12-6-2017

Interviewers' Ratings of Respondents' Health: Predictors and Association With Mortality

Dana Garbarski
Loyola University Chicago, dgarbarski@luc.edu

Nora Cate Schaeffer
University of Wisconsin

Jennifer Dykema
University of Wisconsin

Deborah Carr

Recommended Citation

Garbarski, Dana; Schaeffer, Nora Cate; Dykema, Jennifer; and Carr, Deborah. Interviewers' Ratings of Respondents' Health: Predictors and Association With Mortality. The Journals of Gerontology: Series B, , : 1-9, 2017. Retrieved from Loyola eCommons, Sociology: Faculty Publications and Other Works, http://dx.doi.org/10.1093/geronb/gbx146

Interviewers' Ratings of Respondents' Health: Predictors and Association with Mortality

Dana Garbarski, PhD[1*]
Phone 1(773) 508-3455
Email dgarbarski@luc.edu

Nora Cate Schaeffer, PhD[2,3]
Email schaeffer@ssc.wisc.edu

Jennifer Dykema, PhD[3]
Email dykema@ssc.wisc.edu

[1]Department of Sociology, Loyola University Chicago, Coffey Hall 440
1032 W. Sheridan Rd., Chicago, IL 60660 USA

[2]Department of Sociology, University of Wisconsin-Madison, 475 N. Charter Street
Madison, WI 53706 USA

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

*Corresponding author

Acknowledgments

This work was supported by core funding to the Center for Demography and Ecology from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (P2C HD047873) and core funding to the Center for Demography of Health and Aging from the National Institute on Aging (P30 AG017266) at the University of Wisconsin–Madison. This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin-Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775, AG-21079, AG-033285, and AG-041868), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. Since 1992, data have been collected by the University of Wisconsin Survey Center. A public use file of data from the Wisconsin

Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at http://www.ssc.wisc.edu/wlsresearch/data/. The opinions expressed herein are those of the authors.

Author contributions

Dana Garbarski planned the study, conducted the data analysis, and wrote the manuscript. Nora Cate Schaeffer and Jennifer Dykema contributed to planning the study and writing the manuscript.

Conflict of interest

Dana Garbarski declares that she has no conflict of interest. Nora Cate Schaeffer declares that she has no conflict of interest. Jennifer Dykema declares that she has no conflict of interest.

Abstract

Objectives. Recent research indicates that survey interviewers' ratings of respondents' health (IRH) may provide supplementary health information about respondents in surveys of older adults. Although IRH is a potentially promising measure of health to include in surveys, our understanding of the factors contributing to IRH remains incomplete.

Methods. We use data from the 2011 face-to-face wave of the Wisconsin Longitudinal Study, a longitudinal study of older adults from the Wisconsin high school class of 1957 and their selected siblings. We first examine whether a range of factors predict IRH: respondents' characteristics that interviewers learn about and observe as respondents answer survey questions, interviewers' evaluations of some of what they observe, and interviewers' characteristics. We then examine the role of IRH, respondents' self-rated health (SRH), and associated factors in predicting mortality over a three-year follow-up.

Results. As in prior studies, we find that IRH is associated with respondents' characteristics. In addition, this study is the first to document how IRH is associated with both interviewers' evaluations of respondents and interviewers' characteristics. Furthermore, the association between IRH and the strong criterion of mortality remains after controlling for respondents' characteristics and interviewers' evaluations of respondents.

Discussion. We propose that researchers incorporate IRH in surveys of older adults as a cost-effective, easily implemented, and supplementary measure of health.

Keywords

Interviewer-rated health, self-rated health, surveys, interviewer observations

Interviewers' ratings of respondents' health (IRH)—e.g., "Would you say the respondent's health in general is excellent, very good, good, fair, or poor?"—have the potential to augment the power of survey health measures beyond the ubiquitous measure of self-rated health (SRH). Prior studies show differences between IRH and SRH in their sociodemographic, health, and functioning correlates (Brissette, Leventhal, and Leventhal 2003; Smith and Goldman 2011), indicating that respondents and interviewers draw on different information when assessing respondents' health. Although no single objective measure of "true" health exists with which to examine the validity of measures such as SRH and IRH (Garbarski 2016; Jylhä 2009), mortality is one relevant criterion for physical health, particularly in studies of older adults (Idler and Benyamini 1997). In a study in Taiwan, IRH was associated with mortality, yet including the information gathered during the interview attenuated the association between IRH and mortality such that it was no longer statistically significant (Todd and Goldman 2013). Studies in the US and China found that IRH predicted mortality and that this association was attenuated but still statistically significant when controlling for health covariates from the interview (Brissette et al. 2003; Feng et al. 2016).

IRH may provide information about respondents' health that supplements other measures and is relatively inexpensive to incorporate in a variety of study designs. However, our understanding of what underlies IRH is incomplete in several ways, a shortcoming we address in the current study. We first develop a theoretically informed conceptual model of factors that influence IRH: respondents' characteristics that interviewers learn about and observe as respondents answer survey questions, interviewers' evaluations of some of what they observe, and interviewers' characteristics. Although respondents' characteristics and their relationship with IRH have been explored in prior studies, interviewers' characteristics have not been and are

central to understanding interviewers' response processes when rating respondents' health. Interviewers' evaluations of some of what they observe about respondents during interviews are assumed to inform IRH (Feng et al. 2016; Todd and Goldman 2013), yet prior studies have not examined this. Furthermore, conflicting results from prior studies about whether respondents' answers to health questions completely or partially explain the association between IRH and mortality likely depend in part on details of the study, such as the population under study and the types of health questions and assessments included (Brissette et al. 2003; Feng et al. 2016; Todd and Goldman 2013). Our study joins the small set of studies describing the conditions under which IRH simply summarizes the information provided by and observed about respondents during the interview versus when IRH increases the ability to predict mortality net of these factors. Finally, little research on IRH exists outside of a few studies, so more research is needed in other contexts.

Background

One way to potentially expand our understanding of survey respondents' health is to incorporate IRH in interviewer-administered surveys. Although interviewers' evaluations of respondents' engagement with the survey process have long been collected for a variety of administrative and analytic purposes (Olson and Parkhurst 2013), obtaining interviewers' evaluations about respondents in other domains—such as health—is a relatively recent phenomenon. The continuum model of impression formation suggests that interviewers might form impressions about respondents using various levels of processing, ranging from category-based processing (based on stereotypes associated with immediately salient categories, such as gender, race/ethnicity, age, body size) to individuating processing (piecemeal integration, attribute by attribute, to form an overall impression) (Fiske and Neuberg 1990; Fiske, Lin, and

Neuberg 1999). When making assessments at the end of the interview, as they do in the current and prior studies, interviewers have an opportunity for piecemeal integration of information about respondents' health based on respondents' answers to survey questions and their own observations about respondents' appearance, environment, and physical, psychological, and social functioning during the survey interview; their doing so potentially expands upon our understanding of respondents' health beyond more common health measures such as SRH.

Figure 1 displays a conceptual model of the factors influencing IRH. First are *respondents' characteristics*, which includes a range of information that interviewers ascertain from 1) respondents' answers to survey questions, 2) observations of respondents' living environments, appearance, and functioning, or 3) a combination of the two. Most information about respondents' characteristics probably combine both sources. For example, some sociodemographic characteristics, like gender, can be both observed by interviewers and reported by respondents. In addition, some of the survey tasks in this and other studies involve performance-based measures, such as anthropometric and physical functioning measurements that interviewers observe and collect. Other characteristics are likely observed with error, like age and body mass index (BMI), but are then specified more precisely by respondents' answers to questions.

The surveys in the current and prior studies are quite lengthy, with respondents answering many questions about their health and related factors, allowing interviewers to potentially integrate several pieces of information in forming their assessments of respondents' health. Respondents' sociodemographic characteristics—such as gender, race/ethnicity, socioeconomic status, and age—may also influence how interviewers rate respondents' health beyond respondents' answers to and performance on health survey items. Previous research

demonstrates that differences in evaluative frameworks may influence how respondents rate their own health, leading to systematic differences in SRH across groups defined by race/ethnicity, gender, socioeconomic status, and age among individuals that are otherwise similarly situated with respect to health (Garbarski 2016; Jylha 2009). For example, women tend to rate their own health worse than do men at younger ages but better than do men at older ages (Case and Paxson 2005; Grol-Prokopczyk et al. 2011). We might expect the respondents' gender and age to interact in their effects on IRH if interviewers go through the same response process as respondents do when rating their own health, such that differences in IRH stem from the person being rated (Garbarski 2016).

Respondents' living conditions, appearance, and various forms of physical, psychological, and social functioning during the interview likely influence how interviewers assess respondents' health. Psychologists have noted that people make attributions about others' personality characteristics from their facial features with consensus (although not necessarily accuracy), with implications for outcomes such as voting and criminal sentencing (Todorov et al. 2015). Ratings of perceived age made by strangers using facial photographs were associated with mortality of those in the photographs, indicating that health information relevant to mortality is conveyed in one's facial and bodily features (Christensen et al. 2009). In-person interviewers are able to observe respondents' physical functioning and mobility before and during the interview; prior research shows that IRH is more strongly associated with external physical health issues than is SRH (Brissette et al. 2003; Feng et al. 2016). Indeed, interviewers may notice limitations in physical functioning respondents do not consider when rating their own health, for example, if the respondent has adapted to the limitation and no longer considers it salient. Respondents' attentiveness, performance, concentration, disposition, and cooperation

during the interview task also provide information on respondents' psychological—affective and cognitive—and social functioning that interviewers may incorporate in their assessment of respondents' health. Related to all these types of functioning is respondents' voice clarity and strength, which interviewers are able to observe (Brissette et al. 2003).

Interviewers' evaluations of respondents comprise the second set of factors influencing IRH. Interviewers' evaluations are driven in part by the respondents' characteristics that interviewers learn about and observe during the interview, but are distinct in that they indicate interviewers' perceptions of some of what they have observed (noted by the curved arrow between respondents' characteristics and interviewers' evaluations in Figure 1). How interviewers perceive what they learn and observe about respondents is likely influenced by interviewers' own characteristics (noted by the curved arrow between interviewers' evaluations and interviewers' characteristics in Figure 1). Interviewers' evaluations of what they observe during interviews are assumed to inform IRH (Feng et al. 2016; Todd and Goldman 2013), but have not been examined in previous research. The current study includes interviewers' evaluative observations about respondents: assessments of respondents' cooperativeness, issues with completing the survey, attractiveness, and grooming.

The third set of factors informing how interviewers rate respondents' health is *interviewers' characteristics*, which are unexamined in previous research. At least two categories of characteristics might influence IRH: interviewers' sociodemographic characteristics and their interviewing experience. Differences in evaluative frameworks across interviewers' sociodemographic characteristics may lead to differences in how interviewers rate the health of respondents by influencing how interviewers interpret and integrate what they observe when formulating an assessment (Garbarski 2016; Jylha 2009). For example, older

respondents tend to rate their own health optimistically compared to younger respondents (Idler 1993), so older interviewers might rate the health of respondents more positively than younger interviewers. An additional feature of incorporating interviewers' characteristics is the degree to which the interviewer's sociodemographic characteristics may interact with those of the respondent, extending the notion of differences in health ratings across sociodemographic groups to both the rater and person being rated simultaneously (noted by the curved arrow between respondents' and interviewers' characteristics in Figure 1).

Previous research shows that differences in interviewers' experience are associated with various measures of data quality (West and Blom 2017), although the direction and strength of the relationship depends on the outcome of interest. Yet we know little about how interviewers' experience—prior interviewing experience and the number of interviews completed for the current study—may influence their evaluative observations about respondents such as IRH. For example, interviewers might change how they rate respondents' health as they complete more interviews over the field period, and so access increasingly more relevant and representative referents with which to compare the current respondent's health (Brissette et al. 2003; Feng et al. 2016).[1]

[1] Because training does not vary across interviewers, we cannot include it as a covariate. We learned from the project director for the WLS that interviewers were not trained how to make observations but were informed that the instrument contained questions about the participant, their home, and the interview overall (personal communication with Kerryann DiLoreto, September 28, 2016).

This study examines the interrelationships among characteristics of respondents and interviewers, IRH, and mortality in a longitudinal study of older adults in the US. We examine 1) respondents' characteristics that interviewers ascertain from answers to survey questions and observations about respondents during the interview, 2) interviewers' evaluations of some of what they observe about respondents, and 3) interviewers' characteristics; the latter two sets of factors are unexamined in prior research. We then examine the role of IRH and associated factors in predicting mortality, given the inconsistent empirical findings about the association between IRH and mortality in prior studies (Brissette et al. 2003; Feng et al. 2016; Todd and Goldman 2013). The substantive issue is the extent to which IRH increases the ability to predict mortality or simply summarizes the information provided by and observed about the respondent in this context.

Methods

Data

Data come from the Wisconsin Longitudinal Study (WLS), a one-third random sample of the Wisconsin high school class of 1957 that has been interviewed periodically in the intervening decades along with selected siblings, spouses, and children (Herd, Carr, and Roan 2014). Respondents in the current study include graduates and siblings interviewed face-to-face in 2011 (N=9,138; 5,832 graduates, 3,306 siblings). Most interviews took place in respondents' residences and consisted of several modules of questions and tasks. Sixty-five interviewers completed between 2 and 378 interviews (mean=143.62, SD=109.20). Of the 65 interviewers, complete data on their characteristics are available for 62 interviewers; data on prior interviewing experience is only available for 58 interviewers. WLS gathers mortality data from the following sources: 1) reports from family members informing the WLS of respondents' death

or through tracing efforts by the WLS staff or 2) matching respondents' information with either the Social Security Administration's Death Master File or the National Death Index. WLS staff last updated mortality data in 2014; the last recorded date of death is July 2014.

Table 1 shows the descriptive statistics for SRH, IRH, and mortality. Supplementary Table 1 shows descriptive statistics for other covariates.

Respondents' characteristics. The first question in the health section asked respondents to rate their own health (SRH) (Table 1). (Supplementary Appendix A examines measures of agreement between IRH and SRH in this study and how they compare to prior studies of IRH.) Respondents then answered questions about their functioning across eight domains from the Health Utilities Index Mark 3 (HUI): vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. The HUI (mean=0.78) ranged from -.29 (a health state worse than death) to 1 (perfect health). We converted this continuous measure into tertiles and included a category for missing data. The health section also contained questions about whether the respondent had ever been diagnosed with high blood pressure, high blood sugar, diabetes, cancer, heart problems, stroke, and mental illness; we summed across the conditions to form an index. Questions about activities of daily living included difficulties in six basic (e.g., dressing and eating) and seven instrumental (e.g., shopping for groceries and doing housework) activities; we summed across each of these sets of questions to form indices.

The interview included several cognitive tasks. We examine the letter fluency task which asked respondents to list all of the words they could think of that began with the letter F or L in one minute. This task was asked of all respondents and provides an overt display of cognitive functioning in terms of processing speed and retrieval. Thus, the measure of letter fluency could be both a primary vehicle through which interviewers observe respondents' cognitive processing

and a proxy for cognitive functioning more generally. We standardized scores for each letter to make them comparable, and then divided the range of scores into tertiles and include a category for missing data. In addition, we included a measure of early life cognitive ability (high school IQ) which is associated with future health outcomes and survey participation in prior research (Hauser 2010). We converted this continuous measure into tertiles with a category for missing.

The anthropometric section of the interview included measurements of: height and weight (to compute BMI); waist and hip circumference (to compute a waist-to-hip ratio); lung strength (peak flow liters per minute, best of 3 attempts); grip strength (kilograms, best of 2 with dominant hand); chair rise time (seconds to go from sitting to standing); and walking time (seconds to walk 2.5 meters, best of 2). We split each of these continuous measures into tertiles and included a category for missing data. Respondents' sociodemographic characteristics included their gender, age, education, and marital status.

Interviewers' evaluations of respondents. At the conclusion of each section of the interview, interviewers reported whether they observed respondents receiving help from others during that section; we created a dichotomous variable indicating whether the interviewer rated the respondent as needing any help during any section of the interview. At the conclusion of the interview, interviewers evaluated respondents on the following dimensions: cooperativeness (on a scale of 1 to 7), IRH (Table 1), grooming (on a scale of 0 to 9), and attractiveness (on a scale of 0 to 9) (we split grooming and attractiveness into tertiles and added a category for missing data on these measures). We constructed a measure of respondents' performance issues during the interview as a dichotomous variable (any vs. none) from interviewers' reports about the following: having concerns about the respondent's future participation; whether the respondent

was easily confused, distracted or disrupted; whether the respondent contradicted herself; and whether the respondent had difficulty understanding.

Interviewers' characteristics include gender, age, race/ethnicity, prior interviewing experience, and how many interviews the interviewer completed at the time of the respondent's interview.

As noted above, we converted several continuous variables into tertiles for analysis so that we could include "missing" as a category for these variables, as we expect the data are missing not at random. Alternatives were to drop the cases by listwise deletion or to use multiple imputation to replace the missing data, which is justifiable when data are missing at random but potentially problematic when data are missing not at random or with multilevel data like that used here. Missing data levels were higher for items that were associated with respondents' willingness and ability to complete tasks (HUI, letter fluency cognitive task, and measures from the anthropometric section) and interviewers' willingness to rate respondents' appearance (interviewers' ratings of the respondents' grooming and attractiveness), a task that is potentially more fraught than other sorts of assessments. In addition, high school IQ is not missing at random, as missing data are for siblings of the selected graduates only. Thus, we expect that missingness on these items is associated with IRH and include indicators for missing values for each in our models.

Analytic strategy

We conducted analyses in Stata Version 14.1. We examine the factors from the conceptual model predicting positive IRH ("excellent," "very good," and "good" coded as 1 versus "fair" and "poor" coded as 0) using a mixed effects logistic regression (melogit) that accounts for the nesting of respondents within interviewers with a random intercept for

interviewers.[2] We present the results using a binary dependent variable because 1) the proportional odds assumption is violated with an ordinal logistic regression, 2) the results of the

[2] The lack of random assignment of respondents to interviewers means that the variance component for interviewers is likely overestimated in that it conflates interviewer effects with geographic and other clustering since interviewer assignments are often based on geography, although the impact of geography is likely less here than in an area probability sample that selects clusters.

To estimate the proportion of the variance in IRH that is explained by interviewers, we first computed the intraclass correlation using the random intercept for interviewers from an unconditional mixed effects logistic model regressing IRH on a random intercept for interviewers (variance component σ^2=0.26, 95% CI 0.17 to 0.40). We then calculate the intraclass correlation as $\rho=\sigma^2/(\sigma^2+\pi^2/3)$ (Hedeker 2003). The proportion of variance in IRH that is explained by the interviewers is ρ=0.07, similar to the estimates of interviewer effects of the interviewer ratings of health and sickness in the study by Brissette and colleagues (2003). Thus, most of the variation in IRH is due to factors other than the interviewer.

Interestingly, the proportion of the variance in IRH explained by the random effect of interviewers increases when controlling for the covariates, to ρ=0.18 in Model 1 in Supplementary Table 2 (σ^2=0.74, 95% CI 0.46 to 1.18). This may seem counterintuitive but makes sense in the mixed effects framework. Consider an interviewer that frequently gives answers that are different from what the model with covariates predicts. The more covariates added into the model, the larger her unique effect on IRH—that is, the random intercept—will be.

more complex multinomial logistic regression models are largely similar to the more parsimonious logistic regression models, and 3) modeling health ratings as binary dependent variable is also consistent with the analysis of SRH in numerous studies (Garbarski 2016). We then examine the role of IRH, SRH, and associated factors in predicting the timing of mortality (through July 2014) in a survival analysis using a Cox proportional hazard model (stcox). All models have standard errors that are adjusted for the clustering of respondents within interviewers.

Results

Factors associated with IRH

Supplementary Table 2 shows results from mixed effects logistic regressions of IRH on the predictors. Because higher scores indicate better IRH ("excellent," "very good," or "good"=1, "fair" or "poor"=0), a positive coefficient indicates that an increase in the independent variable is associated with better IRH, and a negative coefficient indicates that an increase in the independent variable is associated with worse IRH.

Many of the respondents' characteristics—SRH, HUI, health conditions, basic and instrumental activity limitations, letter fluency cognitive task, lung strength, grip strength, chair rise time, and walk time—are associated with IRH in the expected directions and net of the other characteristics (Model 1 in Supplementary Table 2). For example, missing data or being in the lowest or middle tertile for lung strength (compared to the highest tertile) is associated with worse IRH, and being in the highest tertile for walking time (compared to lowest tertile) is associated with worse IRH. BMI shows a curvilinear relationship with IRH: being underweight or obese II relative to the "normal" weight category is associated with worse IRH, while being

overweight relative to the "normal" weight category is associated with better IRH. Waist-to-hip ratio shows no association with IRH net of these other factors.[3]

Interviewers' evaluations are overwhelmingly associated with IRH net of other factors—whether the respondent ever needed help during the interview, had problems with the survey task, or was in the lowest tertiles (compared to highest) of grooming and attractiveness—are each associated with worse IRH net of other factors. Only interviewers' evaluations of respondents' cooperativeness are not associated with IRH.[4] Finally, Model 1 shows significant main effects for respondents' gender and age and interviewers' age, but these characteristics show a significant three-way interaction in Model 2 and their effects are discussed below.

[3] The associations of IRH with high school IQ and education appear counterintuitive: being in the lowest tertile for high school IQ (relative to the highest) is associated with better IRH, and having some college relative to a high school diploma is associated with worse IRH. However, these results are likely driven by multicollinearity with each other and other variables (such as the letter fluency cognitive task), as their bivariate associations with IRH are in the expected direction (not shown).

[4] We also examined interviewers' evaluations of how 1) well-kept and 2) clean were respondents' residences, which were only ascertained for respondents who were interviewed in their residence (N=6,710). These were measured on 1 to 7 scale from "not at all" to "extremely," and a higher score on each measure was significantly associated with IRH when replicating Model 1 for this subset of cases. The effect of interviewers' evaluation of respondents' grooming is no longer significant when controlling for these evaluations of respondents' residences.

We next examined a series of interactions among: respondents' gender and age, interviewers' gender and age, interviewers' and respondents' gender, interviewers' and respondents' age, and combinations of interviewers' and respondents' gender and age. A three-way interaction between respondents' age, respondents' gender, and interviewers' age is statistically significant in predicting better IRH (Supplementary Table 2 Model 2) and shows an improvement in model fit over Model 1 and the lower order interactions (using likelihood ratio tests not shown). Figure 2 helps to describe the results of this interaction: the predicted probability of better IRH is similar across respondents' age and gender when the interviewer is age 30 or 40. When the interviewer is age 50, 60, or 70, however, the probability of interviewers reporting better IRH increases with the age of female respondents and decreases with the age of male respondents.

IRH and mortality

We next examine the relationship between IRH and mortality and include the relationship between SRH and mortality for comparison. Overall, 3% of respondents (graduates and siblings) from the 2011 wave of data collection died by July 2014. The probability of having died by July 2014 is remarkably similar for IRH and SRH; for example, 25% of respondents with poor IRH and 24% with poor SRH died, while 8% with fair IRH and 10% with fair SRH died (Table 1). Yet a binary outcome for survival does not indicate whether these categories are associated with the timing of death, which is important to examine as a shorter time to death indicates a higher risk of death. We performed a series of Cox proportional hazard models to examine the associations between ratings of health and the timing of mortality and whether these associations are attenuated when including respondents' characteristics and interviewers' evaluations. We do

not include interviewers' characteristics in these models since we have no reason to expect that these are associated with the respondents' mortality.

Table 2 shows that SRH and IRH each predict age-specific mortality in a dose-response relationship of increasing mortality risk with a worse health rating (relative to "excellent"). Respondents who rated their own health as "poor" had almost 25 times the chance of dying as respondents who rated their health as "excellent" (Model 1), while respondents for whom the interviewer rated their health as "poor" had almost 56 times the chance of dying as respondents for whom the interviewer rated their health as "excellent" (Model 2). These effects are attenuated but still significant once both SRH and IRH are included in the model simultaneously (Model 3). After controlling for respondents' characteristics and interviewers' evaluations of respondents, the effects of SRH and IRH on mortality are further attenuated but are still significant (Model 4). Indeed, a larger reduction in the hazard of mortality upon inclusion of covariates occurred for IRH than SRH, indicating that what interviewers learn and observe about respondents during the interview explains part of the association between IRH and mortality. Yet IRH is still an independent predictor of mortality net of these factors, capturing information about respondents that predicts mortality even beyond the rich set of factors considered here.

--Table 2 about here--

Discussion

This study demonstrates the utility of IRH as an additional measure of health in surveys by extending our understanding of the predictors of IRH and the association between IRH and mortality. As in prior studies, we find that IRH is associated with respondents' characteristics. In addition, this study is the first to document how IRH is associated with both interviewers' evaluations of respondents and interviewers' characteristics. Overall, this study demonstrates the

utility of IRH as a measure of health that 1) appears to summarize in part health information provided by and observed about respondents in the interview and yet 2) increases our ability to predict mortality beyond what is learned and observed about respondents.

To begin, we find that IRH is associated with a range of respondents' characteristics that interviewers learn and observe about the respondent during the course of the detailed face-to-face interview that includes several measures of health, well-being, and functioning. Notably, these effects are significant net of many other factors and the results align with the few prior studies examining IRH (Brissette et al. 2003; Feng et al. 2016; Smith and Goldman 2011; Todd and Goldman 2013).

This study is the first to explicitly examine the role of the rater in IRH by examining interviewers' evaluations of what they observe and interviewers' characteristics. We find that interviewers' evaluations of respondents' competence during the survey interview -- how interviewers perceive respondents' performance and need for help -- as well as how they evaluate respondents' appearance (grooming and attractiveness) are associated with IRH. These relationships hold net of other factors, including the rich set of health information interviewers are privy to during the course of the interview. Rather than viewing interviewers' evaluations as independent predictors of IRH, we might construe these assessments as indicators of a methodological halo effect in which an interviewer's evaluations about a respondent are consistently positive (or negative) across the domains they report on. Future research should contend with this issue and make the associations among various interviewers' evaluations a topic of inquiry to illuminate which are worth gathering.

The multiple dimensions and frameworks through which health is subjectively rated (Garbarski 2016) indicates a complex response process that is likely further complicated through

the lens of a professional data collector like an interviewer. Although prior studies find that interviewers rate female respondents as having worse IRH (Brissette et al. 2003; Smith and Goldman 2011), these studies do not consider the interaction between respondents' gender and age—nor the interactions among characteristics of respondents and interviewers—in predicting IRH. The current study suggests that increasing respondent age is associated with increased probability of better IRH for female respondents and decreased probability for male respondents, following a similar pattern to what is reported for SRH (Case and Paxson 2005; Grol-Prokopczyk et al. 2011)—but only for older interviewers. That interviewers' age is associated with IRH is evidence that interviewers may also have differences across sociodemographic characteristics in the evaluative frameworks through which they rate the health of respondent, much like evaluative framework differences for SRH (Garbarski 2016). Future research should continue to examine the mechanisms underlying evaluative framework differences in interviewers when they are evaluating respondents in survey interviews, such as self-evaluation motives (Sedikides and Strube 1997).

The evaluative framework of the interviewer does not seem to influence the validity of IRH with respect to predicting mortality, as the association between IRH and mortality does not vary across interviewer characteristics (these results are available upon request). In this study, IRH is an independent predictor of mortality even after controlling for covariates, and IRH more strongly predicts mortality than SRH when comparing their hazard ratios. In particular, it appears that IRH in this study is a strong predictor of "early" mortality (Todd and Goldman 2013), indicating that some of what is unmeasured in the IRH-mortality link may be indications of the severity of illness or frailty.

This study contains limitations. First, interviewers rate respondents' health at the end of the interview in this and previous studies. The health information that interviewers are able to ascertain and the conditions fostered by a survey with several health-relevant questions and tasks may elicit (thus far) unmeasured health information that interviewers are using to assess respondents' health, and these conditions might not extend to shorter surveys or those asking for limited health information. We might expect that respondents' sociodemographic characteristics and behavior during the interview would show stronger relationships with IRH in these sorts of studies, as the interviewer would have less health-specific information to draw on and would instead form their assessments based on the limited available information (Fiske et al. 1999; Kirchner, Olson, and Smyth forthcoming). Although we have examined whether IRH is an independent predictor of 3-year mortality net of a rich set of health measures, future studies should examine the predictive validity of IRH in the absence of such survey conditions and with longer mortality follow-up periods. Another limitation to the current and prior studies is that the order of the questions does not vary across respondents, such that order of the items is a constant influence on the associations reported (Brissette et al. 2003). Finally, the homogeneity of the samples of both respondents (mainly white non-Hispanic older adults) and interviewers (mainly white non-Hispanic women) in this study precludes the ability to examine a broader range of respondents' and interviewers' characteristics—and their interactions—with respect to associations with IRH and mortality.

Conclusion

Although IRH is in part summarizing health information from the survey, it also measures something different than SRH and other health measures. Part of this "something different" derives from the interviewer's own characteristics, rendering IRH vulnerable to the

same criticism as SRH in terms of evaluative framework differences in reporting. However, other parts of this "something different," thus far unidentified, lead to IRH predicting mortality net of relevant information ascertained from the interview. Future research should continue to examine the factors that predict IRH and explain the association between IRH and mortality, with a particular focus on whether the utility of IRH extends to other survey conditions. In the meantime, we suggest that researchers and practitioners incorporate IRH in surveys as a cost-effective, easily implemented, and supplementary measure of health.

References

Brissette, I., Leventhal, H., & Leventhal, E. A. (2003). Observer ratings of health and sickness: Can other people tell us anything about our health that we don't already know? *Health Psychology*, 22(5), 471-478. doi:10.1037/0278-6133.22.5.471

Case, A., & Paxson, C. (2005). Sex Differences in Morbidity and Mortality. *Demography*, 42(2), 189-214. doi:10.1353/dem.2005.0011

Christensen, K., Thinggaard, M., McGue, M., Rexbye, H., Hjelmborg, J. V., Aviv, A., Gunn, D., van der Ouderaa, F., Vaupel, J. W. (2009). Perceived age as clinically useful biomarker of ageing: cohort study. *BMJ*, 339, b5262. doi:10.1136/bmj.b5262

Feng, Q., Zhu, H., Zhen, Z., & Gu, D. (2015). Self-Rated Health, Interviewer-Rated Health, and Their Predictive Powers on Mortality in Old Age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 71(3), 538-550. doi:10.1093/geronb/gbu186

Fiske, S. T., Lin, M., & Neuberg, S. (1999). The continuum model. *Dual-process theories in social psychology*, 321-254.

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, 23, 1-74. doi:10.1016/s0065-2601(08)60317-2

Garbarski, D. (2016). Research Synthesis: Research in and Prospects for the Measurement of Health Using Self-Rated Health. *Public Opinion Quarterly*, 80(4), 977-997. doi:10.1093/poq/nfw033

Grol-Prokopczyk, H., Freese, J., & Hauser, R. M. (2011). Using Anchoring Vignettes to Assess Group Differences in General Self-Rated Health. *Journal of Health and Social Behavior*, 52(2), 246-261. doi:10.1177/0022146510396713

Hauser, R. M. (2010). Causes and Consequences of Cognitive Functioning Across the Life Course. *Educational Researcher, 39*(2), 95-109. doi:doi:10.3102/0013189X10363171

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine, 22*(9), 1433-1446. doi:10.1002/sim.1522

Herd, P., Carr, D., & Roan, C. (2014). Cohort Profile: Wisconsin longitudinal study (WLS). *International Journal of Epidemiology*, 43(1), 34-41. doi:10.1093/ije/dys194

Idler, E. L. (1993). Age Differences in Self-Assessments of Health: Age Changes, Cohort Differences, or Survivorship? *Journal of Gerontology*, 48(6), S289-S300. doi:10.1093/geronj/48.6.s289

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. doi:10.2307/2955359

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3), 307-316. doi:10.1016/j.socscimed.2009.05.013

Kirchner, A., Olson, K., & Smyth, J. Do Interviewer Post-survey Evaluations of Respondents' Engagement Measure Who Respondents Are or What They Do? A Behavior Coding Study. Forthcoming in *Public Opinion Quarterly.*

Olson, K., & Parkhurst, B. (2013). "Collecting Paradata for Measurement Error Evaluations." In *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. Frauke Kreuter, pp. 43-72. John Wiley & Sons.

Sedikides, C., & Strube, M. J. (1997). Self-Evaluation: To Thine Own Self Be Good, To Thine Own Self Be Sure, To Thine Own Self Be True, and To Thine Own Self be Better. *Advances in experimental social psychology, 29*, 209-269. doi:http://dx.doi.org/10.1016/S0065-2601(08)60018-0

Smith, K. V., & Goldman, N. (2011). Measuring Health Status: Self-, Interviewer, and Physician Reports of Overall Health. *Journal of Aging and Health*, 23(2), 242-266. doi:10.1177/0898264310383421

Todd, M. A., & Goldman, N. (2013). Do Interviewer and Physician Health Ratings Predict Mortality?: A Comparison with Self-Rated Health. *Epidemiology*, 24(6), 913-920. doi:10.1097/EDE.0b013e3182a713a8

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66(1), 519-545. doi:10.1146/annurev-psych-113011-143831

West, B. T., & Blom, A. G. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175-211. doi:10.1093/jssam/smw024

Tables and Figures

Table 1. Descriptive Statistics for Self-Rated Health (SRH), Interviewers' Ratings of Respondents' Health (IRH), and Mortality by July 2014, 2011 Wave of Wisconsin Longitudinal Study In-Person Interviews

	Health Rating	Percent Died by July
Variable	2011	2014
SRH		
Excellent	19.22 %	0.80 %
Very good	38.53 %	1.05 %
Good	30.38 %	3.42 %
Fair	9.49 %	10.27 %
Poor	2.37 %	23.50 %
Missing	0.01 %	0 %
IRH		
Excellent	17.39 %	0.38 %
Very good	36.57 %	1.02 %
Good	28.09 %	2.38 %
Fair	13.84 %	7.91 %
Poor	3.68 %	25.30 %
Missing	0.43 %	0 %

Notes

N=9,138

Table 2. Hazard Ratios of 2011 Health Ratings for Mortality by July 2014, Wisconsin Longitudinal Study

	Model 1		Model 2		Model 3		Model 4	
SRH Poor	24.61	***			3.68	***	3.61	**
SRH Fair	11.05	***			2.68	**	2.41	*
SRH Good	3.85	***			1.72		1.65	
SRH Very good	1.25				0.88		0.91	
SRH Excellent	Ref.				Ref.		Ref.	
IRH Poor			55.71	***	22.27	***	8.07	***
IRH Fair			17.01	***	8.53	***	4.45	***
IRH Good			5.31	***	3.59	**	2.52	*
IRH Very good			2.49		2.19		1.82	
IRH Excellent			Ref.		Ref.		Ref.	
N	9,127		9099		9098		9017	

Notes

All models are Cox proportional-hazard models. Model 1 predicts mortality by SRH, Model 2 by IRH, Model 3 by SRH and IRH, and Model 4 by SRH, IRH, and covariates (HUI, health conditions, basic activity limitation, instrumental activity limitations, letter fluency cognitive ability, high school IQ, BMI, waist-to-hip ratio, lung strength, grip strength, chair rise time, walk time, interviewers' evaluations [help needed, cooperativeness, grooming, attractiveness, performance issues], respondents' sociodemographic characteristics [gender, age, education, martial status])

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

Figure 1. Predictors of Interviewers' Ratings of Respondents' Health (IRH)

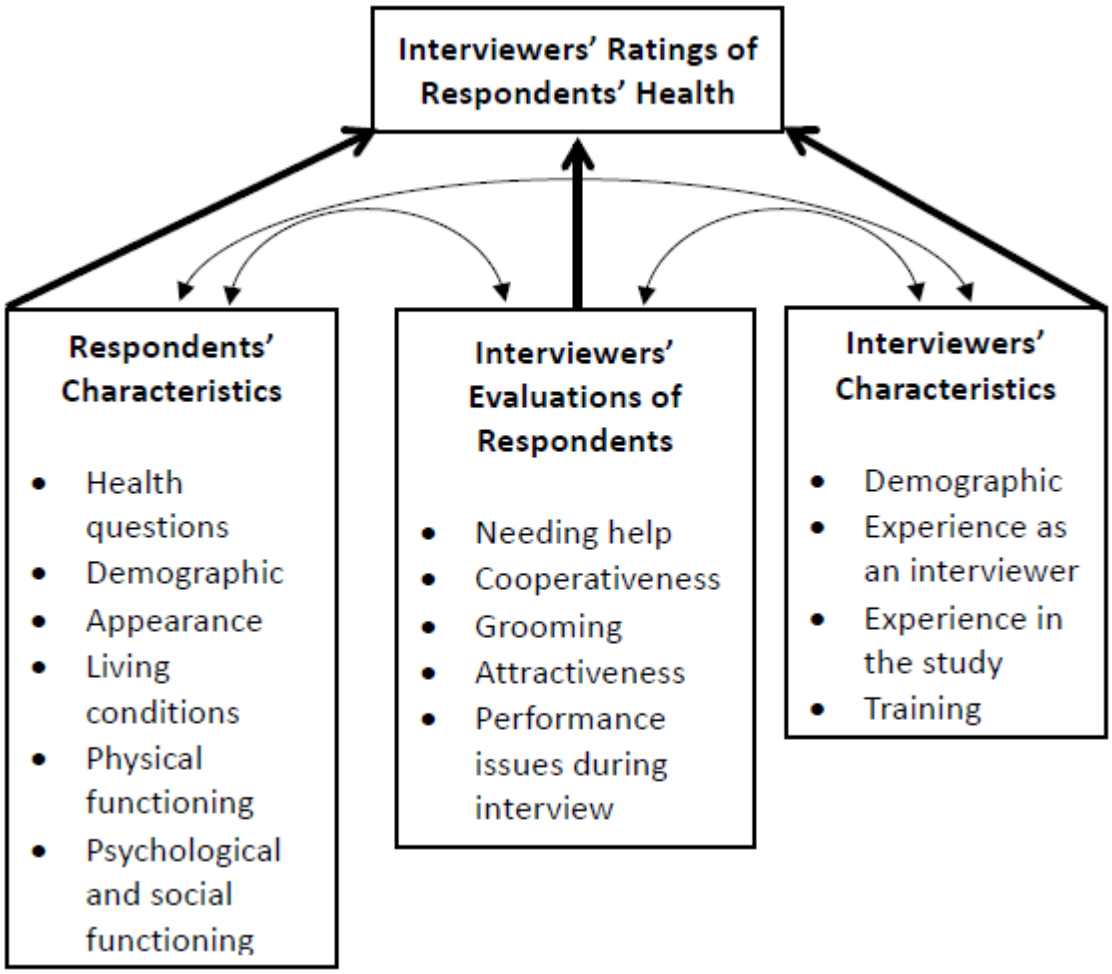
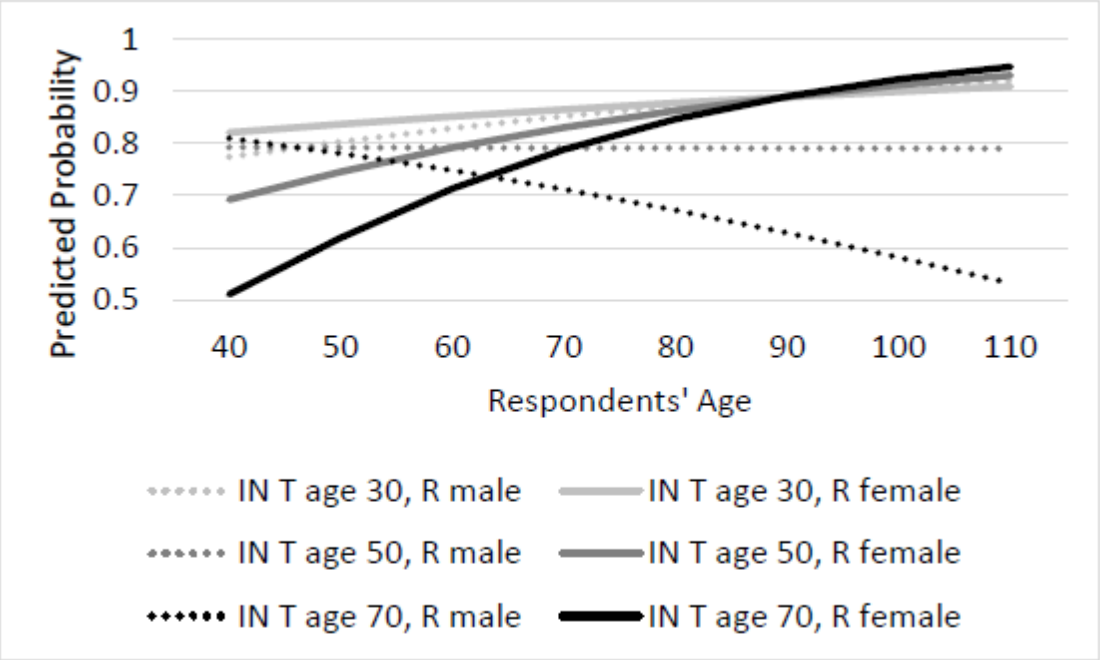


Figure 2. Predicted Probability of Better Interviewers' Ratings of Respondents' Health by Respondents' Age, Respondents' Gender, and Interviewers' Age, 2011 Wisconsin Longitudinal Study



Supplementary Table 1. Descriptive Statistics, 2011 Wave of Wisconsin Longitudinal Study In-Person Interviews

Variable	Mean or Percent		Std. Dev.	Min	Max
Respondents' characteristics					
Health Utilities Index (HUI)	0.78		0.23	-0.29	1
Missing	9.70	%			
Number of health conditions	1.48		1.31	0	6
Missing	.01	%			
Number of basic activity limitations	0.20		0.71	0	6
Number of instrumental activity limitations	0.28		0.69	0	7
Missing	.01	%			
Letter fluency score (unstandardized)	11.31		4.24	0	31
Missing (percent)	8.32	%			
High school IQ	103.06		15.07	61	145
Missing	5.95	%			
Body mass index (BMI)					
Underweight	0.67	%			
Normal weight	18.32	%			
Overweight	35.71	%			
Obese I	25.80	%			
Obese II	16.58	%			
Missing	2.92	%			

Waist-to-hip ratio	0.94		0.08	0.67	1.87
Missing	4.50	%			
Lung strength (peak flow liters per minute)	405.44		133.81	60	880
Missing	3.79	%			
Grip strength (kilograms)	29.80		10.97	0	90
Missing	2.32	%			
Chair rise time (in seconds)	10.11		3.72	0.24	129.58
Missing	8.59	%			
Walking time for 2.5 meters (in seconds)	2.69		3.51	0.01	318
Missing	3.50	%			
R is female (vs. male)	52.79	%			
Rs' age in 2011	71.36		4.28	41	93
Rs' education					
Less than high school	1.50	%			
High school	50.44	%			
Some college	16.77	%			
College	30.46	%			
Missing	0.84	%			
R is married (vs. not)	71.58	%			
Interviewers' evaluations of respondents					

INTs' ratings of whether R ever needed help (vs. no)	4.21	%			
INTs' ratings of Rs' cooperativeness	6.45		0.95	1	7
Missing	0.43	%			
INTs' ratings of Rs' grooming	7.37		1.49	0	9
Missing	19.33	%			
INTs' ratings of Rs' attractiveness	7.01		1.58	0	9
Missing	15.97	%			
INTs' ratings of Rs' performance issue during interview					
Yes	25.21	%			
No	74.34	%			
Missing	0.45	%			
Interviewers' characteristics					
INT gender					
Female	69.23	%			
Male	26.15	%			
Missing	4.62	%			
INTs' age in 2011	43.92		14.25	21	71
Missing	4.62	%			
INT race/ethnicity					
White	87.69	%			

Nonwhite	7.69	%			
Missing	4.62	%			
INT prior interviewing experience					
Yes	40.00	%			
No	49.23	%			
Missing	10.77	%			
Total number of interviews completed by INT	143.62		109.20	2	378

Notes

R=respondent, INT=interviewer

Total analytic sample size is N=9,138 for respondents (variables under "respondents' characteristics" and "interviewers' evaluations of respondents" and N=65 for interviewers (variables under "interviewers' characteristics").

Supplementary Table 2. Mixed-Effects Logistic Regression of IRH on Predictors, 2011 Wisconsin Longitudinal Study

	Model 1			Model 2		
	Coef.	Std. Err.		Coef.	Std. Err.	
Respondents' characteristics						
Self-rated health (SRH)						
Poor	-4.651	0.355	***	-4.713	0.349	***
Fair	-2.809	0.296	***	-2.846	0.305	***
Good	-1.094	0.200	***	-1.120	0.204	***
Very good	-0.516	0.208	*	-0.527	0.212	*
Excellent	Ref.			Ref.		
Health Utilities Index (HUI)						
Missing	-0.237	0.137		-0.238	0.138	
Lowest tertile	-0.652	0.130	***	-0.659	0.131	***
Middle tertile	-0.128	0.146		-0.136	0.149	
Highest tertile	Ref.			Ref.		
Number of health conditions	-0.352	0.037	***	-0.352	0.038	***
Number of basic activity limitations	-0.410	0.061	***	-0.409	0.061	***
Number of instrumental activity limitations	-0.403	0.061	***	-0.402	0.063	***
Letter fluency score						
Missing	-0.074	0.167		-0.097	0.164	

Lowest tertile	-0.353	0.130	**	-0.353	0.131	**
Middle tertile	-0.289	0.133	*	-0.308	0.133	*
Highest tertile	Ref.					
High school IQ						
Missing	0.121	0.186		0.042	0.182	
Lowest tertile	0.251	0.103	*	0.245	0.103	*
Middle tertile	0.204	0.105		0.211	0.105	*
Highest tertile	Ref.					
Body mass index (BMI)						
Missing	-0.386	0.378		-0.397	0.386	
Underweight	-1.599	0.380	***	-1.622	0.379	***
Normal weight	Ref.			Ref.		
Overweight	0.350	0.120	**	0.357	0.119	**
Obese I	0.261	0.149		0.270	0.148	
Obese II	-0.396	0.173	*	-0.403	0.171	*
Waist-to-hip ratio						
Missing	0.319	0.322		0.328	0.319	
Lowest tertile	0.158	0.164		0.166	0.168	
Middle tertile	0.131	0.100		0.144	0.102	
Highest tertile	Ref.			Ref.		
Lung strength (peak flow liters per minute)						
Missing	-0.661	0.328	*	-0.683	0.331	*

Lowest tertile	-0.517	0.152	***	-0.520	0.157	***
Middle tertile	-0.289	0.134	*	-0.260	0.138	
Highest tertile	Ref.			Ref.		
Grip strength (kilograms)						
Missing	-0.202	0.469		-0.149	0.469	
Lowest tertile	-0.508	0.141	***	-0.517	0.143	***
Middle tertile	-0.390	0.133	**	-0.366	0.132	**
Highest tertile	Ref.			Ref.		
Chair rise time (in seconds)						
Missing	-0.713	0.192	***	-0.724	0.196	***
Lowest tertile	Ref.			Ref.		
Middle tertile	-0.156	0.118		-0.169	0.121	
Highest tertile	-0.628	0.121	***	-0.641	0.121	***
Walking time for 2.5 meters (in seconds)						
Missing	-0.571	0.407		-0.592	0.395	
Lowest tertile	Ref.			Ref.		
Middle tertile	-0.101	0.124		-0.101	0.124	
Highest tertile	-0.464	0.142	***	-0.479	0.141	***
R is female (vs. male)	0.624	0.149	***	8.227	5.430	
Rs' age in 2011	0.030	0.011	**	0.099	0.055	
Rs' education						
Less than high school	0.143	0.318		0.146	0.317	

High school	Ref.			Ref.		
Some college	-0.273	0.101	**	-0.269	0.101	**
College	-0.027	0.109		-0.016	0.112	
R is married (vs. not)	0.154	0.081		0.158	0.079	*
Interviewers' evaluations of respondents						
INTs' ratings of whether R ever needed help (vs. no)	-0.451	0.199	*	-0.438	0.198	*
INTs' ratings of Rs' cooperativeness	0.038	0.047		0.037	0.047	
INTs' ratings of Rs' grooming						
Missing	0.079	0.273		0.111	0.274	
Lowest tertile	-0.637	0.233	**	-0.616	0.233	**
Middle tertile	-0.135	0.220		-0.105	0.220	
Highest tertile	Ref.			Ref.		
INTs' ratings of Rs' attractiveness						
Missing	0.105	0.299		0.072	0.294	
Lowest tertile	-0.862	0.283	**	-0.860	0.278	**
Middle tertile	-0.069	0.280		-0.085	0.271	
Highest tertile	Ref.			Ref.		
INTs' ratings of Rs' performance issue during interview (any vs. none)	-1.403	0.116	***	-1.401	0.119	***
Interviewers' characteristics						
INT is female (vs. male)	0.394	0.363		0.401	0.358	

INTs' age in 2011	-0.039	0.009	***	0.092	0.071
INT is nonwhite (vs. white)	-0.676	0.740		-0.707	0.737
INT has prior interviewing experience (vs. none)	-0.269	0.277		-0.261	0.277
Number of prior interviews completed by the interviewer	0.001	0.001		0.001	0.001
Interaction effects (Model 2)					
R female*R age				-0.121	0.077
R female*INT age				-0.234	0.108 *
R age*INT age				-0.002	0.001 *
R female*R age*INT age				0.004	0.002 *
Intercept	5.631	0.890	***	1.223	3.840
Random intercept for interviewers	0.737	0.176		0.729	0.173

Notes

Coef.=coefficient, Std. Err.=standard error, Ref.=reference group, R=respondent, INT=interviewer

N=8,857

Standard errors adjusted for clustering of respondents in interviewers in Models 1 and 2.

IRH coded as excellent, very good, or good health=1 vs. fair or poor health=0

Model 1 regresses IRH on all predictors of interest. Model 2 shows the results from the significant three-way interaction between R female, R age, and INT age

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

Supplementary Appendix A. Agreement between SRH and IRH

In the Wisconsin Longitudinal Study (WLS), mean IRH is lower (mean=3.50, standard deviation [SD]=1.05) than mean SRH (3.63, SD= 0.97), treating the response options as equidistant from poor=1 to excellent=5. In the sample of older Taiwanese adults, mean IRH was higher (mean=3.9, SD=0.92) than SRH (mean=3.2, SD=0.98) (Smith and Goldman 2011). The agreement between IRH and SRH in the WLS appears to be moderate across a variety of agreement measures and larger than in prior studies, although differences across the studies mitigate the ability to make direct comparisons. The polychoric correlation between IRH and SRH is .69, and Cohen's Kappa shows fair agreement at .32 (N=9,098 respondents who have both IRH and SRH measures).[5] In the sample of older adults in Taiwan, the correlation between IRH and SRH is .55 (Todd and Goldman 2013), while in the US sample of older adults, the correlation between SRH and interviewers' ratings of healthiness and sickness (on a scale of "not at all" to "very," reverse coded for sickness) is .49 and .41, respectively (Brissette et al. 2003). In the Taiwanese sample of older adults, Kappa was .13, showing slight agreement (Smith and Goldman 2011).

Table A.1 shows the distribution of IRH within levels of SRH because we subsequently use SRH to predict IRH. The contingency table complements the correlations and Kappa by providing information about the absolute levels of agreement between the measures. Percent agreement among interviewers and respondents is strongest on the diagonal, ranging from 47%

[5] Levels of agreement indicated by the Kappa statistic are as follows: 0 to 0.2 = *slight*; 0.2 to 0.4 = *fair*; 0.4 to 0.6 = *moderate*; 0.6 to 0.8 = *substantial*; and 0.8 to 1.0 = *almost perfect* (Landis and Koch 1977).

to 53% agreement depending on the category. The discordance is usually a one category difference in IRH and SRH; for example, 37.5% of interviewers choosing "very good" for respondents when respondents chose "excellent" for themselves, while 12% of interviewers chose "good," "fair," or "poor" when respondents chose "excellent." Off the diagonals, 29% of respondents had lower IRH than SRH, and 21% had higher IRH than SRH. Overall, based on the polychoric correlation, Kappa, and percent agreement, IRH is correlated with but not a perfect substitute for SRH in this study.

Table A.1. Percent distribution of IRH by SRH, 2011 Wisconsin Longitudinal Study

	SRH Excellent	SRH Very Good	SRH Good	SRH Fair	SRH Poor
IRH excellent	**50.5**	16.8	4.1	0.1	0.5
IRH very good	37.5	**52.5**	28.8	5.7	1.4
IRH good	9.5	24.1	**46.9**	28.7	5.1
IRH fair	2.0	6.1	17.1	**52.5**	40.1
IRH poor	0.5	0.5	3.0	13.1	**53.0**
	N=1747	N=3508	N=2761	N=865	N=217

Notes

Columns sum to 100%

References

Brissette, I., Leventhal, H., & Leventhal, E. A. (2003). Observer ratings of health and sickness: Can other people tell us anything about our health that we don't already know? *Health Psychology*, 22(5), 471-478. doi:10.1037/0278-6133.22.5.471

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Smith, K. V., & Goldman, N. (2011). Measuring Health Status: Self-, Interviewer, and Physician Reports of Overall Health. *Journal of Aging and Health*, 23(2), 242-266. doi:10.1177/0898264310383421

Todd, M. A., & Goldman, N. (2013). Do Interviewer and Physician Health Ratings Predict Mortality?: A Comparison with Self-Rated Health. *Epidemiology*, 24(6), 913-920. doi:10.1097/EDE.0b013e3182a713a8