Does self-reported health bias the measurement of health inequalities in U.S. adults? Evidence using anchoring vignettes from the Health and Retirement Study

Jennifer Dowd, Hunter College, City University of New York (CUNY), CUNY School of Public Health and CUNY Institute for Demographic Research (CIDR)
Megan Todd, CUNY Institute for Demographic Research (CIDR)

Presented at the Annual Meeting of the Population Association of America, Dallas, TX, April 2010

Word count for abstract: 200

Word count for text: 4525

Figures: 1

Tables: 4

Running Head: self-rated health, anchoring vignettes

Abstract:

Objectives: Measurement of health inequalities based on self-reports may be biased if individuals use response scales in systematically different ways. We use anchoring vignettes to test and adjust for reporting differences by level of education and race/ethnicity in self-reported health in six domains (pain, sleep, mobility, memory, shortness of breath, and depression).

Methods: Using data from the U.S. Health and Retirement Study (HRS) 2007 Disability Vignette Study (DVS), we estimated generalized ordered probit models of the respondent's rating of each vignette character's health problem, allowing cut-points to vary by education and race/ethnicity. We then used one-step hierarchical ordered probit (HOPIT) models to jointly estimate the respondent's cut-points from the vignettes and the severity of the respondent's own health problems based on these vignette cut-points.

Results: We found strong evidence of reporting differences by both education and race/ethnicity, with the magnitude depending on the specific health domain. Overall, traditional models not accounting for reporting differences underestimated the magnitude of health inequalities by education and race/ethnicity.

Discussion: These results suggest caution in relying on self-reported health measures to quantify and explain health disparities by socioeconomic status and race/ethnicity in the U.S. The findings support expansion of the use of anchoring vignettes to properly account for reporting differences in self-reports of health.

Key Words: Self-rated health, anchoring vignettes, socioeconomic status, education, racial disparities, HRS

INTRODUCTION

The accurate measurement of population health is crucial for monitoring levels and trends in health across different groups, testing the mechanisms that account for health differences, and evaluating the impact of interventions. There are persistent inequalities in health and mortality by socioeconomic status (SES) and race/ethnicity (Adler, Boyce, Chesney, Folkman, & Syme, 1993; Williams, Yan, Jackson, & Anderson, 1997). General self-rated health (SRH) status on a five-point scale has been used extensively to measure health inequalities, based partly on its ease of collection and consistent associations with subsequent morbidity and mortality (DeSalvo, Bloser, Reynolds, He, & Muntner, 2006; Idler & Benyamini, 1997; Idler, Russell, & Davis, 2000). The use of similar self-reported scales to measure morbidity in specific domains such as mobility, pain, and depression is also common (Schoeni, Martin, Andreski, & Freedman, 2005).

The measurement of health inequalities using self-reported measures may be problematic, however, if individuals from different social groups have systematically different expectations or reporting standards for health which would lead them to use response categories such as “excellent” or “poor” in different ways. Such systematic reporting differences could lead to significant over or under-estimates of health inequalities based on self-reported measures. The problem of reporting differences has alternatively been referred to as "state-dependent reporting bias"(Kerkhofs & Lindeboom, 1995),"response category differential item functioning" (King, Murray, Salomon, & Tandon, 2004; King & Wand, 2007) and finally "reporting heterogeneity" (Lindeboom & van Doorslaer, 2004; Shmueli, 2003)**,** the term we will use in the current paper. Reporting heterogeneity implies the use of different category cut-points when reporting health status, *holding constant* a given level of “true” health.

Reporting heterogeneity may arise in a variety of ways. Culturally or linguistically, groups may understand and use the ordinal response scale in systematically different ways, such as the

propensity to use extreme categories. In Spanish for example, the middle category options "fair" or "average" seem to replace "good" as what is considered a normal anchor point for evaluations (Bzostek, Goldman, & Pebley, 2007). While "excellent" is a commonly-used term in English, Germans consider use of this word in their language an exaggeration and are thus less likely to use this category to rate their health (Jürges, 2007). Individuals may also compare themselves to a reference group of their peers, a phenomenon which has been identified in evaluations across age (Schnittker, 2005). Respondents with more education may compare themselves to their relatively healthy peers for instance, resulting in stricter standards for what is considered "excellent" health. Such standards would lead to systematically worse health reports compared to lower education groups and an *underestimate* of health inequalities. On the other hand, higher health expectations for those with more education could result in more lenient standards for health, a so-called "wishful thinking" scenario (Iburg, 2001). In this case, individuals with more education who believe they *should* be healthy would systematically upgrade their health compared to the same level of "true" health for those with less education, leading to an *overestimate* of health inequalities.

Uniform shifts in the thresholds that define response categories have been referred to as index, or parallel shifts (Lindeboom & van Doorslaer, 2004). In addition to uniformly shifting up or down, it is possible that the relative positions of reporting thresholds may differ across groups, which is referred to as a cut-point, or non-parallel shift (Lindeboom & van Doorslaer, 2004). Figure 1 illustrates a hypothetical example of reporting heterogeneity with respect to a five-point SRH scale by SES, with the cut-points of higher SES individuals shifted in a non-parallel way so that they are more likely to report greater health problems for a given level of objective health.

[Figure 1 about here]

Existing Studies

Thus far, there have been several approaches to testing for reporting heterogeneity in self-reported health measures (Ferraro & Kelley-Moore, 2001; Lee et al., 2007; McGee, Liao, Cao, &

Cooper, 1999). The first strategy uses other types of self-reports as benchmarks for "objective" health such as self-reported illnesses or physical functioning (Delpierre, Lauwers-Cances, Datta, Lang, & Berkman, 2009; Etile & Milcent, 2006; Melzer, Lan, Tom, Deeg, & Guralnik, 2004), or the more detailed self-reported MacMaster Health Utility Index or SF-36 instrument (Humphries & Van Doorslaer, 2000; Shmueli, 2003). These studies have identified evidence for reporting heterogeneity by SES in France, Canada, and Israel (Etile & Milcent, 2006; Humphries & Van Doorslaer, 2000; Shmueli, 2003). Applying this approach to race/ethnic reporting differences in the U.S., black individuals at the same level of physical functioning have been found to report significantly and substantially worse SRH than whites (Boardman, 2004; Spencer et al., 2009), a difference compatible with "health pessimism" among black adults (Ferraro, 1993). Others have found a similar pattern of differences in reporting between whites and Hispanics (Borrell & Crawford, 2006).

Since the use of self-reports to validate SRH cannot rule out reporting differences shared by both sets of self-reports, another strategy has been to test whether the association of SRH with mortality risk varies by SES. Results have varied by country, with evidence that SRH is a stronger predictor of mortality for higher SES groups in the U.S. and the Netherlands, a weaker predictor for higher SES groups in France, and an equal predictor across SES in Sweden and the U.K. (Burstrom & Fredlund, 2001; Dowd & Zajacova, 2007; Huisman, Lenthe, & Mackenbach, 2007; McFadden et al., 2009; Quesnel Vallee, 2007; Singh-Manoux et al., 2007; van Doorslaer & Gerdtham, 2003). Within the U.S., SRH among black adults has been found to be less predictive of mortality than those of whites (Ferraro & Kelley-Moore, 2001; Lee et al., 2007), and less predictive for less acculturated versus more acculturated Hispanics (Finch, Hummer, Reindl, & Vega, 2002). While this approach has advantages over using other self-reports as benchmarks to identify reporting differences, it is nonetheless limited in its ability to isolate heterogeneity in the use of health

categories from other factors such as knowledge of underlying health conditions or differences in the weight given to mental or emotional factors when reporting global self-rated health.

One unique tool for separating "true" health differences from health reporting differences that has been proposed is the use of anchoring vignettes. In this approach, survey respondents rate the health of fictitious individuals with given health conditions or disabilities, and their evaluations of these fictitious individuals are used to estimate reporting thresholds (or cut-points), which may depend on individual characteristics. These thresholds are then used to adjust the rating of the respondent's own health (Kapteyn, Smith, & van Soest, 2007; King & Wand, 2007; Salomon, Tandon, & Murray, 2004). The vignettes represent fixed levels of latent health, so that any differences in ratings of the vignettes can be attributed to reporting differences. Under the assumption that respondents rate the vignettes in the same way as they rate their own health ("response consistency"), it is possible to identify a health measure that is purged of reporting differences (Bago d'Uva, van Doorslaer, Lindeboom, & O'Donnell, 2008; King et al., 2004). Identification of cut-points via vignettes also relies on a second assumption, "vignette equivalence," which requires that the level of health represented by any one vignette is perceived by all respondents in the same way and on the same unidimensional scale (King et al., 2004). Research using anchoring vignettes has identified reporting heterogeneity by education and income in several European and Asian countries (Bago d'Uva, O'Donnell, & van Doorslaer, 2008; Bago d'Uva, van Doorslaer et al., 2008), but thus far no studies have used anchoring vignettes to examine reporting differences by SES or race/ethnicity in the U.S.

This paper tests for reporting differences in self-reported health measures by level of education and race/ethnicity in the U.S. using anchoring vignettes from the Health and Retirement Study in the domains of mobility, sleep, pain, memory, shortness of breath, and depression. We then estimate inequalities in each health domain by education and race/ethnicity before and after

adjustment for reporting heterogeneity to quantify the magnitude of potential biases in health reporting using self-reported measures.

METHODS

Data

Data come from the U.S. Health and Retirement Study (HRS) 2006 Core, and the Disability Vignette Study (DVS), a mail survey conducted in 2007 by HRS. The HRS Core is a nationally representative survey of U.S. adults aged 50 and older, as well as their spouses (regardless of age). DVS respondents were asked to rate their own health in various domains, then read a short vignette and rate the health of the vignette characters. Respondents who participated in an HRS 2006 Core self-interview were eligible for participation in the DVS. Those who had died prior to the start of the DVS, those who requested removal from the sample, and those who were participating in other HRS studies in 2007 were removed. Of 5,678 mailed questionnaires, 4,639 were returned (81.7%). DVS data were combined with HRS 2006 Core data including demographic information on sex, age, race, and education.

Measures

Respondents were asked to rate the severity of their own health problems first, and then the health problems of vignette characters on a five-point scale, where 1=none, 2=mild, 3=moderate, 4=severe, and 5=extreme. Self-rated health in each domain was obtained from the following questions: "Overall, in the last 30 days, how much…."

- "pain or bodily aches did you have?" (pain)
- "difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?" (sleep)
- "of a problem did you have with moving around?" (mobility)
- "difficulty did you have with concentrating or remembering things?" (memory)

- "of a problem did you have because of shortness of breath?" (shortness of breath)
- "of a problem did you have with feeling sad, low, or depressed?" (depression)

Vignette questions were of the form: "Charles has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall, in the last 30 days, how much of bodily aches or pains did Charles have?" Three vignettes were presented for each health domain, representing different levels of severity. Vignette questions are listed in detail in the Supplementary Materials. Age was measured continuously as age at the end of 2006. Race was categorized as white (reference), black, and other. Educational attainment was classified into four categories: less than high school (reference), high school, some college, and college or more. Descriptive statistics for the sample and covariates are presented in Table 1.

[Table 1 about here]

Statistical Analysis

For each health domain, we first estimated traditional ordered probit models of self-reported health on education, race/ethnicity, sex and age, assuming no reporting differences. Next, we estimated a generalized ordered probit model of the respondents' severity rating of vignette characters' health problems to test the hypothesis of homogeneous reporting. Severity ratings depended only on a vignette dummy, with each of the four cut-points depending on age, sex, race, and education. We tested the hypothesis of homogenous reporting with joint and individual Wald tests of the significance of covariates in all four cut-point models; a significant coefficient for a covariate provides evidence that thresholds varied by this covariate and thus reporting was not homogeneous across the covariate. To test the hypothesis of parallel versus non-parallel shifts in cut-points, we performed Wald tests for whether the coefficients for a given covariate were the same in all four cut-point models. Finally, we estimated a one-step hierarchical ordered probit (HOPIT) model (Bago d'Uva, O'Donnell et al., 2008; Bago d'Uva, van Doorslaer et al., 2008; Jones, 2007), which

jointly estimated cut-points based on respondents' ratings of the severity level of vignette characters' health problems and the severity level of the respondent's own health problems based on these cut-points. The thresholds for reporting extreme, severe, moderate, mild or no problem in each health domain were allowed to vary by age, sex, education, and race/ethnicity. Finally, we compared the coefficients for education and race/ethnicity from the HOPIT models of health problem severity in each domain to those from the traditional ordered probit models to assess the direction and magnitude of associations before and after adjustment for reporting heterogeneity. All statistical analyses were performed with Stata 10 (StataCorp, College Station, TX) using the command oprobit for the ordered probit models and code adapted from (Jones, 2007) for the generalized ordered probit and HOPIT models.

RESULTS

Table 2 presents results of tests of homogenous reporting, finding strong evidence of reporting differences across all covariates. The null hypothesis of homogenous reporting was rejected (p-value<0.05) in every health domain when age, gender, education level, and race were jointly considered. Homogenous reporting by gender was rejected in four out of six health domains (all but mobility and memory), and rejected in all domains with respect to age. Homogenous reporting by race was rejected in every health domain for black respondents and in five domains (all but memory) for respondents of "other" race. By educational attainment, homogeneous reporting was also rejected in every domain for respondents with at least some college, and was rejected in five out of six domains (all but memory) for respondents with a high school education.

[Table 2 about here]

Tests for parallel shift versus non-parallel cut-point shifts are presented in Table 3. When all covariates were jointly considered, the null hypothesis of parallel shift was rejected (p-value<0.05) in all six health domains. Parallel shift was rejected in four domains for gender (all but mobility and memory), all domains for age, five domains (all but memory) for black respondents,

and five domains (all but memory) for respondents of "other" race. Parallel shift by educational attainment was also rejected in all health domains for respondents with at least some college, and was rejected in five health domains (all but memory) for respondents with a high school education. These results provide strong evidence that where reporting differences exist, they are not characterized by uniform shifts in thresholds up or down the latent health scale; reporting differences may be stronger at some levels of health compared to others. Taken together, the results in Tables 2 and 3 provide strong evidence of heterogeneous reporting behaviors by age, sex, race, and educational attainment in most health domains, with slightly less evidence of reporting differences in the domains of mobility and memory. The results support the need for a model that accounts for reporting differences behavior in respondents' self-reported health.

[Table 3 about here]

Table 4 illustrates the magnitude and direction of cut-point shifts by each covariate. Stricter health standards are reflected in negative coefficients, such that the cut-point moves downward (towards the milder end) on the objective health scale. This lower threshold translates into a greater likelihood of reporting a more severe rating for a given health state compared to the reference respondent. A positive coefficient indicates that the cut-point is pushed higher along the objective health scale, translating into a less severe rating for a given health problem compared to the reference respondent. Rather than a pattern of uniform reporting differences, the results in Table 4 indicate a wide range of reporting behaviors. Non-parallel shift can be seen in the unequal coefficients for a given covariate from one cut-point model to another. For example, in the domain of sleep, those with a high school education have consistently lower cut-points (negative coefficients in all four cut-point models) compared to those with less than a high school education, so they will consistently rate a given sleep problem as more severe. However, the magnitude of the shift varies for each cut-point. The larger shifts for cut-points 1 and 2 compared to cut-points 3 and 4 indicate that a respondent with a high school education will rate a relatively minor sleep problem

as much more severe than the reference respondent would, while the reporting differences for more serious sleep problems are less pronounced. Cut-point shifts need not all be in the same direction; one cut-point may shift up while another shifts down. This type of shift can be seen in the domain of shortness of breath among black respondents. Cut-points 1 and 2 are shifted up for black respondents, while cut-points 3 and 4 are shifted down. Compared to white respondents, black respondents are thus more likely to rate a relatively minor shortness of breath problem as less severe, while at the same time rating a relatively serious shortness of breath problem as more severe. Black respondents' cut-points are relatively close together such that their ratings tend to be more heavily distributed toward both extremes of the severity scale. The coefficient in cut-point 3 is not significant at the .05 level, meaning that white and black respondents have similar thresholds for separating "moderate" and "severe" shortness of breath problems.

[Table 4 about here]

Table 4 can also shed light on whether reporting differences are applied similarly to all health domains. The pattern of cut-point shifts by covariates is different across health domains, suggesting that reporting differences reflect more than a tendency to use the five-point scale in a uniformly different way. For example, black respondents have a positive coefficient in the model for cut-point 1 in the domain of mobility, but a negative coefficient in the model for cut-point 1 in the domain of pain. This means that, compared to a white respondent, a black respondent is more likely to rate a minor health problem as "none" (versus "mild") if the health problem is related to mobility, but more likely to rate it as "mild" (rather than "none") if the problem is related to pain.

Table 5 presents results for all six health domains in models unadjusted (ordered probit) and adjusted (HOPIT) for reporting heterogeneity. The adjustment had a mixed impact on the magnitude of the estimated coefficients for age, gender, race, and education. The changes in the association of age with each health domain were the least noticeable, with the exception of depression, for which an apparent older age advantage transformed into no significant difference by

age after adjustment. The magnitude of the coefficient for males increased by approximately 50% in the direction of less pain than females after adjustment, but the male-female difference decreased by more than half (and became insignificant) with regard to the severity of sleep problems.

[Table 5 about here]

Adjustment changed many substantive interpretations of black-white health differences across domains. For example, black-white differences in pain changed from an insignificant coefficient indicating increased pain for blacks to a large, significant coefficient indicating *less* severe pain for blacks compared to whites. In the cases of shortness of breath and depression, what looked like no racial difference prior to adjustment became a significant health disadvantage for blacks after adjustment. Sleep worked in a similar direction, such that prior to adjustment there appeared to be a black advantage with regard to severity of sleep problems that disappeared upon adjustment. Memory worked in the opposite direction, such that prior to adjustment there was a black disadvantage that dissipated upon adjustment for reporting differences. Mobility was the only domain for which black-white differences did not significantly change before and after adjustment, although the coefficient moved in the direction of increasing black disadvantage.

In contrast to the changing direction of effects by race, higher levels of education were consistently associated with better health ratings before and after adjustment, though the magnitude of the coefficients sometimes changed dramatically. While the educational differences were approximately the same for the domains of pain and memory, the coefficients on health differences by education more than doubled for sleep and shortness of breath, nearly doubled for depression, and increased by 36% for mobility differences after adjustment for reporting differences. These results were quite consistent across educational categories. Based on these findings, educational inequalities in many health domains may be substantially underestimated when not adjusted for reporting differences.

We tested the robustness of our results with several alternative model specifications. First, we stratified models by sex to see if reporting differences by education and race/ethnicity varied by sex. Overall, reporting differences by education were quite comparable for men and women, but there was some evidence that reporting differences between blacks and whites differed by sex (see Supplementary Table 6S). The reporting differences by race/ethnicity in the domains of pain and sleep in the pooled results were seen only for men in the stratified results. Black women reported more severe mobility limitations than whites after adjustment, while no reporting differences were seen in mobility for men. Black women reported more severe memory problems before adjustment, but there was no significant difference compared to white women after adjustment. For shortness of breath, a disadvantage for black women emerged only after adjustment, while black men reported less severe problems prior to adjustment, with no difference after adjustment. Black women reported more severe depression compared to whites after adjustment, but for men there were no racial differences in depression before or after adjustment. Overall, the direction of reporting differences by race was consistent across sex, with the pooled results demonstrating the relative health optimism of black reporting in most domains.

Many studies that use self-reported health measures collapse the 5-category response scale to two or three categories for ease of analysis and interpretation. In the context of the current paper, such a reduction in categories may reduce the degree of reporting heterogeneity, specifically if the heterogeneity derives from differences in cut-point thresholds between the categories that are being combined. While Table 4 indicated that reporting heterogeneity exists across all cut-points, we ran identical HOPIT models with three categories of self-reported severity (1=none/mild, 2=moderate, 3=severe/extreme) to see if this would reduce the differences in the magnitude of health inequalities by education and race/ethnicity seen before and after adjustment. The reduction in categories did slightly reduce the before-and-after changes in race/ethnic inequalities compared to the 5-category results, but educational inequalities were still much larger in adjusted models compared to

unadjusted models (see Supplementary Table 7S). This suggests that concerns about reporting heterogeneity are not likely easily mitigated by collapsing response categories.

DISCUSSION

This paper used anchoring vignettes to test for reporting heterogeneity in self-reported health measures by level of education and race/ethnicity in middle-aged and older adults from the Health and Retirement Study. We found strong evidence of reporting differences, but the magnitude of this heterogeneity varied depending on the specific health domain. Overall, traditional models that do not account for reporting differences seem to underestimate the magnitude of health inequalities by education, especially for self-reported depression, shortness of breath, mobility, and sleep. These differences reflect the stricter reporting standards of more highly educated groups compared to those with less education in those domains. The magnitude of educational reporting differences is much smaller in the domains of pain and memory. Similarly, for comparisons of black versus white respondents, not accounting for reporting differences in traditional models appears to underestimate inequalities in most domains, suggesting that black respondents are relatively optimistic in their ratings compared to whites, perhaps comparing themselves to their less healthy peers as a reference group. Importantly, in several cases (pain, shortness of breath, and depression), accounting for reporting differences revealed racial/ethnic inequalities that were not evident when looking at results from traditional ordered probit models. Parallel shifts in reporting thresholds was rejected in all but a few cases, suggesting that reporting differences are not uniform across the entire latent health scale, but may be stronger or weaker depending on the level of health.

To our knowledge, this is the first study to test for reporting differences in health measures by SES and race/ethnicity in the U.S. using anchoring vignettes. Previous work in other countries has suggested, not surprisingly, that the degree of reporting heterogeneity by SES differs substantially across contexts. In Indonesia and India, Bago d'Uva et al. found that more educated

respondents generally over-reported their health, leading to a reduction in educational health disparities after adjustment from vignettes (Bago d'Uva, van Doorslaer et al., 2008). In China, in contrast, adjustment raised inequalities by education in most cases, more similar to the results found here. Our results are also consistent with results from eight European countries, where correcting for reporting differences increased inequalities in many cases and resulted in the emergence of inequalities by education that did not previously exist in several cases (Bago d'Uva, O'Donnell et al., 2008). These effects were largest for Belgium, France, Germany, and the Netherlands, while for Sweden and Spain adjusting for reporting differences in education tended to decrease inequalities, underscoring the substantial variation in reporting differences depending on context (Bago d'Uva, O'Donnell et al., 2008).

Our finding that reporting differences by race and education varied by health domain suggests that reporting heterogeneity is not a result of uniform differences in pessimism or optimism that apply equally to all health domains. Consequently, generalizing the results of reporting differences for any one domain to domains that are not measured may be unwise. This may suggest significant challenges in extending anchoring vignettes to general health status, for which many different domains are presumably absorbed into one rating.

The current finding that racial differences in health are generally underestimated when not accounting for reporting differences contrasts with previous research that has described relative "health pessimism" among blacks such that blacks report significantly worse self-rated health when compared to whites with similar levels of self-reported morbidity (Boardman, 2004; Ferraro, 1993; Spencer et al., 2009). The majority of previous work on health pessimism among blacks used other self-reported measures as "objective" benchmarks for global self-reported health, a strategy that might suffer if both sets of reports suffer from the same reporting differences. A recent paper by Spencer and Schulz extended this work by comparing race differences in global SRH after adjustment for objectively measured physical functioning measures, finding that blacks rated their

health lower than whites at similar levels of physical functioning (Spencer et al., 2009). It is possible that in contrast to the self-reported health questions analyzed here that focused a respondent's attention on specific health domains, general self-rated health may encompass broader dimensions such as "enduring self-concept" for which blacks are more pessimistic (Spencer et al., 2009). Previous work suggests that whites may use physical functioning as a frame of reference for global self-rated health while blacks focus more on health problems (Krause & Jay, 1994). Since the health domains examined here focused more on functioning than specific health conditions, this could be consistent with our findings of relative black optimism. In general, the finding of more lenient reporting standards for both blacks and those with less education in the sample would be consistent with a reference group or peer comparison assessment, since on average these groups are in worse health. Future work should attempt to replicate and reconcile these findings with the broader literature on black health pessimism in global SRH.

Overall, these results suggest caution in relying on self-reported health measures to quantify and explain health disparities by socioeconomic status and race/ethnicity in the United States. Additional research using anchoring vignettes to test whether these findings are consistent across samples and across time is warranted. The findings support expansion of the use of anchoring vignettes to properly account for reporting differences in self-reports of health. Failure to account for such differences could lead researchers and policy-makers to the wrong conclusions regarding the magnitude of and trends in health inequalities.

REFERENCES

Adler, N. E., Boyce, W. T., Chesney, M. A., Folkman, S., & Syme, S. L. (1993). Socioeconomic inequalities in health: No easy solution. *Journal of the American Medical Association, 269*(24), 3140-3145.

Bago d'Uva, T., O'Donnell, O., & van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology, 37*(6), 1375-1383.

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics, 17*(3), 351-375.

Boardman, J. D. (2004). Health pessimism among black and white adults: the role of interpersonal and institutional maltreatment. *Social Science & Medicine, 59*(12), 2523-2533.

Borrell, L. N., & Crawford, N. D. (2006). Race, Ethnicity, and Self-Rated Health Status in the Behavioral Risk Factor Surveillance System Survey. *Hispanic Journal of Behavioral Sciences, 28*(3), 387-403.

Burstrom, B., & Fredlund, P. (2001). Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Health, 55*(11), 836-840.

Bzostek, S., Goldman, N., & Pebley, A. (2007). Why do Hispanics in the USA report poor health? *Social Science & Medicine, 65*(5), 990-1003.

Delpierre, C., Lauwers-Cances, V., Datta, G. D., Lang, T., & Berkman, L. (2009). Using self-rated health for analysing social inequalities in health: a risk for underestimating the gap between socioeconomic groups? *Journal of Epidemiology and Community Health, 63*(6), 426-432.

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality Prediction with a Single General Self-Rated Health Question. A Meta-Analysis. *Journal of General Internal Medicine, 21*(3), 267-275.

Dowd, J. B., & Zajacova, A. (2007). Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US? *International Journal of Epidemiology, 36*(6), 1214-1221.

Etile, F., & Milcent, C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from France. *Health Economics, 15*(9), 965-981.

Ferraro, K. F. (1993). Are Black Older Adults Health-Pessimistic? *Journal of Health and Social Behavior, 34*(3), 201-214.

Ferraro, K. F., & Kelley-Moore, J. A. (2001). Self-Rated Health and Mortality Among Black and White Adults: Examining the Dynamic Evaluation Thesis. *Journal of Gerontology: Social Science, 56*(4), S195-205.

Finch, B. K., Hummer, R. A., Reindl, M., & Vega, W. A. (2002). Validity of Self-rated Health among Latino(a)s. *American Journal of Epidemiology, 155*(8), 755-759.

Huisman, M., Lenthe, F. v., & Mackenbach, J. (2007). The predictive ability of self-assessed health for mortality in different educational groups. *International Journal of Epidemiology*, 36(6): 1161-4.

Humphries, K. H., & Van Doorslaer, E. (2000). Income-related health inequality in Canada. *Social Science and Medicine, 50*(5), 663-671.

Iburg, K., JA Salomon, A Tandon CJL Murray. (2001). Cross-Population Comparability of Physician-Assessed and Self-Reported Measures of Health. *World Health Organization*, *Global Programme on Evidence for Health Policy, Discussion Paper Number 14*

Idler, E. L., & Benyamini, Y. (1997). Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior, 38*(1), 21-37.

Idler, E. L., Russell, L. B., & Davis, D. (2000). Survival, Functional Limitations, and Self-rated Health in the NHANES I Epidemiologic Follow-up Study, 1992. *American Journal of Epidemiology, 152*(9), 874-883.

Jones A, R. N., Bago d'Uva T, and Balia S. (2007). *Applied Health Economics*. London: Routledge.

Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics, 16*(2), 163-178.

Kapteyn, A., Smith, J. P., & van Soest, A. (2007). Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. *The American Economic Review, 97*, 461-473.

Kerkhofs, M., & Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics, 4*(3), 221-235.

King, G., Murray, C. J., Salomon, J., & Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review, 98*(1), 191-207.

King, G., & Wand, J. (2007). Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes. *Political Analysis, 15*(1), 46-66.

Krause, N. M., & Jay, G. M. (1994). What do global self-rated health items measure? *Medical Care, 32*(9), 930-942.

Lee, S. J., Moody-Ayers, S. Y., Landefeld, C. S., Walter, L. C., Lindquist, K., Segal, M. R., Covinsky, K.E. (2007). The Relationship Between Self-Rated Health and Mortality in Older Black and White Americans. *Journal of the American Geriatrics Society, 55*(10), 1624-1629.

Lindeboom, M., & van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics, 23*(6), 1083-1099.

McFadden, E., Luben, R., Bingham, S., Wareham, N., Kinmonth, A. L., & Khaw, K. T. (2009). Does the association between self-rated health and mortality vary by social class? *Social Science & Medicine, 68*(2), 275-280.

McGee, D. L., Liao, Y. L., Cao, G. C., & Cooper, R. S. (1999). Self-Reported Health Status and Mortality in a Multiethnic US Cohort. *American Journal of Epidemiology, 149*(1), 41-46.

Melzer, D., Lan, T.-Y., Tom, B. D. M., Deeg, D. J. H., & Guralnik, J. M. (2004). Variation in Thresholds for Reporting Mobility Disability Between National Population Subgroups and Studies. *Journal of Gerontology: Medical Sciences, 59*(12), 1295-1303.

Quesnel Vallee, A. (2007). Self-rated health: caught in the crossfire of the quest for 'true' health? *International Journal of Epidemiology, 36*(6), 1161-1164.

Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *British Medical Journal, 328*(7434), 258-.

Schnittker, J. (2005). When Mental Health Becomes Health: Age and the Shifting Meaning of Self-Evaluations of General Health. *The Milbank Quarterly, 83*(3), 397-423.

Schoeni, R. F., Martin, L. G., Andreski, P. M., & Freedman, V. A. (2005). Persistent and growing socioeconomic disparities in disability among the elderly: 1982-2002. *American Journal of Public Health, 95*(11), 2065-2070.

Shmueli, A. (2003). Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. *Social Science & Medicine, 57*(1), 125-134.

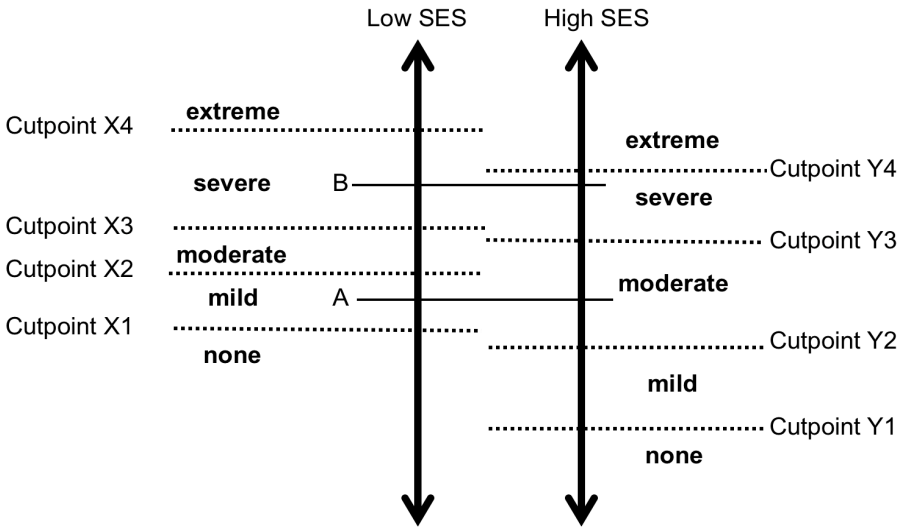
Singh-Manoux, A., Dugravot, A., Shipley, M. J., Ferrie, J. E., Martikainen, P., Goldberg, M., et al. (2007). The association between self-rated health and mortality in different socioeconomic groups in the GAZEL cohort study. *International Journal of Epidemiology, 36*(6), 1222-1228.

Spencer, S., Schulz, R., Rooks, R., Albert, S., Thorpe, R., Brenes, G., Harris, T.B., Koster, A., Satterfield, S., Ayonayon, H.N., Newman, A.B. (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: Social Sciences, 64*(1), 87-94.

van Doorslaer, E., & Gerdtham, U.-G. (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science & Medicine, 57*(9), 1621-1629.

Williams, D. R., Yan, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial Differences in Physical and Mental Health: Socio-economic Status, Stress and Discrimination. *Journal of Health Psychology, 2*(3), 335-351.

Figure 1: Heterogeneous Reporting: Non-parallel Cut-point Shift



Lines represent latent health scale; A and B represent two different levels of "objective health"

Table 1: Descriptive Statistics of Respondents

	Mean or Proportion
Age (as of end of year 2006)	65.17
Age: in 30s	0.28%
Age: in 40s	3.61%
Age: in 50s	33.81%
Age: in 60s	31.73%
Age: in 70s	19.82%
Age: in 80s	9.42%
Age: in 90s	1.28%
Age: 100 +	0.04%
Race: white	83.53%
Race: black	11.39%
Race: other	5.08%
Education: < high school	16.32%
Education: high school	34.67%
Education: some college	23.74%
Education: college +	25.27%
Self rating: pain	2.37
Self rating: sleep	2.28
Self rating: mobility	1.78
Self rating: memory	1.88
Self rating: shortness of breath	1.48
Self rating: depression	1.78
N	4,626

Note: Self-report of severity of health problems is rated on a five point scale where 1=none, 2=mild, 3=moderate, 4=severe, 5=extreme.

Table 2: Wald Test of Reporting Homogeneity in Rating Severity of Vignette Characters' Health Problems

	Chi-Squared	P-Value		Chi-Squared	P-Value
Pain			**Memory**		
All	229.219	<0.001	All	168.251	<0.001
Male	9.879	0.043	Male	7.988	0.092
Age	41.987	<0.001	Age	53.214	<0.001
Race: black	110.440	<0.001	Race: black	40.819	<0.001
Race: other	10.347	0.035	Race: other	7.494	0.112
Education: high school	31.859	<0.001	Education: high school	7.529	0.110
Education: some college	21.344	<0.001	Education: some college	20.319	<0.001
Education: college +	36.153	<0.001	Education: college +	22.213	<0.001
Sleep			**Shortness of Breath**		
All	338.788	<0.001	All	516.411	<0.001
Male	23.536	<0.001	Male	23.694	<0.001
Age	60.862	<0.001	Age	63.524	<0.001
Race: black	55.541	<0.001	Race: black	42.704	<0.001
Race: other	25.734	<0.001	Race: other	30.363	<0.001
Education: high school	38.682	<0.001	Education: high school	92.069	<0.001
Education: some college	69.718	<0.001	Education: some college	124.331	<0.001
Education: college +	121.861	<0.001	Education: college +	243.979	<0.001
Mobility			**Depression**		
All	214.645	<0.001	All	413.297	<0.001
Male	7.804	0.099	Male	18.269	0.001
Age	47.726	<0.001	Age	94.986	<0.001
Race: black	39.980	<0.001	Race: black	67.415	<0.001
Race: other	17.769	0.001	Race: other	21.134	<0.001
Education: high school	40.531	<0.001	Education: high school	52.134	<0.001
Education: some college	30.219	<0.001	Education: some college	69.649	<0.001
Education: college +	68.808	<0.001	Education: college +	146.085	<0.001

Note: Wald tests were performed on a generalized ordered probit model of respondents' ratings of vignette characters' health problems.

Table 3: Wald Test of Parallel Shift in Rating Severity of Vignette Characters' Health Problems

	Chi-Squared	P-Value		Chi-Squared	P-Value
Pain			**Memory**		
All	120.515	<0.001	All	58.197	<0.001
Male	8.619	0.035	Male	4.264	0.234
Age	41.958	<0.001	Age	26.216	<0.001
Race: black	25.623	<0.001	Race: black	0.472	0.925
Race: other	0.304	0.959	Race: other	6.810	0.078
Education: high school	28.365	<0.001	Education: high school	7.484	0.058
Education: some college	14.277	0.003	Education: some college	9.575	0.023
Education: college +	29.802	<0.001	Education: college +	8.229	0.042
Sleep			**Shortness of Breath**		
All	165.653	<0.001	All	420.953	<0.001
Male	15.697	0.001	Male	23.497	<0.001
Age	23.894	<0.001	Age	29.499	<0.001
Race: black	55.384	<0.001	Race: black	38.720	<0.001
Race: other	12.124	0.007	Race: other	29.712	<0.001
Education: high school	10.774	0.013	Education: high school	77.552	<0.001
Education: some college	18.855	<0.001	Education: some college	100.895	<0.001
Education: college +	36.434	<0.001	Education: college +	209.017	<0.001
Mobility			**Depression**		
All	154.769	<0.001	All	266.713	<0.001
Male	7.699	0.053	Male	18.228	<0.001
Age	32.655	<0.001	Age	40.198	<0.001
Race: black	32.467	<0.001	Race: black	59.686	<0.001
Race: other	8.784	0.032	Race: other	19.316	<0.001
Education: high school	25.145	<0.001	Education: high school	27.856	<0.001
Education: some college	19.787	<0.001	Education: some college	40.687	<0.001
Education: college +	53.826	<0.001	Education: college +	84.960	<0.001

Note: Wald tests were performed on a generalized ordered probit model of respondents' ratings of vignette characters' health problems.

Table 4: Estimated Coefficients on Respondent Sex, Age, Education, and Race Dummies in Generalized Ordered Probit Model of Respondents' Rating of Vignette Characters' Health

	Cut-point 1		Cut-point 2		Cut-point 3		Cut-point 4	
	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value
Pain								
Male	-0.063	0.101	-0.049	0.082	0.033	0.222	-0.043	0.279
Age	-0.006	0.001	-0.003	0.044	0.002	0.224	0.010	<0.001
Race: black	-0.291	<0.001	-0.116	0.007	-0.338	<0.001	-0.384	<0.001
Race: other	-0.159	0.070	-0.113	0.072	-0.136	0.023	-0.153	0.066
Education: high school	-0.078	0.169	-0.044	0.283	0.095	0.015	0.282	<0.001
Education: some college	0.017	0.774	-0.002	0.957	0.097	0.020	0.261	<0.001
Education: college +	-0.035	0.563	-0.027	0.533	0.093	0.025	0.347	<0.001
Sleep								
Male	0.230	<0.001	0.077	0.012	0.016	0.482	0.066	0.028
Age	0.009	<0.001	0.003	0.016	0.003	0.003	0.010	<0.001
Race: black	0.393	<0.001	0.198	<0.001	-0.005	0.880	-0.169	<0.001
Race: other	0.424	<0.001	0.223	<0.001	0.159	0.002	0.019	0.778
Education: high school	-0.244	<0.001	-0.246	<0.001	-0.122	<0.001	-0.075	0.099
Education: some college	-0.439	<0.001	-0.345	<0.001	-0.175	<0.001	-0.122	0.011
Education: college +	-0.679	<0.001	-0.420	<0.001	-0.258	<0.001	-0.121	0.011
Mobility								
Male	0.089	0.041	-0.016	0.572	-0.002	0.934	-0.043	0.187
Age	0.001	0.757	-0.001	0.409	0.004	0.002	0.010	<0.001
Race: black	0.187	0.003	-0.015	0.727	-0.080	0.035	-0.250	<0.001
Race: other	0.330	<0.001	0.157	0.009	0.096	0.084	0.016	0.825
Education: high school	-0.347	<0.001	-0.129	0.001	-0.100	0.006	0.021	0.668
Education: some college	-0.325	<0.001	-0.126	0.004	-0.083	0.032	0.031	0.548
Education: college +	-0.520	<0.001	-0.130	0.003	-0.097	0.011	0.065	0.206
Memory								
Male	-0.089	0.011	-0.046	0.076	-0.015	0.595	0.017	0.702
Age	-0.009	<0.001	-0.006	<0.001	-0.003	0.009	0.004	0.059
Race: black	-0.154	0.006	-0.190	<0.001	-0.195	<0.001	-0.201	0.001
Race: other	0.085	0.249	-0.107	0.064	-0.065	0.283	0.034	0.740
Education: high school	-0.044	0.405	-0.043	0.256	0.033	0.399	0.137	0.025
Education: some college	0.079	0.153	0.063	0.124	0.074	0.082	0.283	<0.001
Education: college +	0.110	0.044	0.094	0.020	0.066	0.118	0.266	<0.001
Shortness of Breath								
Male	0.208	<0.001	0.065	0.053	-0.022	0.395	-0.012	0.647
Age	0.007	0.001	0.001	0.343	0.002	0.038	0.009	<0.001
Race: black	0.241	<0.001	0.134	0.007	-0.035	0.372	-0.193	<0.001
Race: other	0.383	<0.001	0.231	<0.001	-0.045	0.428	-0.042	0.471
Education: high school	-0.470	<0.001	-0.377	<0.001	-0.085	0.021	0.052	0.186
Education: some college	-0.709	<0.001	-0.448	<0.001	-0.088	0.026	0.007	0.865
Education: college +	-1.094	<0.001	-0.653	<0.001	-0.107	0.007	0.045	0.275
Depression								
Male	0.188	<0.001	0.010	0.726	-0.034	0.201	-0.015	0.636
Age	0.014	<0.001	0.002	0.142	0.007	<0.001	0.010	<0.001
Race: black	0.355	<0.001	-0.005	0.908	-0.143	<0.001	-0.239	<0.001
Race: other	0.406	<0.001	0.085	0.165	-0.028	0.629	-0.012	0.866
Education: high school	-0.346	<0.001	-0.237	<0.001	-0.107	0.005	0.019	0.682
Education: some college	-0.541	<0.001	-0.244	<0.001	-0.104	0.012	-0.011	0.827
Education: college +	-0.879	<0.001	-0.309	<0.001	-0.189	<0.001	-0.036	0.454

Note: Cut-points are the threshold between severity ratings of health problems. Cut-point 1 is the threshold between none and mild, cut-point 2 is the threshold between mild and moderate, cut-point 3 is the threshold between moderate and severe, and cut-point 4 is the threshold between severe and extreme.

Table 5: Estimated Coefficients on Respondent Sex, Age, Education, and Race Dummies, Before and After Adjustment

	Before		After	
Covariate	Coefficient	P-Value	Coefficient	P-Value
Pain				
Male	-0.076	0.061	-0.112	0.015
Age	0.003	0.082	0.001	0.691
Race: black	0.049	0.434	-0.165	0.019
Race: other	-0.035	0.699	-0.159	0.124
Education: high school	-0.255	<0.001	-0.266	<0.001
Education: some college	-0.367	<0.001	-0.332	<0.001
Education: college +	-0.674	<0.001	-0.671	<0.001
Sleep				
Male	-0.187	<0.001	-0.079	0.092
Age	0.004	0.037	0.007	<0.001
Race: black	-0.136	0.024	0.095	0.182
Race: other	-0.015	0.862	0.217	0.033
Education: high school	-0.153	0.007	-0.375	<0.001
Education: some college	-0.193	0.001	-0.528	<0.001
Education: college +	-0.442	<0.001	-0.947	<0.001
Mobility				
Male	0.052	0.300	0.074	0.200
Age	0.024	<0.001	0.025	<0.001
Race: black	0.043	0.582	0.112	0.203
Race: other	-0.023	0.838	0.217	0.090
Education: high school	-0.319	<0.001	-0.526	<0.001
Education: some college	-0.407	<0.001	-0.611	<0.001
Education: college +	-0.887	<0.001	-1.202	<0.001
Memory				
Male	-0.013	0.722	-0.059	0.184
Age	0.014	<0.001	0.007	0.001
Race: black	0.129	0.027	-0.019	0.782
Race: other	0.085	0.314	0.087	0.377
Education: high school	-0.318	<0.001	-0.368	<0.001
Education: some college	-0.448	<0.001	-0.394	<0.001
Education: college +	-0.618	<0.001	-0.544	<0.001
Shortness of Breath				
Male	0.121	0.009	0.253	<0.001
Age	0.012	<0.001	0.017	<0.001
Race: black	-0.032	0.654	0.172	0.051
Race: other	-0.043	0.681	0.276	0.027
Education: high school	-0.241	<0.001	-0.646	<0.001
Education: some college	-0.324	<0.001	-0.847	<0.001
Education: college +	-0.704	<0.001	-1.523	<0.001
Depression				
Male	-0.267	<0.001	-0.154	0.006
Age	-0.008	<0.001	0.000	0.853
Race: black	0.043	0.555	0.209	0.014
Race: other	0.181	0.082	0.362	0.003
Education: high school	-0.495	<0.001	-0.756	<0.001
Education: some college	-0.414	<0.001	-0.761	<0.001
Education: college +	-0.770	<0.001	-1.333	<0.001

Note: Before adjustment means initial ordered probit model of rating of own health problems on the listed covariates. All coefficients have been multiplied by sigma to ensure comparability with HOPIT results. After adjustment means HOPIT model of rating of own health problems on the listed covariates, where individual cutpoints are allowed to vary based on the covariates listed.

Supplementary Materials

Excerpted Disability Vignette Survey questionnaire, version A.
Version B asks the same questions, but the gender of vignette characters is reversed and the vignette questions are presented in a different order. Self-ratings are always asked first.

The possible responses for all questions are the same: none, mild, moderate, severe, extreme.

1. Overall, in the last 30 days, how much pain or bodily aches did you have?

2. In the last 30 days, how much difficulty did you have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?

3. Overall, in the last 30 days, how much of a problem did you have with moving around?

4. Overall, in the last 30 days, how much difficulty did you have with concentrating or remembering things?

5. In the last 30 days, how much of a problem did you have because of shortness of breath?

6. Overall, in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?

…

8. Paul has a headache once a month that is relieved after taking a pill. During the headache he can carry on with his day-to-day affairs. Overall, in the last 30 days, how much of a problem did Paul have with bodily aches or pains?

9. Maria takes about two hours every night to fall asleep. She wakes up once or twice a night feeling panicked and takes more than one hour to fall asleep again. In the last 30 days, how much difficulty did Maria have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?

10. Henry has pain that radiates down his right arm and wrist during his day at work. This is slightly relieved in the evenings when he is no longer working on his computer. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?

11. Karen wakes up almost once every hour during the night. When he [sic] wakes up in the night, it takes around 15 minutes for her to go back to sleep. In the morning she does not feel well-rested. In the last 30 days, how much difficulty did Karen have with sleeping such as falling asleep, waking up frequently during the night or waking up too early in the morning?

12. Charles has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall, in the last 30 days, how much of bodily aches or pains did Charles have?

13. Alice falls asleep easily at night, but two nights a week she wakes up in the middle of the night and cannot go back to sleep for the rest of the night. In the last 30 days, how much difficulty did

Alice have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?

14. Tom has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy. Overall, in the last 30 days, how much of a problem did Tom have with moving around?

15. Lisa can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within five minutes. Overall, in the last 30 days, how much difficulty did Lisa have with concentrating or remembering things?

16. Kevin does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry the groceries and do some light household work. Overall, in the last 30 days, how much of a problem did Kevin have with moving around?

17. Sue is keen to learn new recipes but finds that she often makes mistakes and has to reread several times before she is able to do them properly. Overall, in the last 30 days, how much difficulty did Sue have with concentrating and remembering things?

18. Rob is able to walk distances of up to 200 metres without any problems but feels tired after walking one kilometre or climbing more than one flight of stairs. He has no problems with day-to-day activities, such as carrying food from the market. Overall, in the last 30 days, how much of a problem did Rob have with moving around?

19. Eve cannot concentrate for more than 15 minutes and has difficulty paying attention to what is being said to her. Whenever she starts a task, she never manages to finish it and often forgets what she was doing. She is able to learn the names of people she meets. Overall, in the last 30 days, how much difficulty did Eve have with concentrating or remembering things?

20. Mark has no problems with walking slowly. He gets out of breath easily when climbing uphill for 20 meters or a flight of stairs. In the last 30 days, how much of a problem did Mark have because of shortness of breath?

21. Anna feels depressed most of the time. She weeps frequently and feels hopeless about the future. She feels that she has become a burden on others and that she would be better dead. Overall, in the last 30 days, how much of a problem did Anna have with feeling sad, low or depressed?

22. Paul suffers from respiratory infections about once every year. He is short of breath 3 or 4 times a week and had to be admitted in hospital twice in the past month with a bad cough that required treatment with antibiotics. In the last 30 days, how much of a problem did Paul have because of shortness of breath?

23. Maria feels nervous and anxious. She worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests her. When she is alone she tends to feel useless and empty. Overall, in the last 30 days, how much of a problem did Maria have with feeling sad, low or depressed?

24. Henry has been a heavy smoker for 30 years and wakes up with a cough every morning. He gets short of breath even while resting and does not leave the house anymore. He often needs to be put on oxygen. In the last 30 days, how much of a problem did Henry have because of shortness of breath?

25. Karen enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities. Overall, in the last 30 days, how much of a problem did Karen have with feeling sad, low or depressed?

…

Table 6S: Estimated Coefficients on Respondent Sex, Age, Education, and Race Dummies, Before and After Adjustment, Stratified by Sex

		Male				Female			
		Before		After		Before		After	
	Covariate	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value
Pain									
	Age	0.002	0.528	0.004	0.254	0.004	0.081	-0.001	0.764
	Race: black	-0.126	0.238	-0.428	<0.001	0.142	0.068	-0.039	0.655
	Race: other	-0.019	0.897	-0.086	0.594	-0.048	0.685	-0.213	0.112
	Education: high school	-0.146	0.125	-0.156	0.145	-0.336	<0.001	-0.335	<0.001
	Education: some college	-0.158	0.124	-0.138	0.229	-0.509	<0.001	-0.454	<0.001
	Education: college +	-0.559	<0.001	-0.683	<0.001	-0.759	<0.001	-0.650	<0.001
Sleep									
	Age	0.006	0.024	0.013	<0.001	0.002	0.476	0.004	0.124
	Race: black	-0.197	0.043	-0.028	0.816	-0.098	0.200	0.163	0.070
	Race: other	-0.030	0.815	0.015	0.924	-0.001	0.994	0.353	0.009
	Education: high school	-0.219	0.011	-0.389	<0.001	-0.112	0.137	-0.365	<0.001
	Education: some college	-0.183	0.046	-0.394	<0.001	-0.198	0.015	-0.618	<0.001
	Education: college +	-0.316	<0.001	-0.867	<0.001	-0.564	<0.001	-1.011	<0.001
Mobility									
	Age	0.023	<0.001	0.029	<0.001	0.026	<0.001	0.023	<0.001
	Race: black	-0.187	0.136	-0.144	0.318	0.167	0.095	0.247	0.028
	Race: other	0.241	0.138	0.329	0.074	-0.249	0.120	0.096	0.589
	Education: high school	-0.164	0.126	-0.354	0.004	-0.441	<0.001	-0.638	<0.001
	Education: some college	-0.200	0.083	-0.483	<0.001	-0.560	<0.001	-0.696	<0.001
	Education: college +	-0.737	<0.001	-1.147	<0.001	-1.008	<0.001	-1.241	<0.001
Memory									
	Age	0.022	<0.001	0.019	<0.001	0.010	<0.001	0.000	0.849
	Race: black	-0.043	0.674	-0.195	0.107	0.218	0.002	0.068	0.412
	Race: other	0.054	0.691	-0.028	0.858	0.100	0.354	0.158	0.208
	Education: high school	-0.179	0.049	-0.254	0.015	-0.397	<0.001	-0.417	<0.001
	Education: some college	-0.275	0.005	-0.287	0.011	-0.551	<0.001	-0.446	<0.001
	Education: college +	-0.559	<0.001	-0.595	<0.001	-0.653	<0.001	-0.490	<0.001
Shortness of Breath									
	Age	0.015	<0.001	0.028	<0.001	0.010	<0.001	0.011	0.004
	Race: black	-0.282	0.017	-0.160	0.276	0.107	0.247	0.357	0.001
	Race: other	0.008	0.958	0.270	0.137	-0.072	0.624	0.281	0.103
	Education: high school	-0.298	0.002	-0.603	<0.001	-0.204	0.022	-0.678	<0.001
	Education: some college	-0.205	0.046	-0.629	<0.001	-0.411	<0.001	-1.013	<0.001
	Education: college +	-0.706	<0.001	-1.613	<0.001	-0.701	<0.001	-1.435	<0.001
Depression									
	Age	-0.010	0.004	0.004	0.354	-0.007	0.020	-0.004	0.285
	Race: black	-0.155	0.220	0.033	0.822	0.146	0.111	0.298	0.005
	Race: other	0.237	0.137	0.377	0.044	0.143	0.302	0.359	0.023
	Education: high school	-0.499	<0.001	-0.744	<0.001	-0.517	<0.001	-0.771	<0.001
	Education: some college	-0.266	0.022	-0.570	<0.001	-0.522	<0.001	-0.886	<0.001
	Education: college +	-0.627	<0.001	-1.363	<0.001	-0.900	<0.001	-1.311	<0.001

Note: Before adjustment means initial ordered probit model of rating of own health problems on the listed covariates. All coefficients have been multiplied by sigma to ensure comparability with HOPIT results. After adjustment means HOPIT model of rating of own health problems on the listed covariates, where individual cut-points are allowed to vary based on the covariates listed.

Table 7S: Estimated Coefficients on Respondent Sex, Age, Education, and Race Dummies, Before and After Adjustment, Trichotomous Self-Reported Health Categories

	Before		After	
Covariate	Coefficient	P-Value	Coefficient	P-Value
Pain				
Male	-0.055	0.243	-0.080	0.129
Age	0.006	0.010	0.004	0.080
Race: black	0.123	0.088	-0.058	0.469
Race: other	0.060	0.565	-0.055	0.640
Education: high school	-0.327	<0.001	-0.335	<0.001
Education: some college	-0.439	<0.001	-0.414	<0.001
Education: college +	-0.851	<0.001	-0.851	<0.001
Sleep				
Male	-0.168	<0.001	-0.116	0.026
Age	0.003	0.104	0.006	0.010
Race: black	-0.091	0.197	0.040	0.618
Race: other	-0.050	0.623	0.149	0.196
Education: high school	-0.173	0.008	-0.380	<0.001
Education: some college	-0.231	0.001	-0.523	<0.001
Education: college +	-0.572	<0.001	-0.947	<0.001
Mobility				
Male	0.027	0.662	0.015	0.823
Age	0.025	<0.001	0.025	<0.001
Race: black	0.097	0.298	0.072	0.471
Race: other	0.050	0.716	0.201	0.171
Education: high school	-0.403	<0.001	-0.524	<0.001
Education: some college	-0.502	<0.001	-0.622	<0.001
Education: college +	-1.008	<0.001	-1.143	<0.001
Memory				
Male	0.088	0.078	0.047	0.393
Age	0.013	<0.001	0.008	0.003
Race: black	0.246	<0.001	0.066	0.422
Race: other	0.207	0.052	0.126	0.290
Education: high school	-0.372	<0.001	-0.409	<0.001
Education: some college	-0.528	<0.001	-0.473	<0.001
Education: college +	-0.773	<0.001	-0.695	<0.001
Shortness of Breath				
Male	0.117	0.104	0.173	0.026
Age	0.014	<0.001	0.016	<0.001
Race: black	-0.004	0.971	0.103	0.394
Race: other	0.112	0.472	0.277	0.099
Education: high school	-0.369	<0.001	-0.677	<0.001
Education: some college	-0.453	<0.001	-0.809	<0.001
Education: college +	-0.963	<0.001	-1.506	<0.001
Depression				
Male	-0.203	<0.001	-0.195	0.003
Age	-0.007	0.006	-0.005	0.104
Race: black	0.126	0.144	0.097	0.304
Race: other	0.038	0.767	0.103	0.455
Education: high school	-0.542	<0.001	-0.755	<0.001
Education: some college	-0.490	<0.001	-0.702	<0.001
Education: college +	-0.908	<0.001	-1.204	<0.001

Note: Before adjustment means initial ordered probit model of rating of own health problems on the listed covariates. All coefficients have been multiplied by sigma to ensure comparability with HOPIT results. After adjustment means HOPIT model of rating of own health problems on the listed covariates, where individual cut-points are allowed to vary based on the covariates listed.