



HHS Public Access

Author manuscript

Qual Life Res. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:

Qual Life Res. 2015 June ; 24(6): 1443–1453. doi:10.1007/s11136-014-0861-y.

The Effects of Response Option Order and Question Order on Self-Rated Health

Dana Garbarski*, Nora Cate Schaeffer, and Jennifer Dykema

Department of Sociology, Loyola University Chicago (Garbarski); Department of Sociology, University of Wisconsin-Madison (Schaeffer); University of Wisconsin Survey Center, University of Wisconsin-Madison (Schaeffer, Dykema)

Abstract

Objectives—This study aims to assess the impact of response option order and question order on the distribution of responses to the self-rated health (SRH) question and the relationship between SRH and other health-related measures.

Methods—In an online panel survey, we implement a 2-by-2 between-subjects factorial experiment, manipulating the following levels of each factor: 1) order of response options (“excellent” to “poor” versus “poor” to “excellent”); and 2) order of SRH item (either preceding or following the administration of domain-specific health items). We use chi-square difference tests, polychoric correlations, and differences in means and proportions to evaluate the effect of the experimental treatments on SRH responses and the relationship between SRH and other health measures.

Results—Mean SRH is higher (better health) and proportion in “fair” or “poor” health lower when response options are ordered from “excellent” to “poor” and SRH is presented first compared to other experimental treatments. Presenting SRH after domain-specific health items increases its correlation with these items, particularly when response options are ordered “excellent” to “poor.” Among participants with the highest level of current health risks, SRH is worse when it is presented last versus first.

Conclusion—While more research on the presentation of SRH is needed across a range of surveys, we suggest that ordering response options from “poor” to “excellent” might reduce positive clustering. Given the question order effects found here, we suggest presenting SRH before domain-specific health items in order to increase inter-survey comparability, as domain-specific health items will vary across surveys.

Keywords

Self-rated health; response option order; question order; assimilation effects; validity; U.S

*Corresponding author: Department of Sociology, Loyola University Chicago, 1032 West Sheridan Road, Chicago, Illinois 60660, dgarbarski@luc.edu, 773-508-3786.

CONFLICT OF INTEREST

None declared.

The opinions expressed herein are those of the authors, and any errors are the sole responsibility of the authors.

The self-rated health (SRH) question – e.g., “would you say your health in general is excellent, very good, good, fair, or poor?” – is one of the most widely used items to study health across a range of disciplines and populations because of its ability to predict morbidity and mortality [1], which has strengthened over time in the U.S. [2]. SRH is related to multiple domains of health including illnesses, symptoms of undiagnosed diseases, judgments about the severity of illness, family history, dynamic health trajectories, complex health histories, health behaviors, and the presence or absence of resources for good health [1; 3–11]. In sum, “a very long list of variables is required to explain the effect of one brief 4- or 5-point scale item...” [1]. Different versions of the SRH question exist, varying in terms of the set of response options used (“excellent” to “poor”; “very good” to “very bad”) and the number of response options used (four or five). However, most surveys present SRH with the response options ordered from the positive to negative end of the scale and preceding rather than following other domain-specific health items. Relatively little research examines the impact of these features of the measurement process on the distribution of responses to SRH and its association with the other domain-specific health items included in the survey.

BACKGROUND

Both theory and research in survey methodology highlight the consequences of the order in which response options are presented. Research on response option order effects indicates that options near the beginning of the scale are more likely to be chosen, particularly the first response option that the respondent perceives to be acceptable [12–15]. Reducing the attractiveness of the first option may have been a reason that Sudman and Bradburn [16] suggested beginning with the least desirable response option. The least desirable options for SRH are likely those that indicate worse health; however, most surveys begin with the most positive category regardless of the mode of administration. There is limited experimental evidence that concurrent validity is better when SRH is administered with the response options ordered from negative to positive [17], although these results require replication because of the small sample size.

The placement of SRH relative to other health items may be consequential for respondents’ answers and the validity of SRH. Keller and Ware [18] recommend asking SRH before questions about more specific aspects of health, so that respondents’ answers to domain-specific health items—questions about specific aspects of health—do not affect their SRH answers. To consider how SRH answers might be affected when SRH follows domain-specific health items rather than precedes them, we review two ways survey researchers think about question order effects: assimilation effects and contrast effects.

With an assimilation effect, the associations between SRH and domain-specific health items would be greater when SRH is administered after domain-specific health items compared to when SRH is administered before. This could occur if 1) the sequence of questions communicates that SRH should summarize or globally assess the more specific health information the respondent previously provided; 2) the sequence of questions provides a common definition of health for all respondents [19]; 3) the sequence of questions activates a memory structure of beliefs, evaluations, and feelings about health which become salient

when formulating an answer to the SRH question [20]; or 4) the sequence of questions helps to define the SRH response scale in a similar way for all respondents [21], reducing random error in responses and thus increasing the strength of estimated relationships. If a contrast effect occurred, the association between SRH and domain-specific health items would be smaller when SRH is administered after domain-specific health items compared to when SRH is administered before [20; 22; 23]. This might occur because respondents infer that SRH must be asking about something different from the health questions previously asked.

While previous studies examine how the placement of SRH with respect to specific questions about health affects the distribution of SRH answers [24–26], these results do not indicate whether placing SRH after domain-specific health items elicits an assimilation or contrast effect. To our knowledge, no study has examined how the association between SRH and domain-specific health items changes depending on whether SRH precedes or follows these health items, yet this type of analysis is needed in order to determine whether assimilation or contrast effects occur. Such effects on the association between SRH and other health items have implications for many types of multivariable analysis in which SRH and other health items from the survey are modeled simultaneously, such as increasing the potential for multicollinearity when SRH and other domain-specific health items are included as independent variables in a model or attenuating the effects of other independent variables when SRH is the dependent variable.

A complication in the study of question order effects for SRH is that such effects may depend on the respondent's health status. For example, when SRH is asked after domain-specific health items, respondents who are generally in better health may report a higher health status after being "reminded" of the various domains in which their health is good; a respondent who repeatedly says "no" when asked about different health conditions, limitations, and poor health behaviors may conclude that they must be in good health for the purposes of the survey. Alternatively, respondents who report many health conditions, limitations, and poor health behaviors may report lower SRH after being reminded of the various domains in which their health is not good.

Overall, it is unclear whether response option order and question order work independently or together to affect SRH. These manipulations are critical to evaluate given the importance of this particular item and the fact that it is typically presented either before or after other questions about health in ways that may not be controlled or understood. Based on a review of the survey methodological literature, we hypothesize the following:

- Hypothesis 1** We expect that the mean values of SRH will be higher (indicating better health) and the proportion in "fair" or "poor" health lower when the response options are ordered from "excellent" to "poor" compared to when they are ordered from "poor" to "excellent."
- Hypothesis 2** Given that the wording of the SRH question uses the phrase "in general," which invites a summary without pointing toward an explicit contrast with preceding questions, we expect assimilation effects to occur when SRH is administered after a set of domain-specific health items compared to when it is administered first. We hypothesize that

the associations between SRH and each of the domain-specific health items will be stronger when SRH follows these more specific health items compared to when it precedes them.

Hypothesis 3 We expect question order effects will depend on the respondent's health status: those in better health will have more positively-rated health when SRH follows a list of domain-specific health items compared to when SRH precedes such a list, while those in worse health will have more negatively-rated health when SRH follows a list of domain-specific health items compared to when SRH precedes domain-specific health items.

METHODS

Data

Data for the study come from Time-sharing Experiments for the Social Sciences (TESS). TESS is funded by the National Science Foundation as a mechanism for investigators to share resources in conducting peer-reviewed population-based experiments (information on TESS is available here: <http://www.tessexperiments.org>). The data for this study were collected by market research institute GfK in the KnowledgePanel online panel study, the target population for which is adults in the U.S. Panel recruitment for GfK's KnowledgePanel is done using random digit dialing telephone methods and address-based sampling (summary available in [27]). A random sample of 4,119 respondents was taken from GfK's KnowledgePanel. There were 2,696 responses to the invitation, yielding a final stage completion rate of 65.5%. The recruitment rate for KnowledgePanel corresponding to the current study was 15.2%, and the profile rate (of recruited households successfully completed a profile survey) was 65.0%, yielding a cumulative response rate with respect to the target population of 6.5% [27]. (Additional information on KnowledgePanel's design is available here: [http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel\(R\)-Design-Summary-Description.pdf](http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel(R)-Design-Summary-Description.pdf).)

The experiment follows a 2-by-2 factorial design in which participants are randomly assigned to one of two levels for each factor. For the first factor, the response options are ordered as "excellent, very good, good, fair, or poor" or "poor, fair, good, very good, or excellent." For the second factor, the administration of SRH either precedes or follows the administration of the domain-specific health items. This leads to four experimental treatment groups: Treatment 1 shows the response options ordered from "excellent" to "poor" and presents SRH first, Treatment 2 shows the response options ordered from "excellent" to "poor" and presents SRH last, Treatment 3 shows the response options ordered from "poor" to "excellent" and presents SRH first, and Treatment 4 shows the response options ordered from "poor" to "excellent" and presents SRH last. In the TESS administration of the survey, the response options are listed vertically on the screen.

Measures

In addition to SRH, each experimental treatment contains several items meant to cover a range of health domains: alcohol use, smoking, exercise, functional ability, health

conditions, and perceived mental health (see Appendix A). An index of current health risks was constructed by summing dichotomies of health risks derived from Questions 3 through 8 (Question 3: exercise less than once a week versus 1 to 2 times per week or more; Question 4: current smoker versus never smoked or former smoker; Question 5: had a work limitation versus not; Question 6: had an activity limitation versus not; Question 7: had a chronic condition versus not; Question 8: felt irritable, anxious or depressed occasionally or more versus rarely or less).¹

Analytic strategy

All analyses were conducted in Stata Version 13.1. We use listwise deletion for analyses in which there is item nonresponse. The first hypothesis is that mean SRH will be higher (better) and proportion in “fair” or “poor” health lower when the response options are ordered from “excellent” to “poor” compared to when they are ordered from “poor” to “excellent.” To examine Hypothesis 1, we examine whether mean SRH or proportion in “fair” or “poor” health varies across 1) the response option order factor (“excellent” to “poor” versus “poor” to “excellent”) and 2) the experimental treatment groups. We treat SRH as a continuous variable with equidistant categories, as a continuous variable with varying distances between categories, and as a dichotomous variable coded as “fair” or “poor” versus “excellent,” “very good,” or “good.” We specify SRH as a continuous variable with varying distances between categories in two ways: first using values averaged across peer-reviewed studies of the scaling of verbal labels as presented by Krosnick [29] (“excellent”=94, “very good”=81, “good”=70, “fair”=51, and “poor”=21), then based on the values derived by Perneger and colleagues [30] (“excellent”=5, “very good”=4.5, “good”=3.7, “fair”=2, and “poor”=1). We examine these various specification of self-rated health given its use in studies as both a continuous variable and a dichotomous variable; varying distances between categories is a potential improvement over equidistant categories and retains more information than the dichotomous specification. As this study is based on experimental data and because generalization about the population is not of interest, we used unweighted analyses to test for differences in means and proportions.

The second hypothesis is that the associations between SRH and each of the domain-specific health items will be stronger when SRH follows these more specific health items compared to when it precedes them. To examine Hypothesis 2, we compute the polychoric correlations between SRH and each of the domain-specific health items listed in Appendix A (using *polychoric* command in Stata) across 1) the question order experimental factor (SRH before the domain-specific health items versus after) and 2) the experimental treatment groups. Dichotomous variables for this analysis include Question 4 (current smoker versus never smoked or former smoker) and Questions 5–7 (yes versus no); categorical variables include SRH (Question 1 with response options coded as listed in Appendix A), Question 2 (0, 1–10, 11 or more days), Question 3 (response options coded as listed in Appendix A), and Question 8 (response options coded as listed in Appendix A, with “almost always” and

¹Question 2, about alcohol consumption, is excluded from the index of current health risks. This question was included as part of the corpus to prime respondents in the conditions in which SRH is presented last to think of a range of health behaviors, conditions, and limitations, but cannot be used to reliably estimate behavioral risk given that the complex relationship with health cannot be assessed without additional data on the number of alcoholic drinks consumed daily [28].

“often” combined into one category given that less than two percent of respondents reported “almost always”). Tests of whether the correlations between SRH and each domain-specific health item are significantly different across the question order experimental factor and experimental treatments were conducted on quantpsy.org using Fisher’s r -to- z transformation [31]. We note that this method for testing the difference in correlations across two independent samples is used for Pearson’s r and is untested in the literature with respect to polychoric correlations. The results of these tests for significant differences in the polychoric correlations should thus be considered preliminary.

The third hypothesis is that question order effects will depend on the respondent’s health status. We examine Hypothesis 3 by analyzing differences in mean SRH and proportion in “fair” or “poor” health within levels of the index of current health risks across 1) the question order experimental factor and 2) the experimental treatment groups. We present unweighted analyses to examine Hypotheses 2 and 3 since the goals of these analyses are not to represent the population but to understand the role of question order and response option order in influencing the relationship between domain-specific health items and SRH.

RESULTS

We examine whether the distribution of responses to SRH varies across the order of the SRH response options (“excellent” to “poor” versus “poor” to “excellent”) and SRH’s placement (before versus after a set of domain-specific health items) in a 2-by-2 factorial experiment. Table 1 shows weighted and unweighted descriptive characteristics for the study sample, as well as the final sample size for SRH and each domain-specific health item. Table 2 presents the unweighted distributions of SRH within the experimental factors and experimental treatments (the weighted distributions of SRH is remarkably similar to the distribution in Table 2). Overall, the distribution of SRH varies across the experimental treatments (likelihood-ratio chi-square ($df\ 12$) = 30.40, $P = 0.002$). Examining the distribution of SRH across each of the experimental factors suggests that “good” and “fair” are more likely to be chosen and “very good” less likely to be chosen when the response options are ordered from “poor” to “excellent” compared to “excellent” to “poor.” “Very good” and “good” appear to be slightly less likely to be chosen and “fair” is more likely to be chosen when SRH is administered after other health items compared to before. (A test of the interaction between experimental factors is not statistically significant.) Across the experimental treatment groups, most of the differences in the distribution occur in the middle three categories, in which “fair” is less likely to be endorsed and “very good” is more likely to be selected in the standard presentation of SRH (treatment 1; response options ordered from “excellent” to “poor” and SRH presented first) compared to the other treatment groups.

The first hypothesis is that mean SRH will be higher (better) and proportion in “fair” or “poor” health lower when the response options are ordered from “excellent” to “poor” compared to when they are ordered from “poor” to “excellent.” We examine whether mean SRH and the proportion of “fair” or “poor” answers depend on the order in which the response options are given in Table 3. Looking first at the response option order experimental factor, mean SRH is slightly higher (e.g., better) and the proportion of “fair” or

“poor” answers slightly lower when SRH is ordered “excellent” to “poor” compared to “poor” to “excellent”; these differences are statistically significant when SRH is treated as an equidistant continuous measure as well as using the varying distances from Perneger and colleagues [30]. Examining the results by experimental treatment group, we see that this pattern is particularly strong when SRH is presented first: Mean SRH is slightly higher and the proportion of “fair” or “poor” answers slightly lower with treatment 1, the standard presentation of SRH (“excellent” to “poor” and before other health items) compared to treatment 3 (“poor” to “excellent” and before other health items); these differences are statistically significant for all operationalizations of SRH, with the exception of proportion in “fair” or “poor” health. Overall, mean SRH is higher and proportion in “fair” or “poor” health lower with the standard presentation of SRH (treatment 1) compared to treatments 2, 3, and 4. The results are partially consistent with Hypothesis 1: SRH is more concentrated at the positive end of the scale when the response options are ordered from “excellent” to “poor” compared to when they are ordered from “poor” to “excellent,” particularly when SRH is presented first.

Hypothesis 2 states that the associations between SRH and each of the domain-specific health items will be stronger when SRH follows these more specific health items compared to when it precedes them. Table 4 shows that the placement of SRH with respect to domain-specific health items (before versus after) plays a role in the association between the domain-specific health items and SRH. The first two columns of Table 4 show the correlation between each question and SRH across the question order experimental factor. Consistent with the expected assimilation effect, many of these correlations are larger when SRH is presented last compared to first, with significant differences in correlations across question order with Questions 6, 7, and 8. Examining the correlations across experimental treatments shows that these question order effects are particularly pronounced when SRH is ordered from “excellent” to “poor” (comparing treatments 1 and 2), with the exception of Question 4. In particular, SRH is more highly correlated with the domain-specific health items asked about immediately before (Questions 5–8) the SRH question when SRH is administered last (treatment 2) compared to first (treatment 1). It is interesting to note that questions 5–8 ask for respondents’ perceptions of whether they fit into a particular health state as opposed to questions about behaviors like questions 2–4, with the former arguably more similar to SRH than the latter. In contrast, there is no discernible question order effect when the response options are ordered “poor” to “excellent” (comparing treatments 3 and 4). Thus, Hypothesis 2 is partially supported, in that the results are consistent with the hypothesized assimilation effects when SRH is ordered from “excellent” to “poor” (but not “poor” to “excellent”).

Hypothesis 3 states that question order effects will depend on the respondent’s health status: those in better health will have more positively-rated health when SRH follows a list of domain-specific health items compared to when SRH precedes such a list, while those in worse health will have more negatively-rated health when SRH follows a list of domain-specific health items compared to when SRH precedes domain-specific health items. Table 5 examines mean SRH (5= “excellent” to 1=“poor”) and proportion in “fair” or “poor” health within groups of current health risks across 1) the question order experimental factor and 2) the experimental treatment groups (results are comparable using the scale values of the

verbal labels and are available upon request). For those with 0, 1, 2, or 3 current health risks, mean SRH does not significantly differ across question order, yet there is a significant difference in mean SRH when SRH is presented first compared to last with 4 or more current health risks; mean SRH is lower (worse) when SRH is presented last compared to first. There is a significant difference in proportion of respondents in “fair” or “poor” health for those with one current health risk, in which the proportion in “fair” or “poor” health is higher when SRH is presented last compared to first. (The difference in proportion “fair” or “poor” health across question order for those with four or more current health risks is marginally significant.)

Examining the differences in mean or proportion across experimental treatment groups within a level of current health risk shows that with little exception, mean SRH is highest and proportion “fair” or “poor” lowest with treatment 1, the standard presentation of SRH (response options ordered “excellent” to “poor” and SRH presented first) compared to the other experimental treatment groups. Among respondents with no current health risks, mean SRH and proportion “fair” or “poor” do not differ significantly across the four experimental treatments. Among respondents with 4 or more current health risks, however, mean SRH is higher (better) and the proportion in “fair” or “poor” health lower with treatment 1 compared to treatments 2, 3, and 4; these differences are statistically significant for all but one comparison. Overall, Hypothesis 3 is partially supported, in that there are conditional effects of question order in the expected direction for those with the highest level of current health risks (but not the lowest). Examining effects within experimental treatment, higher mean SRH and lower “fair” or “poor” health occurs with the standard administration of SRH (presented first and ordered “excellent” to “poor”) compared to the other treatment groups.

DISCUSSION

This study documents how response option order and question order work independently and together to influence SRH, and is the first to do so with an experimental design. Overall, the results depend on the interplay of question order and response option order, with hypotheses about one experimental factor being partially supported depending on the level of the other experimental factor.

With respect to Hypothesis 1 (that mean SRH will be higher and proportion in “fair” or “poor” health lower when the response options are ordered from positive to negative), we find evidence that mean SRH is slightly higher when SRH is ordered from “excellent” to “poor.” When we look across experimental treatments, mean SRH is higher and proportion in “fair” or “poor” health lower with the standard presentation of SRH (treatment 1; response options are ordered from “excellent” to “poor” and SRH is presented first) compared to the other experimental treatments, with many of these differences reaching statistical significance. The pattern of results is consistent across specifications of SRH (continuous and equally spaced, continuous with varying distances between categories, dichotomous), indicating that the implications of question order and response option order for the distribution of SRH are the same regardless of specification.

One interpretation of these results is that ordering the SRH response options from “poor” to “excellent” to increases the likelihood that respondents consider some of the less desirable response options—that is, those that indicate worse health—in making their assessment rather than choosing the first answer that is perceived to be acceptable [14; 15]. While reducing the attractiveness of the first response option has been suggested as desirable for survey questions [16], more research is needed to strengthen a recommendation to do so for SRH. Because the data come from a web survey in which the questions are presented visually, future research should examine which order of response options gives results that are consistent across self-administration and interviewer-administration, because aural presentations of items are associated with recency effects in which respondents are more likely to endorse response options presented at the end of the list [20]. In addition, previous research suggests that ordering options from negative to positive may increase measurement error [32], although this research uses items in which the negative to positive ordering goes against conversational norms (“against or for” compared to “for or against”) in a way that is not comparable to SRH.

Our study also finds evidence that placing SRH after domain-specific health items leads to an assimilation effect for many items (evidenced by correlations between SRH and a domain-specific health item that are larger when SRH is presented last compared to first). These assimilation effects are particularly pronounced when the response options are ordered from “excellent” to “poor” and do not appear consistently when the response options are ordered from “poor” to “excellent.” Finally, there are conditional question order effects for respondents with the highest number of current health risks, consistent with the idea of a priming mechanism in which respondents who have worse health adjust their assessment of their health downward after being reminded of the various domains in which their health is not good. When we look across experimental treatment groups, it appears that mean SRH was highest and proportion in “fair” or “poor” health the lowest in treatment 1 (“excellent” to “poor” and presented first) compared to all other experimental treatments.

These assimilation question order effects have important implications for research practice, in particular with respect to multivariable analyses that incorporate both SRH and other health items. We suggest that researchers use a version of SRH that presents SRH first, because the context provided by domain-specific health items, leading to these question order effects, may vary across studies. For example, the health context in this study consists of seven items meant to prime respondents to think of a range of health behaviors, conditions, and limitations in the experimental treatments in which these health items preceded SRH. This health context is different from that in the California Health Interview Survey used by Lee and colleagues [33], in which SRH is asked after questions about 1) specific health conditions or 2) mental health assessment and service utilization questions. It is interesting to note that with respect to better health, the current study finds that respondents with 4 or more current health risks have significantly higher mean SRH when SRH is presented *first compared to last* (and respondents with 1 current health risk have significantly lower proportion or “fair” or “poor” health when SRH is presented first compared to last), while Lee and colleagues find that those with one (English and Spanish-speaking) or two (Spanish speaking) current comorbidities have a higher proportion of

positive health ratings when SRH is presented *last compared to first* [33]. While the conflicting results in the two studies could be driven by several factors, it raises the question of whether the different health contexts in each of the studies produce different patterns of the association between SRH and current health risks. Varying results could also occur if the distribution of the specific health conditions asked about varies across study populations. In particular, if respondents interpret preceding health items as questions to use in defining overall health, and the health conditions asked about are not those that occur in the study population, the accuracy of SRH as a summary measure may be reduced. These issues are particularly important for comparability of health estimates derived from SRH across studies in which SRH is preceded by different sets of health items.

What do the results of this study indicate with respect to the validity of SRH? Lee and Schwarz [19] find that differences in the ability of SRH to predict mortality across white non-Hispanics and Hispanics/Latinos were attenuated by preceding SRH with other health items. Thus, it is plausible that preceding SRH with other health items provides a common referent for respondents, diminishing differences in how SRH is interpreted and thus increasing the predictive validity of SRH. However, it is unclear whether mortality should be considered a gold standard criterion for SRH given its limited utility as criterion at younger ages, debate as to whether SRH represents an enduring self-concept or spontaneous assessment [34], and debate as to whether a criterion for perceptions of health exists [35–37]. If the goal is for SRH to capture perceptions of health rather than function as a summary measure of more objective health measures, it may be problematic to deliberately influence the health referents used by different groups: doing so may diminish the between-group discrepancies in definitions of health, peer comparisons, and other factors that influence global health assessments that are precisely of interest. Overall, whether SRH is more valid when presented after domain-specific health items depends on the criteria used to examine validity and the stated purpose of SRH.

Conclusion

The results presented here are from one online survey of a panel sample, and more research is needed on the optimal way to present SRH using a range of populations, modes, and criteria for assessing validity. We suggest the following for future research: 1) ordering the SRH response options from “poor” to “excellent” in self-administered questionnaires given the tendency for SRH to cluster toward the positive end of the scale when positive response options are offered first and SRH is presented before other health items, 2) examining the impact of where SRH is placed with respect to domain-specific health items and the impact of the number, content, and order of those items on the distribution of SRH and its association with the domain-specific health items, 3) comparing how varying question order and response option order affects SRH across interviewer-administered and self-administered questionnaires, 4) comparing how the presentation of the response options (vertical or horizontal) affects SRH in self-administered instruments [38], and 5) examining how the effects of response option order and question order on SRH vary across sociodemographic covariates.

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). This study was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Wisconsin–Madison. A previous version of this paper was presented at the 2014 meeting of the American Association for Public Opinion Research in Anaheim, CA. We thank conference participants and the peer reviewers for their insightful comments.

References

1. Idler EL, Benyamini Y. Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*. 1997; 38(1):21–37. [PubMed: 9097506]
2. Schnitker J, Bacak V. The increasing predictive validity of self-rated health. *PLoS ONE*. 2014; 9(1):e84933. [PubMed: 24465452]
3. Benyamini Y, Idler EL, Leventhal H, Leventhal EA. Positive affect and function as influences on self-assessments of health. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2000; 55(2):P107–P116.
4. Benyamini Y, Leventhal EA, Leventhal H. Self-assessments of health. *Research on Aging*. 1999; 21(3):477–500.
5. Benyamini Y, Leventhal EA, Leventhal H. Gender differences in processing information for making self-assessments of health. *Psychosomatic Medicine*. 2000; 62(3):354–364. [PubMed: 10845349]
6. Benyamini Y, Leventhal EA, Leventhal H. Elderly people's ratings of the importance of health-related factors to their self-assessments of health. *Social Science & Medicine*. 2003; 56(8):1661–1667. [PubMed: 12639583]
7. Canfield, B.; Miller, K.; Beatty, P.; Whitaker, K.; Calvillo, A.; Wilson, B. Adult questions on the health interview survey – Results of cognitive testing interviews conducted April–May 2003. Hyattsville, MD: National Center for Health Statistics, Cognitive Methods Staff; 2003. p. 1–41.
8. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine*. 2006; 21(3):267–275. [PubMed: 16336622]
9. Garbarski D, Schaeffer NC, Dykema J. Are interactional behaviors exhibited when the self-reported health question is asked associated with health status? *Social Science Research*. 2011; 40(4):1025–1036. [PubMed: 21927518]
10. Groves, RM.; Fultz, FN.; Martin, E. Direct questioning about comprehension in a survey setting. In: Tanur, JM., editor. *Questions About questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation; 1992. p. 49–61.
11. Krause NM, Jay GM. What do global self-rated health items measure? *Medical Care*. 1994; 32(9): 930–942. [PubMed: 8090045]
12. Carp FM. Position effects on interview responses. *Journal of Gerontology*. 1974; 29(5):581–587. [PubMed: 4853701]
13. Chan JC. Response-order effects in Likert-type scales. *Educational and Psychological Measurement*. 1991; 51(3):531–540.
14. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*. 1991; 5(3):213–236.
15. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*. 1987; 51(2):201–219.
16. Sudman, S.; Bradburn, NM. *Asking questions*. Jossey-Bass; 1982.
17. Means B, Nigam A, Zarrow M, Loftus EF, Donaldson MS. Autobiographical memory for health-related events. *Cognition and Survey Research*. 1989; 6(2):1–38. 6, 2, DHHS (PHS) 89–1077.
18. Keller SD, Ware JE. Questions and answers about SF-36 and SF-12. *Medical Outcomes Trust Bulletin*. 1996; 4(3)

19. Lee S, Schwarz N. 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*. 2014; 104(1):179–185. [PubMed: 23678900]
20. Tourangeau, R.; Rips, LJ.; Rasinski, KA. *The psychology of survey response*. Cambridge University Press; 2000.
21. Hopkins DJ, King G. Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*. 2010; 74(2):201–222.
22. Tourangeau R, Rasinski KA, Bradburn N. Measuring happiness in surveys: a test of the subtraction hypothesis. *Public Opinion Quarterly*. 1991; 55(2):255–266.
23. Schwarz N, Strack F, Mai HP. Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. *Public Opinion Quarterly*. 1991; 55(1):3–23.
24. Bowling A, Windsor J. The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA). *Journal of Epidemiology and Community Health*. 2008; 62(1):81–85. [PubMed: 18079338]
25. Crossley TF, Kennedy S. The reliability of self-assessed health status. *Journal of Health Economics*. 2002; 21(4):643–658. [PubMed: 12146595]
26. Lee S, Grant D. The effect of question order on self-rated general health status in a multilingual survey context. *American Journal of Epidemiology*. 2009; 169(12):1525–1530. [PubMed: 19363097]
27. Callegaro M, DiSogra C. Computing Response Metrics for Online Panels. *Public Opinion Quarterly*. 2008; 72(5):1008–1032.
28. National Institute on Alcohol, A., & Alcoholism. *Helping patients who drink too much*. Bethesda, MD: US Department of Health and Human Services; 2007.
29. Krosnick, JA. Improving question design to maximize reliability and validity. On Conference on the Future of Survey Research; Arlington, VA. 2012.
30. Perneger T, Gayet-Ageron A, Courvoisier D, Agoritsas T, Cullati S. Self-rated health: analysis of distances and transitions between response options. *Quality of Life Research*. 2013; 22(10):2761–2768. [PubMed: 23615958]
31. Preacher, KJ. Calculation for the test of the difference between two independent correlation coefficients [Computer software]. 2002. Available from <http://quantpsy.org>
32. Holbrook AL, Krosnick JA, Carson RT, Mitchell RC. Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions. *Journal of Experimental Social Psychology*. 2000; 36(5):465–494.
33. Lee S, Schwarz N, Goldstein LS. Culture-Sensitive Question Order Effects of Self-Rated Health Between Older Hispanic and Non-Hispanic Adults in the United States. *Journal of Aging and Health*. 2014
34. Bailis DS, Segall A, Chipperfield JG. Two views of self-rated general health status. *Social Science & Medicine*. 2003; 56(2):203–217. [PubMed: 12473308]
35. Huisman M, Deeg DJH. A commentary on Marja Jylhä's "What is self-rated health and why does it predict mortality? Towards a unified conceptual model"(69:3, 2009, 307–316). *Social Science & Medicine*. 2010; 70(5):652–654. [PubMed: 19942333]
36. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*. 2009; 69(3):307–316. [PubMed: 19520474]
37. Jylhä M. Self-rated health between psychology and biology. A response to Huisman and Deeg. *Social Science & Medicine*. 2010; 70(5):655–657.
38. Tourangeau R, Couper MP, Conrad FG. "Up means good": the effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly*. 2013; 77(S1):69–88. [PubMed: 24634546]

APPENDIX A. Survey Questions

Q1	Would you say your health in general is ²		
		Factor 1	
		Excellent, very good, good, fair, poor?	Poor, fair, good, very good, excellent?
	Factor 2	Q1 before Q2–Q8	Treatment 1
		Q1 after Q2–Q8	Treatment 2
		Treatment 3	Treatment 4
Q2	During the last month, on how many days did you drink alcoholic beverages, such as beer, wine, liquor, or mixed alcoholic drinks? ³		
	Type in the number for the answer		
	___ DAYS		
Q3	During an average week, how often do you exercise? ⁴		
	RESPONSE OPTIONS OFFERED: Never, Less than once a week, 1–2 times a week, 3–5 times a week, 6 or more times a week		
Q4	How would you describe your current smoking status? ³		
	RESPONSE OPTIONS OFFERED: Never smoked, Former smoker, Current smoker		
Q5	Are you limited in the kind or amount of work you do because of a physical, mental, or emotional problem? ⁵		
	RESPONSE OPTIONS OFFERED: Yes, No		
Q6	Because of a physical, mental, or emotional problem, do you need the help of other persons in handling routine needs, such as everyday household chores, business, shopping, or getting around for other purposes? ⁶		
	RESPONSE OPTIONS OFFERED: Yes, No		
Q7	Have you had a serious or chronic illness, injury, or disability that has required a lot of medical care in the past 2 years? ⁷		
	RESPONSE OPTIONS OFFERED: Yes, No		
Q8	During the past four weeks, how often did you feel fretful, angry, irritable, anxious or depressed? ⁸		
	RESPONSE OPTIONS OFFERED: Never, Rarely, Occasionally, Often, Almost always		

² Adapted from the National Health Interview Survey.

³ Adapted from Knowledge Networks Health Profile.

⁴ Investigator developed.

⁵ Adapted from the National Longitudinal Study of Youth and the National Health Interview Survey.

⁶ Adapted from the National Longitudinal Study of Youth and the National Health Interview Survey.

⁷ Adapted from Knowledge Networks Health Profile.

⁸ Adapted from the Health Utilities Mark 2 Index and the Wisconsin Longitudinal Study.

Table 1

Sample Descriptive Statistics (Percent), TESS 2013

	Unweighted	Weighted	N
Gender			2,696
Male	51	48	
Female	49	52	
Race/Ethnicity			2,696
White, Non-Hispanic	74	67	
Black, Non-Hispanic	8	12	
Other, Non-Hispanic	3	6	
Hispanic	10	14	
2+ Races, Non-Hispanic	4	1	
Age			2,696
18–29	17	21	
30–44	24	26	
45–59	29	27	
60+	31	26	
Marital Status			2,696
Married	55	50	
Widowed	5	5	
Divorced	11	12	
Separated	2	2	
Never married	20	23	
Living with partner	7	8	
Education			2,696
Less than high school	8	12	
High school	28	30	
Some college	32	29	
Bachelor's degree or higher	31	29	
Household income			2,696
Less than \$25,000	17	19	

	Unweighted	Weighted	N
\$25,000–49,999	23		23
\$50,000–99,999	34		33
\$100,000 or more	26		25
Current employment status			2,696
Not working for pay	44		45
Working for pay	56		55
Region			2,696
Northeast	18		18
Midwest	24		22
South	35		37
West	23		23
Metropolitan Statistical Area Status			2,696
Non-Metro	16		16
Metro	84		84
Q2: Alcohol			2,684
No alcohol in past month	43		45
1–10 days	41		40
11–31 days	16		15
Q3: Exercise			2,652
Never	15		15
Less than once a week	22		22
1–2 times a week	24		23
3–5 times a week	31		32
6 or more times a week	8		8
Q4: Smoking			2,660
Never or former smoker	85		85
Current smoker	15		15
Q5: Work limitation			2,647
No	80		80
Yes	20		20
Q6: Activity limitation			2,626

	Unweighted	Weighted	N
No	94		94
Yes	6		6
Q7: Chronic condition			2,628
No	84		85
Yes	16		15
Q8: Mental Health			2,632
Never	29		30
Rarely	37		36
Occasionally	24		24
Often or almost always	10		10
Index of current health risks			2,690
0	34		35
1	34		34
2	17		15
3	9		9
4 or more	7		8

Notes

TESS: Time-sharing Experiments for the Social Sciences

Columns may not sum to 100 due to rounding

Table 2
 Distribution of Self-Rated Health (Percent) within Experimental Treatments and Factors, TESS 2013

	Factor 1			Factor 2		Treatment 1	Treatment 2	Treatment 3	Treatment 4
	“Excellent” to “Poor”	“Poor” to “Excellent”	Self-Rated Health First	Self-Rated Health Last	“Excellent” to “Poor” and “Poor” and First	“Excellent” to “Poor” and “Poor” and Last	“Poor” to “Excellent” and “Excellent” and First	“Poor” to “Excellent” and “Excellent” and Last	
Poor	3	2	2	3	3	3	3	2	3
Fair	14	16	13	17	11	16	15	15	17
Good	36	41	40	37	37	35	43	43	39
Very good	38	31	35	33	39	36	31	31	30
Excellent	10	10	9	11	10	10	8	8	11
N	1,323	1,328	1,347	1,304	671	652	676	676	652

Notes

TESS: Time-sharing Experiments for the Social Sciences

Columns may not sum to 100 due to rounding

Table 3
 Difference in Mean Self-Rated Health and Proportion in Fair or Poor Health across Response Option Order Experimental Factor and Experimental Treatments, TESS 2013

	Factor 1		Treatment 1		Treatment 2		Treatment 3		Treatment 4	
	“Excellent” to “Poor”	“Poor” to “Excellent”	“Excellent” to “Poor” and First	“Poor” to “Excellent” and Last	“Excellent” to “Poor” and First	“Poor” to “Excellent” and Last	“Poor” to “Excellent” and First	“Excellent” to “Poor” and Last	“Poor” to “Excellent” and First	“Poor” to “Excellent” and Last
Mean self-rated health (5=excellent-1=poor)	3.39	3.30*	3.43	3.34	3.29**	3.30*	3.30*	3.30*	3.30*	3.30*
Mean self-rated health (Krosnick scale values)	72.64	71.57	73.34	71.92	71.68*	71.46*	71.68*	71.46*	71.68*	71.46*
Mean self-rated health (Permejer et al. scale values)	3.83	3.74*	3.89	3.77*	3.75**	3.73**	3.75**	3.73**	3.75**	3.73**
Proportion in “fair” or “poor” health	0.16	0.18	0.14	0.19**	0.17	0.19**	0.17	0.19**	0.17	0.19**

Notes

TESS: Time-sharing Experiments for the Social Sciences

Tests of differences in means or proportion across response option order experimental factor or comparing treatments 2–4 to treatment 1:

* p<.05,

**

p<.01,

p<.001

Table 4
 Polychoric Correlations between Self-Rated Health and Each Domain-Specific Health Item, TESS 2013

Question	Description	Factor 2		Treatment 1	Treatment 2	Treatment 3	Treatment 4
		Self-Rated Health First	Self-Rated Health Last	“Excellent” to “Poor” and First	“Excellent” to “Poor” and Last	“Poor” to “Excellent” and First	“Poor” to “Excellent” and Last
2	Alcohol	0.14	0.15	0.17	0.19	0.12	0.12
3	Exercise	0.31	0.35	0.30	0.33	0.32	0.36
4	Smoking	-0.25	-0.25	-0.26	-0.23	-0.26	-0.28
5	Work limitation	-0.65	-0.68	-0.64	-0.74***	-0.67	-0.61
6	Activity limitation	-0.55	-0.64***	-0.52	-0.66***	-0.59	-0.62
7	Chronic condition	-0.56	-0.63**	-0.57	-0.73***	-0.56	-0.55
8	Mental health	-0.32	-0.41**	-0.30	-0.44**	-0.34	-0.39

Notes

TESS: Time-sharing Experiments for the Social Sciences

Tests of differences in correlation across question order experimental factor, comparing treatments 1 and 2, or comparing treatments 3 and 4:

* p<.05,

** p<.01,

*** p<.001

Table 5
 Mean Self-Rated Health and Proportion Fair or Poor Health across Experimental Treatments and Index of Current Health Risks, TESS 2013

Current Health Risks	Mean Self-Rated Health ("Excellent"=5 to "Poor"=1)				
	Factor 2	Treatment 1	Treatment 2	Treatment 3	Treatment 4
0	3.83	3.86	3.90	3.77	3.83
1	3.47	3.42	3.55	3.38*	3.39*
2	3.06	2.99	3.08	3.03	2.92
3	2.68	2.59	2.85	2.48*	2.60
4	2.24	2.00*	2.34	2.11	1.98*

Current Health Risks	Proportion in "Fair" or "Poor" Health				
	Factor 2	Treatment 1	Treatment 2	Treatment 3	Treatment 4
0	0.02	0.03	0.02	0.02	0.04
1	0.08	0.14*	0.06	0.11	0.14**
2	0.21	0.23	0.22	0.20	0.25
3	0.40	0.48	0.31	0.50*	0.45
4	0.67	0.79	0.58	0.79*	0.80*

Notes

TESS: Time-sharing Experiments for the Social Sciences

Creation of the current health risks index is described in the *Measures* section

Tests of differences in mean or proportion across question order experimental factor or comparing treatment 1 to each of treatments 2, 3, and 4:

* p<.05,

** p<.01,

*** p<.001