goal will be met by increasing life expectancy and improving HRQoL among people and communities. *Healthy People 2010* proposes three surveillance tools for measuring HRQoL: Self-Reported Health, Healthy Days, and Years of Healthy Life.¹⁰

The Behavioral Risk Factor Surveillance System (BRFSS), supported by the Centers for Disease Control and Prevention, is an ongoing, state based survey that measures health behaviours and risk factors of the US non-institutionalised adult population, through random digit dialled telephone interviews.^{11–12} Since 1993, the BRFSS has included four questions designed to monitor HRQoL (HRQoL Core). The BRFSS HRQoL Core includes questions used to measure both Self-Reported Health and Healthy Days mentioned above as recommended tools for monitoring HRQoL. Despite the importance of these tools in determining whether individuals and populations within the US achieve the first goal of *Healthy People 2010*, few studies have examined the psychometric properties of these items.

The objective of this study was to investigate the retest reliability of the BRFSS HRQoL Core, specifically Self-Reported

Health and Healthy Days, among a population of 868 adults from the US state of Missouri. The reliability of these questions was measured within the total group of retest respondents, as well as by subgroups based on gender, age, race, income, and education in order to examine the measurement properties among different segments of the population.

METHODS

Measures

The BRFSS HRQoL Core (also called the Healthy Days module) has standardised survey and scoring methods, and established comparative population norms.^{5 13–16} The HRQoL Core includes a global assessment question that measures Self-Reported Health (that is, "Would you say that in general your health is excellent, very good, good, fair, or poor?"). This single question has a substantial research track for its usefulness in community studies.¹⁷ Also included in the HRQoL Core are two questions reporting the number of days during the previous 30 days in which the respondent's physical or mental health was not good. The sum of these two measures results in the total number of "unhealthy days" (ranging from 0 to 30 days). This summary measure of HRQoL is usually expressed in the positive, that is Healthy Days, and represents the number of recent days in which both the respondent's physical and mental health were good. Healthy Days is calculated by subtracting the number of "unhealthy days" from 30. The fourth question of the HRQoL Core measures the number of days in the previous 30 days in which the respondent experienced activity limitation because of poor physical or mental health, and is asked only of those people who reported at least one day of poor physical or mental health.

Abbreviations: HRQoL, health related quality of life; BRFSS, Behavioral Risk Factor Surveillance System

Table 1 Descriptive characteristics US Missouri BRFSS retest respondents (n=868), 1999

Demographic characteristic	Time 1	
Women, % (n)	59.1 (513)	
Age		
Years, median, mean (SD)	48.0, 49.5 (17.8)	
65 or older, % (n)	23.5 (204)	
Ethnic group, % (n)		
White	93.2 (809)	
African-American	4.4 (38)	
Married, % (n)	57.1 (496)	
Employed, % (n)	60.6 (526)	
Educational level, % (n)		
Less than high school	16.7 (145)	
High school graduate	36.8 (319)	
Some college	23.5 (204)	
College graduate	23.0 (200)	
Annual household income, % (n)		
less than \$15,000	24.7 (194)	
\$15,000–34,999	36.9 (290)	
\$35,000–49,999	20.0 (157)	
\$50,000 or more	18.4 (145)	
Interval between interviews		
Days, median, mean (SD)	13.0, 13.5 (4.6)	
3–13 days, % (n)	51.0 (443)	
BRFSS HRQoL Measure	Time 1	Time 2
Self reported Health, % (n)		
Excellent	18.9 (164)	16.2 (140)
Very good	30.9 (268)	27.8 (241)
Good	30.4 (264)	35.8 (310)
Fair	13.2 (115)	13.9 (120)
Poor	6.6 (57)	6.4 (55)
Poor physical health days		
Days, median, mean (SD)	0.0, 4.5 (8.7)	0.0, 4.4 (8.8)
0 days, % (n)	58.8 (507)	60.5 (515)
30 days, % (n)	7.7 (67)	7.9 (67)
Poor mental health days		
Days, median, mean (SD)	0.0, 3.2 (7.3)	0.0, 3.0 (7.4)
0 days, % (n)	66.0 (573)	71.8 (613)
30 days, % (n)	4.6 (40)	4.7 (40)
Healthy days		
Days, median, mean (SD)	28.0, 23.1 (10.1)	30.0, 23.6 (10.3)
0 days, % (n)	11.2 (95)	12.5 (105)
30 days, % (n)	44.1 (375)	50.5 (425)
Limited activity days		
Days, median, mean (SD)	0.0, 2.2 (6.6)	0.0, 2.1 (6.4)
0 days, % (n)	79.5 (686)	81.9 (705)
30 days, % (n)	3.7 (32)	3.5 (30)
Frequent mental distress, % (n)	9.6 (82)	10.0 (85)
Frequent physical distress, % (n)	14.7 (127)	13.9 (118)

In addition, the mental health days have been used as a dichotomous measure, with 14 or more poor mental health days representing “frequent mental distress”.¹⁵ For this analysis, we added a similar dichotomous measure for physical health days (that is, “frequent physical distress”). Standard BRFSS questions on gender, age, race/ethnicity, marital status, education, employment, and income provided demographic data for this sample.

Subjects

The 1999 Missouri BRFSS included 4277 subjects (adults 18 years of age or older). Data were collected through computer assisted telephone interviews using random digit dialling and disproportionate random sampling.^{18–20} During the months of January, February, and March 1999, 1459 potential respondents were identified and 1114 people were interviewed, for a response of 76.4%. Following standard BRFSS procedures, multiple attempts were made to re-interview 1015 respondents; 99 respondents were excluded from attempts to conduct retest interviews primarily because of the telephone laboratory schedule: respondents who completed interviews at the

beginning of January and end of March were not scheduled for retest. During February, March, and early April 1999, 889 re-interviews were conducted, for a response of 87.6%. Of these, 21 interviews contained missing, incomplete or incorrect information, leaving a total of 868 test-retest subjects. Retest interviews of the HRQoL questions were made about two weeks after the initial call (range 3–34 days; mean 13.5 days, standard deviation (SD) 4.6 days).

Analyses

Primary analyses were conducted using SPSS statistical software.²¹ Retest sample characteristics are reported as either unweighted means and standard deviations or proportions. We report skewness of the continuous measures (that is, floor and ceiling effects); the distribution was considered skewed if 20% or more of subjects were grouped at either extreme.²² In this study, a ceiling effect refers to good health (that is, 0 poor health days or 30 Healthy Days) and floor effect refers to poor or bad health. We also compared the prevalence of BRFSS data from Missouri overall, and the USA overall, to the retest sample using weighted data from the CDC BRFSS data web site.²³

Table 2 Retest reliability and agreement of BRFSS HRQoL measures among a sample of US Missouri retest respondents (n=868), 1999

Sample	Self reported health*	Poor physical health days†	Poor mental health days†	Limited activity days†	Healthy days†	Frequent mental distress‡	Frequent physical distress‡
All (n=868)							
Response number (n)	866	846	845	857	828	845	846
Reliability (95% CI)	0.75 (0.71 to 0.79)	0.71 (0.67 to 0.74)	0.67 (0.63 to 0.71)	0.57 (0.52 to 0.61)	0.75 (0.72 to 0.78)	0.58 (0.49 to 0.67)	0.64 (0.56 to 0.71)
% Agreement or means	62%	4.3, 4.4	3.3, 2.9	2.2, 2.1	23.3, 23.6	93%	91%
Men (n=355)							
Reliability (95% CI)	0.74 (0.67 to 0.80)	0.76 (0.71 to 0.80)	0.72 (0.66 to 0.77)	0.52 (0.44 to 0.60)	0.76 (0.71 to 0.80)	0.68 (0.53 to 0.82)	0.70 (0.59 to 0.82)
% Agreement or means	61%	4.1, 4.2	3.0, 2.3	2.3, 1.9	24.0, 24.6	95%	93%
Women (n=513)							
Reliability (95% CI)	0.76 (0.71 to 0.81)	0.67 (0.62 to 0.71)	0.64 (0.58 to 0.69)	0.61 (0.55 to 0.66)	0.74 (0.70 to 0.78)	0.53 (0.41 to 0.65)	0.59 (0.49 to 0.69)
% Agreement or means	64%	4.5, 4.6	3.4, 3.4	2.2, 2.2	22.7, 22.9	91%	90%
Ages 18–64 (n=661)							
Reliability (95% CI)	0.75 (0.70 to 0.79)	0.73 (0.69 to 0.76)	0.70 (0.66 to 0.74)	0.60 (0.55 to 0.65)	0.76 (0.73 to 0.79)	0.61 (0.50 to 0.71)	0.62 (0.53 to 0.72)
% Agreement or means	65%	3.7, 3.8	3.6, 3.2	2.0, 2.0	23.6, 24.0	93%	93%
Ages 65+ (n=204)							
Reliability (95% CI)	0.67 (0.58 to 0.76)	0.66 (0.57 to 0.73)	0.55 (0.44 to 0.64)	0.50 (0.39 to 0.60)	0.72 (0.64 to 0.78)	0.46 (0.22 to 0.70)	0.63 (0.50 to 0.76)
% Agreement or means	56%	6.5, 6.7	2.2, 2.2	3.2, 2.4	22.1, 22.2	93%	86%
White (n=809)							
Reliability (95% CI)	0.76 (0.71 to 0.79)	0.70 (0.66 to 0.73)	0.67 (0.63 to 0.71)	0.55 (0.50 to 0.60)	0.74 (0.71 to 0.77)	0.57 (0.47 to 0.67)	0.63 (0.55 to 0.71)
% Agreement or means	63%	4.4, 4.5	3.2, 2.9	2.2, 2.1	23.2, 23.6	92%	91%
African American (n=38)							
Reliability (95% CI)	0.68 (0.49 to 0.86)	0.86 (0.74 to 0.92)	0.56 (0.30 to 0.74)	0.82 (0.68 to 0.90)	0.83 (0.69 to 0.91)	0.53 (0.06, 1.0)	0.64 (0.17, 1.1)
% Agreement or means	53%	3.0, 3.4	3.9, 2.8	2.7, 1.7	23.9, 23.6	92%	94%
Education: Less than high school (n=145)							
Reliability (95% CI)	0.72 (0.63 to 0.81)	0.75 (0.67 to 0.82)	0.51 (0.38 to 0.63)	0.67 (0.56 to 0.75)	0.74 (0.66 to 0.81)	0.49 (0.27 to 0.70)	0.72 (0.60 to 0.85)
% Agreement or means	63%	8.3, 8.8	3.7, 4.0	3.6, 3.0	19.7, 19.5	88%	88%
Education: High school graduate/GED (n=319)							
Reliability (95% CI)	0.67 (0.59 to 0.74)	0.70 (0.64 to 0.76)	0.59 (0.52 to 0.66)	0.37 (0.27 to 0.46)	0.71 (0.65 to 0.76)	0.53 (0.36 to 0.69)	0.60 (0.47 to 0.73)
% Agreement or means	55%	4.5, 4.5	3.2, 2.7	1.9, 2.2	22.9, 23.4	92%	91%
Education: Some college (n=204)							
Reliability (95% CI)	0.77 (0.69 to 0.84)	0.65 (0.56 to 0.72)	0.82 (0.78 to 0.86)	0.69 (0.61 to 0.76)	0.81 (0.76 to 0.86)	0.70 (0.54 to 0.85)	0.59 (0.42 to 0.76)
% Agreement or means	68%	3.8, 3.7	3.8, 3.6	2.8, 2.2	23.7, 24.1	94%	91%
Education: College grad (n=200)							
Reliability (95% CI)	0.71 (0.61 to 0.80)	0.49 (0.38 to 0.59)	0.80 (0.75 to 0.85)	0.60 (0.51 to 0.68)	0.69 (0.60 to 0.75)	0.58 (0.34 to 0.83)	0.32 (0.02 to 0.62)
% Agreement or means	70%	1.9, 2.1	2.5, 1.9	1.2, 1.1	25.8, 26.2	95%	94%
Annual income: Less than \$15000 (n=194)							
Reliability (95% CI)	0.75 (0.67 to 0.82)	0.74 (0.66 to 0.80)	0.55 (0.44 to 0.64)	0.72 (0.64 to 0.78)	0.74 (0.66 to 0.80)	0.44 (0.26 to 0.63)	0.70 (0.58 to 0.81)
% Agreement or Means	60%	8.7, 8.8	4.3, 4.3	4.2, 3.7	18.8, 18.7	87%	87%
Annual income: \$15000–\$34999 (n=290)							
Reliability (95% CI)	0.70 (0.62 to 0.78)	0.64 (0.57 to 0.71)	0.67 (0.60 to 0.73)	0.40 (0.30 to 0.49)	0.70 (0.64 to 0.76)	0.60 (0.45 to 0.76)	0.58 (0.41 to 0.74)
% Agreement or means	66%	3.3, 3.3	3.2, 2.8	1.9, 1.5	24.2, 24.7	93%	93%
Annual income: \$35000–\$49999 (n=157)							
Reliability (95% CI)	0.67 (0.57 to 0.78)	0.69 (0.59 to 0.76)	0.90 (0.86 to 0.92)	0.61 (0.50 to 0.70)	0.88 (0.83 to 0.91)	0.76 (0.59 to 0.93)	0.61 (0.39 to 0.83)
% Agreement or means	59%	2.6, 3.5	3.7, 3.2	1.4, 1.3	24.2, 24.2	95%	94%
Annual income: \$50000+ (n=145)							
Reliability (95% CI)	0.72 (0.62 to 0.83)	0.70 (0.60 to 0.77)	0.78 (0.70 to 0.83)	0.68 (0.58 to 0.76)	0.76 (0.68 to 0.82)	0.60 (0.31 to 0.89)	0.46 (0.18 to 0.74)
% Agreement or means	66%	2.5, 2.3	2.3, 1.9	1.3, 1.6	25.5, 26.0	96%	93%
Retest interval: 3–7 days (n=64)							
Reliability (95% CI)	0.83 (0.73 to 0.92)	0.66 (0.49 to 0.79)	0.83 (0.73 to 0.89)	0.50 (0.28 to 0.66)	0.73 (0.58 to 0.83)	0.63 (0.30 to 0.96)	0.68 (0.38 to 0.97)
% Agreement or means	69%	4.5, 4.3	3.0, 2.7	2.7, 3.2	23.1, 23.7	94%	93%
Retest interval: 8–21 days (n=748)							
Reliability (95% CI)	0.74 (0.70 to 0.78)	0.70 (0.66 to 0.74)	0.66 (0.62 to 0.70)	0.60 (0.55 to 0.64)	0.75 (0.72 to 0.78)	0.59 (0.49 to 0.69)	0.63 (0.55 to 0.72)
% Agreement or means	62%	4.2, 4.4	3.2, 3.0	2.2, 2.0	23.4, 23.6	93%	91%
Retest interval: 22+ days (n=56)							
Reliability (95% CI)	0.82 (0.72 to 0.92)	0.81 (0.70 to 0.88)	0.63 (0.44 to 0.76)	0.34 (0.08 to 0.55)	0.73 (0.59 to 0.84)	0.44 (0.07 to 0.82)	0.64 (0.37 to 0.90)
% Agreement or means	61%	6.0, 5.6	4.0, 2.8	2.1, 2.5	21.8, 23.3	89%	89%

*Weighted κ (95% CI) and % agreement. †Intraclass correlation coefficient (95% CI) and means (time 1, time 2). ‡ κ (95% CI) and % agreement.

Descriptive reliability results are presented showing subject mean responses for time one and time two, or per cent agreement for categorical items. Reliability of continuous measures (that is, responses to “days” questions) was evaluated by intraclass correlation coefficients (ICC) using a two way random effects model.²⁴ For dichotomous measures (for example, frequent mental distress) we used κ statistics.²⁵ A weighted κ statistic was computed for the Self-Reported Health question to account for partial agreement among ordered categories.²⁶ We classified ICC and κ statistics above 0.75 as excellent agreement and below 0.40 as poor agreement.²⁵ Reliability is reported for the entire sample and for demographic subgroups that are often reported to demonstrate differing HRQoL (age, gender, ethnicity, education, and income). Because of the small number of respondents who reported they were other than white or African American, analyses of potential ethnic differences were restricted to only these two groups. We also report reliability for groups based on the time interval between test and retest.

RESULTS

Participants of the retest study were 59.1% women; the average age was 49.5 years old (range 18 to 93), with 23.5% of the sample aged 65 or older. Most (93.2%) was white. Table 1 shows the demographic characteristics of the sample in more detail. In general, retest subjects were not different in a systematic manner from the entire 1999 BRFSS sample.²³ For example, among all Missouri respondents to the BRFSS in 1999, 14.5% reported they had less than a high school education, 52.6% were women, and 19.6% were aged 65 and older. Table 1 also lists the descriptive results for the four BRFSS HRQoL Core questions and three computed measures. For the four “days” measures, the percentage of people who reported their HRQoL to be either at the minimum or maximum health (that is, floor or ceiling effects) is also shown. None of the measures demonstrated floor effects for poor health, however, all four measures showed ceiling effects (that is, 44% reported a maximum of 30 Healthy Days and 80% reported zero limited activity days at time 1). The skewed distributions also are evident in that the three “days” questions from the HRQoL Core result in a median of zero days, with means ranging between 2.2 and 4.5 days. Missing data were not a large problem with the BRFSS questions. Response was 97% or higher for all questions with the exception of baseline household income question (9.4% missing). When data were paired for the reliability estimates the response remained above 97%, although the Healthy Days measure was missing for 4.6% of subjects because it relies on two questions (table 2).

The Missouri sample was compared with previously published HRQoL data from other US states and the nation.^{5 16 23} HRQoL was somewhat better in US samples than in the Missouri sample, for example in 1999, American states reported that 22.7% of BRFSS respondents were in excellent health and 3.3% in poor health, compared with 18.9% and 6.6% for the Missouri retest sample.

Retest reliability of the seven measures ranged from 0.58 (frequent mental distress) to 0.75 (Self-Reported Health and Healthy Days) for the entire sample. Reliability was lower for the computed categorical measures of mental and physical distress compared with the original continuous “days” version of these questions. As shown in table 2, the level of reliability varied somewhat among subgroups. However, the only consistent group difference was for lower reliability among older adult respondents, and there was a slight tendency for lower reliability for longer time intervals between test and retest. There was no consistent pattern of different reliability by gender, ethnic group, education, or income. In a post hoc analysis, we examined if there were differences for time intervals among demographic subgroups and found no relation between the interval and any variable.

Key points

- Public health surveillance activities often focus on specific events, like mortality or cancer. However, health related quality of life (HRQoL) is included as an overarching aspect of the American public health plan.
- Surveillance questions on HRQoL can be used to examine different outcomes for specific conditions, like arthritis or heart disease, and to detect general health disparities.
- Retest reliability of the BRFSS measures for HRQoL seems to be moderate to strong. There are no major systematic demographic differences in reliability except for lower estimates for older adults.

DISCUSSION

The retest reliability of the four Core HRQoL questions of the BRFSS and three computed measures (that is, Healthy Days) was moderate to excellent. This reinforces the positive earlier results of retest reliability for these questions from a sample of people with disability,²⁷ and for retest reliability of the BRFSS in general.^{18 28 29} Not surprisingly, reliability decreased somewhat for older adults, and with longer intervals between tests. Longer retest times are naturally associated with more real health changes, and older adults have more chronic conditions and variable health.³⁰ However, there was no evidence that reliability varied systematically by other demographic characteristics of typical BRFSS samples. The Healthy Days summary measure had slightly higher reliability than each of its component measures, physical health and mental health. This suggests that the summary item is a more consistent measure and that both components are important elements of HRQoL.

The κ values for the dichotomous categories of “days” measures (mental health and physical health) were low compared with per cent agreement; however, this is partly due to the low prevalence of respondents categorised with “frequent distress,” which biases κ values.³¹ In addition, the reliability of the dichotomous measures was lower than that of the corresponding continuous measures, suggesting that information was lost in the categorisation and misclassification may have occurred near the cut off value. Considering the brief surveillance nature of these questions, it is not surprising that the “days” questions result in somewhat skewed distributions. Future research will need to examine the merits of categorical scaling of these questions. Also important to the future use of the questions is the issue of what constitutes meaningful group differences. If an acceptable criterion measure for HRQoL states can be adopted, ROC analyses would be useful in developing thresholds for determining, for example, how many poor physical health days constitutes an important decrement in HRQoL.

The BRFSS HRQoL Core questions have been subjected to other tests of their measurement characteristics within the US. These include generally positive cognitive testing,³² and positive qualitative reports by Native American respondents in New Mexico.⁶ Research among adults with disability also has been good, including the finding that the questions are acceptable and pose very low burden on respondents.^{27 33} There is some evidence that acceptability of BRFSS surveys, including HRQoL, differs by interviewer type: Gilliland and colleagues found that Native American subject response to Native American interviewers was 77% compared with 69% for other interviewer ethnic groups.⁶ On the other hand, the response for the 1999 BRFSS baseline study was higher than that for similar but smaller studies previously conducted within the US states of Missouri (76.4% *v* 66%),¹⁸ Massachusetts (65%),²⁹ or New York (59.7%),²⁸ while the retest response was comparable to the earlier Missouri study (87.6% *v* 88%) and higher than the other two studies (70% and 55.6%, respectively).

This study is based on BRFSS retest respondents from the single state of Missouri; the sample reported somewhat lower

HRQoL than the general US population, a finding that has been reported previously.^{3,5} However, there is no known evidence to think that slight differences in population levels of health would affect the reliability patterns reported here. The slightly larger older age group of the Missouri retest sample could produce lower reliability overall. For samples with disability or chronic health problems, reliability might vary more than reported here. Importantly, the level of reliability was lower than the 0.75 cut off used to rate excellent agreement for all but two measures, suggesting that there will be some error in population estimates based on the HRQoL questions. Retest reliability of the HRQoL Core should be assessed in other US states and population subgroups, as well as internationally. If results are consistent with our findings, then we can be assured that the BRFSS HRQoL surveillance core is strong.

ACKNOWLEDGEMENTS

We are grateful to Missouri BRFSS Project staff (Dr. Theophile Murayi, Steve Kilfoil, Jane Heath, Jacqui Turner and Pat Nelson) and to telephone interviewers from Kelly Services. We extend our special thanks for the support and advice of our CDC colleagues, Dave Moriarty and Matt Zack.

Authors' affiliations

E M Andresen, T K Catlin, K W Wyrwich, Saint Louis University School of Public Health, Department of Community Health, Salus Center, St Louis, USA

K W Wyrwich, Saint Louis University College of Public Service, Department of Research Methodology, St Louis, USA

J Jackson-Thompson, Missouri Department of Health, Office of Surveillance, Research and Evaluation, Jefferson City, USA

Funding: this work was supported, in part, with funding from the Health Care and Aging Studies Branch, Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion (CDC; R13/CCR717041-01), and the Behavioral Surveillance Branch (CDC; U58/CCU700950-14). Support also was provided by the CDC through the Saint Louis University Prevention Research Center (CDC; U48/CCU710806).

Conflicts of interest: none.

REFERENCES

- 1 **Testa M**, Simonson D. Current concepts: assessment of quality of life outcomes. *N Engl J Med* 1996;**334**:835–40.
- 2 **Andresen EM**, Brownson RC. Disability and health status: ethnic differences among women in the United States. *J Epidemiol Community Health* 2000;**54**:200–6.
- 3 **Andresen EM**, Prince-Caldwell A, Akin F, et al. The Missouri Disability Epidemiology and Health Project. *Am J Prev Med* 1999;**16** (suppl):63–71.
- 4 **Centers for Disease Control and Prevention**. Factors associated with self-perceived excellent and very good health among blacks—Kansas, 1995. *MMWR Morb Mortal Wkly Rep* 1996;**45**:906–11.
- 5 **Centers for Disease Control and Prevention**. State differences in reported healthy days among adults—United States, 1993–1996. *MMWR Morb Mortal Wkly Rep* 1998;**47**:239–43.
- 6 **Gilliland FD**, Mahler R, Davis SM. Health-related quality of life for rural American Indians in New Mexico. *Ethn Health* 1998;**3**:223–9.
- 7 **Fouts BS**, Andresen EM, Hagglund K. Disability and satisfaction with access to health care. *J Epidemiol Community Health* 2000;**54**:770–1.
- 8 **Centers for Disease Control and Prevention**. Factors associated with prevalent self-reported arthritis and other rheumatic conditions—United States, 1989–1991. *MMWR Morb Mortal Wkly Rep* 1996;**45**:487–91.
- 9 **Campbell VA**, Crews JE, Moriarty DG, et al. Surveillance for sensory impairment, activity limitation, and health-related quality of life among older adults—United States, 1993–1997. *MMWR CDC Surveill Summ* 1999;**48**:131–56.
- 10 **US Department of Health and Human Services**. *Healthy People 2010* (Conference edition, in two volumes). Washington, DC: US Department of Health and Human Services, 2000. (available at <http://www.health.gov/healthypeople>).
- 11 **Gentry EM**, Kalsbeek WD, Hogelin GC, et al. The behavioral risk factor surveys: II. Design, methods, and estimates from combined state data. *Am J Prev Med* 1985;**1**:9–14.
- 12 **Remington PL**, Smith MY, Williamson DF, et al. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–87. *Public Health Rep* 1988;**103**:366–75.
- 13 **Centers for Disease Control and Prevention**. Quality of life as a new public health measure—Behavioral Risk Factor Surveillance System, 1993. *MMWR Morb Mortal Wkly Rep* 1994;**43**:375–80.
- 14 **Centers for Disease Control and Prevention**. Health-related quality-of-life measures—United States, 1993. *MMWR Morb Mortal Wkly Rep* 1995;**44**:195–200.
- 15 **Centers for Disease Control and Prevention**. Self-reported frequent mental distress among adults—United States, 1993–1996. *MMWR Morb Mortal Wkly Rep* 1998;**47**:326–31.
- 16 **Hennessy CH**, Moriarty DG, Zack MM, et al. Measuring health-related quality of life for public health surveillance. *Public Health Rep* 1994;**109**:665–72.
- 17 **Idler EL**, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;**38**:21–37.
- 18 **Brownson RC**, Jackson-Thompson J, Wilkerson JC, et al. Reliability of information on chronic disease risk factors collected in the Missouri Behavioral Risk Factor Surveillance System. *Epidemiology* 1994;**5**:545–9.
- 19 **Jackson-Thompson J**, Hagan R, Wilkerson J, et al. From basic BRFSS to special surveys: the collection, analysis and use of state, area and local data. In: *Proceedings of the 1991 Public Health Conference on Records and Statistics*. Hyattsville, MD: US Department of Health and Human Services, National Center for Health Statistics, 1992; DHHS publication no (PHS) 92–1214:337–342.
- 20 **Siegel PZ**, Brackbill RM, Frazier EL, et al. Behavioral risk factor surveillance, 1986–1990. *MMWR CDC Surveill Summ* 1991;**40**:1–23.
- 21 **SPSS**. *SPSS Release 10.0*. Chicago: SPSS, 1999.
- 22 **Andresen EM**. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000;**81** (suppl 2):S15–20.
- 23 **Behavioral Risk Factor Surveillance System website**. *Interactive prevalence data*. (<http://www.cdc.gov/brfss>) (Missouri 1999 BRFSS data). Accessed 23 July 2002.
- 24 **Fleiss JL**. *The design and analysis of clinical experiments*. New York, NY: Wiley, 1986.
- 25 **Fleiss JL**. *Statistical methods for rates and proportions*. New York, NY: Wiley, 1981.
- 26 **Armstrong BK**, White E, Saracci R. *Principles of exposure measurement in epidemiology*. New York, NY: Oxford University Press, 1992.
- 27 **Nanda U**, Andresen EM. Performance of measures of health-related quality of life and function among disabled adults. [Abstract]. *Qual Life Res* 1998;**7**:644.
- 28 **Shea S**, Stein AD, Lantigua R, et al. Reliability of the behavioral risk factor survey in a triethnic population. *Am J Epidemiol* 1991;**133**:489–500.
- 29 **Stein AD**, Lederman RI, Shea S. The Behavioral Risk Factor Surveillance System questionnaire: its reliability in a statewide sample. *Am J Public Health* 1993;**83**:1768–72.
- 30 **Andresen EM**, Gravit GW, Aydelotte ME, et al. Limitations of the SF-36 in a sample of nursing home residents. *Age Ageing* 1999;**28**:562–6.
- 31 **Agresti A**, Ghosh A, Bini M. Ranking kappa: describing potential impact of marginal distributions on measures of agreement. *Biomed J* 1995;**37**:811–20.
- 32 **Schechter S**, Beatty P, Willis GB. Asking survey respondents about health status: judgement and response issues. In: Schwartz N, Park DC, Knauper B, et al., eds. *Cognition, aging, and self-reports*. Philadelphia, PA: Psychology Press, 1999.
- 33 **Andresen EM**, Fouts BS, Romeis JC, et al. Performance of health-related quality-of-life instruments in a spinal cord injured population. *Arch Phys Med Rehabil* 1999;**80**:877–84.