Can we do better than self-rated health?

Improving mortality prediction with external health ratings

Megan A. Todd

Noreen Goldman

April 2012

DRAFT: Please do not cite

Abstract

A large literature on self-rated health (SRH) has found a robust link between SRH and subsequent mortality. The reasons for this link, however, remain poorly understood. Prior research comparing the health ratings provided by survey respondents, interviewers, and physicians has found that these three evaluators place different weights on particular health factors when rating survey respondents' overall health status. This suggests that health assessments made by interviewers and physicians may contribute valuable additional information regarding a respondent's health, beyond what is incorporated in SRH.

The contribution of this study is in determining whether this additional information from external health evaluators improves mortality prediction. Using data from the Social Environment and Biomarkers of Aging Study (SEBAS), we evaluate the predictive power of self, interviewer, and physician health assessments for subsequent mortality among older Taiwanese adults. We find that interviewer health ratings outperform both physician and self ratings in prediction mortality. This predictive power persists in a number of models that include demographic and health covariates; however, interviewer ratings are unable to add predictive power beyond performance-based functioning covariates.

Background

Reliable and comparable indicators of health status are critical for measuring population health, determining the impact of health interventions, and understanding health disparities. Self-rated health (SRH), wherein survey respondents are asked to rate their own overall health, typically on a five-point scale ranging from poor to excellent, is a widely used measure of general health status. SRH is an attractive measure of overall health both because it is inexpensive and easy to collect in a survey, and because it has been shown to predict many future health outcomes, including morbidity, health care utilization, physical functioning and mortality (Benyamini et al. 2003). The relationship between SRH and mortality is particularly robust in a variety of settings, and typically remains even after controlling for other measures of health status, such as biological measures and chronic conditions (Idler and Benyamini 1997; Benyamini and Idler 1999; DeSalvo et al. 2006). The persistence of SRH's independent predictive power indicates an incomplete understanding of the exact health factors that contribute to SRH.

The SRH literature suggests that self ratings encapsulate a range of information across many health domains (Krause and Jay 1994), including physical health and functioning, mental health, a sense of well-being, and health behaviors, as well as other determinants of health and mortality, such as family history and demographic factors (Benyamini, Leventhal, and Leventhal 2003; Jylhä, Volpato, and Guralnik 2006). In recent studies of Asian and U.S. populations, biomarkers are independently associated with SRH, meaning that SRH may be picking up biological aspects of health beyond those of typical physical and mental evaluations (Goldman, Glei, and Chang 2004; Jylhä, Volpato, and Guralnik 2006). Interestingly, several of the biomarkers used in these studies are unlikely to be known by participants, and thus unlikely to be explicitly considered in their self assessments, as opposed to, for example, a physician's diagnosis of hypertension[1] (Goldman, Glei, and Chang 2004; Jylhä, Volpato, and Guralnik 2006). This could indicate that biomarkers enter the respondents' consideration of overall health unconsciously as a sense of well-being; however, it could also mean that the interview simply failed to capture complete information about known health conditions, which, rather than the biomarker measures per se, are driving SRH (Jylhä, Volpato, and Guralnik 2006).

Several studies have found evidence that population subgroups differ in their reporting styles, with one group more or less likely to report poor health than another, even given similar "objective" health levels. Evidence of reporting differences have been found by socioeconomic status (SES) (Dowd and Zajacova 2007; Etile and Milcent 2006), race/ethnicity (Spencer et al. 2009; Boardman 2004), gender (Benyamini et al. 2003; Lindeboom and van Doorslaer 2004), and age (Idler 1993; Schnittker 2005). These reporting differences between population subgroups could be due to differences in overall health optimism or pessimism (Spencer et al. 2009), or they could result from subgroups placing different weights on particular domains of health, such that the perceived importance to general health of a given chronic condition may differ by group (Finch et al. 2002; Benyamini, Leventhal, and Leventhal 2000). In addition, individuals may assess their overall health primarily by comparing themselves to their

[1] Biomarkers are not always completely distinct from physician diagnoses of health conditions, such as for hypertension, high cholesterol, or high blood sugar. Other biomarkers, such as measures of cortisol or dopamine, are typically unknown by respondents and unrelated to a diagnosis.

peers, which could result in surprisingly positive health ratings among older or disadvantaged groups as they rate their own health in relation to the poor health of those around them (Schnittker 2005).

In a handful of older studies, SRH is compared to overall health assessments made by physicians or nurses (Suchman, Phillips, and Streib 1958; Friedsam and Martin 1963; Markides et al. 1993; Valanis and Yeaworth 1982), but to our knowledge, the only research examining interviewer health assessments uses the same data we use in the current study (Smith and Goldman 2011). There is reason to believe that, despite their lack of medical training, survey interviewers may be able to make informed assessments of respondents' overall health based on the batteries of health questions in many interviews and their observations of respondents' physical functioning. Nevertheless, interviewers are rarely asked to evaluate respondents' health.

Using the Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan, Smith and Goldman (2011) compare survey respondents' self assessments of overall health status to assessments of the same respondents made by interviewers and physicians. Using SEBAS's considerable information on self-reported health indicators, physical functioning measures, and biomarkers, they conclude that respondents, interviewers and physicians place different weights on different aspects of health. Interviewers place more weight on physical functioning than respondents, while physicians place more weight on clinical risk factors such as BMI and smoking status. The fact that external evaluators place more weight on certain health factors than respondents does suggest that interviewer and physician assessments provide new information on health status not captured in SRH. But does this additional information lead to better prediction of subsequent mortality?

In the present study, we analyze this question, examining the predictive power of self, interviewer, and physician health assessments for future mortality. Our objectives are to determine 1) which of self-rated health (SRH), interviewer-rated health (IRH) and physician rated health (PRH) best predicts mortality in the follow-up period, 2) the incremental predictive content of each assessment when all three are included in a multivariate specification, and 3) whether any predictive power persists after controlling for a variety of demographic and health information.

Our analysis of these questions sheds light on the value of these assessments as predictors of future mortality, and may help researchers decide whether interviewer and physician assessments should be included in a survey. Of particular interest is whether specialized medical training renders physicians' assessments superior to interviewers' assessments, or whether interviewers benefit from the extended time period talking with and observing respondents. Interviewer assessments of overall health are relatively simple and inexpensive to incorporate into face-to-face surveys; if considering interviewer assessments improves mortality prediction, it may be worthwhile for surveys to obtain these assessments.

Data and Method

Data used in this analysis are from the 2006 wave of SEBAS and mortality follow-up through June 2011. SEBAS is a subsample from the Study of Health and Living Status of the Elderly in Taiwan, also called the Taiwan Longitudinal Study of Aging (TLSA). TLSA is a national longitudinal study that began in 1989 with a sample of Taiwanese adults aged 60 and above. Follow-up waves have been conducted every 3-4 years, with additional samples of adults aged 50-66 being added in 1996 and 2003. The 2011 wave of TLSA is currently being fielded. The first wave of SEBAS in 2000 consists of a subsample of the respondents in the 1999 wave of TLSA; SEBAS's second wave in 2006 includes survivors from the 2000 wave as well as an additional sample of participants who were added to TLSA in 2003. Further details on TLSA and SEBAS can be found elsewhere (Chang et al. 2007).[2]

The 2006 wave of SEBAS consists of a home interview (n=1,284, 87% response rate) and a hospital-based physical exam (n=1,036, 81% of those interviewed). The home interview includes questions on self-reported health conditions (e.g. chronic disease diagnoses) and interviewer-administered physical performance measures (e.g. timed walks and chair stands). The hospital-based exam typically takes place several weeks after the home interview. The exam includes collection of a blood sample (obtained by a phlebotomist in the hospital) and a 12-hour overnight urine specimen (collected at home by respondents the morning of the exam); these samples allow the measurement of a variety of biomarkers, including cholesterol, blood sugar, and neuroendocrine markers. The exam also yields information on certain physical abnormalities detected by the physician, based on a physical examination and abdominal ultrasound. Exclusion criteria for the exam include living in an institution, and having a serious health condition. As a result of these criteria and the pattern of voluntary participation, respondents who did not participate in the hospital exam are older, have more ADL limitations, and are more likely to have died than participants. Exam participants and non-participants are not significantly different in terms of sex, education, employment status, residential arrangements, participation in social activities, SRH, or depression measures.

Follow-up mortality data were collected in preparation for TLSA's 2008 wave and 2011 wave. Deaths identified in TLSA are verified by linkage with the Taiwanese Ministry of the Interior's Household

[2] See also the following unpublished manuscripts available from the authors: Chang MC, Lin HS, Chuang YL, et al. Social Environment and Biomarkers of Aging Study in Taiwan (SEBAS 2000 and SEBAS 2006): main documentation for SEBAS longitudinal public use data.

Registration file and the Department of Health's death registration records. As of June 2011, 159 SEBAS respondents (12.4% of those interviewed) died; of these, 62 did not complete the hospital exam.

Respondents, interviewers, and physicians rate respondents' health using identical scales. SEBAS respondents are asked to rate their own overall health with the following question: "Regarding your current state of health, do you feel it is excellent (5), good (4), average (3), not so good (2), or poor (1)?" This question is asked early in the home interview, before detailed questions about health conditions are asked and physical assessments are conducted. Interviewers and physicians are asked: "Regarding the respondent's current state of health, do you (interviewer/physician) feel it is excellent (5), good (4), average (3), not so good (2), or poor (1)?" Interviewers are asked this question at the conclusion of the home interview; that is, after the respondent has answered questions about health conditions and performed physical functioning tasks. Physicians assess respondents' health after performing a physical exam and reviewing a medical history form; they do not have access to information collected during the interview. Neither interviewers nor physicians know the results of any laboratory tests or biomarker measures at the time of assessment.

We consider a range of covariates that may mediate the relationship between the three health assessments and subsequent mortality. All covariates are measured as of SEBAS's 2006 wave. Sociodemographic variables include age, sex, urban/rural residence, educational attainment, marital status, participation in social activities, socioeconomic index (SEI) of occupation, and self-perceived SES. Participation in social activities is measured as a count of social organizations, such as neighborhood associations or religious groups, in which the respondent reports participating in. The SEI occupation code measures occupational prestige based on respondents' lifetime occupation, ranging from 55.1 for farm laborers to 76.1 for doctors. The codes were constructed following Tsai and Chiu (1991). To measure self-perceived SES, respondents are asked to identify their relative socioeconomic position in Taiwan by imagining where they would be located on a ten-rung ladder.

Self-reported chronic conditions are included as dichotomous variables for whether the respondent has high blood pressure, diabetes, heart disease, cancer, respiratory disease, ulcer, liver disease, kidney disease, and gout. Smoking is also dichotomous for whether the respondent reports smoking daily.

We consider three indices of self-reported functioning. First is the number of activities of daily life (ADLs) that the respondent reports having any difficulty with. Six ALDs are included in the measure: bathing, dressing, eating, getting out of bed/standing up/sitting in a chair, moving around the house, and toileting. The measure ranges from 0 to 6. The second variable is the number of the following nine mobility activities the respondent reports having difficulty with: standing for 15 minutes, standing for 2 hours, squatting, reaching over one's head, grasping with fingers, lifting/carrying 11-12 kg, running 20-30m, walking 200-300m, and climbing 2-3 flights of stairs. The final self-reported functioning variable is a version of the second variable: all of the items except for standing for two hours are coded as 0=no difficulty, 1=a little/some difficulty, 2=a lot of difficulty, and 3=unable. The eight items are summed, a constant of 0.5 is added, and the log is taken.

Finally, four interviewer-administered performance-based functioning tests are considered. Hand grip strength (in kg) and peak lung flow (in L/sec) are measured as the maximum value of three trials. Normal walking speed (m/sec) is measured as the faster of two trials, and chair stand speed (chair stands/sec) is measured based on completion time for five chair stands.

Results

Tables 1-2 show descriptive statistics for the analysis sample. Table 1 lists the mean and standard deviation for covariates considered in our models. Table 2 shows the distribution of health ratings provided by respondents, interviewers, and physicians among respondents alive as of 6/30/2011 (Panel 1), and among those who died by 6/30/2011 (Panel 2). Among respondents still alive on 6/30/2011, interviewers give the most positive ratings (mean: 3.89), followed by physicians (mean: 3.37); respondents' self ratings are lowest (mean: 3.17). This pattern holds among the subset of respondents who went on to die by mid 2011: interviewers provide the most positive ratings (mean: 3.14), followed by physicians (mean: 3.07) and respondents (2.63). Interviewers show the largest difference between the mean rating given to respondents alive by 6/30/2011 and the mean rating given to respondents who die; physicians show the smallest difference.

We use logistic regression models to estimate the explanatory power of 2006 self, interviewer, and physician health ratings for subsequent mortality. All analyses are conducted in Stata (version 11.2).

Four logistic regression models of mortality as a function of 2006 health ratings are shown in Table 3. In Models 1-3, mortality depends only on the rating given by each of the three assessors in turn. The reference group in each model consists of those respondents whose health was rated 3 (average). As expected, in all three models, the relative odds of dying generally increase as the health rating decreases from average to poor, and decrease as the health rating increases from average to excellent. Chi squared tests reject the null hypothesis that mortality is independent of the health ratings of all three assessors (p-value < 0.001 for respondents and interviewers; p-value = 0.002 for physicians).

The magnitude of the change in odds of death with changing ratings, however, varies considerably between the three assessments. The odds of dying for a respondent whose health is rated 1 (poor) by a physician is 2.34 times the odds for a respondent whose health is rated 3 (average). However, this difference in odds is not statistically significant at the 5% level (p-value = 0.453). If the same rating is provided by the respondent, the odds of mortality is 4.39 times that of the reference group (p-value < 0.001); if an interviewer provides the rating, the odds ratio is 9.01 (p-value <0.001). A respondent whose physician rating is 5 (excellent) has half the odds of dying compared to a respondent with physician rating 3 (average); a respondent with a self rating of 5 has odds of dying 60% lower than that of a respondent with self rating 3. Neither of these differences are statistically significant, however (p-value = 0.359 and 0.134, respectively). Respondents with an interviewer rating of 5 have 83% lower odds of dying compared to respondents with an interviewer rating of 3 (p-value <0.001). The odds of mortality change most dramatically with changing interviewer ratings. The model of mortality based on interviewer ratings (Model 2) has the highest pseudo R^2 of the three models, indicating that interviewer ratings provide the best univariate prediction of mortality over the next five years.

Model 4 (Table 2) includes self, interviewer, and physician ratings as explanatory variables, allowing mortality odds to vary with all three assessments. In this model, the coefficients on interviewer ratings remain large in magnitude, and a chi squared test of the joint significance of all rating levels of interviewer assessments show that interviewer ratings are statistically significant in predicting mortality (p-value < 0.001). The ratings given by respondents and physicians become jointly insignificant (and, with the exception of a self rating of 2, individually insignificant).

Table 4 shows logistic regression models that add sociodemographic characteristics, self-reported chronic conditions, self-reported functioning, and performance-based functioning as covariates. Table 4(a) adds sociodemographic covariates and smoking status. The magnitude of the coefficients on the health ratings declines after adding these covariates, indicating that sociodemographic characteristics account for some of the assessments' predictive power. Self ratings (Model 1) and interviewer ratings (Model 2) remain jointly significant in predicting mortality beyond these sociodemographic characteristics, but physician ratings (Model 3) become jointly insignificant in the face of the covariates. When all three assessments are considered in the same model (Model 4), interviewer ratings remain jointly significant, but self ratings and physician ratings are insignificant.

We consider models with health assessments and self-reported chronic conditions in Table 4(b). As with sociodemographic characteristics, adding chronic conditions as predictors of mortality reduces the magnitude of the odds ratios on self, interviewer and physician health ratings. Self and interviewer ratings remain jointly significant (Models 1-2); physician ratings do not. When all ratings are considered together in addition to the chronic conditions, only interviewer ratings are jointly significant in predicting subsequent mortality.

In Table 4(c), we add self-reported functioning variables to the base models. These variables reduce the magnitude of the odds ratios associated with self, interviewer, and physician health ratings. When these functioning variables are considered, self ratings and physician ratings become insignificant (Models 1 and 3); only interviewer ratings remain jointly significant in predicting subsequent mortality (Model 2). This predictive power persists when all three assessments are considered together with the self-reported functioning variables (Model 4).

Table 4(d) shows models with performance-based functioning added to the base models. Physician and self assessments become insignificant predictors of mortality at the 5% level (Models 1 and 3), while interviewer assessments remain significant (Model 2). When all three assessments are considered together, none of the assessments remain jointly significant as a predictor of mortality at the 5% level, though interviewer ratings are jointly significant at the 10% level (Model 4).

Discussion and Conclusion

In this study, we investigate whether self, interviewer, and physician health ratings predict subsequent mortality. In base models with only the health assessments considered as predictors, interviewers outperform self and physician ratings in predicting mortality over the five year follow-up period. We consider a number of additional models that in turn add information on sociodemographic characteristics, self-reported chronic conditions, self-reported physical functioning, and performance-based physical functioning. Physician ratings fail to add predictive power beyond the covariates in any of the models. Self ratings are jointly significant as predictors of mortality beyond sociodemographic characteristics and self-reported chronic conditions, but they fail to improve mortality prediction beyond self-reported functioning or performance-based functioning. Interviewer ratings are jointly significant as predictors of subsequent mortality beyond all four sets of considered covariates.

When self, interviewer, and physician ratings are considered in the same model, only interviewer ratings maintain predictive power of subsequent mortality, and this predictive power persists when covariates are added to the model. The exception is performance-based functioning covariates, the addition of which make interviewer assessments insignificant at the 5% level (though still significant at the 10% level).

There is something special about performance-based functioning measures, but it remains unclear whether these measures are the mechanism by which interviewers make health ratings that are more predictive of mortality. For example, it could be that interviewers obtain the same information from casual observation of respondents moving about during the interview as from the formal performance-based functioning tests. If this were the case, then the predictive power of interviewers' health assessments could simply be due to observation during the lengthy interview. This hypothesis is currently untestable, as the current data, which include performance-based functioning measures, are the only data of which we are aware that include interviewer assessed health measures.

The poor performance of physician health ratings in predicting subsequent mortality is a puzzle. Since they are medical experts, their ratings might be expected to be highly correlated with mortality. Physicians place greater weight than respondents or interviewers on clinical measures and risk factors in assessing respondents' health (Smith and Goldman 2011), but these factors are not strong predictors of future mortality.

While SRH is nearly always collected in large social and health surveys, external health ratings almost never are. It is impossible to say whether respondents, interviewers, or physicians are most accurate in assessing health status, but we can say that interviewers add information that improves prediction of subsequent mortality. Our results suggest that interviewer assessments should be added to health surveys. Adding interviewer health ratings to a survey is relatively easy and inexpensive, making this measure even more attractive.

Future research should focus on determining the mechanism by which interviewer ratings are better predictors of subsequent mortality than respondents and physicians. Is it the length of time interviewers spend with respondents or the specific performance-based functioning measures that make interviewer ratings a good predictor? Is this finding unique to Taiwan? Would the health ratings of medically trained interviewers perform more like physicians or like untrained interviewers? Answering these questions will help researchers better understand how health assessments are made, and design survey questions to more easily and efficiently measure and collect information related to health status and mortality risk.

References:

Benyamini, Yael, Tzvia Blumstein, Ayala Lusky, and Baruch Modan. 2003. “Gender Differences in the Self-Rated Health-Mortality Association: Is It Poor Self-Rated Health That Predicts Mortality or Excellent Self-Rated Health That Predicts Survival?” *The Gerontologist* 43 (3): 396–405.

Benyamini, Yael, and Ellen L. Idler. 1999. “Community Studies Reporting Association Between Self-Rated Health and Mortality: Additional Studies, 1995 to 1998.” *Research on Aging* 21 (3) (May): 392–401. doi:10.1177/0164027599213002.

Benyamini, Yael, Elaine A. Leventhal, and Howard Leventhal. 2000. “Gender Differences in Processing Information for Making Self-assessments of Health.” *Psychosomatic Medicine* 62 (3): 354–364.

———. 2003. “Elderly People’s Ratings of the Importance of Health-related Factors to Their Self-assessments of Health.” *Social Science & Medicine* 56 (8) (April): 1661–1667. doi:10.1016/S0277-9536(02)00175-2.

Boardman, Jason D. 2004. “Health Pessimism Among Black and White Adults: The Role of Interpersonal and Institutional Maltreatment.” *Social Science & Medicine* 59 (12): 11.

Chang, Ming-Cheng, Dana A. Glei, Noreen Goldman, and Maxine Weinstein. 2007. “The Taiwan Biomarker Project.” In *Biosocial Surveys: Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Sicence Surveys*, 60–77. The National Academies Press.

DeSalvo, Karen B., Nicole Bloser, Kristi Reynolds, Jiang He, and Paul Muntner. 2006. “Mortality Prediction with a Single General Self-Rated Health Question. A Meta-Analysis.” *Journal of General Internal Medicine* 21 (3): 267–275. doi:doi:10.1111/j.1525-1497.2005.00291.x.

Dowd, Jennifer Beam, and Anna Zajacova. 2007. “Does the Predictive Power of Self-rated Health for Subsequent Mortality Risk Vary by Socioeconomic Status in the US?” *Int. J. Epidemiol.* 36 (6): 1214–1221. doi:10.1093/ije/dym214.

Etile, Fabrice, and Carine Milcent. 2006. “Income-related Reporting Heterogeneity in Self-assessed Health: Evidence from France.” *Health Economics* 15 (9): 17.

Finch, Brian Karl, Robert A. Hummer, Maureen Reindl, and William A. Vega. 2002. “Validity of Self-rated Health Among Latino(a)s.” *Am. J. Epidemiol.* 155 (8): 755–759. doi:10.1093/aje/155.8.755.

Friedsam, H J, and H W Martin. 1963. “A Comparison of Self and Physicians’ Health Ratings in an Older Population.” *Journal of Health and Human Behavior* 4: 179–183.

Goldman, Noreen, Dana A. Glei, and Ming-Cheng Chang. 2004. “The Role of Clinical Risk Factors in Understanding Self-rated Health.” *Annals of Epidemiology* 14 (1) (January): 49–57. doi:10.1016/S1047-2797(03)00077-2.

Idler, Ellen L. 1993. “Age Differences in Self-Assessments of Health: Age Changes, Cohort Differences, or Survivorship?” *Journal of Gerontology: Social Sciences* 48 (6): S289–S300.

Idler, Ellen L., and Yael Benyamini. 1997. “Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies.” *Journal of Health and Social Behavior* 38 (1): 17.

Jylhä, Marja, Stefano Volpato, and Jack M. Guralnik. 2006. “Self-rated Health Showed a Graded Association with Frequently Used Biomarkers in a Large Population Sample.” *Journal of Clinical Epidemiology* 59 (5) (May): 465–471. doi:10.1016/j.jclinepi.2005.12.004.

Krause, Neal M., and Gina M. Jay. 1994. “What Do Global Self-rated Health Items Measure?” *Medical Care* 32 (9): 930–942.

Lindeboom, Maarten, and Eddy van Doorslaer. 2004. “Cut-point Shift and Index Shift in Self-reported Health.” *Journal of Health Economics* 23 (6): 17.

Markides, K S, D J Lee, L A Ray, and S A Black. 1993. “Physicians’ Ratings of Health in Middle and Old Age: a Cautionary Note.” *Journal of Gerontology* 48 (1) (January): S24–27.

Schnittker, Jason. 2005. “When Mental Health Becomes Health: Age and the Shifting Meaning of Self-Evaluations of General Health.” *The Milbank Quarterly* 83 (3): 397–423. doi:doi:10.1111/j.1468-0009.2005.00407.x.

Smith, Kimberly V., and Noreen Goldman. 2011. "Measuring Health Status: Self-, Interviewer, and Physician Reports of Overall Health." *Journal of Aging and Health* 23 (2) (March): 242–266. doi:10.1177/0898264310383421.

Spencer, S. Melinda, Richard Schulz, Ronica N. Rooks, Steven M. Albert, Roland J. Thorpe, Gretchen A. Brenes, Tamara B. Harris, et al. 2009. "Racial Differences in Self-Rated Health at Similar Levels of Physical Functioning: An Examination of Health Pessimism in the Health, Aging, and Body Composition Study." *Journal of Gerontology: Psychological Ans Social Sciences.* 64B (1): 87–94. doi:10.1093/geronb/gbn007.

Suchman, Edward A., Bernard S. Phillips, and Gordon F. Streib. 1958. "An Analysis of the Validity of Health Questionnaires." *Social Forces* 36 (3) (March): 223–232.

Tsai, S-L, and H-Y Chiu. 1991. "Constructing Occupational Scales for Taiwan." In *Research in Social Stratification and Mobility: A Research Annual*, ed. R Althauser and M Wallace, 229–253. Greenwich, CT: JAI Press Inc.

Valanis, B G, and R Yeaworth. 1982. "Ratings of Physical and Mental Health in the Older Bereaved." *Research in Nursing & Health* 5 (3) (September): 137–146.

Table 1: Descriptive statistics

	Mean or Percent	Standard Deviation
Sociodemographic covariates		
Female	46.1%	0.50
Age	66.34	10.05
SEI occupation code	62.06	5.12
Urban	58.8%	0.49
SES ladder	6.30	13.63
Completed education	1.55	1.29
Number of social activities	0.82	1.10
Married or has companion	75.8%	0.43
Smokes daily	16.6%	0.37
Self-reported chronic conditions		
High blood pressure	33.5%	0.47
Take meds for high BP	560.1%	3.15
Diabetes	17.4%	0.38
Heart disease	16.7%	0.37
Cancer	1.5%	0.12
Respiritory disease	6.6%	0.25
Ulcer	14.2%	0.35
Liver disease	8.4%	0.28
Kidney disease	5.1%	0.22
Gout	8.0%	0.27
Self-reported functioning		
Num ADL difficulties	0.21	0.90
Num mobility difficulties	1.92	2.42
Difficulty functioning scale	0.38	1.26
Performance-based functioning		
Missing grip strength	2.8%	0.17
Grip strength (kg)	26.96	11.38
Missing peak flow	2.0%	0.14
Peak flow (L/min)	325.90	149.44
Missing walk	3.1%	0.17
Walk speed (m/sec)	0.83	0.32
Missing chair stand	7.7%	0.27
Chair stand (stands/sec)	0.48	0.23
N	953	

Table 2: Frequency of health ratings assigned by respondents, interviewers, and physicians

	Panel 1: Respondents alive as of 6/30/2011			Panel 2: Respondents who died by 6/30/2011		
	Self	Interviewer	Physician	Self	Interviewer	Physician
Excellent (5)	10.9%	27.2%	4.3%	3.1%	7.2%	2.1%
Good (4)	22.3%	44.3%	39.7%	16.5%	35.1%	29.9%
Average (3)	43.1%	19.7%	44.7%	30.9%	30.9%	42.3%
Not so Good (2)	20.4%	8.2%	10.7%	39.2%	18.6%	24.7%
Poor (1)	3.3%	0.6%	0.5%	10.3%	8.2%	1.0%
Mean Rating	3.17	3.89	3.37	2.63	3.14	3.07
N[a]	856	856	856	97	97	97

[a] N is the number of obsevations with a valid value for all three health assessment.

Table 3: Logistic regression models of mortality by June 2011

Ratings		Model 1 OR	Model 1 P-value	Model 2 OR	Model 2 P-value	Model 3 OR	Model 3 P-value	Model 4 OR	Model 4 P-value
Self	Poor (1)	4.39	<0.001					1.44	0.469
	Not so good (2)	2.67	<0.001					1.85	0.029
	Average (3)	1.00	.					1.00	.
	Good (4)	1.03	0.926					1.39	0.336
	Excellent (5)	0.40	0.134					0.88	0.838
Interviewer	Poor (1)			9.01	<0.001			7.45	0.004
	Not so good (2)			1.45	0.262			1.16	0.688
	Average (3)			1.00	.			1.00	.
	Good (4)			0.51	0.011			0.59	0.062
	Excellent (5)			0.17	<0.001			0.22	0.001
Physician	Poor (1)					2.34	0.453	1.47	0.762
	Not so good (2)					2.44	0.002	1.76	0.073
	Average (3)					1.00	.	1.00	.
	Good (4)					0.80	0.371	0.98	0.928
	Excellent (5)					0.51	0.359	0.81	0.785
N[a]		953		953		953		953	
Pseudo R^2		0.05		0.09		0.02		0.10	
Chi squared, full model		30.32		47.16		16.69		56.38	
P-value		<0.001		<0.001		0.002		<0.001	
DF		4		4		4		12	
Chi squared, self								5.27	
P-value								0.260	
DF								4	
Chi squared, interviewer								22.75	
P-value								<0.001	
DF								4	
Chi squared, physician								3.73	
P-value								0.444	
DF								4	

[a] The sample includes all respondents who participated in both the home interview and the hospital-based exam.

Table 4(a): Logistic regression models of mortality by June 2011 with sociodemographic covariates

		Model 1		Model 2		Model 3		Model 4	
Ratings		OR	P-value	OR	P-value	OR	P-value	OR	P-value
Self	Poor (1)	3.60	0.021					1.80	0.342
	Not so good (2)	2.38	0.003					1.74	0.095
	Average (3)	1.00	.					1.00	.
	Good (4)	0.95	0.903					1.07	0.868
	Excellent (5)	0.39	0.247					0.69	0.641
Interviewer	Poor (1)			5.54	0.022			4.23	0.073
	Not so good (2)			1.62	0.206			1.28	0.550
	Average (3)			1.00	.			1.00	.
	Good (4)			0.53	0.041			0.62	0.134
	Excellent (5)			0.22	0.003			0.29	0.020
Physician	Poor (1)					1.68	0.517	1.19	0.842
	Not so good (2)					1.73	0.089	1.25	0.533
	Average (3)					1.00	.	1.00	.
	Good (4)					1.01	0.970	1.13	0.711
	Excellent (5)					0.94	0.924	1.39	0.628
Female		0.38	0.008	0.33	0.002	0.41	0.010	0.34	0.004
Age		0.86	0.447	0.87	0.486	0.93	0.718	0.85	0.420
Age squared		1.00	0.239	1.00	0.276	1.00	0.442	1.00	0.231
SEI occupation code		0.99	0.780	1.01	0.858	0.99	0.666	1.00	0.973
Urban		1.22	0.476	1.36	0.256	1.21	0.493	1.34	0.292
SES ladder		1.01	0.192	1.01	0.345	1.01	0.168	1.01	0.293
Completed education		1.02	0.903	0.96	0.807	0.98	0.894	1.00	0.974
Number of social activities		0.87	0.338	0.94	0.674	0.86	0.270	0.94	0.649
Married or has companion		0.53	0.040	0.53	0.037	0.49	0.015	0.55	0.051
Smokes daily		1.24	0.510	1.22	0.567	1.13	0.713	1.22	0.570
N[a]		834		834		834		834	
Pseudo R^2		0.211		0.225		0.184		0.233	
Chi squared, full model		93.42		107.83		95.85		119.04	
P-value		<0.001		<0.001		<0.001		<0.001	
DF		14		14		14		22	
Chi squared, self		16.78						3.48	
P-value		0.002						0.481	
DF		4						4	
Chi squared, interviewer				25.41				11.88	
P-value				<0.001				0.018	
DF				4				4	
Chi squared, physician						3.62		0.56	
P-value						0.461		0.967	
DF						4		4	

[a] The sample includes all respondents who participated in both the home interview and the hospital-based exam.

Table 4(b): Logistic regression models of mortality by June 2011 with self-reported chronic conditions covariates

		Model 1		Model 2		Model 3		Model 4	
Ratings		OR	P-value	OR	P-value	OR	P-value	OR	P-value
Self	Poor (1)	3.07	0.017					1.23	0.703
	Not so good (2)	2.36	0.003					1.76	0.064
	Average (3)	1.00	.					1.00	.
	Good (4)	1.16	0.653					1.45	0.277
	Excellent (5)	0.49	0.251					0.92	0.902
Interviewer	Poor (1)			7.77	0.001			6.96	0.008
	Not so good (2)			1.36	0.360			1.15	0.700
	Average (3)			1.00	.			1.00	.
	Good (4)			0.59	0.056			0.65	0.134
	Excellent (5)			0.22	0.001			0.25	0.004
Physician	Poor (1)					1.23	0.843	1.19	0.878
	Not so good (2)					2.11	0.014	1.72	0.093
	Average (3)					1.00	.	1.00	.
	Good (4)					1.02	0.945	1.09	0.770
	Excellent (5)					0.77	0.737	0.98	0.982
High blood pressure		1.70	0.359	1.42	0.557	1.67	0.369	1.54	0.491
Take meds for high BP		1.04	0.649	1.03	0.757	1.04	0.621	1.04	0.665
Diabetes		1.73	0.035	1.39	0.218	1.83	0.021	1.34	0.295
Heart disease		1.26	0.421	1.38	0.258	1.46	0.164	1.28	0.403
Cancer		2.27	0.203	2.21	0.193	2.93	0.118	2.08	0.212
Respiritory disease		1.44	0.368	1.55	0.242	1.76	0.155	1.45	0.358
Ulcer		1.05	0.876	1.14	0.650	1.20	0.552	1.03	0.924
Liver disease		0.92	0.837	0.91	0.841	0.81	0.630	0.89	0.791
Kidney disease		1.74	0.184	1.48	0.338	1.85	0.121	1.52	0.301
Gout		0.92	0.833	0.89	0.760	1.00	0.993	0.91	0.801
N[a]		953		953		953		953	
Pseudo R^2		0.071		0.099		0.055		0.111	
Chi squared, full model		46.14		59.57		38.22		67.36	
P-value		<0.001		<0.001		<0.001		<0.001	
DF		14		14		14		22	
Chi squared, self		14.75						4.22	
P-value		0.005						0.377	
DF		4						4	
Chi squared, interviewer				30.78				18.43	
P-value				<0.001				0.001	
DF				4				4	
Chi squared, physician						6.69		2.86	
P-value						0.153		0.582	
DF						4		4	

[a] The sample includes all respondents who participated in both the home interview and the hospital-based exam.

Table 4(c): Logistic regression models of mortality by June 2011 with self-reported functioning covariates

		Model 1		Model 2		Model 3		Model 4	
Ratings		OR	P-value	OR	P-value	OR	P-value	OR	P-value
Self	Poor (1)	2.04	0.116					1.28	0.633
	Not so good (2)	1.64	0.083					1.57	0.120
	Average (3)	1.00	.					1.00	.
	Good (4)	1.29	0.437					1.48	0.249
	Excellent (5)	0.61	0.439					0.96	0.950
Interviewer	Poor (1)			6.42	0.017			6.21	0.040
	Not so good (2)			0.94	0.868			0.84	0.657
	Average (3)			1.00	.			1.00	.
	Good (4)			0.66	0.134			0.68	0.170
	Excellent (5)			0.33	0.015			0.35	0.029
Physician	Poor (1)					1.04	0.972	1.23	0.856
	Not so good (2)					1.80	0.051	1.71	0.088
	Average (3)					1.00	.	1.00	.
	Good (4)					1.16	0.591	1.17	0.598
	Excellent (5)					1.02	0.981	1.09	0.909
Num ADL difficulties		1.02	0.873	0.90	0.409	1.03	0.768	0.88	0.330
Num mobility difficulties		0.98	0.853	0.96	0.748	0.97	0.797	0.96	0.725
Difficulty functioning scale		1.82	0.013	1.79	0.016	1.96	0.006	1.76	0.027
N[a]		947		947		947		947	
Pseudo R^2		0.108		0.123		0.105		0.133	
Chi squared, full model		67.38		66.04		67.69		73.11	
P-value		<0.001		<0.001		<0.001		<0.001	
DF		7		7		7		15	
Chi squared, self		5.39						3.09	
P-value		0.249						0.543	
DF		4						4	
Chi squared, interviewer				13.37				10.57	
P-value				0.010				0.032	
DF				4				4	
Chi squared, physician						3.89		2.92	
P-value						0.421		0.571	
DF						4		4	

[a] The sample includes all respondents who participated in both the home interview and the hospital-based exam.

Table 4(d): Logistic regression models of mortality by June 2011 with performance-based functioning covariates

		Model 1		Model 2		Model 3		Model 4	
Ratings		OR	P-value	OR	P-value	OR	P-value	OR	P-value
Self	Poor (1)	2.43	0.048					1.46	0.475
	Not so good (2)	1.95	0.020					1.66	0.091
	Average (3)	1.00	.					1.00	.
	Good (4)	1.34	0.384					1.58	0.190
	Excellent (5)	0.55	0.339					0.91	0.888
Interviewer	Poor (1)			4.71	0.049			4.13	0.101
	Not so good (2)			1.17	0.679			1.00	0.995
	Average (3)			1.00	.			1.00	.
	Good (4)			0.63	0.112			0.68	0.205
	Excellent (5)			0.30	0.007			0.33	0.025
Physician	Poor (1)					1.16	0.884	1.17	0.892
	Not so good (2)					1.98	0.031	1.83	0.063
	Average (3)					1.00	.	1.00	.
	Good (4)					1.01	0.985	1.00	0.990
	Excellent (5)					0.86	0.845	0.97	0.972
Missing grip strength		0.23	0.144	0.21	0.147	0.21	0.142	0.20	0.142
Grip strength (kg)		0.98	0.315	0.99	0.420	0.98	0.224	0.99	0.504
Missing peak flow		3.73	0.112	2.88	0.254	3.34	0.213	2.44	0.324
Peak flow (L/min)		1.00	0.383	1.00	0.469	1.00	0.225	1.00	0.352
Missing walk		1.05	0.951	1.01	0.992	0.90	0.883	1.04	0.960
Walk speed (m/sec)		0.46	0.283	0.58	0.443	0.48	0.333	0.70	0.618
Missing chair stand		0.80	0.704	0.71	0.549	0.93	0.895	0.90	0.858
Chair stand (stands/sec)		0.32	0.282	0.33	0.252	0.35	0.351	0.39	0.337
N[a]		930		930		930		930	
Pseudo R^2		0.097		0.106		0.089		0.119	
Chi squared, self		9.41						3.84	
P-value		0.052						0.428	
DF		4						4	
Chi squared, interviewer				12.99				8.96	
P-value				0.011				0.062	
DF				4				4	
Chi squared, physician						5.14		3.81	
P-value						0.273		0.432	
DF						4		4	

[a] The sample includes all respondents who participated in both the home interview and the hospital-based exam.