

Forthcoming in *Quality of Life Research*

The Effect of Response Option Order on Self-Rated Health: A Replication Study

Dana Garbarski¹
Nora Cate Schaeffer^{2,3}
Jennifer Dykema³

¹Department of Sociology*
Loyola University Chicago
Coffey Hall 440
1032 W. Sheridan Rd.
Chicago, IL 60660
U.S.

dgarbarski@luc.edu

TEL: 1(773) 508-3445

FAX: 1(773) 508-7099

* Corresponding author

²Department of Sociology
University of Wisconsin Survey Center
University of Wisconsin-Madison
475 N. Charter Street
Madison, WI 53706

schaeffe@ssc.wisc.edu

³University of Wisconsin Survey Center
University of Wisconsin-Madison
475 N. Charter Street
Madison, WI 53706

dykema@ssc.wisc.edu

Acknowledgments

This research was supported in part by funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development grants to the Center for Demography and Ecology (T32 HD007014) and the Health Disparities Research Scholars training program (T32 HD049302), and from core funding to the Center for Demography and Ecology (R24 HD047873) at the University of Wisconsin–Madison. The data used in this study were collected by GfK with funding from Time-sharing Experiments for the Social Sciences (NSF Grant SES-0818839, Jeremy Freese and James Druckman, Principal Investigators) and RTI International as part of their 2012 Research Challenge. This study was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Wisconsin–Madison.

Opinions expressed here are those of the authors and do not necessarily reflect those of the sponsors or related organizations.

Abstract

Purpose. Following calls for replication of research studies, this study documents the results of two studies that experimentally examine the impact of response option order on self-rated health (SRH).

Methods. Two studies from an online panel survey examined how the order of response options (positive to negative vs. negative to positive) influences the distribution of SRH answers.

Results. The results of both studies indicate that the distribution of SRH varies across the experimental treatments, and mean SRH is lower (worse) when the response options start with “poor” rather than “excellent.” In addition, there are differences across the two studies in the distribution of SRH and mean SRH when the response options begin with “excellent,” but not when the response options begin with “poor.”

Conclusion. The similarities in the general findings across the two studies strengthen the claim that SRH will be lower (worse) when the response options are ordered beginning with “poor” rather than “excellent” in online self-administered questionnaires, with implications for the validity of SRH. The slight differences in the administration of the seemingly identical studies further strengthen the claim and also serve as a reminder of the inherent variability of a single permutation of any given study.

Introduction

Replication of research studies adds credibility to previous findings and nuance to the conditions under which findings remain constant or change. Although funding constraints and publication bias privilege novelty over replication-based social science, there is increasing interest in direct replication: attempting to replicate a previous study using the same method or a method that differs from the original in ways that should not affect the results [1; 2]. The current study examines whether the effect of response option order on the distribution of self-rated health (SRH) is replicated across two studies that experimentally manipulate the ordering of response options.

Background and Study Descriptions

The SRH question—e.g., “would you say your health in general is excellent, very good, good, fair, or poor?”—is widely used to study health in social, behavioral, and medical research because of its ability to predict morbidity and mortality [3]. Surveys typically present SRH with the response options ordered from “excellent” to “poor”; relatively little research examines the impact of this practice on the distribution and validity of responses.

Survey methodological research and theory highlight the consequences that response option ordering has on outcomes. Research on response option ordering indicates that options located near the beginning of the scale, particularly the first response option that the respondent perceives as acceptable, are more likely to be chosen [4-8]. Thus, we expected that levels of SRH would be higher (better) when the response options are ordered “excellent” to “poor” compared to when they are ordered “poor” to “excellent” [9; 10].

We sought to examine whether response option ordering affects the distribution of SRH in two different studies. Both studies were conducted in early 2013 on samples drawn from a

larger sample of US adults in the online panel survey KnowledgePanel collected by market research institute GfK. Recruitment for GfK's KnowledgePanel uses random digit dialing telephone methods and address-based sampling [11].

[http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel\(R\)-Design-Summary-Description.pdf](http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel(R)-Design-Summary-Description.pdf).)

Study 1 was supported by Time-sharing Experiments for the Social Sciences (TESS), funded by the National Science Foundation (<http://www.tessexperiments.org>). Study 2 was supported by RTI International as part of their 2012 Research Challenge. Each study included an experiment randomly assigning SRH response options to be ordered as “excellent” to “poor” or “poor-excellent.” (Additional information about each of these studies is available from the authors upon request.) Table 1 compares various design features across the two studies, and Table 2 shows differences in the distribution of respondents' characteristics across the studies and experimental treatments. There are several notable differences:

- 1) While the recruitment rate from the target population is the same across studies, the completion rate is lower for Study 2 than Study 1. Differences in the results could be due to differing levels of nonresponse if the distribution of characteristics related to nonresponse differs across the studies and the characteristics are related to health ratings. Table 2 shows a few minor differences in the distribution of respondents' characteristics across studies (comparing the “Total” columns for Study 1 and Study 2). Compared to Study 1, Study 2 has more respondents who are white or college educated; fewer respondents who are Hispanic, in the normal weight BMI category, or have no chronic conditions; and more missing data on the prior health characteristics (measured when respondents first enroll in KnowledgePanel). These small (4 to 5 percentage point) and relatively few differences--and the fact that they have different

relationships with health--minimize concerns of potential bias due to differing completion rates across studies.

2) Another difference across studies is the extent to which random assignment produced samples with similar sociodemographic and health characteristics in each experimental treatment (Table 2, comparing first and second columns then fourth and fifth columns). In Study 1, the distribution of sample characteristics is comparable across experimental treatments. In Study 2, however, the distribution of characteristics varies across treatments: Compared to the “poor” to “excellent” group, the “excellent” to “poor” group has more respondents who are white, married, or in the obese II BMI category; and fewer respondents who are college educated or in the normal weight BMI category (differences range from 4 to 8 percentage points).¹ The differences in the distribution of respondents’ characteristics does not necessarily imply that there was a problem with the random assignment for Study 2, as randomization does not guarantee that covariates will be similarly distributed across treatment group [12]. However, the differences in the distribution of characteristics across experimental treatments could influence the results if those characteristics are associated with health and reports of health are influenced by the experimental treatment.

3) The sample size is larger for Study 1 than Study 2.

4) The SRH questions differ in their placement of “in general,” although we expect the effect of this difference to be minimal.

5) SRH was embedded in different contexts in the studies. Study 1 presented SRH as the first question in the survey. In Study 2, SRH is preceded by 35-47 questions (depending on skip patterns) about political identity, police, intergroup relations, and (for half the sample) food

¹ BMI in this study is calculated from self-reported height and weight, and is thus subject to measurement error.

expenditures. Previous research finds evidence of question order effects for SRH when it is preceded by health-based question [9; 13], and it is plausible that such results extend to non-health context effects.

Results

The results are substantively the same across the two studies. Table 3 shows that the distribution of SRH significantly differs by treatment in Study 1. Descriptively, the differences occur in the middle three categories, in which “fair” and “good” are more likely to be endorsed and “very good” is less likely to be endorsed when the categories begin with “poor” rather than “excellent.” In Study 2, the difference in the distribution across experimental treatments is marginally significant, likely due to the smaller sample size. The descriptive differences across treatments for Study 2 are the same as for Study 1, except that “excellent” is less likely to be endorsed when the categories begin with “poor” rather than “excellent.” The lower portion of Table 3 presents the differences in mean SRH and proportion in “fair/poor” health across treatments within each study. In each study, mean SRH is higher when the response options begin with “excellent” rather than “poor”; these differences were statistically significant. The proportion of “fair/poor” answers is larger when the categories begin with “poor” rather than “excellent,” although the differences were not statistically significant.

We also examined differences in the distribution of SRH answers, mean SRH, and the proportion in “fair/poor” health across the two studies within each experimental treatment (e.g., comparing Study 1 and Study 2 “excellent” to “poor”). For the “poor” to “excellent” treatment, there are no differences in the distribution of SRH, mean SRH, and proportion in “fair/poor” health across Study 1 and Study 2. For the “excellent” to “poor” condition, the difference in the distributions across studies is marginally significant ($p=.084$), and the difference in mean SRH

across studies is significant ($p=.046$). Descriptively, these differences across studies for the “excellent” to “poor” treatment seems to be driven by the 6 percentage-point increase in respondents who report “excellent” health in Study 2 compared to Study 1. We speculate that this may be due to the larger proportion of respondents who are white, married, or college educated (i.e., factors associated with better SRH) in Study 2 compared to Study 1 in the “excellent” to “poor” group—although there are potentially countervailing factors in the larger proportions of respondents in the normal weight or no chronic conditions categories in Study 1 compared to Study 2 (first and fourth columns of Table 2; differences of 4 to 8 percentage points).

Discussion

The general findings are similar across the two studies, strengthening the claim that SRH is lower (worse) when the response options begin with the negative end of the scale in an online self-administered questionnaire [9; 10]. As we have argued elsewhere [9], these findings have implications for the administration of SRH in surveys. We expect that starting with the negative end of the scale induces respondents to consider some of the less desirable response options, but whether doing so increases the validity of SRH must be examined in future studies using a criterion for health such as mortality [9; 10]. In addition, future studies should examine whether starting with the negative end of the scale affects the distribution of SRH and its validity when SRH is presented in other languages [14] and whether response order effects influence other key measures of health and quality of life.

Further strengthening the claim that starting with the negative end of the scale for SRH leads to lower (worse) SRH is that these studies are “seemingly” identical on the surface, but actually vary in ways that could contribute to discrepant results. The comparison of these two

studies highlights the potential impact of sampling variability and differences in survey administration (nonresponse, the distribution of sample members across experimental treatments, sample size, question wording, and question context) on results. Overall, while the reproduction of the general findings strengthens the claim that SRH is lower (worse) when the negative response options are presented first, the slight differences in the administration of the study serves as a reminder of the inherent variability of a single permutation of any given study.

Table 1. Comparison of study design features

	Study 1 (TESS)	Study 2 (RTI)
Field period	1/31/13-2/8/13	3/6/13-3/18/13
Panel recruitment rate	15%	15%
Study completion rate	66%	58%
Sample size (who answered SRH)	1347	499
Wording of SRH	Would you say your health in general is...	In general, would you say your health is...
Location of SRH in the questionnaire	First question	After 35-47 questions (depending on skip patterns)

Table 2. Percentage distribution of respondents' characteristics by study and experimental treatments^a

	Study 1 (TESS) ^b			Study 2 (RTI) ^c		
	Response Option Ordering		Total	Response Option Ordering		Total
	Excellent to Poor	Poor to Excellent		Excellent to Poor	Poor to Excellent	
<u>Sociodemographic characteristics</u>						
Gender						
Female	50	50	50	48	50	49
Male	50	50	50	52	50	51
Race/Ethnicity						
White	73	74	74	81	77	79
Black	9	8	8	8	10	9
Other race	3	3	3	2	4	3
Hispanic	10	11	11	7	6	7
Two or more races	5	4	4	2	3	3
Marital status						
Married	55	55	55	60	52	57
Not married	45	45	45	40	48	43
Education						
Less than high school	9	8	9	7	6	7
High school	28	28	28	27	27	27
Some college	32	31	31	31	28	30
Bachelor's degree or more	31	33	32	35	39	37
Age						
18-29	17	17	17	19	18	18
30-44	24	25	25	23	21	22
45-59	28	27	28	29	28	29
60 and older	32	31	31	30	33	31
<u>Prior health characteristics^d</u>						
Self-rated health						
Poor	3	3	3	3	3	3
Fair	12	12	12	13	10	11
Good	36	36	36	32	34	33
Very good	39	38	38	38	37	37
Excellent	10	11	11	11	11	11
Missing	<1	1	1	4	6	5
Body mass index category						
Normal weight	34	34	34	27	35	30
Overweight	33	33	33	36	35	36
Obese I	17	19	18	18	15	17
Obese II	14	13	13	13	9	11
Missing	2	2	2	6	6	6

Chronic condition ^e						
No	83	82	83	79	76	78
Yes	16	18	17	17	18	17
Missing	<1	1	1	4	6	5

^aDistributions for each variable may not sum to 100 due to rounding.

^bN=1,347 (671 in the “excellent” to “poor” treatment and 676 in the “poor” to “excellent” treatment)

^cN=499 (264 in the “excellent” to “poor” treatment and 235 in the “poor” to “excellent” treatment)

^dThese are health characteristics measured when respondents first join the Knowledge Panel, prior to their participation in Studies 1 and 2. There are missing values for these variables, unlike the basic sociodemographic variables presented above that have no missing data.

^eThe question reads “Have you had a serious or chronic illness, injury, or disability that has required A LOT of medical care in the past 2 years?”

Table 3. Percentage Distribution of Self-Rated Health, Mean of Self-Rated Health, and Proportion in Fair/Poor Health across Experimental Treatments

	Study 1 (TESS) ^a			Study 2 (RTI) ^b		
	Excellent to Poor	Poor to Excellent		Excellent to Poor	Poor to Excellent	
Percentage distribution SRH ^c			**		+	
Poor	3	2		3	3	
Fair	11	15		9	13	
Good	37	43		34	39	
Very good	39	31		38	35	
Excellent	10	8		16	9	
Mean SRH (S.D.) ^d	3.43 (0.90)	3.29 (0.89)	**	3.56 (0.96)	3.35 (0.93)	*
Proportion fair/poor health ^e	0.14	0.17	+	0.12	0.16	

+ p<.1, * p<.05, **p<.01, ***p<.001

^aN=1,347 (671 in the “excellent” to “poor” treatment and 676 in the “poor” to “excellent” treatment)

^bN=499 (264 in the “excellent” to “poor” treatment and 235 in the “poor” to “excellent” treatment)

^cDistributions for each variable may not sum to 100 due to rounding. Differences in distribution tested using Pearson’s chi-square.

^dTreating SRH response options as equidistant categories such that “excellent”=5, “very good”=4, “good”=3, “fair”=2, and “poor”=1. S.D. = standard deviation. Differences in means tested using an independent samples t-test with unequal variances.

^eDifferences in proportions tested using a two-proportion z-test.

References

1. Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
2. Nosek, B. A., & Lakens, D. (2014). Registered Reports: A Method to Increase the Credibility of Published Results. *Social Psychology*, 45(3), 137-141.
3. Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38(1), 21-37.
4. Carp, F. M. (1974). Position effects on interview responses. *Journal of Gerontology*, 29(5), 581-587.
5. Chan, J. C. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement*, 51(3), 531-540.
6. Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
7. Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201-219.
8. Sudman, S., & Bradburn, N. M. (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
9. Garbarski, D., Schaeffer, N., & Dykema, J. (2015). The effects of response option order and question order on self-rated health. *Quality of Life Research*, 24(6), 1443-1453.
10. Means, B., Nigam, A., Zarrow, M., Loftus, E. F., & Donaldson, M. S. (1989). *Autobiographical Memory for Health-Related Events (No. DHHS (PHS) 89-1077)*. Rockville, MD: National Center for Health Statistics.
11. Callegaro, M., & DiSogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*, 72(5), 1008-1032.
12. Mutz, D. C. (2011). *Population-based survey experiments*: Princeton University Press.
13. Lee, S., & Schwarz, N. (2014). Question context and priming meaning of health: effect on differences in self-rated health between Hispanics and non-Hispanic whites. *American Journal of Public Health*, 104(1), 179-185.
14. Sanchez, G. R., & Vargas, E. D. (2015). Language bias and self-rated health status among the Latino population: evidence of the influence of translation in a wording experiment. *Quality of Life Research*, 1-6.