

The Impact of Alternative Response Scales
on Measuring Self-ratings of Health

Tom W. Smith

NORC/University of Chicago

GSS Methodological Report No. 103

July, 2005

Introduction

Following the First Law of Studying Societal Change, the General Social Survey (GSS) strives for consistent measurement over time by employing constant measures (Smith, 2005). However, in certain cases measures have been changed for various reasons. When such alterations have been introduced, the GSS has introduced the revised version in a controlled manner, typically using some combination of across-subjects experiments and within-subjects repetition (Smith, 2005). This report considers a possible change in the GSS measure of self-rated health (HEALTH).

Self-rated Health

Since 1972, the GSS has included a general measure of self-rated health (HEALTH - Would you say your own health, in general, is excellent, good, fair, or poor?). This simple measure is widely used in health studies and has proven to be a notable predictor of mortality and other health outcomes even with other variables such as specific health history and medical evaluations controlled for (Ferraro and Farmer, 1999; Hardy and Pavalko, 1986; Idler and Angel, 1990; Idler and Benyamini, 1997; Perry, et al., 1996; Remle, 2004; Siegel, 1994).

The specific wording used by the GSS was adopted from Gallup surveys in 1941 and 1950 (Davis, Smith, and Stephenson, 1980). In the 1970s about half of the major, national studies using the self-rated health item employed a 4-category, response scale and half used a 5-category version (Danchik and Drury, 1986). When the National Health Interview Survey was redesigned in 1982, it switched from a 4-category version to a 5-category format (Kovar and Poe, 1985). Consistent with that decision, virtually all health surveys conducted for the US government now use 5-category versions (e.g. the National Health Examination Survey, the Health and Retirement Study, the Study of Assets and Health Dynamics, the Behavioral Risk Factors Surveillance Study), as do most other health scales (e.g. SF-36). Besides the GSS, relatively few studies continue to employ a 4-category version (see e.g. Ratner, Johnson, and Jeffery, 1998).¹

The NHIS study switched to five categories in order "to improve the ability to differentiate among people" (Kovar and Poe, 1985) and others have preferred it for similar reasons (Danchik and Drury, 1986). The unarticulated expectation of the finer measurement was that the item would more accurately measure

¹For methodological studies of the meaning of the self-rated health measure and how evaluations are done by respondents see Groves, Fultz, and Martin, 1992; Mallinson, 2002; Schechter, 1993; Schulster, 1994; Singer, 1994.

health status and produce stronger associations with related health variables and demographics (Alwin, 1992).

Table 1 examines the impact of the 4- and 5-category response scales on marginals. Table 1A looks at non-experimental comparisons and Table 1B includes experiments. Of course the adding of the fifth "very good" category takes responses from the more positive "excellent" option and the less positive "good" option and reduces both. The declines in "excellent" ranges from 4.9 percentage points to 16.8 points and "good" decreases from between 15.4 points to 21.2 points. There is considerable difference as to whether most of the "very good" responses appear to come from "excellent" or "good". The decline in "excellent" apparently contributes as little as 19.5% of the "very good" responses (Table 1B-2) to as much as 61.5% (Table 1BA-3). The differences are even notable within just the experimental studies. There is little impact on the distribution of "fair" and "poor" response across response scales. Levels in the 4-category items differ from the 5-category items by only from -1.9 points to +2.0 points and average just +0.5 points. An within-subjects design among employed adults on the 2002 GSS confirms the very limited impact on these two more negative responses. Only 3.7% of those giving "very good" on the 5-category scale gave responses of "fair" or "poor" on the 4-category version.

Table 2 compares the correlations that the 4- and 5-category health measures have with other variables. Overall there is no meaningful difference in the strength or statistical significance of associations. In 14 cases correlations were higher for the 4-category measure and in 14 cases for the 5-category wording. The average absolute correlations were .130 for the former and .132 for the latter. In only two cases do notable differences appear. In 2004, but not in 2002, the 4-category measure was more highly associated with an assessment of on how many days in the last month ones "mental health was not good" ($r=.285$) than the 5-category scale was ($r=.061$). In the second case the interviewers evaluation of whether the respondent was over- or underweight was correlated more highly with the 5-category questions ($r=.230$) than with 4-categories ($r=.130$). Thus, the two notable differences out of 28 comparisons are in off-setting directions and the difference in correlations with poor, mental-health days was notable in only one of the two years compared.

The lack of any meaningful and consistent difference in correlations is not surprising based on the fact that several previous GSS studies showed little or no impact on associations of using response scales with more categories (Peterson, 1985;

Smith, 1994a; 1994b).² It is also expected from the fact that on the 2002 GSS the correlation between 4- and 5-category health items is 0.85 and if Excellent on the 4-category scale is considered consistent with Excellent or Very Good on the 5-category scale and likewise Good with Very Good or Good, that 93.6% of the cases are on the diagonal when crosstabulating the two items. Also, as indicated above, there is little impact on the bottom two categories and at least one researcher argues that the "predictive value of self-rated health is driven by ratings of fair or good health" (Singer, 1994).

Summary

This evaluation of the 4- and 5-category, self-rated, health items indicates that 1) there is no discernable difference in the explanatory power of the two scales, 2) there are major shifts in the distributions at the positive end of the distribution, but little or none at the negative end, and 3) the variation in the contributions from Excellent and Good to the added Very-Good option would not allow trends in these categories to be reliably estimated across scales and, as a result, would restrict trend analysis combining both 4- and 5-category data points to comparing the bottom two responses with the combined top two or three categories.

²Alwin (1992) found a slight increase in reliability as one moved from 4 to more than 4 categories, but Davis, Wellens, and DeMaio (1996) found no gains from moving from 4 categories to 5 or 6 categories.

Table 1

Comparisons of the Distributions of Self-Rated Health
Using 4 or 5 Response Options

A. Non-Experimental

1. 1981 and 1982 National Health Interview Survey (NHIS)

	1981	1982
Excellent	42.0%	32.2%
Very Good	----	25.4
Good	41.2	25.8
Fair	12.7	11.5
Poor	4.1	5.1

Source: Danchik and Drury, 1986

2. 1979 NHIS and 1979 Fourth Quarter Evaluation Study (FQES)

	NHIS	FQES
Excellent	42.8%	30.6%
Very Good	----	28.8
Good	40.3	24.7
Fair	12.8	11.4
Poor	4.1	4.4

Source: Danchik and Drury, 1985

3. 1976 NHIS and National Health and Nutrition Examination Survey II (ages 20-74)

	NHIS	NHANES
Excellent	43.9%	27.1%
Very Good	----	27.3
Good	40.1	27.9
Fair	11.9	12.5
Poor	3.7	5.0
Missing	0.4	0.2

Source: Forthofer, 1983

Table 1 (continued)

B. Experimental

1. NHIS Experiments, 1979

	Standard	Variant
Excellent	48.0%	36.0% ^a
Very Good	----	28.0
Good	39.0	21.0
Fair	10.0	8.0
Poor	3.0	3.0

Source: Kovar and Poe, 1985

2. General Social Survey, 2002 (Employed People)

	Standard	Variant
Excellent	35.9%	31.0%
Very Good	----	25.1
Good	48.7	30.0
Fair	13.3	12.1
Poor	2.1	1.8
	(1193)	(1186)

Source: GSS, see Davis, Smith, and Marsden, 2005

3. General Social Survey, 2004

	Standard	Variant
Excellent	35.7%	26.3%
Very Good	----	30.6
Good	47.8	26.5
Fair	12.2	11.4
Poor	4.3	5.3
	(466)	(517)

Source: GSS, see Davis, Smith, and Marsden, 2005

^aTotal adds up to only 96% in original source.

Table 2

Correlates of 4-Category and 5-Category Health Self-Ratings

(Pearson's r/probability)

A. 2002 GSS (Employed People)

	4-Category	5-Category
Age (AGE)	.027/.355	.044/.129
Gender (SEX)	-.023/.425	.013/.644
Race (RACE)	.064/.026	.040/.160
Education (EDUC)	-.196/.000	-.177/.000
Occ. Prestige (PRESTGE80)	-.149/.000	-.165/.000
Attend Church (ATTEND)	-.091/.002	-.075/.009
Frequency of Praying (PRAY)	.028/.497	.005/.902
Happiness (HAPPY)	.258/.000	.234/.000
Life Exciting (LIFE)	.223/.000	.201/.000
Physical Health (PHYSHLTH)	.316/.000	.313/.000
Mental Health (MNTLHLTH)	.224/.000	.213/.000
Health Days, Month (HLTHDAYS)	.178/.000	.188/.000
Feel Used Up by Job (USEDUP)	-.140/.000	-.165/.000
Suffer Back Pain (BACKPAIN)	-.154/.000	-.175/.000
Pain in Arms (PAINARMS)	-.126/.000	-.154/.000
Hurt at Work (HURTATWK)	.050/.034	.043/.137
Gov Health Spending (NATHEAL)	-.057/.047	-.069/.017
Medical Confidence (CONMEDIC)	.125/.028	.117/.039

B. 2004 GSS (All Adults)

Age (AGE)	.198/.000	.181/.000
Gender (SEX)	.008/.868	-.078/.076
Race (RACE)	.028/.530	.043/.718
Education (EDUC)	-.274/.000	-.328/.000
Occ. Prestige (PRESTGE80)	-.184/.000	-.218/.000
Attend Church (ATTEND)	-.035/.447	.015/.732
Mental Health (MNTLHLTH)	.285/.000	.061/.256
Jon Stress (WRKSTRESS)	-.050/.049	-.122/.000
Gov Health Spending (NATHEAL)	.021/.338	-.041/.773
Respondent's Weight Judged by Interviewer (INTRWGHT)	.131/.051	.230/.000

Note: Variables names are in parentheses and these items can be found in Davis, Smith, and Marsden, 2005.

References

- Alwin, Duane F., "Information Transmission in the Survey Interview: Number of Response Categories and the Reliability of Attitudes Measurement," in Sociological Methodology, 1992. edited by Peter V. Marsden. Washington, DC: Blackwell, 1992.
- Danchik, Kathleen M. and Drury, Thomas F., "Evaluating the Effects of Survey Design and Administration on the Measurement of Subjective Phenomena: The Case of Self-Assessed Health Status," Proceedings of the Survey Methods Research Section, 1985. Washington, DC: American Statistical Association, 1986.
- Davis, James A.; Smith, Tom W.; and Marsden, Peter V., General Social Surveys, 1972-2004: Cumulative Codebook. Chicago: NORC, 2005.
- Davis, James A.; Smith, Tom W.; and Stephenson, C. Bruce., General Social Surveys, 1972-1980: Cumulative Codebook. Chicago: NORC, 1980.
- Davis, Wendy; Wellens, Tracy R.; and DeMaio, Theresa J., "Designing Response Scales in an Applied Setting," Census Working Paper in Survey Methodology, 96/7, US Bureau of the Census, 1996.
- Ferraro, Kenneth F. and Farmer, Melissa M., "Utility of Health Data from Social Surveys: Is There a Gold Standard for Measuring Morbidity?" American Sociological Review, 64 (1999), 303-315.
- Forthofer, Ronald N., "Investigation of Nonresponse Bias in NHANES II," American Journal of Epidemiology, 117 (1983), 507-515.
- Groves, Robert M.; Fultz, Nancy H.; and Martin, Elizabeth, "Direct Questioning about Comprehension in a Survey Setting," in Questions about Questions: Inquiries into the Cognitive Bases of Surveys, edited by Judith M. Tanur. New York: Russell Sage Foundation, 1992.
- Hardy, Melissa and Pavalko, Eliza K., "The Internal Structure of Self-Reported Health Measures Among Older Male Workers and Retirees," Journal of Health and Social Behavior, 27 (1986), 346-357.

- Idler, Ellen L. and Angel, R.J., "Self-Reported Health and Mortality in the NHANES-I Epidemiological Follow-up Study," American Journal of Public Health, 80 (1990), 446-452.
- Idler, Ellen, L. and Benyamini, Yael, "Self-rated Health and Mortality: A Review of Twenty-Seven Community Studies," Journal of Health and Social Behavior, 38 (1997), 21-37.
- Kovar, Mary Grace and Poe, G.S., "The National Health Interview Design, 1973-1984 and Procedures, 1975-1983," Vital and Health Statistics, Series 1, No. 18. Washington, DC: GPO, 1985.
- Mallinson, S., "Listening to Respondents: A Qualitative Assessment of the Short-Form 36 Health Status Questionnaire," Social Science Medicine, 54 (1982), 11-21.
- Perry, M. et al., "Factors Associated with Self-Perceived Excellent and Very Good Health Among Blacks," MMWR, 45 (Oct. 25, 1996), 906-911.
- Petersen, Bruce L., "Confidence: Categories and Confusion," GSS Methodological Report No. 31. Chicago: NORC, 1985.
- Ratner, P.A.; Johnson, J.L.; and Jeffery, B., "Examining Emotional, Physical, Social, and Spiritual Health as Determinants of Self-rated Health Status," American Journal of Health Promotion, 12 (1998), 275-282.
- Remle, R. Corey, "Self-rated Health Trajectories: Alternative Measures of Perceived Change in Health as a Predictor of Mortality," Paper presented to the American Sociological Association, San Francisco, August, 2004.
- Schechter, Susan, "Investigation into the Cognitive Process of Answering Self-Assessed Health Status Questions," CDC Working Paper No. 2. Washington, DC: National Center for Health Statistics, 1993.
- Schulster, Jerome R., "Health and Self: Paths for Exploring Cognitive Aspects Underlying the Self-Report of Health Status," Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status, edited by Susan Schechter. NCHS Working Paper Series, No. 10, 1994.
- Siegel, Paul Z., "Self-Reported Health Status: Public Health Surveillance and Small-Area Analysis," Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-

Reported Health Status, edited by Susan Schechter. NCHS Working Paper Series, No. 10, 1994.

Singer, Eleanor, "Self-Rated Health Status: How Are Judgements Made?" Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status, edited by Susan Schechter. NCHS Working Paper Series, No 10, 1994.

Smith, Tom W., "A Comparison of Two Confidence Scales," GSS Methodological Report No. 80. Chicago: NORC, 1994a.

Smith, Tom W., "A Comparison of Two Government Spending Scales," GSS Methodological Report No. 80. Chicago: NORC, 1994b.

Smith, Tom W., "The Laws of Studying Societal Change," Survey Research, (2005), forthcoming.