

## **A broad sense measure of health**

### **A broad sense measure of health and its properties in the HRS**

Phillip Cantu 1\*  
Connor M. Sheehan 1  
Benjamin W. Domingue 2\*

1. Population Research Center and Department of Sociology, University of Texas at Austin
2. Stanford Graduate School of Education

#### **\*Corresponding authors:**

Phillip Cantu, Department of Sociology and Population Research Center, University of Texas at Austin. Population Research Center, G1800, University of Texas at Austin, Austin, TX 78712-0544; E-mail: CantuPhillip@utexas.edu ; Telephone (512) 471-5514; Fax: (512) 471-4886

Ben Domingue, Graduate School of Education, 508 CERAS Stanford CA 94304, bdomingue@stanford.edu

*Acknowledgements:* We thank the University of Texas Population Research Center (Grant R24 HD42849) for administrative and computing support; the NICHD Ruth L. Kirschstein National Research Service Award (T32 HD007081-35) for training support; the University of Michigan and Rand Corporation for making the data available to the public, and the members of the Population Health Lab for their helpful suggestions. The contents of this manuscript are solely the responsibility of the authors and do not represent the official views of NICHD, Rand Corporation, Stanford, or the University of Texas at Austin.

## **A broad sense measure of health**

### **Abstract**

Measuring health is a crucial component of much social research. Two approaches are typical. Health may be measured via either narrowly targeted questions about aspects of disability or chronic conditions. Alternatively, individuals may be asked to self-report their health in some global sense. Both approaches have potential drawbacks. We consider a broad sense measure of health constructed by items from five different batteries related to physical and mental wellbeing. We demonstrate that this measure predicts time until death better than self-reported health, especially for females. Although this measure has promise, we argue that future surveys on health would benefit from the inclusion of additional items focusing on issues salient to younger individuals or other non-disabled respondents.

## **A broad sense measure of health**

### **1. Introduction**

Understanding health and its age-associated dynamics is a key objective of the Health and Retirement Study (HRS) and related studies (e.g., ELSA, MHAS, etc.). Crucial to this effort is a critical investigation of what researchers mean by “health” and how health is measured. Previous research has tended to either (a) look at specific, narrow domains of health (i.e. activities of daily living) or (b) ask respondents to self-report their overall health status. The first approach is common for the study of disability. For example, the battery of questions which comprise the Activities of Daily Living ask if a respondent has the ability to conduct basic daily tasks of functioning (1). While these measures are effective in measuring inability to perform tasks, they cover relatively restrictive domains and presumably do not provide information about relative differences between individuals who are not disabled and who may still have different amounts of “health.” There are likely substantial differences in overall health between respondents with the exact same reports or levels of ADLs, but this heterogeneity is not captured in research focusing on disability. The second approach, asking respondents to self-report their general or overall health, has been used in a vast amount of research (e.g., 2,3). Research has found that self-reported health corresponds with physician evaluations and accurately predicts mortality (4). This measure takes a more sweeping view of “health” than, say, the Activities of Daily Living battery, but this comes at the potential expense of introducing subjectivity into the measurement process. Individuals may respond differently to this question partially as a function of their age (5), sex, and ethnicity (6,7) which is undesirable.

While both of these approaches have been fruitful and will continue to be used to operationalize health, we ask whether a broader, yet still objective, consideration of what constitutes “health” can yield new insight. Our approach is atheoretical in the sense that we are

## **A broad sense measure of health**

endeavoring to make better use of valuable extant data. Our lack of a strict theoretical foundation for what is meant by “health” is similar to the manner in which self-reports of “health” are based on the respondents’ subjective ideas about what this might mean (e.g., does the respondent consider their mental health when they are asked generically about “health”). We comment on this purely utilitarian viewpoint in Section 4 after first demonstrating that it offers potential utility. We use Item Response Theory (IRT; 8) to combine multiple batteries of questions related to functioning, chronic conditions, and psychological health to form a set of health-related items that we use to construct a broad sense measure of health on a continuous scale. We focus on two questions. *First*, what does the IRT analysis tell us about the ways in which our inquiries regarding health could be improved? *Second*, how does our broad sense measure of health potentially improve upon specific batteries and the standard self-reported health measure? Given the graying of the US population (9), these are timely questions.

### **1A. Current operational definitions of “health”?**

In much demographic inquiry, researchers have attempted to measure health as a broad construct regarding overall wellbeing. Studies examining the relationship between educational attainment and “health” are frequently utilize such an indicator (10). Researchers use self-reported health SRH as a way to quickly measure global general health due to its simplicity, convenience, and also because it has a well-documented history as a strong predictor of mortality and aligns well with doctor assessments of health (4,11,12). There are open questions about the validity of SRH, however. For example, there may be period effects related to how people think about health (13). Other research illustrates difficulty in making cross-ethnic comparisons, especially with relation to differences in linguistic backgrounds (6,7). The

## **A broad sense measure of health**

subjective nature of SRH is thus a cause for concern in certain types of inquiry (e.g., inquiry based on diverse populations).

Other types of research have focused on health as measured by different batteries of items inquiring about narrowly defined aspects of respondent wellbeing and functioning and have provided valuable information to researchers about the nature of physical decline (14–22). Broadly speaking, this research conceptualizes health as living without disability. The disablement process (1) offers a framing through which researchers can understand the multiple dimensions of health and wellbeing; specifically the disablement process shows that disability as measured by survey items are the downstream result of pathological conditions upstream. The disablement process allows researchers to conceptualize health in an objective manner (i.e. an individual responds that he or she is cannot walk across a room). Such a focus on disability is less appropriate for use in broader populations that contain younger respondents or healthier (non-disabled) older respondents.

### **1B. The Broad Sense Health approach**

Given the potential subjectivity of SRH and the narrowness of other measures based on functional limitations and behavioral modifications, we think there is room for alternative operational definitions of health. Survey questions are a useful tool in health-related research, in part due to the fact that they are relatively cheap to administer (c.f., clinical assessments) and are included in many studies of human behavior. Survey questions on health can cover a broad variety of ground, from function (23) to diagnosed chronic conditions (24) and mental health (25). We combine information from across these survey items into a single IRT-based composite: broad sense health (BSH). Item response theory was originally developed for educational measurement, but these methods can also be used to study health and functioning.

## **A broad sense measure of health**

Previous work has utilized these techniques to study differences in response processes as a function of ethnicity on indicators of disability (26,27). IRT is a useful tool in this context in that it allows for both a holistic summary of health *and* a critical appraisal of measurement quality. We examine both issues in this paper. In a time with increasing non-response (28) and decreasing funds for surveys, knowing which questions can quickly differentiate the “healthy” from “unhealthy” is critically important.

## **2. Data & Methods**

### **2A. Sample**

We utilize the Health and Retirement Survey (24,29). The HRS is a nationally representative of the non-institutionalized US population longitudinal survey of older Americans. We use data from wave 4, collected in 1998, for our analysis because in 1998 two new cohorts were added which made the HRS nationally representative for those aged 50+ for the first time. The sample contained 21,384 respondents, all those respondents with at least some available information in 1998 on the health indicators. The respondents had a mean birth year of 1932 (SD=11.2; average age of 66 as of 1998). The majority of the sample (58.1%) was female. The sample was 76.3% non-Hispanic white, 13.9% non-Hispanic black, and 6.1% Hispanic. We compare our BSH measure to SRH which is measured in HRS as a 5-category scale (1-excellent, 5-poor). In our analytic sample 12.3% of respondents self-reported excellent health, 25.6% very good health, 30.6% good health, 20.6% fair health, and 10.9% poor health.

### **2B. Indicators**

We use health indicators from the following batteries: Activities of Daily Living, Instrumental Activities of Daily Living, Functional Limitations, Depression, and Chronic

## A broad sense measure of health

Conditions. Additional information on the wording of individual items is contained in the Supplemental Information (SI). Items are coded so that “0” indicated that the respondent reported having a condition or limitation measured (“1” indicates better health in all cases).<sup>1</sup> Information on the content of the utilized items can be found in the SI. Note that we focus on items related to both physical and mental health. We argue that SRH is also likely to reflect the influence of aspects of both physical and mental health. Here, we have the capacity to inquire about the relationship between different indicators (e.g., physical functioning and mental health) and the respondent’s overall health as measured by BSH.

Descriptive statistics for the indicators can be found in the first two columns of Table 1. For the average indicator, 83% of individuals did not have the health condition, physical limitation, depressive symptom, etc. of interest. Some items captured almost no variation in respondent health or behavior. Even the most prevalent of conditions/limitations was not prevalent in more than half of our sample (minimum prevalence of 0.52). This variation in prevalence is also shown in Figure S1A. Note that missingness was heavily concentrated in certain batteries (IADLs and CESD).

## 2C. Methods

We estimate two parameter logistic (2PL) models. The 2PL model assumes that an individual’s response to an item  $i$ ,  $Y_i$ , is a function of the individual’s health,  $\theta$ , and two item parameters,  $a_i$  and  $b_i$ :

$$P(Y_i = 1|\theta) = \frac{\exp(a_i\theta + b_i)}{1 + \exp(a_i\theta + b_i)}. \text{ (Eqn 1)}$$

The probability of responding affirmatively (i.e., not having the limitation or health impediment of interest) is determined by the respondent’s level of “health” (represented as  $\theta$ ) and the items

---

<sup>1</sup> We reverse-coded two CESD items (whether the respondent reported being happy or enjoying life).

## A broad sense measure of health

parameters  $b_i$  (the item easiness) and  $a_i$  (the discrimination). Note that if the respondent's latent health ( $\theta$ ) was observed, this would reduce to a standard logistic regression problem.<sup>2</sup> For identification purposes, we assume that health is distributed as Normal[0,1]. Estimates were obtained via the mirt package in R (30). We estimate our models using multiple group maximum-likelihood estimation using an expectation-maximization algorithm and ability estimates expected a-posteriori factor score estimation.

We compare the predictive utility of SRH and BSH using a number of descriptive and non-parametric models. We also consider a series of models related to the prediction of time until death. We consider two baseline models:

$$\text{Time}_i = b_0 + b_1 * \text{SRH}_i + b_2 * \text{birthdate}_i + e_i \text{ (Eqn 2)}$$

$$\text{Time}_i = b_0 + b_1 * \text{BSH}_i + b_2 * \text{birthdate}_i + e_i \text{ (Eqn 3)}$$

where our general measures of health (BSH in Eqn 2; SRH in Eqn 3) are used to predict time until mortality for those who have recorded deaths as of the wave 11 of HRS data collection. We then consider a model that includes both health predictors:

$$\text{Time}_i = b_0 + b_1 * \text{SRH}_i + b_2 * \text{BSH}_i + b_3 * \text{birthdate}_i + e_i \text{ (Eqn 4)}$$

Interest is in the attenuation of the parameters in Eqn4 relative to their baseline values in Eqns 2-3.

## 3. Results

### 3A. The relationship between BSH and the different indicators

Figure 1 focuses on the item characteristic curves. These curves describe the probability of not having some form of ill health (a physical limitation, a chronic condition, or a mental

---

<sup>2</sup> We write Eqn 1 in a slightly non-standard way to emphasize this similarity to logistic regression. Most IRT texts would consider  $a_i(\theta + b_i)$ .



## **A broad sense measure of health**

health problem) for a given value of BSH as given by Eqn 1. [Further information on the item parameter estimates is also shown in final two columns of Table 1; individual item characteristic curves are shown in the SI]. We discuss these curves by item battery. ADL items and IADL items tended to be the least prevalent (easiest items). This means that the respondent had to be quite low on the BSH scale before they were unable to perform the ADL tasks. IADL items were fairly similar (i.e., there is substantial within-battery in item functioning for these two batteries).

The functional limitations discriminate over a broad span of health. The items range in easiness from walking one block, on par with the IADL items and something that only those with the least health could not do, to climbing a set of stairs, which was one of the most prevalent indicators of less than optimal health. Taken together, these two items exemplify the broad range of health impediments addressed by this battery. Identifying someone who has a functional limitation as “unhealthy” can mask very large differences between someone who cannot walk up a flight of stairs and someone who cannot walk 1 block.

The CESD items are the least discriminating battery of items, presumably due to the fact that most of the others cover physical functioning and other embodied conditions. Finally, chronic diseases such as cancer and high blood pressure do not do a good job differentiating between the healthy and unhealthy. This is perhaps intuitive, in a sense, since the onset of some disease (e.g., some cancers) are essentially random events (or at least driven by impersonal environmental forces, such as pollutants, rather than human behavior (31)).

Turning to the behavior of the entire set of indicators, we make the following observation. As previously mentioned, there are no items with prevalences less than 0.5. Figure S1B gives an alternative representation. Here, we examine the collective focus of the set of indicators via the test characteristic curve, or the expected number of affirmative items for a

## **A broad sense measure of health**

given health. We effectively have no indicators that allow us to discriminate amongst the “health” of roughly half of our respondents. Once someone is in the top half of the distribution of health, we are not administering any items inquiring about mild forms of poor health that these respondents are likely to experience. This is a substantial limitation. In the jargon of IRT, this set of indicators has a substantial “ceiling effect” (32).

### **3B. BSH as an improvement on SRH**

This ceiling effect is again apparent when we examine individual estimates of health. Figure 2 considers some characteristics of the SRH and (standardized to have  $M=0$ ,  $SD=1$ ) BSH measures. Figure 2A shows estimated densities of BSH split by sex in the HRS sample. As in earlier research (33), females have more observable health problems leading to lower BSH. Note that the distributions both show substantial skew due to the “ceiling”. Figure 2B begins to compare the BSH and SRH measures via a boxplot of BSH values for each value of SRH at wave 4. While there is clearly an association between the two measures of health, there is also substantial variation in BSH for each value of SRH. Figure 2C shows the proportion of respondents reporting in each category of SRH (1-excellent through 5-poor) when split by sex. Unlike our BSH measure, there is no sex gradient in the reporting of SRH, a potential indicator of variation in the quality (or at least comparability) of the SRH measure across sex.

As humans have known for millennia, increases in age are associated with declines in health and we now probe the association of our BSH measure with the aging process (Figure 3). Panels 3A and 3B focus on health as a function of age for those who have died as of the 2012 (or Wave 11) of HRS data collection and those who have not. As expected, growing old is associated with declines in the BSH. However, note that the deceased have relatively flat trajectories up until approximately age 70 and then show fairly dramatic decreases. Those who

## A broad sense measure of health

have not died as of the latest wave of data collection tend to be, not surprisingly, much younger, and with better health. Panel 3C focuses explicitly on those who have died (N=8,318) and examines BSH as a function of time before death (defined as time from date of Wave 4 (1998) interview to date of death). As individuals near death, we see steady declines in BSH.

We ask now about the association with time until death (the x-axis of Figure 3C) and our two measures of health (BSH and SRH). The distributions of times until death are shown in Figure 4A. Amongst the deceased in our sample, females tend to have slightly less remaining lifespans than do males. We first consider models where BSH *or* SRH is included as a predictor of time until mortality net of date of birth (Eqns 2-3). These are the darker bars in Figure 4B. At baseline, a one SD increase in BSH predicts a 1.2 year increase in time until death. Compared to the base case of poor health, all of the estimates categories of SRH predict increases in time until death of 1.5, 2.5, 3.0, and 3.1 years (respectively for fair, good, very good, and excellent). Table 2 contains  $r^2$  statistics for these models and suggests that BSH is a slightly superior predictor of time until mortality compared to SRH at baseline.

We then consider a combined model with both predictors. There is a decrease in BSH, but it is still a strong predictor net of date of birth and self-reported health. Even net of one's SRH, a one SD increase in the BSH is associated with an additional 8/10 of a year of life. There is, however, a sex gradient to these findings that is explored in panels 4C and 4D. In models without SRH, BSH predicts time until mortality fairly consistently across sex. A 1-SD increase in the measure is associated with an extra 1.2 years of life in males and 1.3 years in females. However, net of SRH, a one SD increase only predicts an extra 0.8 years for males but over a year of extra life for females. Moreover, the effects of the various SRH categories for females are substantially reduced in the combined model. Additionally, the various category effects are

## **A broad sense measure of health**

similar relative to the effect of poor health (there is little difference between being in fair health versus excellent health). This can be further observed in Table 2 which compares the  $r^2$  statistics for the models in Figure 4. Note that there is little benefit in terms of  $r^2$  increase of including SRH for females once BSH is included at baseline.

## **4. Discussion**

Using the data from the Health and Retirement Study, we utilize Item Response Theory to combine multiple batteries of health-related indicators (ADLs, IADLs, functional limitations, CESD, and Chronic Conditions) into a single measure of “health” which we describe as broad-sense health (BSH). Looking at the battery-level averages in Table 1, our BSH measure is driven largely by the behavioral modifications (ADLs, IADLs, functional limitations) reported by respondents. Not surprisingly, the depressive symptoms and chronic conditions tend to be both less prevalent and less associated with the overall total. One major finding here is that current questions are quite limited in the sense that they are overwhelmingly measure poor health.

For general populations, researchers may want to consider including items that have higher prevalence (e.g., the existence of back pain; additional items on sleep quality; ability to engage in physical activities that are more strenuous than those in the functional limitations indicators). Developing such indicators is crucial given the simultaneous rise in the number of elderly people who are in relatively good health (34) and the interest in measuring the aging process at in younger populations such as adolescents (35,36). While research focusing on biomarkers will be of keen interest (35,37), survey-based research will continue to be of interest given its cost effectiveness. Although development of such indicators may not be easy, this difficulty is due in part to gaps in current knowledge about the aging process (i.e., what early warning signs at midlife are associated with greater declines in health later in the lifecourse?).

## **A broad sense measure of health**

Thus, the development and validation of such items may not only offer additional information about the overall health of peoples, but will also potentially help us better understand the nascent health problems which will one day become the types of functional limitations which are the focus of the indicator batteries in HRS.

Despite the limited focus on poor health of the batteries that go into our BSH measure, the proposed measure yielded increased predictive power of time until death relative to SRH. It also reflected a sex-specific story to health declines that earlier research has uncovered. relative improved measurements of global health relative to SRH in some cases. Indeed, we found our measure, Broad Sense Health (BSH) to be a stronger predictor of time until mortality than Self-Rated Health (SRH), especially for females. The results of Figure 3C, suggesting that women are in worse health prior to death (despite the fact that they tend to live longer), has been observed elsewhere (38). The fact that the BSH seems to more reliably uncover sex differences in health and aging vis-à-vis SRH is an important fact.

Research has typically regarded the disablement process as a progression between multiple stages from chronic condition, to symptom, to limitation, to disability and ultimately death (1). Research on the disability has reinforced this viewpoint by showing a near perfect overlap of IADL disability with ADL disability but not the other way around. However, this hierarchy could potentially be scaled via IRT such that we have more information about the “amount” of disablement that must occur to move from an IADL type disability to an ADL type disability. For example, the indicators with the highest prevalence were eating, bathing, and walking across a room (very few HRS respondents could not perform these tasks relative to other indicators of poor health in the HRS). All three of these are ADL items and would, from our theoretical model, imply the most severe disabling limitations. The fourth least easy (most

## **A broad sense measure of health**

difficult) item in the HRS is ability to prepare a meal, an IADL measurement, which is followed by ability to use a toilet, an ADL measurement. This finding is very surprising and calls into question the precision of ADL and IADL items as currently defined. It is not hard to imagine a world where technological advancement has changed the context of IADL and ADL disability. That is to say the contextual meaning of ADL and IADL measurement were so different in 1998 from the ideal type they were created under that ability to prepare a meal may be more closely reflective of ADL disability than IADL disability. What is certain is that IRT allows us to better understand the severity of different survey disability items. Furthermore, difficulty parameters from functional limitation items reveal a great deal of heterogeneity in what we learn about health from a battery of items that are often combined together. Our results gives estimates of the difference in information about health by items in each battery and suggests that some batteries represent a more unified conceptualization of health while others explore wider ranges of health.

### **4A. Future Research**

A crucial methodological limitation is that we are applying unidimensional IRT models to a construct, health, which is presumably multidimensional. Multidimensional IRT models exist (39) and their application may lead to refined BSH measures. Future research examining the utility of multidimensional IRT models would clearly be of interest. Testlet models (40), which allow for dependences between indicators within bundles of items (i.e., “testlets”), are particularly appealing given the fact that the our items are clustered into several well defined batteries (ADLs, chronic conditions, etc). A second area of clear interest would be in cross-cultural studies of health. The MHAS is a parallel study to the HRS focusing on Mexicans. Many of the same batteries are used. The key question in this comparison would be whether the BSH measure is invariant (41) to the cultural context and linguistic changes in the items. Given the

## **A broad sense measure of health**

sex differences observed here, other measurement variation may well exist. If variation is found, then modifications to more closely align the items, perhaps focusing on their language, may be useful. If no variation is found, then BSH could be used to examine the potential for subjective differences in SRH as a function of cultural context.

## **4B. Conclusion**

In this paper we discuss a global health measure based on combining information from many different health batteries. In terms of survey design, this measure suggests that we are generating little information about those individuals who are not in poor health. Given the current state of indicators in HRS, we are not readily able to differentiate between those who are in “fair” and “excellent” health. This is partially a side effect of the focus on the disability process and specific limitations and conditions. Including these in a central manner in studies of disability is rightfully important, but broader measures may be worth considering in an era of increasingly healthy elderly. Despite this shortcoming, we still find our composite indicator to be a superior measure of overall health (e.g., in terms of mortality prediction) than self-reported health. This is especially true for females. Taken together, we think that there is room for improvement in both how we develop indicators asking about health and how we go about operationalizing measures of health once we have asked.

What do we mean by “health”? Those who study disability have a definition of fairly narrow scope. Broader social research largely evades deep thought on the issue by simply asking the respondent to self-report their “health”, whatever that might mean to the respondent. Although the approach we take here was data-driven, we think it illuminates shortcomings in our current conceptions of health (at least to the extent that these conceptions are being operationalized in large-scale surveys). As medical advances allow individuals to live longer and

## **A broad sense measure of health**

in relatively better health, we think it prudent to begin to expand upon what precisely it is we mean by “health”.



## A broad sense measure of health

### References

1. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med.* 1994;38(1):1–14.
2. Siegel M, Bradley EH, Kasl SV. Self-rated life expectancy as a predictor of mortality: evidence from the HRS and AHEAD surveys. *Gerontology.* 2003;49(4):265–271.
3. Latham K, Peek CW. Self-rated health and morbidity onset among late midlife US adults. *J Gerontol B Psychol Sci Soc Sci.* 2012;gbs104.
4. Idler EL, Benyamini Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *J Health Soc Behav.* 1997;38(1):21–37.
5. Ferraro KF. Self-ratings of health among the old and the old-old. *J Health Soc Behav.* 1980;377–383.
6. 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.
7. Angel RJ. After Babel: Language and the fundamental challenges of comparative aging research. *J Cross-Cult Gerontol.* 2013;28(3):223–238.
8. Hambleton RK, Swaminathan H. Item response theory: Principles and applications. Springer Science & Business Media; 2013.
9. Ortman JM, Guarneri CE. United States population projections: 2000 to 2050. *U S Census Bur.* 2009;1–19.
10. Montez JK, Friedman EM. Educational attainment and adult health: Under what conditions is the association causal? *Soc Sci Med.* 2015;(127):1–7.
11. Levy BR, Slade MD, Kunkel SR, Kasl SV. Longevity increased by positive self-perceptions of aging. *J Pers Soc Psychol.* 2002;83(2):261.
12. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med.* 2009;69(3):307–316.
13. Mitchell R. Commentary: The decline of death—how do we measure and interpret changes in self-reported health across cultures and time? *Int J Epidemiol.* 2005;34(2):306–308.
14. Bowen ME. Childhood socioeconomic status and racial differences in disability: evidence from the Health and Retirement Study (1998–2006). *Soc Sci Med.* 2009;69(3):433–441.
15. Rutters F, Pilz S, Koopman AD, Rauh SP, Te Velde SJ, Stehouwer CD, et al. The association between psychosocial stress and mortality is mediated by lifestyle and chronic diseases: The Hoorn Study. *Soc Sci Med.* 2014;118:166–172.
16. Haas S. Trajectories of functional health: the “long arm” of childhood health and socioeconomic factors. *Soc Sci Med.* 2008;66(4):849–861.

## A broad sense measure of health

17. Warner DF, Brown TH. Understanding how race/ethnicity and gender define age-trajectories of disability: An intersectionality approach. *Soc Sci Med.* 2011;72(8):1236–1248.
18. Henretta JC. Early childbearing, marital status, and women’s health and mortality after age 50. *J Health Soc Behav.* 2007;48(3):254–266.
19. Himes CL. Obesity, disease, and functional limitation in later life. *Demography.* 2000;37(1):73–82.
20. Case A, Paxson C. Causes and consequences of early-life health. *Demography.* 2010;47(1):S65–S85.
21. Elder TE. The predictive validity of subjective mortality expectations: Evidence from the Health and Retirement Study. *Demography.* 2013;50(2):569–589.
22. Hayward MD, Gorman BK. The long arm of childhood: The influence of early-life social conditions on men’s mortality. *Demography.* 2004;41(1):87–107.
23. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc.* 1983;31(12):721–727.
24. Wallace RB, Herzog AR. Overview of the health measures in the Health and Retirement Study. *J Hum Resour.* 1995;S84–S107.
25. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: Evaluation of a short form of the CES-D. *Am J Prev Med.* 1994;
26. Chan KS, Kasper JD, Brandt J, Pezzin LE. Measurement equivalence in ADL and IADL difficulty across international surveys of aging: findings from the HRS, SHARE, and ELSA. *J Gerontol B Psychol Sci Soc Sci.* 2012;67(1):121–132.
27. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. *J Gerontol B Psychol Sci Soc Sci.* 2002;57(5):S275–S284.
28. Brick JM, Williams D. Explaining rising nonresponse rates in cross-sectional surveys. *Ann Am Acad Pol Soc Sci.* 2013;645(1):36–59.
29. Juster FT, Suzman R. An overview of the Health and Retirement Study. *J Hum Resour.* 1995;S7–S56.
30. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. *J Stat Softw.* 2012;48(6):1–29.
31. Phelan JC, Link BG, Diez-Roux A, Kawachi I, Levin B. “Fundamental causes” of social inequalities in mortality: a test of the theory. *J Health Soc Behav.* 2004;45(3):265–285.

## A broad sense measure of health

32. Koedel C, Betts J. Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education*. 2010;5(1):54–81.
33. Crimmins EM, Kim JK, Solé-Auró A. Gender differences in health: results from SHARE, ELSA and HRS. *Eur J Public Health*. 2010;ckq022.
34. Fries JF. The compression of morbidity. *Milbank Mem Fund Q Health Soc*. 1983;397–419.
35. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. *Proc Natl Acad Sci*. 2015;112(30):E4104–E4110.
36. Swallen KC, Reither EN, Haas SA, Meier AM. Overweight, obesity, and health-related quality of life among adolescents: the National Longitudinal Study of Adolescent Health. *Pediatrics*. 2005;115(2):340–347.
37. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):1.
38. Mathers CD, Sadana R, Salomon JA, Murray CJ, Lopez AD. Healthy life expectancy in 191 countries, 1999. *The Lancet*. 2001;357(9269):1685–1691.
39. Reckase M. Multidimensional item response theory. Vol. 150. Springer; 2009.
40. Bradlow ET, Wainer H, Wang X. A Bayesian random effects model for testlets. *Psychometrika*. 1999;64(2):153–168.
41. Millsap RE. Statistical approaches to measurement invariance. Routledge; 2012.

## A broad sense measure of health

Table 1. Item statistics for 21384 respondents with health indicators from 1998.

	Mean	r(item,total)	Easiness	Discrimination	# NA
adl.r4walkra	0.927	0.612	6.811	4.404	43
adl.r4dressa	0.897	0.615	4.553	3.146	32
adl.r4batha	0.918	0.633	6.719	4.553	34
adl.r4eata	0.962	0.486	6.773	3.453	39
adl.r4beda	0.927	0.587	5.478	3.374	40
adl.r4toilta	0.938	0.549	5.747	3.343	50
<i>ADL MEAN</i>	<i>0.928</i>	<i>0.58</i>	<i>6.014</i>	<i>3.712</i>	<i>40</i>
iadl.r4mapa	0.873	0.445	2.365	1.354	2865
iadl.r4phonea	0.947	0.463	4.729	2.348	150
iadl.r4moneya	0.931	0.522	4.563	2.561	654
iadl.r4medsa	0.959	0.447	5.507	2.606	1352
iadl.r4shopa	0.895	0.669	5.797	4.351	758
iadl.r4mealsa	0.928	0.604	6.135	3.891	1140
<i>IADL MEAN</i>	<i>0.922</i>	<i>0.525</i>	<i>4.849</i>	<i>2.852</i>	<i>1153</i>
fl.r4walksa	0.723	0.703	2.333	3.54	540
fl.r4walk1a	0.864	0.671	4.623	3.867	224
fl.r4sita	0.817	0.441	1.93	1.282	245
fl.r4chaira	0.635	0.585	0.9	1.909	81
fl.r4climsa	0.579	0.629	0.56	2.598	2470
fl.r4clim1a	0.829	0.68	3.581	3.466	842
fl.r4stoopa	0.593	0.598	0.638	2.144	398
fl.r4lifta	0.77	0.677	2.47	2.979	1024
fl.r4dimea	0.93	0.423	3.576	1.693	102
fl.r4armsa	0.841	0.514	2.36	1.632	150
fl.r4pusha	0.752	0.672	2.141	2.786	1716
<i>FL AVERAGE</i>	<i>0.758</i>	<i>0.599</i>	<i>2.283</i>	<i>2.536</i>	<i>708</i>
cesd.r4depres	0.825	0.468	1.858	1.216	2066
cesd.r4effort	0.737	0.51	1.257	1.272	2066
cesd.r4sleepr	0.653	0.418	0.685	0.9	2062
cesd.r4whappy	0.865	0.377	2.089	0.993	2068
cesd.r4flone	0.821	0.412	1.732	1.012	2062
cesd.r4fsad	0.798	0.427	1.565	1.019	2066
cesd.r4going	0.766	0.465	1.422	1.208	2074
cesd.r4enlife	0.917	0.341	2.754	1.105	2074
<i>CESD AVERAGE</i>	<i>0.798</i>	<i>0.427</i>	<i>1.67</i>	<i>1.091</i>	<i>2067</i>
chron.r4hibpe	0.569	0.286	0.294	0.497	22
chron.r4diabe	0.868	0.235	2.003	0.586	23
chron.r4cancre	0.898	0.117	2.214	0.296	26
chron.r4lunge	0.93	0.244	2.89	0.872	13
chron.r4hearte	0.798	0.325	1.526	0.747	16

## A broad sense measure of health

chron.r4stroke	0.928	0.339	3.116	1.207	12
chron.r4psyche	0.898	0.359	2.566	1.049	13
chron.r4arthre	0.516	0.405	0.081	0.994	19
<i>CHRON</i>					
<i>AVERAGE</i>	<i>0.801</i>	<i>0.289</i>	<i>1.836</i>	<i>0.781</i>	<i>18</i>

---

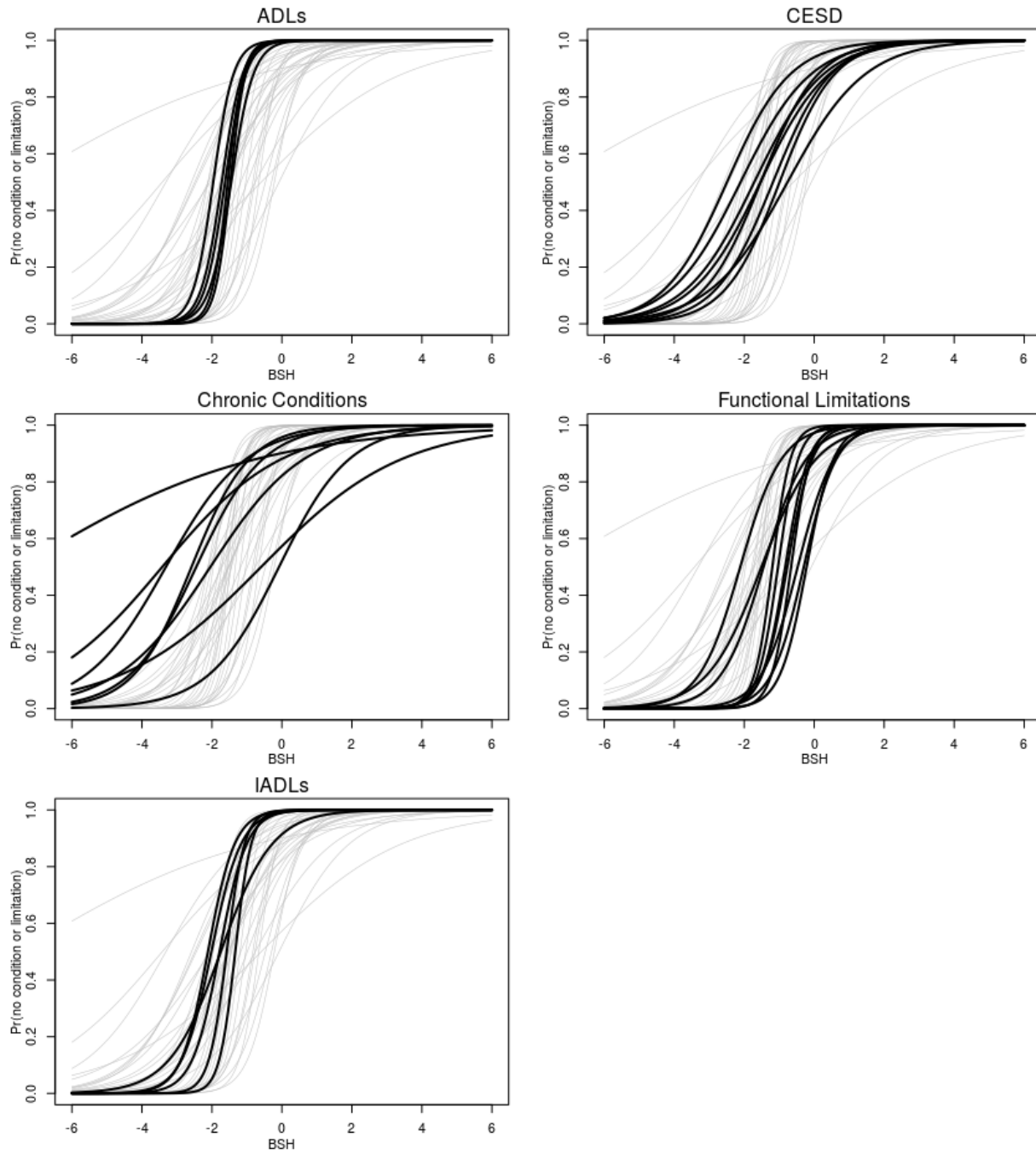
## A broad sense measure of health

Table 2. Mortality model  $r^2$  values.

	Birthdate only	SRH baseline	BSH baseline	Combined
All	0.0613	0.1302	0.1413	0.1547
Males	0.0513	0.1318	0.1351	0.1537
Females	0.0767	0.1400	0.1724	0.1789

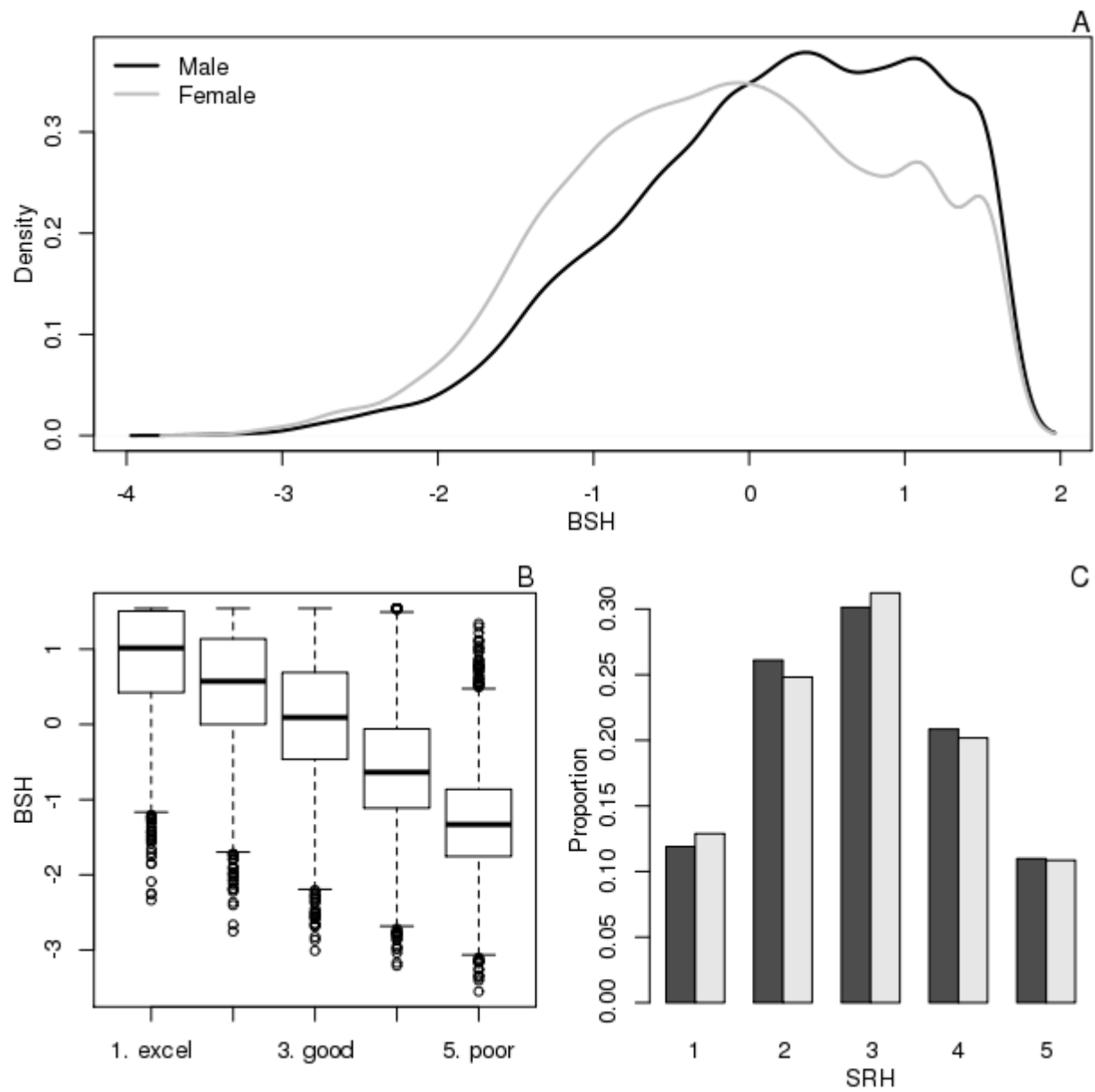
## A broad sense measure of health

Figure 1. Item characteristic curves for different item batteries.



## A broad sense measure of health

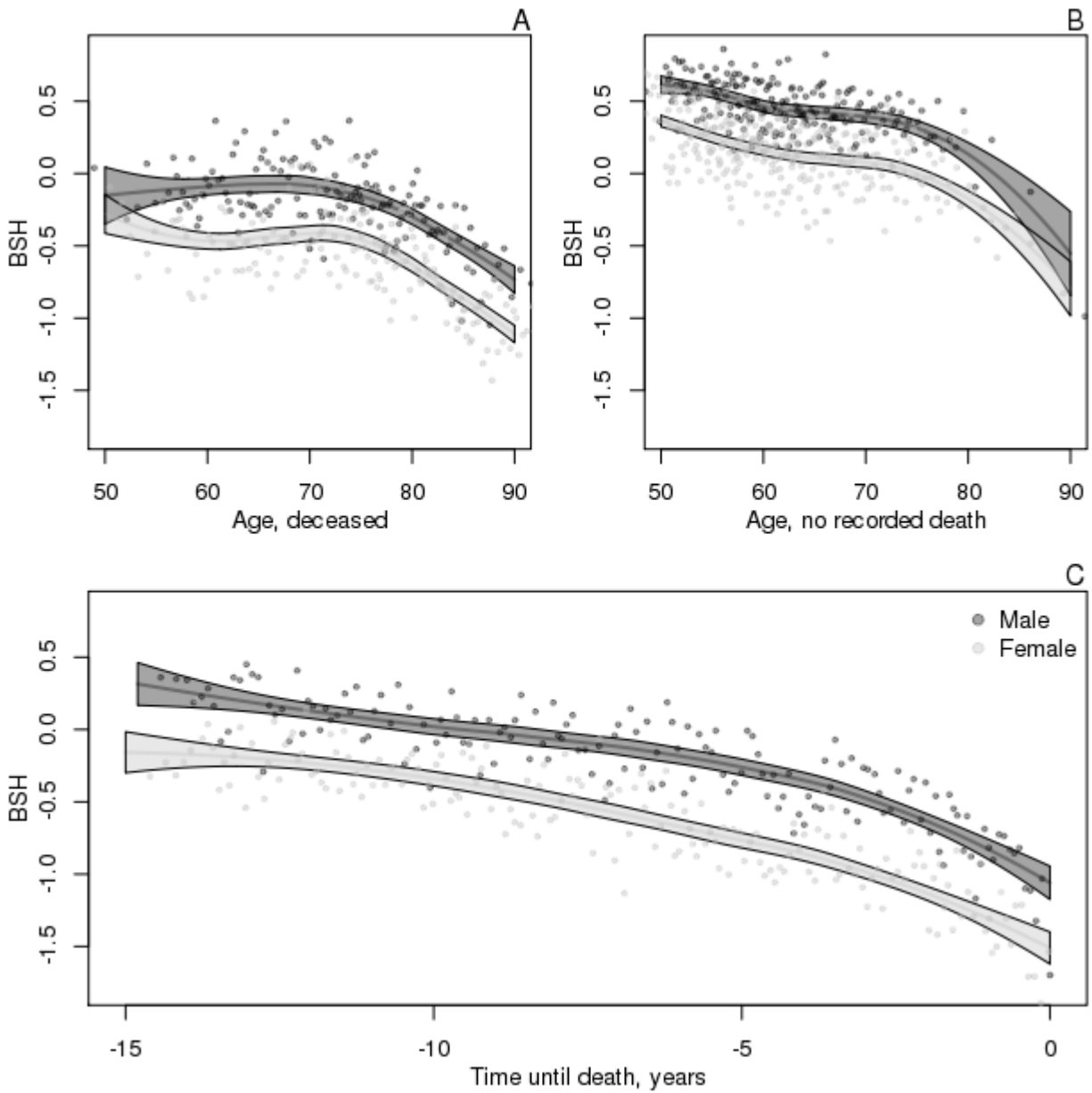
Figure 2. (A) Density of BSH by sex, (B) a comparison to SRH, and (C) proportions of SRH by sex.





## A broad sense measure of health

Figure 3. BSH as a function of age (A-for those who die; B-for those with no recorded death) and time before death (C) with binned points each representing approximately 25 people.



## A broad sense measure of health

Figure 4. Predictions of time until death using BSH and SRH. (A) Densities of time until death for males and females (N=8318), (B-D) comparison of estimates for all respondents and then split by sex in models of time until death.

