Do interviewer and physician health ratings predict mortality? A comparison with self-rated health

Megan A. Todd and
Office of Population Research and Woodrow Wilson School of Public and International Affairs

Noreen Goldman
Office of Population Research and Woodrow Wilson School of Public and International Affairs

Abstract

Background—Despite the serious biases that characterize self-rated health, researchers rely heavily on these ratings to predict mortality. Using newly collected survey data, we examine whether simple ratings of participants' health provided by interviewers and physicians can markedly improve mortality prediction.

Methods—We use data from a prospective cohort study based on a nationally representative sample of older adults in Taiwan. We estimate proportional hazard models of all-cause mortality between the 2006 interview and 30 June 2011 (mean 4.7 years follow-up).

Results—Interviewer ratings were more strongly associated with mortality than physician or self-ratings, even after controlling for a wide range of covariates. Neither respondent nor physician ratings substantially improve mortality prediction in models that include interviewer ratings. The predictive power of interviewer ratings likely arises in part from interviewers' incorporation of information about the respondents' physical and mental health into their assessments.

Conclusions—The findings of this study support the routine inclusion of a simple question at the end of face-to-face interviews, comparable to self-rated health, asking interviewers to provide an assessment of respondents' overall health. The costs of such an undertaking are minimal and the potential gains substantial for demographic and health researchers. Future work should explore the strength of the link between interviewer ratings and mortality in other countries and in surveys that collect less detailed information on respondent health, functioning, and well-being.

In an effort to assess a person's health, researchers often rely on a survey question that asks respondents to rate their overall health using four or five ordered adjectives ranging from poor to excellent. This widely used measure, called self-rated health, has been shown to predict health outcomes including morbidity, health care utilization, physical functioning and mortality, even after controlling for objective measures of health.[1,2] The utility of this simple question results from its encapsulation of information from various health domains, family history, socio-demographic variables, biological factors and clinical measurements.[3,4]

Corresponding author: Megan A. Todd, mtthree@princeton.edu. Office of Population Research, Wallace Hall, 2nd floor, Princeton University, Princeton, NJ 08544. Ph: 617-821-5077, F: 609-258-1039.

Conflicts of Interest: None declared.

Nevertheless, self-rated health suffers from biases that limit its value. Reported variation in self-rated health by socioeconomic status, race, ethnicity, sex, and age may reflect actual differences in health, but may also reflect differences in how respondents think about and describe their health. For example, reporting may be affected by personality, social environment, and language, and sub-populations may use distinct reference groups when assessing their health. [5–8] These differences in reporting style make it difficult to directly compare self-rated health across population groups. In addition, respondents' health reports may deemphasize factors known to be predictors of health and survival, such as smoking and functional limitations.[9]

Despite these problems with self-rated health, researchers have rarely collected global health ratings from external evaluators. The exception is several older studies that collected health ratings from physicians or nurses, typically as "objective" measures with which to validate "subjective" self-rated health measures.[10–14] This gap in research is surprising given two recent findings that suggest non-health personnel may provide valuable health assessments. First, Christensen and colleagues[15] found that when strangers used facial photographs to estimate the age of elderly respondents, this perceived age was as strong a predictor of dying in the follow-up period as actual age, indicating that health information was conveyed by simply observing respondents' faces. Undoubtedly, more insights could be gleaned from directly observing not only the respondent's appearance, but also speech, movement and functioning. Second, a recent study in Taiwan compared self-rated health with corresponding health assessments made by physicians and interviewers, concluding that these external evaluators placed different weight on health-related variables than did respondents.[9] This suggests that external health assessments may provide additional health information not reflected in self-rated health.

We analyzed data from the same survey in Taiwan to determine whether health assessments provided by physicians and interviewers improve mortality prediction. Interviewer ratings would be particularly promising if inclusion of this simple, essentially cost-free question in household surveys were to enhance forecasts of survival and future health. To the best of our knowledge, no previous study has examined links between interviewer health assessments and mortality.

Methods

Data

Data are from the second wave (2006) of the Social Environment and Biomarkers of Aging Study, with mortality follow-up through June 2011 (4.7 years, on average). The first wave (2000) of the Social Environment and Biomarkers of Aging Study was based on a random subsample from the 1999 wave of the Taiwan Longitudinal Study of Aging, a national longitudinal study of persons 60+ initiated in 1989. Follow-up waves of the Taiwan Longitudinal Study of Aging have been conducted every 3-4 years with refresher samples aged 50-66 in 1996 and 2003. The 2006 Social Environment and Biomarkers of Aging Study sample is based on: (1) respondents aged 60+ who participated in the medical exam component of the 2000 wave of the study; and (2) respondents aged 53-60 first sampled in the 2003 wave of the Taiwan Longitudinal Study of Aging. All protocols for the Social Environment and Biomarkers of Aging Study were approved by human subjects committees in Taiwan, and at Georgetown University and Princeton University. All participants gave informed consent before taking part in the study. eFigure 1(http://links.lww.com/EDE/A715) illustrates the two studies; further details are provided elsewhere.[16,17]

The 2006 wave of the Social Environment and Biomarkers of Aging Study consists of a home interview (n=1,284, 87% response rate) and a hospital-based exam (n=1,036, 81% of

those interviewed). Written informed consent was obtained for both components. The interview includes extensive questions on the social and economic environment and health of the respondents and interviewer-administered physical performance measures. Several weeks after the interview, a physical exam was given to respondents at a nearby hospital. The exam included: (1) collection of a venous blood sample, overnight urine collection, and anthropometric measurements; and (2) reports of abnormalities based on a physical examination and abdominal ultrasound. Exams were conducted by physicians affiliated with hospitals participating in the Social Environment and Biomarkers of Aging Study rather than by respondents' personal physicians. Other analyses document predictors of exam participation;[18,19] in the presence of control variables, average self-rated health was almost identical for exam participants and non-participants. Reasons for non-participation in the exam are provided in eFigure 1(http://links.lww.com/EDE/A715) (footnote).

Deaths identified in the Social Environment and Biomarkers of Aging Study were verified in the Taiwanese Ministry of the Interior's Household Registration file and the Department of Health's death registration records. As of June 30, 2011, 159 respondents (12% of those interviewed in 2006) had died.

Respondents, interviewers, and physicians used identical 5-point scales to rate respondents'health. Respondents were asked (in Chinese): "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 was asked early in the interview, before questions about health conditions were asked and performance assessments were conducted. Interviewers and physicians were asked (in Chinese): "Regarding the respondent's current state of health, do you feel it is excellent (5), good (4), average (3), not so good (2), or poor (1)?" Interviewers were asked this question at the conclusion of the interview. Physicians' assessments occurred after they conducted their exam and reviewed a medical history form filled out by the respondent. Physicians did not have access to information collected during the interview, or to results of laboratory tests or biomarker measures other than blood pressure and anthropometry. Respondents, interviewers, and physicians were not given any special guidance regarding how to assess respondents' health.

Variables

In an effort to understand what factors underlie differences in the predictive power of the three health assessments, we consider a range of covariates – sociodemographic factors, self-reports of health conditions and functioning, psychological well-being, and interviewer-administered performance measures – that may mediate the relationship between health assessments and age-specific mortality. Sociodemographic variables include sex, urban/rural residence, educational attainment, marital status, and participation in social activities (measured as a count of social organizations, such as neighborhood associations or religious groups, in which the respondent reports participation).

Chronic conditions are measured as 10 dichotomous variables: whether the respondent reports high blood pressure, taking medication for high blood pressure, heart disease, cancer, respiratory disease, ulcer, liver disease, kidney disease, and gout at the time of interview and whether the respondent reports ever having had diabetes. Smoking is measured as whether the respondent reports daily smoking.

Psychological well-being is captured by three variables. A 10-item version of the Center for Epidemiologic Studies Depression scale measures depressive symptoms experienced in the week prior to interview.[19] Perceived stress measures whether and to what degree the respondent feels stress due to personal or family finances, job, or relationships. An index (range 0-12) is calculated as the sum of these six items, each coded as 0 (no stress), 1 (some

stress), or 2 (a lot of stress). Cronbach's alpha for the six items is 0.71. A five-item version of the Pittsburgh Sleep Quality Index measures sleep quality (range 0-15), based on sleep duration, time to fall asleep, and feeling sleepy during the day; high values indicate poor sleep quality.[20] Cronbach's alpha for the sleep items is 0.77.

Self-reported functioning is assessed by limitations in activities of daily living and mobility. Limitations in activities of daily living are measured as the number of the following six activities with which the respondent has any difficulty: bathing, dressing, eating, getting out of bed/standing up/sitting in a chair, moving around the house, and toileting. The measure of mobility limitation counts the number of the following nine activities with which the respondent has difficulty: standing for 15 minutes, standing for two 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.[9,21]

Finally, there are four interviewer-administered performance-based functioning tests: peak lung flow (L/min), hand grip strength (kg), walking speed (m/sec, with a walking aid if needed), and chair-stand speed. Chair-stand speed is measured as the completion time for five chair stands, adjusted for the height of the chair and person, and the respondent's age and sex (details in Cornman et al.[22]). Indicator variables identify respondents unable to perform each task for any reason (e.g., unable to perform or complete the task, felt unsafe, failed to understand the instructions, refused, or they satisfied the exclusion criteria for the test, such as by having a recent injury or illness).

Analytic Strategy

The analysis is based on two samples. The interview sample includes all respondents completing the home interview, and is used for analyses comparing only the respondent and interviewer reports (n=1,197). The exam sample restricts the interview sample to include only those respondents participating in the medical exam, and is used for analyses that incorporate physician reports (n=919). Of the 1,284 respondents interviewed, 32 had missing self-rated health information and an additional 55 respondents had missing values on at least one covariate (primarily the depression scale or sleep questions), leaving 1,197 respondents in the interview sample. Because 211 of these respondents did not participate in the exam and an additional 67 were missing physician assessments, the exam sample includes 919 respondents. eFigure 1(http://links.lww.com/EDE/A715) illustrates the construction of both samples.

We estimate a series of proportional-hazard models of age-specific mortality over the follow-up period (about five years) to evaluate the explanatory power of self, interviewer, and physician health ratings. To understand whether ratings retain predictive power in the presence of information collected during the interviews, we sequentially add covariates to the models. Our models use the Gompertz distribution, which assumes that the hazard is an exponential function of age and generally provides a close fit to observed death rates at older ages. [23] All analyses are conducted in Stata (version 11.2; StataCorp., College Station, TX).

Results

Table 1 presents descriptive statistics for the interview sample, for the exam sample, and for those respondents excluded from the exam sample but included in the interview sample. On average, the self-rating is slightly higher (i.e., better) for respondents excluded from the exam sample compared with those in the exam sample (3.3 vs. 3.1 on a scale of 1 to 5). Interviewer ratings are similar among those included and excluded from the exam sample. Respondents excluded from the exam sample have somewhat worse performance-based

functioning than exam participants; this is not surprising, as some disabled respondents were intentionally excluded from the exam.

The correlation coefficients between assessments are modest: 0.55 between respondent and interviewer ratings, 0.31 between respondent and physician ratings and 0.29 between interviewer and physician ratings. A prior analysis found only slight inter-rater agreement between the different assessments (unweighted kappa statistics ranged between 0.09 and 0.13, weighted kappa statistics ranged between 0.15 and 0.27; see Smith and Goldman, 2011 [9] for details). Differences between the evaluators are evident in Table 2. For respondents who survive the follow-up period, interviewers favor the rating "good" (4), whereas respondents and physicians are most likely to choose the rating "average" (3). Among survivors, interviewers provide more positive ratings (mean: 3.9) than physicians (mean rating: 3.4) or respondents (mean rating: 3.2). For decedents, both external assessors give more positive ratings (mean rating: 3.1) than respondents (mean rating: 2.6). All evaluators provide more positive assessments for survivors than decedents, with interviewers having the largest difference and physicians the smallest difference.

Table 3 shows the assessments' predictive power for mortality among the exam sample (n=919). The first three proportional-hazard models include only a single set of assessments (self, interviewer or physician) whereas the final model includes all three. The hazard ratios (HRs) indicate the mortality rate associated with the health rating shown relative to "excellent". We use chi-square tests to measure the joint significance of a given assessment in the model; the P-value from this joint Wald test is shown at the bottom of the table. We also test for linearity in ratings by replacing categorical ratings variables with a single linear variable ranging from 1 to 5. The estimated coefficients and confidence intervals associated with this test for linear trend are provided in the appendix (e Table 1, http://links.lww.com/EDE/A715).

Self and interviewer assessments (Models 1-2 of Table 3) are each predictive of age-specific mortality. Hazard ratios (HRs) vary substantially (and generally monotonically) across ratings in these two models; in addition, the coefficients associated with the continuous rating variable indicate a linear relationship between ratings and mortality. By contrast, physician assessments are not associated with mortality (Model 3), and there is no indication of a linear trend in the ratings. The range of HRs was larger for interviewers than for respondents and physicians (i.e., the HR for "poor" relative to "excellent" is 11.2 among interviewers compared to 6.0 for respondents and 1.5 for physicians). This suggests the superiority of interviewer ratings. This inference is consistent with the estimates in Model 4: either respondent or physician assessments can be dropped without significant loss of fit, whereas eliminating interviewer ratings results in a significant loss (P=0.018). In fact, both respondent and physician assessments can be excluded (P=0.590, not shown).

Given the poor performance of physician ratings in predicting age-specific mortality in these initial models, subsequent models examine the predictive power of only interviewer and selfassessments using the larger interview sample (n=1,197). Re-estimation of Models 1 and 2 on this sample produces estimates similar to those in Table 3 (e Table 2, http://links.lww.com/EDE/A715). Re-estimation of Model 4 excluding physician assessments indicates that in the presence of interviewer-rated health, self-rated health does little to improve mortality prediction, whereas interviewer-rated health substantially improves mortality prediction in the presence of self-rated health (Table 4, Model 1).

Table 4 presents coefficients for respondent and interviewer ratings from eight nested proportional-hazard models that sequentially add sociodemographic, health, and physical functioning variables to models of age-specific mortality. The attenuation of HRs for a given

evaluator's assessments upon the introduction of a covariate (e.g., smoking) into the model suggests that: (1) the evaluator uses information conveyed by this covariate (or information associated with this covariate) when forming the health rating, and (2) the introduced covariate is associated with mortality. The HRs for interviewer ratings stay largely unchanged in the first four models as sex, social environment variables, and smoking are sequentially added as covariates. The HRs then progressively attenuate as self-reported chronic conditions, psychological well-being, self-reported functioning, and performance-based functioning variables are added in Models 5-8. Interviewer ratings are jointly associated with mortality in Models 1-5; respondent ratings are not.

With the addition of variables denoting psychological well-being, self-reported functioning, and, to a lesser extent, performance tests in Models 6-8, the predictive power of interviewer-rated health wanes and neither interviewer nor respondent ratings are jointly significant. Models incorporating continuous ratings (e Table 3, http://links.lww.com/EDE/A715) support this result: the linear relationship between interviewer ratings and mortality in the early models attenuates as these final covariates are added.

In Model 8, where all considered covariates are included, the hazard ratios associated with selfrated health are generally larger than those associated with interviewer-rated health. This suggests that, in the presence of extensive information on the physical and mental well-being of the respondent, self-assessments may capture some additional unobservable health information such that these assessments provide a modest improvement in mortality prediction — perhaps more so than interviewer assessments.

As a robustness check, we reestimated all of the hazard models in Tables 3 and 4 using Cox models, which make no assumptions about the baseline hazard function (as opposed to the Gompertz assumption of an exponentially increasing baseline hazard). The estimates resulting from the Cox models are very similar to those from the Gompertz models, and our substantive conclusions remain the same (available on request).

Discussion

Questions on self-rated health are routinely included in surveys and clinical studies. In contrast, few researchers have solicited global health ratings from observers that examine or interact with participants. To the best of our knowledge, none appears to have evaluated the utility of interviewer health ratings for mortality prediction. Our study's major strength is the richness of the Taiwan data, which include health ratings from physicians and interviewers based on standard self-rated health scales. The Taiwan survey also comprises a broad range of covariates, providing insight into the information assessors may use when forming ratings.

Two unexpected results emerge. One is that physicians' ratings are weak predictors of mortality: they are not associated with mortality even in a model with no additional covariates (Model 3, Table 3). This finding calls into question the implicit assumption made in several previous studies that "objective" health ratings provided by health personnel are superior to "subjective" self-ratings.[10,11,13,14] The physicians performed a medical exam equivalent to an annual physical exam offered through Taiwan's national health insurance plus an abdominal ultrasound, and have specialized knowledge regarding the presence, severity, and relative importance of health conditions. Given physicians' access to this information, we might have anticipated a stronger association between their assessments and patient survival. However, the physicians met respondents for the first time at these exams, perused a medical history filled out by respondents themselves, and did not have access to the results of blood and urine tests. A previous study examining determinants of these

physician ratings revealed that physicians weigh certain clinical factors more heavily than interviewers or respondents – most notably respondents' smoking status and medical abnormalities – and are less likely to incorporate physical functioning and psychological well-being (perhaps because they had fewer opportunities to glean this information[9]). As suggested by the results in Table 4, the lack of emphasis given to physical and psychological functioning likely weakened the predictive power of physicians' ratings.

The second unanticipated finding is that interviewers' ratings are considerably more powerful than self-ratings: Interviewer-rated health substantially enhances mortality prediction even in the presence of self-rated health and a large number of covariates. The attenuating HRs for interviewer assessments in Models 6-8 of Table 4 suggest that the strength of these ratings arises largely from interviewers' consideration of respondents' physical and mental health in their assessments. Analyses shown in eTable 4(http://links.lww.com/EDE/A715) indicate that interviewer ratings are superior to respondent ratings for predicting mortality for "early" deaths (Panel 1, http://links.lww.com/EDE/A715), but the two sets of ratings are similarly predictive for "late" deaths (Panel 2, http://links.lww.com/EDE/A715). This finding was ascertained by splitting the five-year follow-up period roughly in half and estimating hazard models separately for each of the two periods. Interviewers may rely more heavily than respondents on information or cues indicating very ill or frail respondents, which are likely strongly indicative of "early" deaths. This interpretation is consistent with the very high HRs associated with an interviewer rating of "poor" health.

The predictive strength of interviewer-rated health may be related in part to respondents' perceived age, a variable examined by Christensen and colleagues.[15] In their study, external assessors estimated respondents'age based on passport-style photographs. These perceived ages are only modestly correlated with actual age, but provide as strong predictions of seven-year mortality as actual age. These findings, replicated by Dykiert et al.,[24] suggest that our study's interviewers may draw not only upon information explicitly collected but also on observations throughout a lengthy interview (76 minutes on average) of facial and bodily features, responsiveness, mobility, and disposition. Moreover, given that interviewer-rated health is a stronger predictor than self-rated health in the presence of most of the control variables, it seems plausible that interviewer-rated health would be as strongly associated with five-year mortality as self-rated health in a more modest interview than the Social Environment and Biomarkers of Aging Study – e.g., a survey that excludes functioning tests and has fewer health questions. Interviewer-rated health may also have greater test-retest reliability than self-rated health, since interviewers are less likely than respondents to be influenced by day-to-day shifts in the respondent's health. It is also possible that interviewers rely on a more representative reference group—i.e., the study population—than respondents when making their assessments – but there is no evidence to support or refute these hypotheses.

This study underscores the utility of interviewer-rated health for improving mortality prediction. The findings from this analysis support inclusion of a simple question at the end of face-to-face household surveys asking interviewers to assess respondents' overall health. If other surveys replicate our findings, this question should become as ubiquitous as self-rated health. The costs of such an undertaking are minimal and the potential gains substantial for demographic and health researchers. Nevertheless, the study has some limitations, most notably modest statistical power and restriction to a single ethnically homogenous country. Although there is no reason to suspect that our results are limited to Taiwan – an industrialized nation with a life expectancy on par with wealthy Western societies – future work could ascertain whether the predictive strength of interviewer-rated health varies with cultural, ethnic, and socioeconomic factors, or with the detail of the

interview and the training of interviewers. In addition, qualitative studies could provide more detailed information regarding how respondents, interviewers, and physicians make their health assessments. Answers to these questions would help researchers understand how interviewers' assessments arise and also provide insights into whether additional or alternative survey questions could further enhance the utility of interviewer assessments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Benyamini Y, Leventhal EA, Leventhal H. Elderly people's ratings of the importance of health-related factors to their self-assessments of health. Social Science & Medicine. 2003; 56(8):1661–1667.10.1016/S0277-9536(02)00175-2 [PubMed: 12639583]

2. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality Prediction with a Single General Self-Rated Health Question. A Meta-Analysis. Journal of General Internal Medicine. 2006; 21(3): 267–275.10.1111/j.1525-1497.2005.00291.x [PubMed: 16336622]

3. Krause NM, Jay GM. What do global self-rated health items measure? Medical care. 1994; 32(9): 930–942. [PubMed: 8090045]

4. Jylhä M, Volpato S, Guralnik JM. Self-rated health showed a graded association with frequently used biomarkers in a large population sample. Journal of Clinical Epidemiology. 2006; 59(5):465–471.10.1016/j.jclinepi.2005.12.004 [PubMed: 16632134]

5. Finch BK, Hummer RA, Reindl M, Vega WA. Validity of Self-rated Health among Latino(a)s. Am J Epidemiol. 2002; 155(8):755–759.10.1093/aje/155.8.755 [PubMed: 11943694]

6. Bzostek S, Goldman N, Pebley A. Why do Hispanics in the USA report poor health? Social Science & Medicine. 2007; 65(5):990–1003. [PubMed: 17574713]

7. Schnittker J. When Mental Health Becomes Health: Age and the Shifting Meaning of Self-Evaluations of General Health. The Milbank Quarterly. 2005; 83(3):397–423.10.1111/j. 1468-0009.2005.00407.x [PubMed: 16201998]

8. Spencer SM, Schulz R, Rooks RN, et al. 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 and Social Sciences. 2009; 64B(1):87–94.10.1093/geronb/gbn007

9. Smith KV, Goldman N. Measuring Health Status: Self-, Interviewer, and Physician Reports of Overall Health. Journal of Aging and Health. 2011; 23(2):242–266.10.1177/0898264310383421 [PubMed: 21041293]

10. Suchman EA, Phillips BS, Streib GF. An Analysis of the Validity of Health Questionnaires. Social Forces. 1958; 36(3):223–232.

11. Friedsam HJ, Martin HW. A Comparison of Self and Physicians'Health Ratings in an Older Population. J Health Hum Behav. 1963; 4:179–183. [PubMed: 14066576]

12. Markides KS, Lee DJ, Ray LA, Black SA. Physicians'ratings of health in middle and old age: a cautionary note. J Gerontol. 1993; 48(1):S24–27. [PubMed: 8418152]

13. Valanis BG, Yeaworth R. Ratings of physical and mental health in the older bereaved. Res Nurs Health. 1982; 5(3):137–146. [PubMed: 6923442]

14. Larue A, Bank L, Jarvik U, Hetland M. Health in Old Age: How Do Physicians'Ratings and Self-ratings Compare? J Gerontol. 1979; 34(5):687–691.10.1093/geronj/34.5.687 [PubMed: 572838]

15. Christensen K, Thinggaard M, McGue M, et al. Perceived age as clinically useful biomarker of ageing: cohort study. BMJ. 2009; 339(dec11 2):b5262–b5262.10.1136/bmj.b5262 [PubMed: 20008378]

16. Chang, MC.; Glei, DA.; Goldman, N.; Weinstein, M. Biosocial Surveys: Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. The National Academies Press; 2007. The Taiwan Biomarker Project; p. 60-77.

17. Chang, MC.; Lin, HS.; Chuang, YL., et al. Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan, 2000 and 2006: main documentation for SEBAS longitudinal public use data. 2012. Available at: www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3792

18. Goldman N, Lin IF, Weinstein M, Lin YH. Evaluating the quality of self-reports of hypertension and diabetes. Journal of Clinical Epidemiology. 2003; 56(2):148–154.10.1016/S0895-4356(02)00580-2 [PubMed: 12654409]

19. Goldman N, Glei DA, Lin YH, Weinstein M. The serotonin transporter polymorphism (5-HTTLPR): allelic variation and links with depressive symptoms. Depression and Anxiety. 2010; 27(3):260–269.10.1002/da.20660 [PubMed: 20196101]

20. Dowd JB, Goldman N, Weinstein M. Sleep Duration, Sleep Quality, and Biomarkers of Inflammation in a Taiwanese Population. Annals of Epidemiology. 2011; 21(11):799–806.10.1016/j.annepidem.2011.07.004 [PubMed: 21889359]

21. Seplaki CL, Goldman N, Weinstein M, Lin YH. Measurement of Cumulative Physiological Dysregulation in an Older Population. Demography. 2006; 43(1):165–183. [PubMed: 16579213]

22. Cornman JC, Glei D, Rodriguez G, Goldman N, Hurng BS, Weinstein M. Demographic and Socioeconomic Status Differences in Perceptions of Difficulty With Mobility in Late Life. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2010; 66B(2):237–248.10.1093/geronb/gbq087

23. Horiuchi S, Coale AJ. A Simple Equation for Estimating the Expectation of Life at Old Ages. Population Studies. 1982; 36(2):317–326.10.2307/2174203 [PubMed: 22077276]

24. Dykiert D, Bates TC, Gow AJ, Penke L, Starr JM, Deary IJ. Predicting Mortality From Human Faces. Psychosom Med. 2012

Table 1

Descriptive Statistics, the Social Environment and Biomarkers of Aging Study, 2006.

	Interview Sample n=1,197		Exam Sample n=919		Included in Interview Sample, Excluded from Exam Sample n=278	
Health ratings and mortality						
Died by 6/30/2011; %	11		10		13	
Self-rated health (scale 1-5); mean (SD)	3.2	(1.0)	3.1	(1.0)	3.3	(1.1)
Interviewer-rated health (scale 1-5); mean (SD)	3.8	(1.0)	3.8	(0.9)	3.9	(1.1)
Physician-rated health (scale 1-5); mean (SD)	--		3.3	(0.8)	--	
Sociodemographic covariates						
Women; %	47		46		51	
Age (years); mean (SD)	65.7	(10.2)	66.1	(9.9)	64.2	(10.8)
Urban (versus rural); %	59		59	--	58	
Completed education (years); mean (SD)	6.9	(4.8)	7.0	(4.9)	6.9	(4.7)
Number of social activities; mean (SD)	0.8	(1.1)	0.8	(1.1)	0.7	(1.1)
Married or has companion; %	76		76		74	
Smokes daily; %	17		16		19	
Self-reported chronic conditions						
High blood pressure; %	34		34		35	
Take meds for high blood pressure; %	32		32		33	
Diabetes; %	17		18		15	
Heart disease; %	17		17		15	
Cancer; %	2		1		3	
Respiratory disease; %	6		7		6	
Ulcer; %	13		14		8	
Liver disease; %	7		8		4	
Kidney disease; %	5		5		6	
Gout; %	8		8		7	
Psychological well-being						
Center for Epidemiologic Studies Depression scale (scale 0-10); mean (SD)	4.8	(5.7)	4.8	(5.6)	4.9	(6.2)
Perceived stress index (scale 0-12); mean (SD)	0.2	(0.3)	0.2	(0.3)	0.2	(0.3)
Poor sleep quality index (scale 0-15); mean (SD)	4.0	(3.3)	4.1	(3.4)	3.8	(3.1)

	Interview Sample n=1,197		Exam Sample n=919		Included in Interview Sample, Excluded from Exam Sample n=278	
Self-reported functioning						
Number of difficulties with activities of daily living (scale 0-6); mean (SD)	0.2	(1.0)	0.2	(0.9)	0.3	(1.2)
Number of mobility limitations (scale 0-9); mean (SD)	1.8	(2.4)	1.8	(2.4)	1.7	(2.6)
Performance-based functioning						
Unable to perform grip strength; %	3		3		4	
Grip strength (mean kilograms); mean (SD)	27.1	(11.5)	27.2	(11.2)	26.5	(12.2)
Unable to perform peak flow; %	4		3		8	
Peak flow (mean L/min); mean (SD)	323	(150)	328	(149)	305	(153)
Unable to perform walk speed; %	5		3		9	
Walk speed (mean m/sec); mean (SD)	0.8	(0.3)	0.8	(0.3)	0.8	(0.4)
Unable to perform chair stand; %	10		8		18	
Chair stand (mean stands/sec); mean (SD)	0.5	(0.2)	0.5	(0.2)	0.4	(0.3)

SD indicates standard deviation.

Table 2

Frequency of Health Ratings Reported by Respondents, Interviewers, and Physicians in the Social Environment and Biomarkers of Aging Study, 2006.

	Respondents alive as of 30 June 2011			Respondents who died by 30 June 2011		
	Self %	Interviewer %	Physician %	Self %	Interviewer %	Physician %
Excellent (5)	13	29	4	3	8	2
Good (4)	22	43	40	13	30	31
Average (3)	43	19	45	33	33	42
Not so Good (2)	18	8	10	36	21	24
Poor (1)	3	1	0	14	8	1
Mean Rating	3.2	3.9	3.4	2.6	3.1	3.1
	(n=1,071)[a]	(n=1,071)[a]	(n=830)[a]	(n=126)[a]	(n=126)[a]	(n=89)[a]

[a] The number of observations with a valid value for the relevant health assessment.

Table 3

Hazard Ratios from Proportional Hazard Models of Dying by June 2011, the Social Environment and Biomarkers of Aging Study, 2006. (n=919, the exam sample; see text for details)

Ratings	Model 1		Model 2		Model 3		Model 4	
	HR	(95% CI)	HR	(95% CI)	(HR	95% CI)	HR	(95% CI)
Self								
Poor (1)	6.00	(1.59 – 22.60)					2.33	(0.52 – 10.46)
Not so good (2)	4.37	(1.34 – 14.23)					2.26	(0.61 – 8.35)
Average (3)	2.33	(0.71 – 7.65)					1.46	(0.41 – 5.25)
Good (4)	2.76	(0.80 – 9.57)					1.89	(0.52 – 6.88)
Excellent (5)[a]	1.00						1.00	
Interviewer								
Poor (1)			11.16	(3.70 – 33.65)			7.06	(2.03 – 24.63)
Not so good (2)			3.58	(1.45 – 8.85)			2.53	(0.91 – 7.04)
Average (3)			3.45	(1.51 – 7.89)			2.96	(1.20 – 7.28)
Good (4)			2.03	(0.89 – 4.61)			1.85	(0.78 – 4.39)
Excellent (5)[a]			1.00				1.00	
Physician								
Poor (1)					1.53	(0.14 – 17.08)	0.83	(0.07 – 10.03)
Not so good (2)					1.39	(0.32 – 6.03)	0.86	(0.19 – 3.84)
Average (3)					0.82	(0.19 – 3.42)	0.59	(0.14 – 2.56)
Good (4)					1.01	(0.24 – 4.27)	0.81	(0.19 – 3.48)
Excellent (5)[a]					1.00		1.00	
Joint test of ratings[b]								
Self	p=0.005						p=0.465	
Interviewer			p=<0.001				p=0.018	
Physician					p=0.431		p=0.641	

[a] Reference category.

[b] The p-value is from a joint Wald test of the four coefficients for the indicated set of ratings.

Table 4

From Proportional Hazard Models of Dying by June 2011, the Social Environment and Biomarkers of Aging Study, 2006.

erviewer sample; see text for details)

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)
	3.54	(1.03 – 12.14)	3.60	(1.05 – 12.33)	3.64	(1.06 – 12.48)	3.65	(1.07 – 12.51)	3.12	(0.89 – 10.95)	3.23	(0.90 – 11.59)	3.32	(0.93 – 11.86)	3.83	(1.06 – 13.85)
	2.68	(0.89 – 8.10)	2.69	(0.89 – 8.12)	2.70	(0.89 – 8.16)	2.73	(0.90 – 8.25)	2.61	(0.85 – 8.05)	2.57	(0.82 – 8.04)	2.59	(0.83 – 8.07)	2.77	(0.89 – 8.57)
	1.66	(0.56 – 4.90)	1.67	(0.57 – 4.93)	1.70	(0.58 – 5.02)	1.71	(0.58 – 5.05)	1.59	(0.53, 4.73)	1.63	(0.54 – 4.89)	1.64	(0.54 – 4.92)	1.72	(0.58 – 5.15)
	1.89	(0.62 – 5.80)	1.91	(0.62 – 5.87)	1.96	(0.64 – 6.04)	1.98	(0.64 – 6.10)	2.05	(0.66 – 6.34)	2.13	(0.69 – 6.61)	2.20	(0.71 – 6.83)	2.40	(0.77 – 7.47)
	1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
	5.98	(2.08 – 17.21)	6.89	(2.39 – 19.90)	5.95	(2.05 – 17.29)	5.99	(2.06 – 17.38)	4.69	(1.54 – 14.31)	3.46	(1.06 – 11.33)	2.38	(0.68 – 8.33)	2.22	(0.61 – 8.13)
	2.49	(1.08 – 5.75)	2.77	(1.19 – 6.46)	2.46	(1.05 – 5.79)	2.47	(1.05 – 5.79)	2.26	(0.95 – 5.37)	2.11	(0.87 – 5.11)	1.60	(0.63 – 4.05)	1.37	(0.54 – 3.47)
	2.84	(1.34 – 6.02)	3.10	(1.46 – 6.62)	2.91	(1.36 – 6.22)	2.92	(1.37 – 6.25)	2.65	(1.22 – 5.74)	2.62	(1.20 – 5.74)	2.40	(1.09 – 5.32)	2.12	(0.96 – 4.67)
	1.67	(0.81 – 3.46)	1.79	(0.86 – 3.71)	1.69	(0.81 – 3.52)	1.69	(0.81 – 3.52)	1.57	(0.75 – 3.28)	1.53	(0.73 – 3.22)	1.49	(0.70 – 3.14)	1.30	(0.62 – 2.76)
	1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
			×		×		×		×		×		×		×	
					×		×		×		×		×		×	
...nditions[c]							×		×		×		×		×	
...g[d]									×		×		×		×	
...g[e]											×		×		×	
...tioning[f]															×	
	p=0.087		p=0.084		p=0.099		p=0.097		p=0.137		p=0.164		p=0.145		p=0.090	
	p=0.005		p=0.002		p=0.005		p=0.005		p=0.024		p=0.069		p=0.119		p=0.139	

[b] Social environment variables: urban/rural residence, educational attainment, participation in social activities, and marital status.

[c] Self-reported chronic conditions: high blood pressure, takes medicine or has prescription for high blood pressure, diabetes, heart disease, cancer, respiratory disease, ulcer, liver disease, kidney disease, gout.

[d] Psychological well-being: Center for Epidemiologic Studies Depression score, perceived stress, sleep quality.

[e] Self-reported functioning: number of activities of daily living difficulties, number of mobility difficulties.

[f] Performance-based functioning: grip strength, peak lung flow, walk speed, chair stand speed, and indicators for whether the respondent was unable to perform each of these four tasks. Grip strength (in kg) is measured with a dynamometer as the maximum value of three trials per hand; we consider the maximum of both hands. Walking speed (in m/sec) is measured as the faster of two trials for a walk of 3m at normal speed (with a walking aid if needed. Chair stand speed (chair stands/sec) is the completion time for five chair stands.

[g] The p-values for the set of self or interviewer ratings are from a joint Wald test of the four coefficients for the indicated set of ratings.