

Measuring health status: An examination of self-, interviewer-, and physician-assessments of overall health

PAA Abstract

Kimberly V. Smith, Princeton University

September 2008

Self-assessed health status (SAH) is one of the most frequently used summary measures of health in empirical research. The measure is based on a survey question that asks respondents to rate their overall health on a four- or five-point scale that typically runs from excellent to poor. The resulting ordinal variable has shown to be a good predictor of a range of health outcomes, including morbidity, use of health services, and mortality (Idler & Benyamini, 1997; Benyamini & Idler, 1999; Idler & Kasl, 1995). However, while the association between SAH and other health outcomes holds within populations, the comparability of health ratings across groups of individuals has come into question.

It is now generally recognized that a given level of health may result in different SAH ratings depending on individuals' norms and expectations. For example, studies have found that among adults with a similar level of chronic disease or functional limitations, older people tend to rate their health as better than comparable younger people (Ferraro, 1980). This has been attributed to lower self-expectations for health and physical functioning at older ages. In contrast, Hispanics, despite having lower levels of morbidity and mortality, systematically report worse overall health than non-Hispanics; this finding has not been explained by differences in socioeconomic, demographic, and health-related variables (Menec, Shooshtari, & Lambert, 2007). Studies comparing SAH across ethnic and racial groups, as well as cross-national studies, suggest that, in addition to differences in referents used, there may be cultural or language differences in the interpretation of ratings categories and in reporting styles, such as tendencies to use/avoid extreme values in a rating scale (Sen, 2002, Jurges, 2007; Bzostek, Goldman, & Pebley,

2007). Another possible explanation is that what is factored into SAH ratings varies across individuals (Krause & Jay, 1994).

These explanations have been discussed in several recent studies that also reveal differences in the validity (predictive power) of SAH by respondent characteristics, such as education, ethnicity, race, and sex (Dowd & Zajacova, 2007; Melzer et al, 2004; Ferraro et al., 2001; Finch et al, 2002; Benyamini et al., 2003; Lyra et al., 2008). However, the self-evaluation process used to arrive at health ratings and the associated biases in the SAH measure are still not well-understood. Given the widespread use of SAH in empirical research, it is important to gain a better understanding of what factors play a role in self-evaluations and to develop ways of adjusting for possible biases arising from the evaluation process. These are two very active areas of research upon which this paper builds.

The former area of research, which focuses on the determinants of SAH, suggests that SAH ratings are based on evaluations of multiple aspects of health status, including physical and mental health and utilization of health services, as well as non-health factors, such as socioeconomic status, family characteristics, and social networks (e.g., Goldman et al., 2003). However, even after controlling for a host of health indicators and other explanatory variables, studies still find that SAH remains an independent predictor of mortality. A reason for this finding may be that most studies rely on self-reported health measures, which are imperfect indicators of health and depend on factors such as access to medical care, ability to understand and desire to share health information, and recall accuracy (Goldman et al., 2003). Recent studies have found a significant association between various biomarkers known to be related to health conditions and SAH, providing further evidence of a biological basis of SAH (Goldman, Gleib, & Chang, 2003). The addition of biomarkers, however, does not lead to a full account of the relationship between SAH and mortality (Jylha, Volpato, & Guralnik, 2006; Idler, Russel, & Davis, 2000).

The second area of research has focused on testing and developing adjustments for measurement error in SAH (Crossley & Kennedy, 2002; Baker, Stabile, & Deri, 2004). Studies have tried to determine what portion of SAH differentials is due to true differences in health status and what portion is due to differences in health rating behavior. Several methods have been used to this end. One method uses vignettes, where respondents rate their own health and the health of fictitious individuals. The use of vignettes in surveys, however, is time-consuming and may not be feasible in all cases; moreover, few existing surveys contain responses to vignettes. Another method is to compare SAH to a set of less subjective health measures, such as those obtained from medical records (Lindeboom & van Doorslaer, 2004; Baker, Stabile, & Deri, 2004). Due to the limited availability of linked survey and medical record data, the use of less subjective measures of health status is often based, at least in part, on self-reports of selected health conditions and is, therefore, subject to self-reporting biases (Lindeboom & van Doorslaer, 2004; Jurges, 2007).

This paper extends this earlier work to better understand SAH and improve its accuracy in measuring ‘true’ health. We examine SAH and two alternative measures of respondents’ health in survey data, interviewer- and physician-assessments of overall health status (IAH and PAH, respectively). We use 2006 data from a nationally representative survey of older adults in Taiwan, which includes extensive self-reported, clinical and biomarker data. The interviewers administering the home-based survey questionnaire and physicians conducting the respondent’s physical exam were each asked to rate the respondent’s health using the same excellent (1) to poor (5) scale used by the respondent to rate his/her own health. Theoretically, such alternative measures may be useful in reducing measurement error in SAH ratings, or in deriving a more comprehensive assessment of a respondent’s health. Interviewers’ and physicians’ ratings are likely not to suffer from the same biases as SAH, such as changes in self-expectations with age or adaption to health conditions. Physicians and interviewers, however, may use different reference levels or reporting styles than respondents, but depending on the direction and magnitude of the

bias, such measures may have the potential to be combined with, or used as instruments for, SAH to obtain a more reliable health measure.

We first investigate what factors are important in self- and external evaluations of health. We find that several key indicators of physical health are important in health ratings across all three types of evaluators, including chronic illness, mobility limitations, blood pressure, hospital days, level of depression, and sleep quality. However, our analysis also reveals notable differences. Age, SES, and levels of stress and pain are significantly associated with health in SAH ratings only. In addition, we find significant coefficients for many of the biomarkers in the PAH model only. We hypothesize that these biomarkers are picking up unmeasured aspects of health that were observed by the physician during the respondent's physical examination and/or the severity of an illness or condition that the physician was better able to determine than the respondent or interviewer.

Next, we examine the level of agreement across health ratings and the extent to which respondent characteristics affect agreement in health ratings by different sources. Our initial finding is that perfect agreement occurs in only 33% - 37% of inter-evaluator comparisons. We also find statistically significant differences in the means and distributions of health ratings across evaluators. On average, respondents tend to rate their health lower than interviewers and physicians. Our analysis further reveals differences in the cut-points used by the different evaluator types, and that certain factors, such as age, SES, mobility limitations, and aspects of psychological well-being, predict the extent and direction of disagreement across evaluators.