WPS3698

Searching for the Economic Gradient in Self-Assessed Health

Michael Lokshin and Martin Ravallion[1]

Development Research Group, World Bank
1818 H Street NW, Washington DC, 20433 USA

Abstract: Can self-assessments of health reveal the true health differentials between "rich" and "poor"? The potential sources of bias include psychological adaptation to ill-health, socioeconomic covariates of health reporting errors and income measurement errors. We propose an estimation method to reduce the bias, by isolating the component of self-assessed health that is explicable in terms of objective health indicators and allowing for broader dimensions of economic welfare than captured by current incomes. On applying our method to survey data for Russia we find a pronounced (nonlinear) economic gradient in health status that is not evident in the raw data. This is largely attributable to the health effects of age, education and location.

JEL: D63, I12, I31

Keywords: Self-assessed health, subjective welfare, gradient, measurement errors, Russia

Correspondence: M. Ravallion, World Bank, 1818 H Street NW, Washington DC, 20433 USA.
mravallion@worldbank.org

World Bank Policy Research Working Paper 3698, September 2005; Revised July 2006

[1] The authors are grateful to Mingshan Lu, Vassilis Hajivassilou and participants at the Fifth World Congress of the International Health Economics Association for their comments on an earlier version of this paper.

1. Introduction

The economic gradient in health status — the extent to which poorer people tend to be less healthy — has been of long-standing interest in health policy discussions and the social sciences more generally.[2] An important strand of the literature has used data on self-assessed health (*SAH*) to study the socioeconomic covariates of health status, including its economic gradient. Respondents assess their overall health on (typically) a five-point scale from "very good" to "very bad" and the responses are combined with data on incomes or expenditures for the same individuals to estimate the gradient, possibly with controls. Using *SAH* data for the US, Smith (1999) found income and wealth gradients, though dampened at high incomes. Also using data for the US, Case et al. (2002) found an income gradient in children's *SAH* (as rated by a parent). Currie et al. (2004) also report an economic gradient in adult assessments of children's *SAH* in England, although the gradient appears to be smaller than in the US. Kunst et al., (1995) report economic gradients in *SAH* for eight developed countries.

How reliable are *SAH* data? A number of studies have reported that people with specific ailments tend to report lower *SAH* than people in good health. Idler and Benyamini (1997) review research indicating that answers to the *SAH* question help predict subsequent mortality.[3] On the basis of such evidence, Franzini et al., (2005, 789) claim that: "Twenty years of empirical evidence indicates that *SAH* is a powerful and reliable predictor of clinical outcomes and mortality." Others have taken a more critical view of *SAH* data, arguing that responses on *SAH* are influenced by various emotional, psychological and knowledge-dependent factors. Amartya Sen's critique of self-reported morbidity data has been influential:

---

[2] Using micro data, economic gradients in health status have been reported by Marmot et al. (1991), Pappas et al., (1993), Sorlie et al. (1995), Mackenbach et al. (1997), Adams et al. (2003), Deaton (2003), Finch (2004). An economic gradient in health indicators is also found across countries; see, for example, Preston (1980), Anand and Ravallion (1993), Pritchett and Summers (1996) and Deaton (2003).

[3] Also see Appels et al., (1996) and Groot (2000).

> "People's perception of illness varies with what they are used to, and also with their medical knowledge. In places where medical care is widespread and good, people often have <u>higher</u> perception of morbidity, even though they may be in much better general health. …In contrast, a population that has little experience of medical care, and which has widespread health problems as a standard condition of existence, can have very low perception of being medically ill." (Sen, 1998, p.18)

A number of studies have emphasized the potential for reporting errors (Butler, et al., 1987; Kerkhofs and Lindboom, 1995; Groot, 2000). The work of Butler et al. (1987) should warn users of *SAH* data against assuming that the measurement errors (relative to objective health status) are uncorrelated with the socioeconomic characteristics of respondents. Groot's (2000) results are suggestive of psychological adaptation to declining health. Adaptation has also been used to explain the gap found between the *SAH* of sick people and those who are asked to rate their health as if they were sick (Menzel et al., 2002). Recognizing the potential for reporting biases, Thomas and Frankenberg (2000) recommend that surveys should combine questions on *SAH* with more objective indicators of health status.

What does all this mean for estimates of the economic gradient based on *SAH* data? Even if we accept that *SAH* contains health-relevant information, adaptive behaviors to poor health and correlated differences in economic circumstances and health knowledge can generate <u>non-ignorable measurement errors</u>, by which we mean errors that are correlated with income or other socioeconomic controls such as education. For example, poorer people may tend to think they are healthier than they really are, such that standard methods of regressing *SAH* on income will underestimate the true economic gradient in health.

There are also longstanding concerns about how well current incomes reveal the true gradient in health. Even using high quality surveys (in which an income aggregate is built up from many detailed questions), current incomes are unlikely to reflect well either past or expected future incomes, which are likely to matter to current health. Expenditure on current consumption may do a better job in this respect, but will still be an imperfect welfare indicator given that inter-temporal markets do not work perfectly. There are also uncertainties about how best to normalize for heterogeneity in consumption needs, such as stemming from demographic differences between households (Pollak, 1991).[4] And even if we had an ideal measure of average economic welfare for a household, health status is an individual attribute, which will depend directly on personal economic welfare. Individual income is a flawed metric of individual command over commodities given some degree of income pooling within households. And household expenditure may well be equally flawed as an indicator of individual economic welfare given that pooling is almost certainly incomplete.[5]

In short, heterogeneity in relevant individual and household circumstances, inter-temporal consumption smoothing and inter-personal income sharing may entail that neither measured current income nor consumption are particularly good proxies for economic welfare, as relevant to health. This too will lead one to underestimate the true economic gradient in health using the methods commonly found in the literature.

This paper proposes a method of estimating the economic gradient in health status that is likely to be more robust to non-ignorable errors in *SAH* data and the measurement errors in data on incomes or expenditures. The essential idea is to use largely independent covariates of both *SAH* and perceived economic welfare to attempt to purge the raw data of the non-ignorable errors, prior to testing for the economic gradient in health status. To calibrate a broader measure of economic welfare we draw on subjective data, though recognizing that this too contains measurement errors when used as an indicator of objective economic welfare. Our estimation

---

4 For example, the poverty lines used as deflators may not correctly weight differences in household size or demographic composition.

5 It is instructive that, in the same setting, Ravallion and Loskhin (2001) find evidence that many "non-income" factors at the individual and household levels impinge on perceived economic welfare in Russia at given current incomes or expenditures on consumption deflated by standard poverty lines.

method essentially generalizes the idea of an instrumental variables estimator to a situation in which the dependent variable contains non-ignorable errors.

We illustrate the method using high-quality socioeconomic survey data for Russia in 2002. We first show that applying standard methods from past research to these data suggests little sign of an economic gradient in the self-assessed health of Russian adults. This might not be too surprising if one was talking about Russia some 20 years earlier, at which time it might be argued that a reasonably well-funded universal health care system meant that the quality of health care varied little with income. However, the lack of a gradient in data for Russia around 2000 is surprising. Worsening public health services, increasing reliance on the market for health care, and rising poverty and inequality, would lead one to expect an emerging economic gradient in health status. Psychological adaptation to the stresses of the economic transition and health crisis could well entail that the structure of errors in *SAH* data for Russia hides the true gradient. One can expect that it will be the poor who are least able to protect their health care during Russia's health crisis, and they may well adapt to the situation by downplaying the health risks they are facing, at least for some time. Then the true economic gradient will be attenuated in the raw data on *SAH*. It may well be difficult to identify the gradient in this setting by standard empirical methods, casting doubt on numerous claims found in the literature.[6]

Applying our preferred estimation method, we show that much stronger signs of the gradient emerge from a more nuanced empirical model — emphasizing non-income factors relevant to perceived economic welfare at the micro level, and allowing for heterogeneity in *SAH* data. This model reveals a marked tendency for the poorest to be less healthy, with dampened health gains as economic welfare rises. In short, while there is little or no income effect evident

[6] For example, Gavrilova et al. (2002) and Brainerd and Cutler (2005) argue that the "psychosocial stresses" of the transition in Russia are a more important cause of the health crisis than poverty; however, the same (unobserved) stresses may well make it hard to identify the importance of poverty.

in the raw data on *SAH*, we find that the deterministic component of *SAH* has a marked gradient with respect to a broader concept of economic welfare in which the role of "non-income" factors is identified using self-assessments of economic welfare. We also show that this (unconditional) gradient is largely attributable to other covariates of health, notably age and education; our preferred estimation method reveals only a small conditional gradient in health.

2. Biases and methods

We want to see how health status (*H*) varies with economic welfare (*W*) but we do not observe *H* and *W* directly. We must rely on proxies for each variable, entailing measurement errors. In the literature (referred to in the introduction), the proxy for *H* is self-assessed health (*SAH*) while for *W* it is "income" (*Y*). The measurement errors may well be uncorrelated with *H* and *W*, although we need not assume that this is so; in particular, we can readily allow the measurement error in *Y* to be positively correlated with *W*, as would happen if there is inter-temporal variability of incomes that can be at least partly smoothed through savings or borrowing to make economic welfare less variable. We also allow the possibility that the measurement error in *SAH* is correlated with *Y*.

There will be two sources of bias in estimating the economic gradient of health by regressing *SAH* on *Y* (with or without controls). The first is the well-known attenuation bias that arises when a regressor is measured with error. The existence of a positive correlation between the measurement error and actual economic welfare, due to inter-temporal smoothing behavior, will strengthen the standard attenuation bias due to classical measurement errors in a regressor.

The second source of bias is intrinsic to the use of subjective data. It stems from the fact that the errors in using self-assessed health to indicate actual health status are unlikely to be statistically ignorable ("white noise"), but rather will be correlated with incomes and other observables (Butler et al., 1987). The direction of this second bias depends on the sign of the correlation between the error in *SAH* (as an indicator of true health status) and *Y*. We would conjecture that poorer people (in terms of *Y*) tend to over-state their health, as a form of psychological adaptation to their relative ill-health. If that is the case then the two biases work in the same direction, leading to an underestimation of the true gradient in health status.

There are two standard approaches to reducing attenuation bias. The first is to find a better measure (with lower error variance) for the regressor and the second is to find an instrumental variable (IV) that is correlated with the miss-measured regressor but uncorrelated with the error term in the main regression. (The IV can be measured with errors, but these must be uncorrelated with the errors in the regressor.) We use both approaches. Our alternative indicator of *W* draws on respondents' perceptions of their own economic welfare, denoted by *SAW* (for "self-assessed welfare"). This appears likely to better reflect the personal, health-relevant, aspects of economic welfare that are missing from current incomes.

Some bias will undoubtedly remain. Latent individual personality traits may jointly influencing *SAW* and *SAH* — personality traits that are of little or no relevance to the true economic gradient of health. For example, a happy disposition may lead simultaneously to higher perceived economic wellbeing and a more positive health assessment at any given level of health. To address these concerns, we shall also try using objective welfare-covariates of *SAW* as instrumental variables, under the assumption that the covariates are uncorrelated with the errors implied by using *SAW* as a measure of economic welfare. (In a linear model, identifying the gradient conditional on control variables will also call for exclusion restrictions.)

What about the second bias, arising from a correlation between the measurement errors in *SAH* (as an indicator of *H*) and economic welfare? The IV estimation method will also help reduce this bias if it is valid to assume that the IVs are uncorrelated with the errors in *SAH*. However, that seems a difficult assumption to defend. Why would the errors generated by using *SAH* as a measure a health status be only correlated with the component of *SAW* that cannot be attributed to the data on observable covariates that one happens to have available for use as IVs? For example, income *Y* might be considered a good IV for *SAW* in identifying the unconditional economic gradient except for the fact that (as we have already argued) the measurement error in *SAH* could well be correlated with *Y*, given income-dependent psychological biases in health assessments. To address this concern about correlations between the error in *SAH* and the covariates of *SAW* we use as the dependent variable <u>predicted</u> *SAH* based on health covariates (physical abilities and medical histories). This works if the errors in proxying actual health by predicted *SAH* (based on objective health covariates) are uncorrelated with the IVs for *SAW*.

To see more clearly what is involved, it is convenient for expository purposes to confine attention to a linear model:

$$H_i = \beta W_i + \gamma Z_i + \varepsilon_i \quad (i=1,..,n) \tag{1}$$

for individual *i*, Z_i is a vector of control variables and ε_i is a zero-mean innovation error term, with $E(\varepsilon \mid W, Z) = 0$. Self-assessed health is $SAH_i = H_i + \varepsilon_i^H$ and self-assessed economic welfare is $SAW_i = W_i + \varepsilon_i^W$ where $E(\varepsilon^k \mid k) = 0$ for *k*=*H,W*. Thus (1) can be written in terms of the observables as:

$$SAH_i = \beta SAW_i + \gamma Z_i + \varepsilon_i + \varepsilon_i^H - \beta \varepsilon_i^W \tag{2}$$

The two sources of bias discussed above are the attenuation bias stemming from the fact that $Cov(\varepsilon^W, SAW) = Var(\varepsilon^W) \neq 0$ and that $Cov(\varepsilon^H, SAW) \neq 0$. In addressing both sources of biase we assume that data are available on vectors of covariates, X_i^H and X_i^W for *H* and *W* respectively such that we can write:

$$H_i = \pi^H X_i^H + \nu_i^H \tag{3.1}$$

$$W_i = \pi^W X_i^W + \nu_i^W \tag{3.2}$$

where $E(\varepsilon^k \mid X^k) = E(\nu^k \mid X^k) = 0$ for *k=H,W* but $Cov(\varepsilon^H, X^W) \neq 0$ and $Cov(\varepsilon^H, \nu^W) \neq 0$. We then have:

$$SAH_i = \beta(\pi^W X_i^W) + \gamma Z_i + \varepsilon_i + \varepsilon_i^H + \beta \nu_i^W \tag{4}$$

As usual, π^W can be replaced by consistent estimates, $\hat{\pi}^W$ from the first-stage regression:

$$SAW_i = \pi^W X_i^W + \nu_i^W + \varepsilon_i^W \tag{5}$$

Identification requires that there is at least one variable in X_i^W that is not in Z_i.

This 2SLS method eliminates the attenuation bias (since ε_i^W no longer appears in the composite error term), but the second source of bias remains (given that ε_i^H still appears there). To address this bias, we simply subtract ε_i^H from both sides of (4) so that the dependent variables becomes the deterministic component of *SAH*, i.e., we re-write (4) as:

$$\pi^H X_i^H = \beta(\pi^W X_i^W) + \gamma Z_i + \varepsilon_i + \beta \nu_i^W - \nu_i^H \tag{6}$$

This is estimated by regressing $\hat{\pi}^H X_i^H$ on $\hat{\pi}^W X_i^W$ and Z_i.

The above formulation makes a number of simplifying assumptions that we will relax in our empirical work. We shall allow for nonlinearity in the economic gradient of health using both non-parametric and more conventional nonlinear parametric methods. For the non-parametric regressions we use locally smoothed scatter plots (Fan, 1992), as programmed in STATA. To allow for control variables we also use a Partial Linear Model (PLM). This treats the *SAH* as a smooth but unknown function of real expenditure with the control variables entering linearly in parameters.[7] We use the STATA program for estimating the PLM written by Lokshin (2006). While this method allows a relatively flexible representation of the income effect, it has the disadvantage that it ignores the fact that our dependent variables come in the form of ordinal qualitative variables for which it cannot be assumed that the difference between (say) "very good" and "good" is the same in terms of health status as the difference between "good" and "average". To address this concern, we follow past work in the literature in interpreting the answers to the subjective health question as an ordinal, categorical, summary of unobserved true health status, which is assumed to be a continuous variable that is determined by a vector of specific health attributes and an i.i.d. error term. In modeling the income effect on health we include a cubic function of real expenditure. Assuming level comparability of *SAH* (conditional on covariates) across persons and that the error term is normally distributed we can use an Ordered Probit (OP) to model the responses. We estimate similar models for *SAW*, where the latent continuous variable is interpreted as individual utility. We will allow for level non-comparability of subjective scales by adding controls for likely socio-economic covariates of reporting bias, following Thomas and Frankenberg (2000). We also allow for a latent individual effect using panel data, though this requires modifications to the econometric specification, for which we use the method of Ravallion and Lokshin (2001).

Our use of objective indicators of economic welfare, based on household expenditures and personal incomes, as regressors raises concerns about their possible endogeneity. The most likely way this could happen appears to be through work effort, based on the argument that latent

[7] The parameters on the control variables are estimated by first ordering the data in terms of real expenditures and then taking differences to eliminate the nonparametric sub-function; then one runs a nonparametric regression of the *SAH* scores net of the controls on real expenditure. (We used Fan regressions for the nonparametric function.) For a review of PLMs and related topics see Yatchew (1998).

determinants of *SAH* influence work effort and hence personal incomes. To help alleviate these concerns we estimate regressions in which personal income is dropped. We also test sensitivity to dropping the labor force activity variables, given the endogeneity concerns.

3. Setting and data

Russia experienced a severe deterioration in overall health status in the 1990s — dubbed a "health crisis" by Field (1995). Male life expectancy fell from 65 years in 1988 to 59 years in 2001, while for women it fell from 74 years to 72 years (Heleniak, 2002).[8] This was in a period of overall economic contraction and rising inequality (Lokshin and Popkin, 1999), the latter stemming in large part from the diverse labor market outcomes of the period (Milanovic, 1999). There is also evidence of considerable "churning" in living standards, with both gainers and losers at any given level of living (Lokshin and Ravallion, 2000). There are many ways these economic changes could impinge on health. Deteriorating public health services, combined with reduced purchasing power over private health inputs, can be expected to take a toll on health status, and probably more so for the poor. Other channels include the psychosocial stresses of the transition from a time (prior to 1989) in which there was relatively little economic insecurity at the individual level to a period of much uncertainty. New opportunities emerged, but so did new stresses on individuals, who naturally had different abilities to respond to those opportunities (Gavrilova et al., 2002; Brainerd and Cutler, 2005).[9] Using a nationally representative survey for Russia in 1995, Cockerham (2000) found no statistically significant

[8] Field (1995) reports a number of other signs of worsening health in the early 1990s. Further evidence and (sometimes differing) interpretations of Russia's "health crisis" can be found in Ellman (1994), DaVanzo and Adamson (1997) and Brainerd and Cutler (2005).

[9] This has also been a theme of the literature on socio-economic inequalities in health; Wilkinson (1996) argues that inequality generates health-impairing stresses. Also see the critical assessments of this view in Elstad (1998), Lynch et al., (2000) and Deaton (2003).

income gradient in the *SAH* of adults, although he did find that *SAH* tends to be higher for men, to fall with age, to rise with education and to be higher for the employed.

We use data from two rounds, 2002 and 2000, of the Russian Longitudinal Monitoring Survey (RLMS).[10] This is a comprehensive socio-economic survey for a nationally representative sample. The 2002 round of the RLMS included an unusually detailed module on health. The sampled households included 9,100 adults (3,900 men and 5,200 women). For aspects of the analysis we also exploit the less complete data available for the 6,000 adults (2,600 men and 3,600 women) in the 2000-02 panel that can be formed from the RLMS. The RLMS sample is re-designed at each survey round to assure that it remains representative.

The survey included the following *SAH* question, asked of each adult in the sampled household: "*How would you evaluate your health? Very good, good, average (not good, not bad), bad, very bad, don't know/refuse.*" Figure 1 gives distributions of the answers for 2002. We find that 16.5% of adults rated their health as "very bad" or "bad;" this was true of 20.0% of women and 11.7% for men. Men rate their overall health higher than women. However, this is not reflected in some other health indicators. Women tend to live much longer: in 2001, male life expectancy was 59 years, versus 72 for women (Heleniak, 2002). And Russian men tend to have worse "health lifestyles" (higher alcohol consumption, smoking more, fattier diets).[11] The gender difference in *SAH* appears to reflect a psychological difference.

The 2002 survey included questions on functional abilities, specific ailments and recent disease histories. The questions covered the level of difficulty incurred in performing various physical activities (running, walking, lifting, crouching etc), weight, height, reported ailments (problems in chest, lungs, liver and so on) and diseases diagnosed (diabetic, heart attack,

---

[10] The RLMS data are obtainable from: http://www.cpc.unc.edu/projects/rlms/rlms home.html.

[11] See Cockerham (2000). Also see the gender differences in vital statistics for Russia reported DaVanzo and Adamson (1997) and Brainerd and Cutler (2005).

hypertension, stroke etc). (The questions on functional ability were only asked of pensioners in 2000, although all adults were asked these questions in 2002.)

We use both objective and subjective measures of economic welfare. The RLMS includes a wide range of socioeconomic characteristics including consumption expenditures and incomes (which included imputed values for consumption and income in kind). Our main objective welfare indicator is total household expenditure normalized by date and region-specific poverty lines, interpreted as cost-of-living deflators. We use established poverty lines for Russia. We also use data on incomes which have the advantage that they are individual-specific.

However, we recognize that these widely used income-based metrics give an incomplete picture of economic welfare, as discussed in section 2. We follow a strand of the literature in economics that has employed subjective assessments of economic welfare as extra identifying information.[12] The RLMS includes the following question on self-assessed (economic) welfare (*SAW*) asked of each adult: *"Please imagine a 9-step ladder where on the bottom, the first step, stand the poorest people, and on the highest step, the ninth, stand the rich. On which step are you today?"* Being individual-specific, the answers to this question arguably give more complete information about the underlying economic welfare of that individual; *SAW* will naturally reflect both household resources and individual command over those resources. However, we must also recognize that answers to this question are likely to be influenced by many factors that one might not want to include in a measure of economic welfare; personality traits, for example, can influence subjective welfare but not be deemed relevant to assessing whether a person is "poor" or not. We thus use the predicted values of *SAW* based on relevant covariates as the preferred welfare metric (section 2). The data on *SAW* establish the weights on various multiple objective dimensions of economic wellbeing.

[12] Kapteyn (1994) provides an overview of this approach and antecedents in the literature.

4. Descriptive results

Figure 2 gives the non-parametric regression of *SAH* against real expenditure in 2002.[13] We find only a weak relationship between *SAH* and household expenditures. Figure 2 also includes the nonparametric regression when we use the PLM to control for individual and household characteristics (age, age-squared, education attainments, household size and demographic composition and regional dummy variables). On adding the controls, the gradient is roughly halved. A steeper gradient is found for subjective economic welfare (panel (b) in Figure 2). Subjective economic welfare rises as household real expenditure rises, with a marked nonlinearity evident at the lower end of the expenditure distribution; mean subjective economic welfare (conditional on expenditure) falls sharply for the poorest.

Table 1 compares subjective health with subjective economic welfare in 2002. There is a positive association, with a Cramer's V statistic of 0.12, but it is clearly small.[14] Indeed, for the whole sample, the most common *SAH* response ("average, neither good nor bad") is invariant to *SAW*, although for men the conditional mode switches to "good" at high *SAW*. The second most common answer to the *SAH* question switches from "bad" to "good" as one goes from the second to the third rung of the welfare ladder. (For women the switch occurs between the third and fourth rungs, while for men it is between the first and second.) Amongst the 18% of Russian adults who placed themselves on two lowest rungs of the *SAW* ladder, 29% rated their health as "bad" or "very bad." By contrast, only 11% of those on rungs six or higher of the *SAW* ladder (the upper 9%) rated their health as bad or very bad. However, the off-diagonal density in the

---

13 The contrast with similar data for the U.S. is striking. Using a very similar five-rung ladder for self-assessed health — where 5="excellent" through to 1="poor" — results reported in Case et al., (2002) imply that the mean rating for middle-aged U.S. adults rises from 2.5 for the poorest in terms of household income to about 4 for the richest. The corresponding gradient for Russians in Figure 2 is from 3.1 to about 3.4. Mean *SAH* is roughly "average" independently of economic welfare.

14 Cramer's V measures the correlation between two (nominal) categorical variables; the statistic varies from 0 to 1, and the higher its value the stronger the association between the two variables.

data is sizeable. A majority of those who think they have poor health do not rate themselves as poor. Amongst the 17% of adults who rated their health as “very bad” or “bad,” only 33% placed themselves on the lowest two rungs of the economic welfare ladder. And of the 27% of adults who rated their health as “good” or “very good,” 60% put themselves on the lower four rungs of the welfare ladder.

Table 2 exploits the panel structure of the RLMS to compare the changes over time in *SAH* with changes in *SAW*. This reduces the impact of any time-invariant individual effects (such as due to latent ill-health or personality traits) that jointly influence subjective health and welfare. (Although it can also reduce the signal-to-noise ratio in the presence of time-varying measurement errors or extraneous transient effects.) We find that 18% of adults in the panel thought that their health had improved, while 20% thought it had worsened. There is little difference by gender, though a slightly higher proportion of men reported that their health had worsened than did women. The panel data suggest an even lower statistical association between the two categorical variables, with an overall Cramer’s V of 0.05. Of the 32% of the panel sample who felt that they had moved down the welfare ladder, 21% felt that their health had declined (by one or more categories in the *SAH* question). By contrast, this was true of 17% of the 38% who felt that they had risen on the *SAW* ladder. Changes in perceived health are almost orthogonal to changes in perceived economic welfare. This holds for both men and women.

However, we find stronger signs of the gradient for a number of the more specific health indicators in the data, as can be seen in Figure 3. Most of the physical activity variables tend to show a deterioration at low incomes (though the causal interpretation is of course unclear). A number of the reported ailments (gastro-intestinal disorders, kidney problems, hypertension) tend to have higher incidence at middle expenditure levels. Adjusting for the other covariates has more effect in some cases than others, and the controls do change the regression function on expenditure; in a number of cases, the controls attenuate the gains at high expenditure levels, even producing an inverted-U relationship.

How well does *SAH* reflect objective health circumstances? We give in Table 3 the cross-sectional OP for 2002 in which we regress *SAH* on all measures of specific abilities and ailments reported in the survey. We included all the variables in the RLMS 2002 round that we felt could reasonably be treated as exogeneous to overall health status. These assumptions can always be questioned, and we cannot rule out the possibility that errors in the self-reporting of even truly exogeneous predictor can introduce a correlation with the error term. We decided not to include two variables that are likely to be correlated with *SAH*, namely whether the respondent is a smoker and alcohol consumption, on the grounds that the concern about endogeneity was too severe in these cases.

For both men and women, roughly half the variance in *SAH* is explicable in terms of age and the specific health attributes identified in the data set; for women a slightly higher share of the variance is explained by these variables. The signs are generally what one would expect. *SAH* tends to fall for both men and women over the bulk of the data.[15] Amongst the physical activity variables, the strongest predictors of *SAH* are ability to run, climb stairs, crouch and eat. The effects of the physical activity variables are similar for men and women, though with generally larger coefficients for men. Body-mass index does not have a strong effect, though it does exhibit the expected inverted U shape. Most of the reported ailments and diseases have the expected signs. Particularly strong and significant effects are evident for lung problems, liver

---

[15] The turning point for men is at age 50, while for women it is over 100 years. The relatively low turning point for men (implying that men over 50 start to perceive that their health as improving with age) could well stem from selective mortality (recalling that male life expectancy is about 60, more than 10 years lower than for women). Adaptive behavior amongst older men may also be playing a role, along the lines of Groot (2000).

problems for men, chronic gastro-intestinal (CGI) problems (particularly for men), chest pains, wearing glasses (for men) and "other problems" (men and women). Amongst the diseases recorded, the strongest predictors of overall *SAH* are heart attacks and high blood pressure.

We also tested two extended versions of these regressions. In the first, we included controls for socio-economic variables that could influence the structure of reporting biases in *SAH* or pick up omitted health attributes. We added controls for household expenditures, personal incomes, education, workforce participation, demographics and regional effects. The results are also given in Table 3. The controls do not add much to the explained variance. However, a number of the controls were highly significant, notably education and the regional effects; we return to discuss these later. We find that the coefficients on the specific health attributes tend to increase when we add the controls. This is most dramatic for the cluster of variables, "had surgery last year" through to "other problems" (Table 3). Adding the controls makes the coefficient on past surgery more negative, and it becomes significant at the 1% level. The controls have a similar effect on the coefficients on "wearing a hearing aid" which only diminishes *SAH* significantly when the controls are added. By contrast, the coefficients on "feeling chest pain" and wearing glasses (for men) become insignificant when the controls are added. The overall pattern is suggestive of adaptive behavior, whereby people with worse objective health status tend to adapt by overstating their subjective health. Thus the resulting negative correlation between objective health and the error term in the *SAH* regressions without controls tend to bias downward the regression coefficients.

The second extension added health attributes from 2000 using the panel, to test for lagged effects consistent with adaptation. (Recall that we cannot do this for the physical activity data, which were only asked of pensioners in 2000.) This means a lower sample size, with 2,600 and

3,600 adult men and women having complete data in the panel (as compared to 3,900 and 5,200.) There were clear signs of lagged effects. Liver, kidney and CGI problems in 2000 affected male *SAH* in 2002, though the effects were only significant at the 10% level. Spinal problems in 2000 had an effect on women's (but not men's) *SAH* in 2002, and the effect was significant at the 5% level. Stroke and TB in 2000 lowered men's *SAH* in 2002, as did high blood pressure for women, though these effects were only significant at the 10% level.

The predicted values of *SAH* scores from Table 3 can be interpreted as a weighted mean of the various specific health indicators in our data, in which the weights are chosen to give best fit with overall health-status. (We refer to the product of the probit regression coefficients for *SAH* and the regressors as the "*SAH* score" and similarly for *SAW*.) Figure 4 plots this weighted mean against normalized expenditure. We find only a small positive income effect.

We assume that *SAH* and *SAW* are functions of household expenditures, respondent income and a range of individual and household socio-economic characteristics that are likely to shift preferences, budget constraints and/or health production functions. In the case of *SAH*, one can interpret this as a reduced form model, in which the specific health indicators in Table 3 are taken to be functions of these variables. In the case of *SAW*, we interpret the "non-income" covariates as proxies for two main factors: differences in "needs" at given current income and differences in past or future living standards not captured in current incomes or expenditures.

Table 4 gives the OP for 2002 using as regressors a cubic polynomial of real expenditure, individual income, household size and demographic composition, age, education, unemployment and regional dummy variables. Similar results for all other variables of interest were obtained on dropping individual income and employment (given the aforementioned endogeneity concerns).

We also found evidence of lagged effects of past incomes on *SAW*. (Ravallion and Lokshin, 2002, report regressions including two years lags, using an earlier round of the RLMS.) However, given that we lose many observations, and including lagged effects did not change our main results, we confine attention here to the cross-sectional regressions for 2002.

Confirming the results of Figure 2, and consistently with the findings of Cockerham (2000) using the 1995 RLMS, we find no sign of an income gradient in *SAH* controlling for other socio-economic covariates. However, a strong income effect for *SAW* (for both men and women) is evident in Table 4.

We find generally more agreement between *SAH* and *SAW* in the effects of the "non-income" variables. Both subjective variables tend to fall with age (though over the range of the data for health, and only up to around 50 years for welfare).[16] Both variables tend to be higher for adults living in larger households, though other demographic effects are in less agreement. Education tends to raise *SAH* (though more so for women than men), as it does for *SAW*. The region with the highest *SAH* (controlling for our other covariates) tends to have the highest *SAW*, though there are discrepancies in ranking amongst other regions. The similarity in the non-income covariates generates a strong positive correlation between the predicted values for *SAH* and *SAW*; the correlation coefficient between the estimated conditional expectations is 0.64.

There are some notable gender differences in the socio-economic covariates of health and welfare. Education has a much stronger effect on female *SAH* than for men, while demographic effects are stronger for men. For men, being in the workforce (either with a job or unemployed) has a strong positive effects on *SAH* (relative to not being in the workforce) but this effect is not evident for women. This could reflect a reporting bias whereby male leisure is seen to be more

[16] It might be conjectured that the gradient is stronger for younger people, with little experience of the older regime of universal health care. This does not appear to be the case; when we added the product of age and expenditure to the regressions in Table 4, the interaction effect was not significant.

socially acceptable if it is due to ill-health. It could also reflect an endogeneity problem, whereby discouraged workers tend to be un-healthy; again one should be careful in giving these regressions a causal interpretation.

It is of interest to compare the socio-economic covariates for *SAH* in Table 4 with the corresponding controls for reporting bias when used in the regressions on specific health attributes (Table 3). We give this comparison in Table 5. Education has a much stronger effect (particularly for men) when it was used as a control variable for dealing with reporting effects than in the reduced form models in Table 4. Better educated individuals tended to have higher *SAH* at given values of the specific health attributes in the survey. Demographic effects tend to be stronger in the reduced form than when used as controls. Male employment status is not significant when used as a control for reporting bias, though it is significant in the reduced form model. Regional effects are similar, however. This might suggest that the regional effects in Table 4 are largely reporting bias, though that assumes that there are no omitted regional effects on health in the regressions on specific health attributes.

We tested for lagged expenditure and income effects using the panel sub-sample. When these were added to the regressions in Table 5 they did not emerge as significant predictors, so we confined attention to the 2002 cross-section. We also tested a panel data model allowing for latent individual effects in the continuous subjective health variable, following the econometric method for modeling subjective qualitative data outlined in Ravallion and Lokshin (2001). Again, we found no significant effect of consumption expenditure on *SAH*.

5. Test for the gradient

We can now provide a test for the gradient that is robust to the biases in *SAH* data, as discussed in section 2. Figure 5 gives predicted *SAH* scores based solely on the health indicators from Table 3, plotted against predicted *SAW* scores based on the covariates in Table 5. (We do not use the specification with socio-economic controls to assure that our health measure is not based on the same variables as our welfare measure. The only common covariate is age.) We give the nonparametric regression with and without controls, for which we use age, age squared, education and regional dummy variables. Table 6 provides the estimates parameters on these control variables.[17] Note that the nonlinearity of the economic welfare effect implies that the usual exclusion restrictions (variables in X^W but no in Z) for linear models are not required. However, we shall assume that household demographics and employment status (unemployed or not) only affect health through their impact on economic welfare.

We focus first on the unconditional gradient. The correlation coefficient between the predicted scores of *SAH* and *SAW* is 0.46 (0.50 for women and 0.36 for men). However, a marked nonlinearity is evident in Figure 5, with a low gradient at low values and high values; concavity is found beyond low values of *SAW*. Comparing the predicted scores with the relevant cut-off points in Tables 3 and 5 (as indicated in Figure 5), we can see that predicted *SAH* score rises from "average health" ("neither good nor bad") for the poorest stratum to "good health" for the richest. This is marked contrast to the unadjusted data, which essentially showed a mean *SAH* of "average health" independently of economic welfare. The gradient is steeper for women.

The economic gradient in health is greatly reduced when we add the control variables. The more robust unconditional gradient that we find using our preferred estimation method appears to be largely accountable to age, education and location. From Table 6 it is clear that the effect of age is statistically important; health declines with age with a concave curvature (so that the marginal decline in health as one ages rises with age). Education and the regional effects are

---

[17] The standard errors are calculated using bootstrap method, given that predicted values of economic welfare are used in the nonparametric sub-function.

also quite strong. On testing alternative specifications for the controls we found that the gradient largely vanished if one conditioned only on age, age-squared and education.

6. Conclusions

We have investigated the economic gradient in health status in Russia. Echoing past research, we find that the self-assessed health of Russian adults shows very little gradient with respect to household consumption or individual income. A steeper gradient emerges between subjective health predicted on the basis of specific objective health indicators and self-rated economic welfare, conditional on largely independent covariates. The nonparametric regression of predicted subjective health scores on predicted economic welfare scores exhibits a plausible S-shape, being flat at low economic welfare, then rising steeply through the bulk of the data, and flattening out for the relatively rich. We also find that the bulk of the gradient can be attributed to a relatively small number of characteristics related to the respondent's age, education and location; the gradient largely vanishes when we add controls for these variables.

Our results are consistent with the view that the health gap between the rich and the poor is underestimated if one ignores the biases stemming from non-ignorable measurement errors in self-assessed health and the existence of factors that influence economic welfare independently of current incomes or expenditures. While our estimation method offers some hope of revealing the true socioeconomic determinants of health status from subjective data, it can only do so by exploiting a large number of objective covariates. The idea that self-assessed health is a reliable short-cut for health measurement in surveys finds little support in our results.

References

Adams, Peter, Michael Hurt, Daniel McFadden, Angela Merrill and Tiago Ribeiro, 2003, "Health, Wealthy and Wise? Tests for Direct Causal paths Between Health and Socioeconomic Status," *Journal of Econometrics* 112:3-56.

Anand, Sudhir and Martin Ravallion, 1993, "Human Development in Poor Countries: On the Role of Private Incomes and Public Services," *Journal of Economic Perspectives*, 7 (Winter): 133-150.

Appels, A., H. Bosma, V. Grabauskas, A. Gostautas and F. Sturmans, 1996, "Self-assessed Health and Mortality in a Lithuanian and a Dutch Population," *Social Science and Medicine* 42(5): 681-689.

Brainerd, Elizabeth and David Cutler, 2005, "Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union," IZA DP 1472, Institute for the Study of Labor, Bonn, Germany.

Butler, J.S., Richard Burkhauser, Jean Mitchell and Theodore Pincus, 1987, "Measurement Error in Self-Reported Health Variables," *Review of Economics and Statistics* 69(4): 644-650,

Case, Anne, Darren Lubotsky and Christina Paxson, 2002, "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review* 92(5): 1308-34.

Cockerham, William, 2000, "Health Lifestyles in Russia," *Social Science and Medicine* 51(9): 1313-24.

Currie, Alison, Michael Shields, Stephen Wheatley Price, 2004, Is the Child Health/Family Income Gradient Universal? Evidence from England. IZA DP 1328, Institute for the Study of Labor, Bonn, Germany.

DaVanzo, Julie and David Adamson, 1997, "Russia's Demographic Crisis: How Real is it?" RAND Issues Paper, Center for Russian and Eurasian Studies, Labor and Population Program, http://www.rand.org/publications/IP/IP162/

Deaton, Angus. 2003. "Health, Inequality and Economic Development," *Journal of Economic Literature* 41(March): 113-158.

Ellman, Michael, 1994, "The Increase in Death and Disease under 'Katastroika'," *Cambridge Journal of Economics* 18(4): 329-355.

Elstad J.I. , 1998, "The Psycho-social Perspective on Social Inequalities in Health," *Sociology of Health and Illness*, 20(5): 598-618.

Fan, Jianqing, 1992, "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87: 998-1004.

Field, Mark, 1995, "The Health Crisis in the Former Soviet Union: A Report from the Post-War Zone," *Social Science and Medicine* 41(11): 1469-78.

Finch, Brian, 2004, "Early Origins of the Gradient: The Relationship Between Socioeconomic Status and Infant Mortality in the United States," *Demography* 40(4): 675-699.

Franzini, Luisa, Margaret Caughy, Willima Spears and Maria Esquer, 2005, "Neighborhood Economic Conditions, Social Processes and Self-assessed Health in Low-Income Neighborhoods in Texas: A Multilevel Latent Variables Model," *Social Science and Medicine* 61(6): 788-804.

Gavrilova, Natalia, Victoria Semyonova and Galina Evokushkina, 2002, "Mortality Crisis in Russia: New Health Trends," Paper Presented at the 2002 Annual Meeting of the Population Association of America, Atlanta, Georgia, USA.

Groot, Wim, 2000, "Adaptation and Scale of Reference Bias in Self-Assessments of Quality of Life," *Journal of Health Economics* 19: 403-420.

Heleniak, Timothy, 2002, "Russia's Demographic Decline Continues," Population Reference Bureau, http://www.prb.org/Template.cfm?Section=PRB&template=/ContentManagement/ContentDisplay.cfm&ContentID=6506 .

Idler, Ellen and Yael Benyamini, 1997, "Self-assessed Health and Mortality: A Review of Twenty Seven Community Studies," *Journal of Health and Social Behavior* 38(1): 21-37.

Kapteyn, Arie, 1994, "The Measurement of Household Cost Functions. Revealed Preference Versus Subjective Measures," *Journal of Population Economics* 7: 333-350.

Kerkhofs, Marcel and Maarten Lindeboom, 1995, "Subjective Health Measures and State-Dependent Reporting Errors," *Health Economics* 4: 221-235.

Kunst, A.E., J.J. Geurts and J. van den Berg, "International Variation in Socioeconomic Inequalities in Self Reported Health," *Journal of Epidemiology and Community Health*, 49(2): 117-123.

Lokshin, Michael, 2006, "Semi-Parametric Difference-Based Estimation of Partial Linear Regression Models." *Stata Journal*, forthcoming.

Lokshin, Michael and Barry Popkin, 1999, "The Emerging Underclass in the Russian Federation: Income Dynamics, 1992-1996," *Economic Development and Cultural Change* 47(4):

803-29.

Lokshin, Michael and Martin Ravallion, 2000, “Welfare Impacts of Russia’s 1998 Financial Crisis and the Response of the Public Safety Net,” *Economics of Transition*, 8(2): 269-295.

Lynch, J.W., Davey Smith, G. Kaplan and J.S. House, 2000, “Income Inequality and Mortality: Importance to Health of Individual Income, Psychosocial Environment, or Material Conditions,” *British Journal of Medicine* 320: 1200-1204.

Mackenbach, Johan, Anton Kunst, Adrienne Cavelaars, Feikje Groenhof and Jose Guerts, 1997, “Socioeconomic Inequalities in Morbidity and Mortality in Western Europe,” *Lancet*, 349: 1655-59.

Marmot, Michael, George Smith, Stephen Stansfeld, Chandra Patel, Fiona North, J. Head, Ian White, Eric Brunner and Amanda Feeny, 1991, “Health Inequalities among British Civil Servants: The Whitehall II Study,” *Lancet,* 337: 1387-93.

Menzel, Paul, Paul Dolan, Jeff Richardson and Jan Abel Olsen, 2002, “The Role of Adaptation to Disability and Disease in Health State Valuation: A preliminary Normative Analysis,” *Social Science and Medicine* 55: 2149-2158.

Milanovic, Branko, 1999, “Explaining the Increase in Inequality during the Transition,” *Economics of Transition* 7(2): 299-341.

Pappas, Gregory, Susan Queen, Wilbur Hadden and Gail Fisher. 1993. “The Increasing Disparity in Mortality Between Socioeconomic Groups in the United States, 1960 and 1986,” *New England Journal of Medicine*, 329(2): 103-109.

Pollak, Robert, 1991, “Welfare Comparisons and Situation Comparisons”, *Journal of Econometrics* 50: 31-48.

Preston, Samuel, 1980, “Mortality Declines in Less developed Countries,” In Richard Easterlin (ed.) *Population and Economic Change in Developing Countries*, Chicago: Chicago University Press.

Pritchett, Lant and Lawrence Summers, 1996, “Wealthier is Healthier,” *Journal of Human Resources* 31(4): 841-68.

Ravallion, Martin and Michael Lokshin, 2001, “Identifying Welfare Effects from Subjective Questions,” *Economica*, 68:. 335-357.

______________ and ______________, 2002, “Self-assessed Economic Welfare in Russia.”

European Economic Review 46, 1453-1473.

Sen, Amartya, 1998, "Mortality as an Indicator of Economic Success and Failure," *Economic Journal* 108: 1-25.

Smith, James P., 1999, "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status," *Journal of Economic Perspectives* 13(2): 145-166.

Sorlie, Paul D., Eric Backlund and Jacob Keller. 1995. "US Mortality by Economic, Demographic and Social Characteristics: The National Longitudinal Mortality Study," *American Journal of Public Health* 85(7): 949-956.

Thomas, Duncan and Elizabeth Frankenberg, 2000, "The Measurement and Interpretation of Health in Social Surveys," Working Paper 01-06, Labor and Population Program, Rand Corporation.

Wilkinson, R.G., 1996, *Unhealthy Societies: The Afflictions of Inequality*. London: Routledge.

Yatchew, Adonis, 1998, "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature* 36(June): 669-721.

Table 1: Self-assessed health and subjective economic welfare, Russia 2002

		Self-assessed health					
		1 (Very bad)	2 (Bad)	3	4(Good)	5 (Very good)	Total
Subjective economic welfare	Cramer's V = 0.123 (0.006)			*Whole sample*			
	1 (Poorest)	50	138	291	70	5	554
	2	56	248	626	187	15	1132
	3	72	349	1304	523	15	2263
	4	34	251	1382	623	28	2318
	5	22	188	1182	581	25	1998
	6	3	33	245	196	13	490
	7+ (Richest)	20	38	154	130	18	360
	Total	257	1245	5184	2310	119	9115
	Cramer's V = 0.111 (0.008)			*Men*			
	1 (Poorest)	8	41	131	33	3	216
	2	17	77	211	106	10	421
	3	19	97	521	296	9	942
	4	15	85	558	357	18	1033
	5	6	69	476	314	17	882
	6	1	8	101	110	7	227
	7+ (Richest)	5	10	64	80	12	171
	Total	71	387	2062	1296	76	3892
	Cramer's V = 0.129 (0.008)			*Women*			
	1 (Poorest)	42	97	160	37	2	338
	2	39	171	415	81	5	711
	3	53	252	783	227	6	1321
	4	19	166	824	266	10	1285
	5	16	119	706	267	8	1116
	6	2	25	144	86	6	263
	7+ (Richest)	15	28	90	50	6	189
	Total	186	858	3122	1014	43	5223

Note: Figures in parentheses are bootstrapped standard errors for Cramer's V statistic.

Table 2: Change in self-assessed heath and subjective economic welfare, panel for 2000-02

Change in subjective economic welfare	Change in self-assessed health						
	-3 (worse)	-2	-1	0	1	2 (better)	Total
Cramer's V = 0.048 (0.008)			*Whole sample*				
-6 (Declined)	0	0	2	4	1	0	7
-5	0	0	3	16	4	0	23
-4	0	1	17	54	7	2	81
-3	0	5	43	132	36	2	218
-2	0	6	124	329	82	5	546
-1	0	12	218	719	171	20	1140
0	0	18	323	1216	289	33	1879
+1	3	15	242	802	219	18	1299
+2	1	4	132	389	117	8	651
+3	1	4	42	145	52	11	255
+4	1	1	23	70	24	4	123
+5	0	1	4	19	6	1	31
+6 (Improved)	0	0	4	11	3	2	20
Total	6	67	1177	3906	1011	106	6273
Cramer's V = 0.078 (0.011)			*Men*				
-6 (Declined)	0	0	1	3	0	0	4
-5	0	0	1	5	0	0	6
-4	0	0	11	17	3	0	31
-3	0	2	21	46	19	0	88
-2	0	2	53	137	31	3	226
-1	0	2	102	309	69	7	489
0	0	8	136	478	109	18	749
+1	1	10	110	320	90	9	540
+2	1	2	55	168	61	5	292
+3	1	4	19	57	24	8	113
+4	0	1	12	31	13	1	58
+5	0	0	2	7	2	0	11
+6 (Improved)	0	0	3	2	1	1	7
Total		31	526	1580	422	52	2614
Cramer's V = 0.063 (0.014)			*Women*				
-6 (Declined)	0	0	1	1	1	0	3
-5	0	0	2	11	4	0	17
-4	0	1	6	37	4	2	50
-3	0	3	22	86	17	2	130
-2	0	4	71	192	51	2	320
-1	0	10	116	410	102	13	651
0	0	10	187	738	180	15	1130
+1	2	5	132	482	129	9	759
+2	0	2	77	221	56	3	359
+3	0	0	23	88	28	3	142
+4	1	0	11	39	11	3	65
+5	0	1	2	12	4	1	20
+6 (Improved)	0	0	1	9	2	1	13
Total	3	36	651	2326	589	54	3659

Table 3: Ordered probits for *SAH* as a function of age and specific health indicators

	Men		Women		Men		Women	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*
Controls	*NO*		*NO*		*YES*		*YES*	
Age	-0.055***	*0.007*	-0.017**	*0.006*	-0.063***	*0.009*	-0.022**	*0.007*
$Age^2/100$	0.057***	*0.008*	0.007	*0.006*	0.068***	*0.001*	0.015*	*0.007*
Difficulty in tasks (1=not difficult)								
Running 1 km	0.214***	*0.022*	0.154***	*0.017*	0.210***	*0.023*	0.164***	*0.018*
Walking 1 km	0.034	*0.042*	0.095**	*0.027*	0.039	*0.043*	0.092**	*0.028*
Walking 200m	0.058	*0.065*	0.126**	*0.040*	0.038	*0.067*	0.122**	*0.041*
Climb stairs	0.094*	*0.035*	0.087***	*0.024*	0.108**	*0.036*	0.103***	*0.024*
Climb stairs	0.048	*0.057*	0.055	*0.035*	0.035	*0.058*	0.043	*0.035*
Lift 5 kb	0.095*	*0.046*	0.091***	*0.024*	0.094*	*0.047*	0.090***	*0.024*
Crouch	0.097**	*0.034*	0.035	*0.023*	0.075*	*0.035*	0.046*	*0.024*
Walk around the room	0.137	*0.109*	-0.036	*0.062*	0.239*	*0.117*	-0.045	*0.064*
Seat for 2 hours	0.049	*0.059*	0.049	*0.037*	0.037	*0.062*	0.040	*0.038*
Get up	-0.084	*0.095*	0.048	*0.052*	-0.115	*0.097*	0.029	*0.053*
Stand	0.080	*0.069*	-0.083*	*0.042*	0.079	*0.072*	-0.079*	*0.043*
Eating	0.260*	*0.142*	0.169*	*0.091*	0.352*	*0.168*	0.191*	*0.100*
Dressing	0.061	*0.110*	0.119	*0.074*	0.081	*0.117*	0.135*	*0.078*
Toilet	-0.011	*0.115*	0.020	*0.074*	-0.019	*0.128*	0.069	*0.079*
Body Mass Index								
BMI	0.016	*0.018*	-0.008	*0.006*	0.022	*0.036*	0.012	*0.007*
BMI2/100	-0.082*	*0.036*	-0.002	*0.008*	0.002	*0.066*	-0.002	*0.009*
Reported that:								
Has heart problems	-0.028	*0.076*	-0.075	*0.052*	-0.041	*0.078*	-0.128*	*0.053*
Has lung problems	-0.240**	*0.080*	-0.322***	*0.071*	-0.195*	*0.081*	-0.314***	*0.072*
Has liver problems	-0.261**	*0.085*	-0.132*	*0.054*	-0.206*	*0.087*	-0.108*	*0.055*
Has kidney problems	-0.186*	*0.084*	-0.346***	*0.053*	-0.195*	*0.085*	-0.364***	*0.055*
Has CGI problems	-0.329***	*0.056*	-0.273***	*0.045*	-0.355***	*0.057*	-0.290***	*0.046*
Has spine problems	-0.197**	*0.057*	-0.107*	*0.046*	-0.188**	*0.058*	-0.132**	*0.047*
Had surgery last year	-0.008	*0.111*	-0.092	*0.086*	-0.388***	*0.055*	-0.320***	*0.041*
Feels chest pain	-0.215***	*0.055*	-0.273***	*0.046*	-0.015	*0.114*	-0.099	*0.088*
Wears hearing aid	-0.001	*0.227*	0.090	*0.230*	-0.189**	*0.056*	-0.239***	*0.047*
Wears glasses	-0.195***	*0.051*	-0.057	*0.042*	0.054	*0.239*	0.126	*0.233*
Other problems	-0.366***	*0.054*	-0.295***	*0.040*	-0.202***	*0.053*	-0.072	*0.044*
Diagnosed with:								
Diabetics	-0.117	*0.139*	-0.128*	*0.075*	-0.218	*0.145*	-0.151*	*0.077*
Heart attack	-0.331**	*0.116*	-0.042	*0.120*	-0.363**	*0.118*	-0.134	*0.124*
High blood pressure	-0.321***	*0.047*	-0.205***	*0.040*	-0.305***	*0.048*	-0.199***	*0.041*
Stroke	-0.064	*0.145*	-0.095	*0.121*	-0.030	*0.149*	-0.160	*0.126*
Anemia	0.036	*0.191*	-0.006	*0.072*	-0.096	*0.198*	-0.010	*0.073*
TB	-0.281*	*0.152*	0.573**	*0.203*	-0.261*	*0.156*	0.642**	*0.210*
Hepatitis	-0.170*	*0.068*	-0.059	*0.064*	-0.182*	*0.069*	-0.062	*0.065*
c1	-0.074	*0.612*	0.052	*0.346*	1.697*	*0.805*	0.714*	*0.421*
c2	1.590*	*0.621*	1.751***	*0.352*	3.377***	*0.813*	2.458***	*0.426*
c3	3.914***	*0.621*	4.304***	*0.352*	5.742***	*0.814*	5.069***	*0.427*
c4	5.930***	*0.622*	6.192***	*0.356*	7.805***	*0.816*	7.020***	*0.431*

# of observations	3892	5223	3844	5148
Aldrich-Nelson R^2	0.512	0.563	0.523	0.577
Log-likelihood	-3161.460	-4009.667	-3067.766	-3859.115

Note: * means significant at 10%, ** at 5%, and *** at 1% level.

Table 4: Ordered probits for *SAH* and *SAW* as functions of socio-economic characteristics

	Self-assessed health				Self-assessed economic welfare			
	Men		Women		Men		Women	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. Error*
Expenditure	0.001$^{\Diamond}$	*0.015*	0.005$^{\Diamond}$	*0.012*	0.124***	*0.013*	0.115***	*0.010*
Expenditure^2 /10^2	0.039	*0.083*	0.025	*0.057*	-0.528***	*0.075*	-0.392***	*0.051*
Expenditure^3/10^3	-0.005	*0.009*	-0.002	*0.005*	0.044***	*0.008*	0.028***	*0.004*
Individual income/10^5	-0.003	*0.047*	0.142**	*0.081*	0.205***	*0.042*	0.338***	*0.072*
Household size	0.044**	*0.014*	0.040**	*0.013*	0.055***	*0.013*	0.050***	*0.011*
Share of children 0-6	0.420**	*0.209*	0.154	*0.177*	-0.146	*0.189*	0.042	*0.158*
Share of children 7-15	0.245	*0.157*	0.021	*0.130*	-0.117	*0.142*	0.187	*0.115*
Share of pensioners	0.209**	*0.119*	0.065	*0.103*	-0.187**	*0.109*	0.027	*0.092*
Share of adult women	0.321**	*0.174*	-0.110	*0.134*	0.269**	*0.158*	0.261**	*0.119*
Age	-0.038***	*0.008*	-0.028***	*0.006*	-0.061***	*0.007*	-0.049***	*0.005*
Age2/100	0.009	*0.009*	-0.010**	*0.006*	0.059***	*0.008*	0.045***	*0.005*
Primary and less	0.061	*0.107*	0.287**	*0.079*	0.121	*0.099*	0.084	*0.073*
Vocational	0.011	*0.109*	0.240**	*0.102*	0.027	*0.101*	0.038	*0.093*
Technical Vocational	0.116	*0.094*	0.307***	*0.073*	0.153**	*0.087*	0.197**	*0.067*
Secondary	0.156	*0.098*	0.403***	*0.074*	0.203**	*0.090*	0.183**	*0.068*
Incomplete Higher	0.261**	*0.100*	0.409***	*0.078*	0.199**	*0.092*	0.173**	*0.071*
Married	-0.076	*0.079*	-0.047	*0.066*	0.209**	*0.072*	0.186**	*0.059*
Divorced	0.001	*0.105*	-0.117	*0.077*	-0.154	*0.096*	-0.034	*0.069*
Live together	-0.146	*0.088*	-0.082	*0.080*	0.014	*0.079*	0.010	*0.071*
Widowed	0.029	*0.127*	-0.059	*0.078*	0.166	*0.117*	-0.099	*0.070*
Unemployed	0.394***	*0.057*	0.051	*0.047*	0.044	*0.051*	-0.041	*0.042*
Has job	0.285**	*0.098*	0.037	*0.088*	-0.137	*0.089*	-0.177**	*0.078*
Northern and North Western	-0.071	*0.092*	-0.094	*0.080*	0.296**	*0.084*	0.092	*0.071*
Central and Central Chernozem	-0.030	*0.069*	-0.099	*0.058*	0.139**	*0.062*	0.046	*0.052*
Volgo-Vyatskiy	0.076	*0.067*	0.021	*0.059*	0.179**	*0.061*	0.187**	*0.052*
North Caucasian	0.304***	*0.073*	0.261***	*0.064*	0.453***	*0.066*	0.354***	*0.057*
Ural	-0.059	*0.073*	-0.056	*0.063*	0.319***	*0.066*	0.160**	*0.056*
Western Siberian	-0.184**	*0.082*	-0.209**	*0.073*	0.160**	*0.075*	0.110**	*0.065*
Eastern Siberian	-0.241**	*0.079*	-0.088	*0.070*	0.166**	*0.072*	0.133**	*0.062*
c1	-3.303***	*0.206*	-3.492***	*0.185*	-2.032***	*0.184*	-1.821***	*0.161*
c2	-2.189***	*0.202*	-2.200***	*0.183*	-1.362***	*0.182*	-1.091***	*0.160*
c3	-0.286	*0.199*	-0.032	*0.179*	-0.563**	*0.181*	-0.315**	*0.159*
c4	1.621***	*0.202*	1.779***	*0.185*	0.161	*0.181*	0.361**	*0.159*
c5					1.036***	*0.182*	1.244***	*0.160*
c6					1.498***	*0.184*	1.694***	*0.162*
# of observations	3863		5186		3870		5191	
Aldrich-Nelson R^2	0.310		0.404		0.128		0.123	
Log-likelihood	-3660.830		-4616.902		-6505.013		-8767.503	

Note: * denotes significant at 10%, ** at 5%, and *** at 1% level; ◊ denotes jointly insignificant at the 5% level.

Table 5: Ordered probits for *SAH* as functions of socio-economic characteristics

	Model without health indicators				Model with both health and socio-economic indicators			
	Men		Women		Men		Women	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. Error*
Health indicators	*NO*		*NO*		*YES*		*YES*	
Expenditure	0.001$^{\lozenge}$	*0.015*	0.005$^{\lozenge}$	*0.012*	0.009	*0.016*	0.015	*0.012*
Expenditure^2 /10^2	0.039	*0.083*	0.025	*0.057*	0.012	*0.088*	-0.018	*0.060*
Expenditure^3/10^3	-0.005	*0.009*	-0.002	*0.005*	-0.002	*0.010*	0.001	*0.005*
Individual income/10^5	-0.003	*0.047*	0.142**	*0.081*	-0.038	*0.050*	0.081	*0.085*
Household size	0.044**	*0.014*	0.040**	*0.013*	0.027*	*0.015*	0.033*	*0.014*
Share of children 0-6	0.420**	*0.209*	0.154	*0.177*	0.259	*0.218*	0.013	*0.187*
Share of children 7-15	0.245	*0.157*	0.021	*0.130*	0.070	*0.166*	-0.159	*0.138*
Share of pensioners	0.209**	*0.119*	0.065	*0.103*	0.027	*0.127*	-0.018	*0.111*
Share of adult women	0.321**	*0.174*	-0.110	*0.134*	0.222	*0.183*	-0.058	*0.143*
Age	-0.038***	*0.008*	-0.028***	*0.006*	-0.063***	*0.009*	-0.022**	*0.007*
Age2/100	0.009	*0.009*	-0.010**	*0.006*	0.068***	*0.001*	0.015*	*0.007*
Primary and less	0.061	*0.107*	0.287**	*0.079*	0.310*	*0.117*	0.373***	*0.086*
Vocational	0.011	*0.109*	0.240**	*0.102*	0.186	*0.118*	0.320**	*0.109*
Technical Vocational	0.116	*0.094*	0.307***	*0.073*	0.287**	*0.104*	0.381***	*0.080*
Secondary	0.156	*0.098*	0.403***	*0.074*	0.278*	*0.107*	0.481***	*0.081*
Incomplete Higher	0.261**	*0.100*	0.409***	*0.078*	0.376**	*0.110*	0.445***	*0.086*
Married	-0.076	*0.079*	-0.047	*0.066*	-0.059	*0.083*	-0.022	*0.070*
Divorced	0.001	*0.105*	-0.117	*0.077*	-0.018	*0.111*	-0.068	*0.082*
Live together	-0.146	*0.088*	-0.082	*0.080*	-0.131	*0.092*	-0.047	*0.084*
Widowed	0.029	*0.127*	-0.059	*0.078*	0.032	*0.137*	0.059	*0.083*
Unemployed	0.394***	*0.057*	0.051	*0.047*	0.124*	*0.061*	-0.065	*0.050*
Has job	0.285**	*0.098*	0.037	*0.088*	0.150	*0.103*	-0.048	*0.092*
Northern and North Western	-0.071	*0.092*	-0.094	*0.080*	-0.131	*0.098*	-0.124	*0.086*
Central and Central Chernozem	-0.030	*0.069*	-0.099	*0.058*	-0.039	*0.073*	-0.156*	*0.062*
Volgo-Vyatskiy	0.076	*0.067*	0.021	*0.059*	-0.035	*0.071*	-0.123*	*0.063*
North Caucasian	0.304***	*0.073*	0.261***	*0.064*	0.256**	*0.077*	0.314***	*0.068*
Ural	-0.059	*0.073*	-0.056	*0.063*	-0.196*	*0.077*	-0.179**	*0.067*
Western Siberian	-0.184**	*0.082*	-0.209**	*0.073*	-0.269**	*0.087*	-0.347***	*0.078*
Eastern Siberian	-0.241**	*0.079*	-0.088	*0.070*	-0.251**	*0.084*	-0.068	*0.074*
C1	-3.303***	*0.206*	-3.492***	*0.185*	1.697*	*0.805*	0.714*	*0.421*
C2	-2.189***	*0.202*	-2.200***	*0.183*	3.377***	*0.813*	2.458***	*0.426*
C3	-0.286	*0.199*	-0.032	*0.179*	5.742***	*0.814*	5.069***	*0.427*
C4	1.621***	*0.202*	1.779***	*0.185*	7.805***	*0.816*	7.020***	*0.431*
# of observations	3863		5186		3844		5148	
Aldrich-Nelson R^2	0.310		0.404		0.523		0.577	
Log-likelihood	-3660.830		-4616.902		-3067.766		-3859.115	

Note: * denotes significant at 10%, ** at 5%, and *** at 1% level; ◊ denotes jointly insignificant at the 5% level.

Table 6: Estimated parameters on controls in partial linear regression

	Total		Males		Females	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*
Age	-0.014***	*0.003*	-0.021***	*0.006*	-0.020***	*0.004*
Age Squared/100	-0.033***	*0.003*	-0.023***	*0.006*	-0.028***	*0.004*
Primary and less	-0.054	*0.049*	0.040	*0.081*	-0.104*	*0.059*
Vocational	-0.114**	*0.043*	-0.201**	*0.070*	-0.120*	*0.053*
Technical Vocational	-0.013	*0.047*	-0.125*	*0.067*	-0.179**	*0.067*
Secondary	-0.056*	*0.027*	-0.058	*0.043*	-0.110**	*0.035*
Incomplete Higher	-0.091**	*0.029*	-0.053	*0.047*	-0.088*	*0.035*
University and higher			*Reference*			
Moscow and St. Petersburg			*Reference*			
Northern and North Western	-0.215***	*0.046*	-0.170*	*0.073*	-0.219***	*0.057*
Central and Central Chernozem	-0.136***	*0.033*	-0.097*	*0.053*	-0.220***	*0.042*
Volgo-Vyatskiy	-0.033	*0.033*	0.036	*0.052*	-0.088*	*0.042*
North Caucasian	0.248***	*0.038*	0.239***	*0.060*	0.217***	*0.048*
Ural	-0.147***	*0.036*	-0.195***	*0.058*	-0.107*	*0.045*
Western Siberian	-0.301***	*0.041*	-0.303***	*0.065*	-0.342***	*0.052*
Eastern Siberian	-0.246***	*0.039*	-0.368***	*0.061*	-0.204***	*0.050*

Note: Bootstrapped standard errors.

Figure 1: Self-assessed health in Russia, 2002

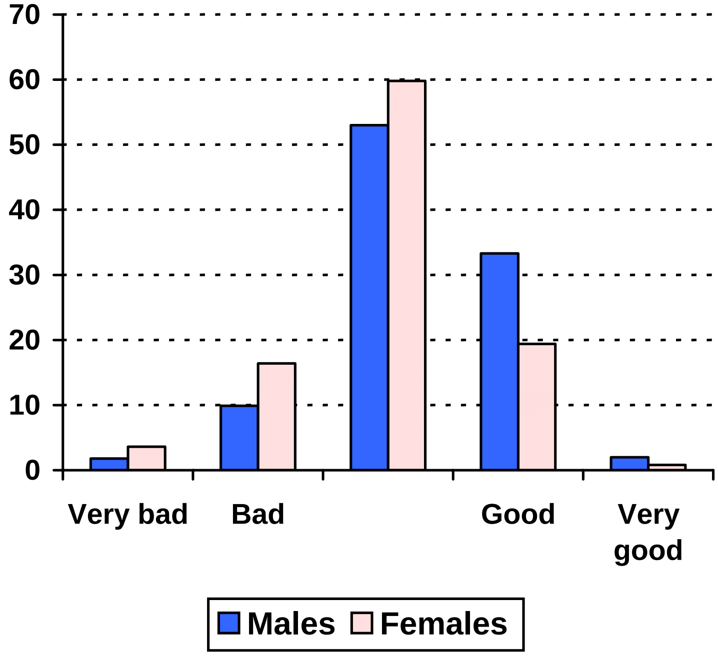
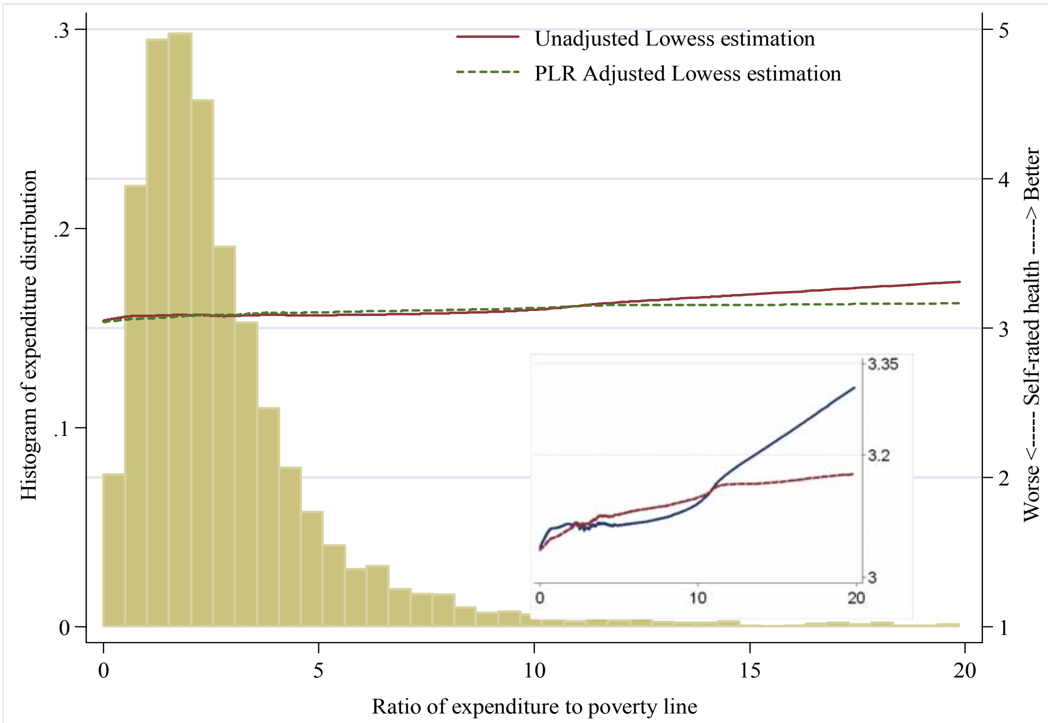


Figure 2: Self-assessed health against household expenditure as a proportion of poverty line, both unadjusted and adjusted for covariates using a partial linear model

(a) Self-assessed health



(b) Self-assessed economic welfare

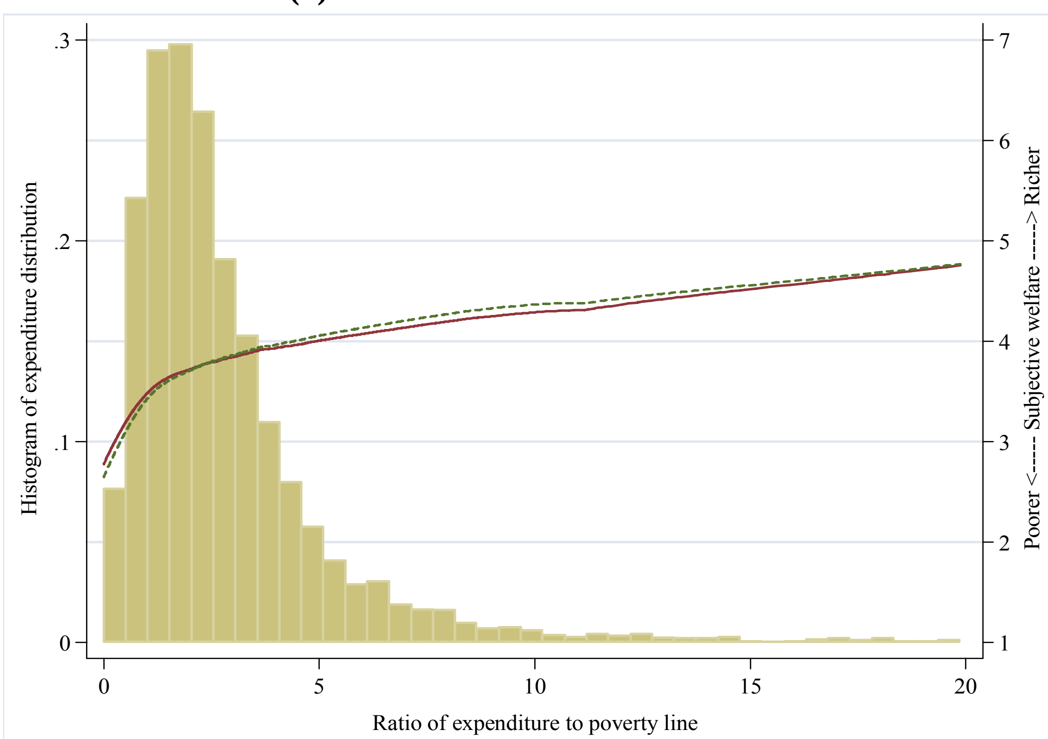


Figure 3: Specific health indicators regressed on household expenditure as a proportion of poverty line both unadjusted and adjusted for covariates using a partial linear model

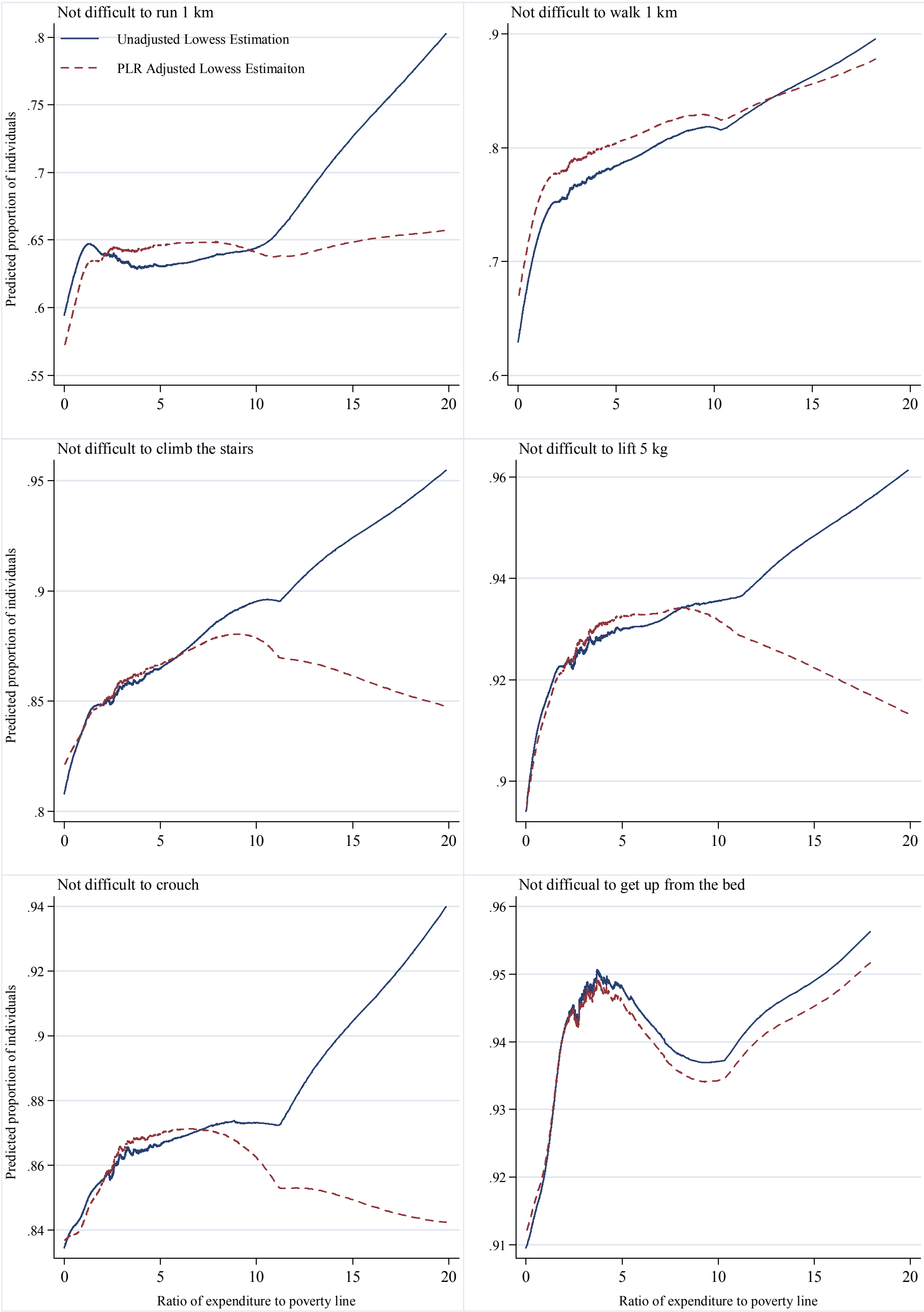


Figure 3 continued

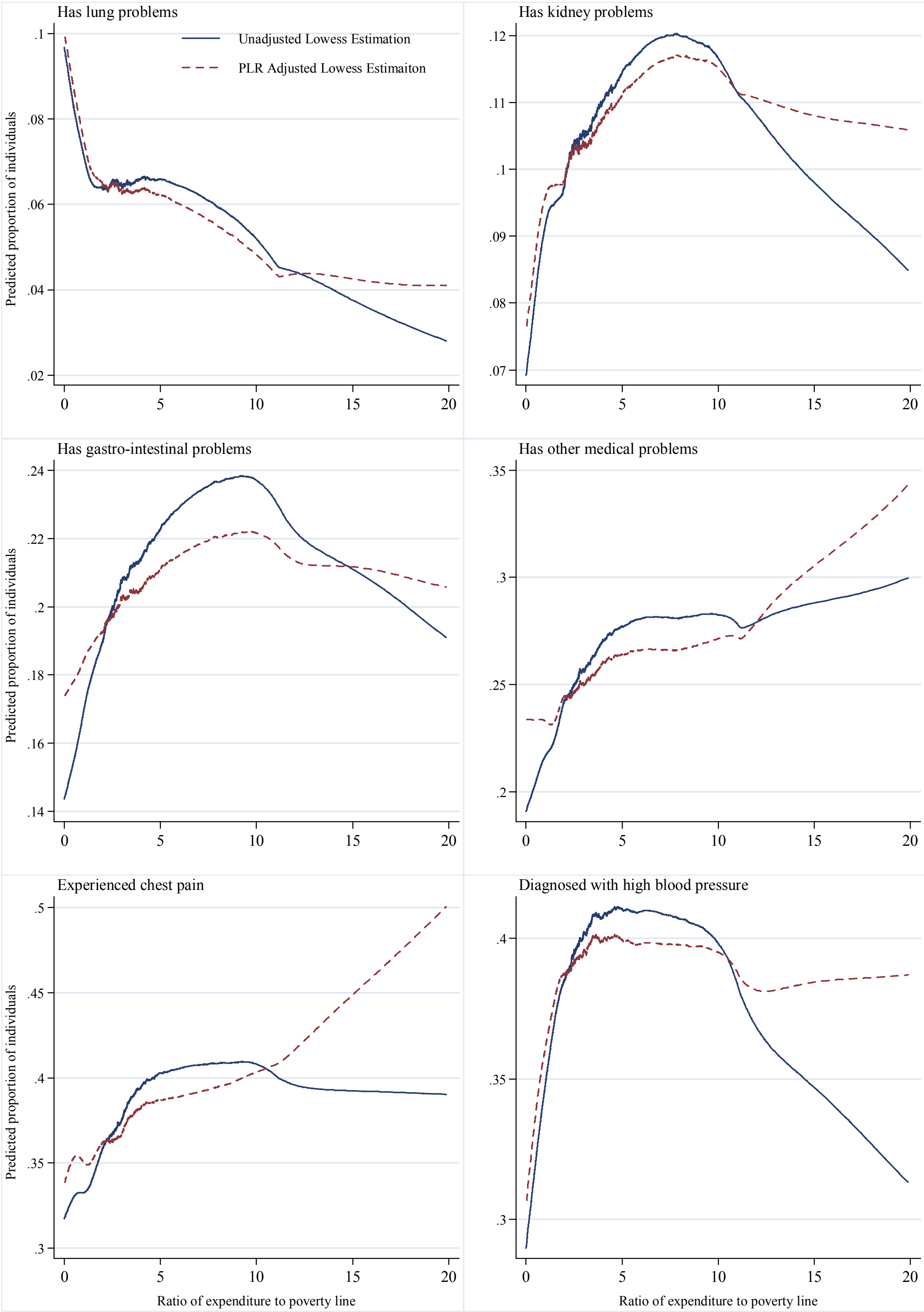


Figure 4: Predicted health perception based on specific health indicators plotted against real household expenditure per person

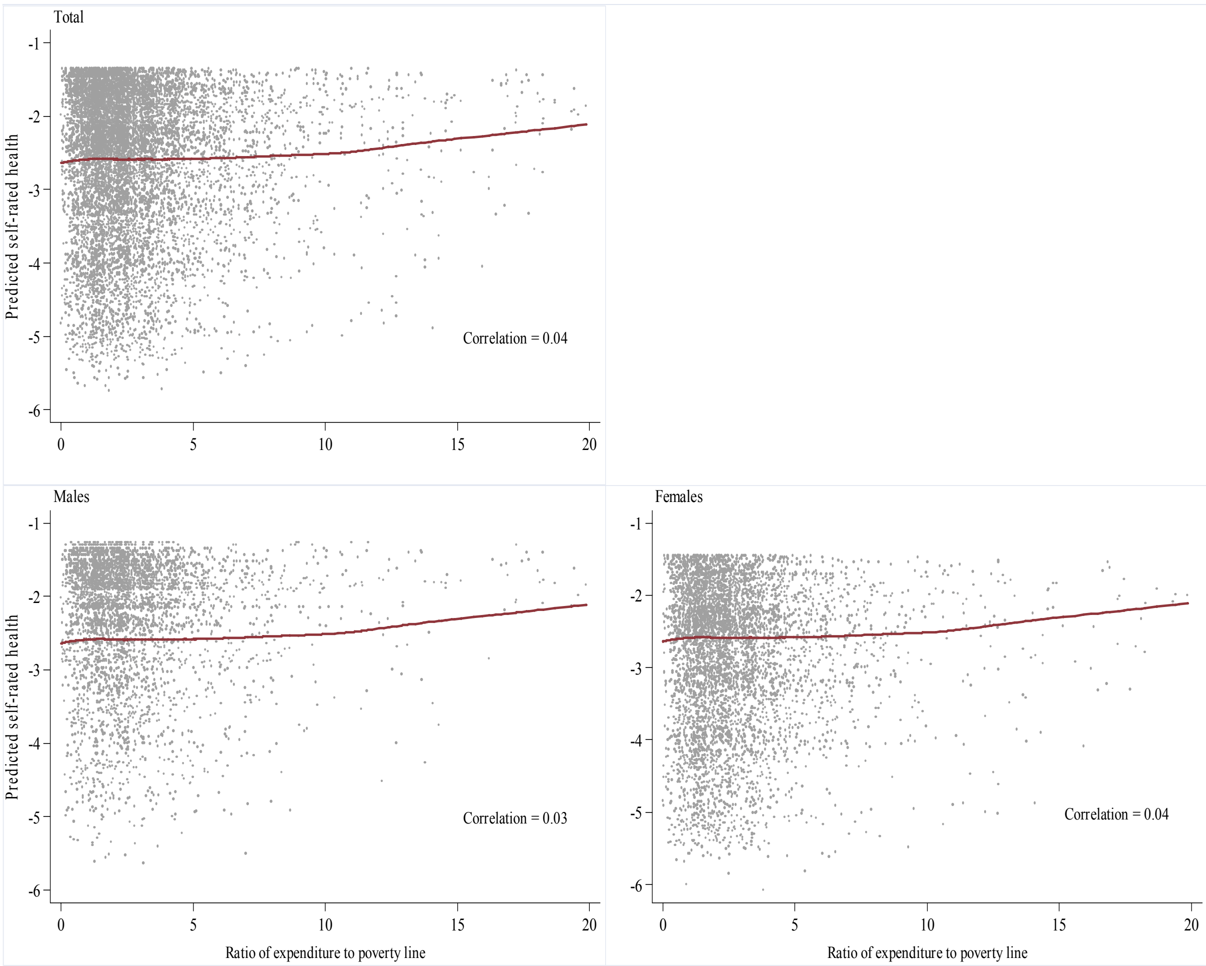


Figure 5: Predicted *SAH* based on specific health indicators (Table 3) as a function of predicted economic welfare based on socio-economic covariates (Table 5) with and without controls