

ARTICLES

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

Tom W Smith¹

¹ NORC, University of Chicago (IL)

Keywords: survey practice

<https://doi.org/10.29115/SP-2008-0004>

Survey Practice

Vol. 1, Issue 1, 2008

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

Introduction

Following the First Law of Studying Societal Change, the General Social Survey (GSS) strives for consistent measurement over time by employing constant measures.¹ However, in certain cases measures have been changed for various reasons. When such alterations occur, the GSS has introduced the revised version in a controlled manner, typically using some combination of across-subjects experiments and within-subjects repetition. This procedure is important so that variation due to measurement effects is not confounded with studying true change. This report considers a possible change in the GSS measure of self-rated health.

Self-rated Health

Since 1972, the GSS has included a self-rated health measure (HEALTH – Would you say your own health, in general, is excellent, good, fair, or poor?). This simple item is widely used in health studies and is a notable predictor of mortality and other health outcomes, even controlling for other variables such as specific health history and medical evaluations.² The GSS wording came from Gallup surveys in 1941 and 1950. In the 1970s about half of major US national studies measuring self-rated health employed a 4-category response scale and half used a 5-category version (Danchik and Drury 1986). When the National Health Interview Survey (NHIS) was redesigned in 1982, it switched from a 4-category version to a 5-category format.³ Consistent with that decision, virtually all US governmental, health surveys now use 5-category versions (e.g. the National Health Examination Survey, the Health and Retirement Study, the Study of Assets and Health

¹ The GSS is the largest and longest-term project of the Sociology Program of the National Science Foundation. It has conducted 26 national, in-person, full-probability surveys of adults living in US households between 1972 and 2006 (Davis et al. 2007).

² See Hardy and Pavalko 1986; Idler and Angel 1990; Siegel 1994; Perry et al. 1996; Idler and Benyamini 1997; Ferraro and Farmer 1999; Remle 2004.

³ The NHIS is the main, continuous health monitoring study of the household population conducted by the US government. For more information see www.cdc.gov/nchs/about/major/nhis/hisdesc.htm. On the switch see Kovar and Poe (1985).

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

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.⁴

As Kovar and Poe (1985) note, the NHIS study switched to five categories in order “to improve the ability to differentiate among people” and others have preferred it for similar reasons. The unarticulated expectation was that the finer measurement would more accurately measure health status and produce stronger associations with health variables and demographics.

On the GSS and other studies a variety of comparisons between the different response scales used for the subjective health measures exists. These include non-experimental comparisons and experiments using both inter-subject and intra-subject designs. Table 1 examines the impact of the 4- and 5-category response scales on marginals. Table 1A looks at non-experimental comparisons in which different surveys of similar populations were conducted at approximately the same time and Table 1B covers intra- and inter-subjects experiments. In the intra-subjects design people were asked both versions of the self-rated health question in different parts of the survey. In the inter-subjects design, different random samples were given 4 or 5 categories versions.

Table 1 Comparisons of the Distributions of Self-Rated Health
Using 4 or 5 Response Options

A. Non-Experimental

B. Experimental (Intra- and Inter-Subject Designs)

Adding 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” range 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”

⁴ On the meaning of the self-rated health measure and how evaluations are done by respondents see Groves et al. 1992; Mallinson 2002; Schechter 1993; Schulster 1994; Singer, 1994.

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 1986

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

1. NHIS Inter-Subjects 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

B.1.Variant total adds up to only 96% in original source.

Source: Kovar and Poe 1985

2. General Social Survey, 2002 (Intra-Subjects Experiment; 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

apparently contributes as little as 19.5% of the “very good” responses (Table 1A-2) to as much as 61.5% (Table 1A-3). The differences are even notable within the experimental studies. There is little impact on the distribution of “fair” and “poor” response across response scales. An intra-subjects design among employed adults on the 2002 GSS confirms the very limited impact

3. General Social Survey, 2004 (Inter-subjects experiment)

	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

on these two more negative responses. The impact of the changes in response scales on distributions is large, but variable, making any simple comparison across the response scales difficult.

Next, the associates of health are examined (Table 2). This examines whether the two items reveal the same structural relationships, and tests the hypothesis that the finer scale yields stronger correlations. Overall there is no meaningful difference in the strength or statistical significance of associations. The average absolute correlations were 0.130 for the former and 0.132 for the latter.

Table 2 Correlates of 4-Category and 5-Category Health Self-Ratings (Pearsons r /probability)

The lack of any meaningful and consistent difference in correlations is not surprising since several previous GSS studies showed little or no impact on associations of using response scales with more categories.⁵ It is also expected because 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 means that 93.6% of the cases are on the diagonal when crosstabulating the items. Also, as indicated above, there is little impact on the bottom two categories and Singer (1994) argues that the “predictive value of self-rated health is driven by ratings of fair or poor health”.

Summary

This evaluation of self-rated, health items indicates that 1) no discernable difference in the explanatory power of the two scales occurs, 2) major shifts in the distributions happen at the positive end, but little at the negative end, 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

⁵ On the GSS, see Peterson 1985; Smith 1994a,b. Alwin (1992) found a slight increase in reliability moving from 4 to more than 4 categories, but Davis et al. (1996) found no gains between 4 categories and 5–6 categories.

A. 2002 GSS (Employed People)

	4-Category	5-Category
Age (AGE)	0.027/0.355	0.044/0.129
Gender (SEX)	-0.023/0.425	0.013/0.644
Race (RACE)	0.064/0.026	0.040/0.160
Education (EDUC)	-0.196/0.000	-0.177/0.000
Occ. Prestige (PRESTGE80)	-0.149/0.000	-0.165/0.000
Attend Church (ATTEND)	-0.091/0.002	-0.075/0.009
Frequency of Praying (PRAY)	0.028/0.497	0.005/0.902
Happiness (HAPPY)	0.258/0.000	234/0.000
Life Exciting (LIFE)	0.223/0.000	0.201/0.000
Physical Health (PHYSHLTH)	0.316/0.000	0.313/0.000
Mental Health (MNTLHLTH)	0.224/0.000	0.213/0.000
Health Days, Month (HLTHDAYS)	0.178/0.000	0.188/0.000
Feel Used Up by Job (USEDUP)	-0.140/0.000	-0.165/0.000
Suffer Back Pain (BACKPAIN)	-0.154/0.000	-0.175/0.000
Pain in Arms (PAINARMS)	-0.126/0.000	-0.154/0.000
Hurt at Work (HURTATWK)	0.050/0.034	0.043/0.137
Gov Health Spending (NATHEAL)	-0.057/0.047	-0.069/0.017
Medical Confidence (CONMEDIC)	0.125/0.028	0.117/0.039

B. 2004 GSS (All Adults)

	4-Category	5-Category
Age (AGE)	0.198/0.000	0.181/0.000
Gender (SEX)	0.008/0.868	-0.078/0.076
Race (RACE)	0.028/0.530	0.043/0.718
Education (EDUC)	-0.274/0.000	-0.328/0.000
Occ. Prestige (PRESTGE80)	-0.184/0.000	-0.218/0.000
Attend Church (ATTEND)	-0.035/0.447	0.015/0.732
Mental Health (MNTLHLTH)	0.285/0.000	0.061/0.256
Job Stress (WRKSTRESS)	-0.050/0.049	-0.122/0.000
Gov Health Spending (NATHEAL)	0.021/0.338	-0.041/0.773
Respondent's Weight Judged by Interviewer (INTRWGHT)	0.131/0.051	0.230/0.000

Note: Variables names are in parentheses and these items can be found in Davis et al. 2005.

responses with the combined top two or three categories, and 4) correlations across studies using the 4- and 5-category scale might be compared since they do not produce different estimates.

The large impact of the shift in response scales over part of the distribution and the unexpected nil impact on correlations underscores that survey researchers must be careful whenever changing methods. Changing methods should always be presumed to muddy, if not eviscerate, valid comparisons.

Additionally, changes will often not yield the improvements expected. When modifications are introduced, experiments and other rigorous designs should be utilized and any expected improvements need to be verified.

REFERENCES

- Alwin, D.F. 1992. "Information Transmission in the Survey Interview: Number of Response Categories and the Reliability of Attitudes Measurement." In *Sociological Methodology*, edited by P.V. Marsden. Washington, DC: Blackwell.
- Danchik, K.M., and T.F. Drury. 1986. *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. Washington, DC: American Statistical Association.
- Davis, J.A., T.W. Smith, and P.V. Marsden. 2007. *General Social Surveys, 1972-2006: Cumulative Codebook*. Chicago: NORC.
- Davis, W., T.R. Wellens, and T.J. DeMaio. 1996. "Designing Response Scales in an Applied Setting." *Working Paper in Survey Methodology*, 96/7, US Bureau of the Census.
- Ferraro, K.F., and M.M. Farmer. 1999. "Utility of Health Data from Social Surveys: Is There a Gold Standard for Measuring Morbidity?" *American Sociological Review* 64:303–15.
- Forthofer, R.N. 1983. "Investigation of Nonresponse Bias in NHANES II." *American Journal of Epidemiology* 117:507–15.
- Groves, R.M., N.H. Fultz, and E. Martin. 1992. "Direct Questioning about Comprehension in a Survey Setting." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, edited by J.M. Tanur. New York: Russell Sage.
- Hardy, M., and E.K. Pavalko. 1986. "The Internal Structure of Self-Reported Health Measures among Older Male Workers and Retirees." *Journal of Health and Social Behavior* 27:346–57.
- Idler, E.L., and R.J. Angel. 1990. "Self-Reported Health and Mortality in the Nhanes-i Epidemiological Follow-up Study." *American Journal of Public Health* 80:446–52.
- Idler, E.L., and Y. Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38:21–37.
- Kovar, M.G., and G.S. Poe. 1985. "The National Health Interview Design, 1973-1984 and Procedures, 1975-1983." *Vital and Health Statistics Series 1* (18).
- Mallinson, S. 1982. "Listening to Respondents: A Qualitative Assessment of the Short-Form 36 Health Status Questionnaire." *Social Science Medicine* 54:11–21.
- Perry, M. et al. 1996. "Factors Associated with Self-Perceived Excellent and Very Good Health among Blacks." *MMWR* 45:906–11.
- Petersen, B.L. 1985. "Confidence: Categories and Confusion." GSS Methodological Report No. 31. Chicago: NORC.
- Ratner, P.A., J.L. Johnson, and B. Jeffery. 1998. "Examining Emotional, Physical, Social, and Spiritual Health as Determinants of Self-Rated Health Status." *American Journal of Health Promotion* 12:275–82.
- Remle, R.C. 2004. "Self-Rated Health Trajectories: Alternative Measures of Perceived Change in Health as a Predictor of Mortality." In *The American Sociological Association*. San Francisco.
- Schechter, S. 1993. "Investigation into the Cognitive Process of Answering Self-Assessed Health Status Questions." CDC Working Paper No. 2. Washington, DC: NCHS.
- Sehulster, J.R. 1994. "Health and Self: Paths for Exploring Cognitive Aspects Underlying the Self-Report of Health Status." In *The 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status*, edited by S. Schechter. Vol. NCHS Working Paper No. 10.
- Siegel, P.Z. 1994. "Self-Reported Health Status: Public Health Surveillance and Small-Area Analysis." Edited by S. Schechter. *The 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status* NCHS Working Paper Series No. 10.

- Singer, E. 1994. "Self-Rated Health Status: How Are Judgements Made?" Edited by S. Schechter. *The 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status* NCHS Working Paper No. 10.
- Smith, T.W. 1994a. "A Comparison of Two Confidence Scales." GSS Methodological Report No. 80. Chicago: NORC.
- . 1994b. *A Comparison of Two Government Spending Scales*. GSS Methodological Report.