

The measurement and interpretation of health in social surveys

Duncan Thomas
RAND & UCLA

and

Elizabeth Frankenberg
RAND

First version: November, 1999
This version: December, 2000

1. Introduction

Health status is hard to measure. It is widely recognized that health is multi-dimensional reflecting the combination of an array of factors that include physical, mental and social well-being, genotype and phenotype influences as well as expectations and information. A multitude of health indicators have been used in scientific studies drawing on data from both the developed and developing world. Understanding what those indicators measure is central if the results reported in the studies are to be interpreted in a meaningful way. Whether one is interested in summarizing the health of a population or understanding the links between health and other measures of well-being at the individual level, poor measurement will likely yield poor inferences.

There is a large literature that discusses the validity and limitations of different health measures. Murray and Chen (1992) and Sadana (2001) provide excellent reviews and discussion. Some of the most insightful empirical studies have compared indicators of specific morbidities reported by respondents in health interview surveys with indicators based on health examinations of the same individuals conducted by trained health workers. Other very influential studies have examined the extent to which self-reported health predicts health problems later in life. A third class of studies contrasts prevalence rates based on health interviews with other sources.

This paper attempts to provide insights into the meaning of a set of relatively general health indicators that are commonly collected in health interview surveys. These self-reports are contrasted with a battery of physical assessments conducted on the same respondents in the same survey, usually a day or two after the interview.

The key complexity that arises from these comparisons is that "true" health status is seldom known. To side-step this complexity, we begin with indicators of health (or, more precisely, nutritional) status which can be measured without any ambiguity: height and weight. Contrasting measurements with self-reports for the same person, we find that self-reports are subject to systematic biases. For example, men are inclined to report themselves as being significantly taller than they actually are; women are inclined to report themselves as significantly lighter than they are. If these were the only differences, it would be simple to adjust reports so they map into measurements. They are not. There is evidence that people are inclined to report their health status as being closer to an "ideal" than it is. This example highlights the complexity of deriving adjustment factors to offset the effects of reporting biases.

We proceed to examine a series of self-reported health indicators collected in the second wave of the Indonesia Family Life Survey (IFLS2). In these cases, truth is not known and so we turn to a comparison of the relationships between health status and socio-economic status (SES) to provide some insights into the meaning of these measures. Contrasting the link between SES and physical assessments with the link between SES and self-reported health, we suggest that the more general the health indicator, the more difficult it is to disentangle "true" health status from "reporting effects."¹ However, an examination of the relationships between self-reports and physical assessments suggests that self-reports, particularly more general questions, do contain a good deal of information about the respondent's health.

We conclude that, given the state of knowledge at this time, it would be prudent to invest in the collection of both self-reported health indicators and physical assessments of health status in individual and household surveys. Recent technological advances in health measurement render it feasible to conduct a broad array of assessments in face-to-face surveys at relatively low costs. The benefits are likely to far outweigh the costs.

2. Self-reports and measured health

Methods that have been used to measure health in social surveys span a wide spectrum, from one or two direct questions asked of individuals about their perception of their overall health to a comprehensive and thorough battery of clinical assessments conducted by physicians or other trained personnel who examine each respondent. For a general review of issues associated with health interview survey implementation, see Kroeger (1985) and Ross and Vaughan (1986); Fisher, Pappas and Limb (1996) provide a discussion in the context of health examination surveys.

The vast majority of surveys in both developed and developing countries have relied on individual responses to a series of questions about their health posed in an interview setting.² As a starting point,

¹Iburg (2000) presents related evidence from Denmark and the United States in which observed health is contrasted with both self-assessments and physician-assessments.

²The content and quantity of those questions depends on the number of different dimensions of health that are of interest and, within each dimension, on the desired sensitivity to different levels of well-being. Standard health assessment instruments include, for example, the Sickness Impact Profile, the Quality of Well-being Scale, and the Nottingham Health Profile. A well-known study (and resulting instrument) is the RAND Medical Outcomes Study (MOS). The full version of the MOS questionnaire contains 149 items and requires an estimated 30-40 minutes to administer (Stewart and Ware, 1992). A reduced instrument, SF-36, has been widely adopted and is advocated as a good basic tool for multi-purpose surveys. For a good discussion of issues about wording, recall periods etc., see Murray and Chen (1992).

let self-reported health, θ^S , indicate underlying health, θ^* , with adjustments that are, for simplicity of exposition, assumed to be additive and comprise three parts: an individual idiosyncratic component, ε_i , a factor associated with the specific observable indicator, ε_j , and a factor that varies with both the individual and the indicator, ε_{ij} :

$$\theta_{ij}^S = \theta_i^* + \varepsilon_i + \varepsilon_j + \varepsilon_{ij} \quad [1]$$

where θ^* is multi-dimensional and hence vector-valued. As a convention, we will treat higher levels of θ as denoting better health status. It is useful to provide concrete examples for the three types of adjustments. First, consider two individuals with the same level of intrinsic health, θ^* ; if one is more inclined to report a health problem than the other, then this difference will be reflected in ε_i . Second, say a health state is difficult to detect because it is asymptomatic; self-reports will tend to be noisy indicators of the prevalence of the state as reflected in a relatively large variance component, ε_j . Third, if a health state is better known among some people, relative to others, then these differences will be reflected in the component ε_{ij} . This might arise, for example, if we were to focus on diagnosed health problems; if two people have the same level of intrinsic health, a respondent who has had more contact with health services is more likely to report a diagnosis relative to a respondent who has never seen a trained physician.³

Two self-reported indicators, θ_{i1} and θ_{i2} , reported by the same individual will share the common component, ε_i , which, intuitively, may be thought of as the propensity to report oneself as ill. As a result, the two will be positively correlated even if neither is related to θ^* . Clearly, "internal validity" of self-reported measures, based on their being correlated with one another, does not imply they are valid indicators of θ^* . In practice, ε_i does provides information about the respondent's perception of his or her own health which may be an important dimension of health status. Recognizing this, we refer to ε_i as a "(self)-reporting effect" for want of a better shorthand.

There are relatively few examples of examination surveys that have been conducted on a large scale. This presumably reflects the costs and complexity of fielding these sorts of surveys. They are, however, a very rich source of information. One of the best examples, perhaps, is the United States National Health and Nutrition Examination Survey (NHANES) which was started in the late 1950s and

³It is straightforward to allow the adjustments to depend on the level of intrinsic health, in which case the effects become multiplicative. This would make sense if respondents are inclined to report their own health as close to the norm (average health) so that the adjustment would be largest for those in poorest (or best) health. We will allow θ and ε to depend on a common characteristic, x , below which also unties the additivity assumption.

is now run on a continuing basis. There are a small number of similar surveys that have been successfully conducted in low income settings which suggest that examination surveys are not only feasible but may also be desirable under certain circumstances (see, for example, Fisher, Pappas and Limb, 1996.)

In contrast with self-reports, health measurements conducted by a trained nurse or physician will not be affected by reporting error⁴ and for those measures that require no respondent interaction (such as height) there may be no individual-specific measure-specific errors ($\xi_{ij}=0$).⁵ Some measures do involve respondent interaction (such as blowing into a tube to measure lung capacity) and, like self-reports, they are not perfect indicators of θ^* ($\xi_j \neq 0$). Thus, in general, measured health, θ^M is:

$$\theta_{ij}^M = \theta_i^* + \xi_j + \xi_{ij} \quad [2]$$

Since interviews are a good deal cheaper to administer and self-reports have been included in a large number of household surveys, it would be very convenient if the relationships between θ^S , θ^M and intrinsic health, θ^* , were mapped out.

Several very important studies have sought to validate self-reports against clinical measurements; typically these studies have contrasted θ_{ij}^M with θ_{ij}^S and assessed the specificity and sensitivity of the self-report. These comparisons tend to be made with specific health indicators such as morbidities associated with particular diseases. It is usually assumed that ξ_j and ξ_{ij} are both zero so that the difference between self-reported and measured health is attributed to the reporting effects, ε . Typically, self-reports tend to understate the incidence of a health problem ($E\varepsilon_i < 0$) and this difference is greater in the case of diseases that are asymptomatic ($E\varepsilon_j < 0$).

Using data from the United States, for example, Krueger (1957) and National Center for Health Statistics (1965) indicate there is a wide gap between respondent perceptions of health and evaluations by a physician and conclude that self-reports are unlikely to serve as a tool for the measurement of diagnosable disease. Belcher et al. (1976) provide an early example in a developing country setting. Respondent evaluations of morbidities were compared with physician reports of the same people one to four days later in a study conducted in Ghana. They treat the physician assessment as truth and conclude

⁴There may be healthworker-specific measurement error although that can be minimized by good training, field practice and supervision.

⁵This restriction will not apply if the amount of information about health contained in any measure varies with individual characteristics.

there is a tendency for self-reports to understate the incidence of health problems and that the extent of understatement is greater the less readily observed the ailment.

There are, however, several problems that are difficult to detect in a clinic setting (such as problems that emerge intermittently) in which case ξ_j will not be zero. Moreover, even in a clinic, many health problems involve eliciting information from the respondent (such as pain) in which case ξ_{ij} may not be zero. More generally, it is reasonable to argue that clinical measures are not perfect and that in some instances, one's perception of one's health is of legitimate concern to health professionals and policy makers. (See, for example, Zurayk et al., 1993a, 1993b, 1995, for an illuminating discussion in the context of reproductive morbidity.) In general, it is going to be very difficult to separately distinguish the components ε and ξ in [1] and [2].

To illustrate some of the complexities, we begin with a very simple case in which θ^* and θ^M are identical and so ξ is zero; we consider height and weight of adults which are considered to be indicators of nutritional status. (Since our attention is focussed on measurement, we put aside the question of whether anthropometrics are indicative of health status.) The difference, [1]-[2],

$$\theta_{ij}^S - \theta_{ij}^M = \varepsilon_i + \varepsilon_j + \varepsilon_{ij} \quad [3]$$

is reported in Figure 1 using measures and self-reports for the same individuals age 20 through 90 in the third round of NHANES (conducted between 1988 and 1994).⁶ For prime age adults, height is fixed and so any difference between reported and measured height must be due to reporting effects. As seen in the upper panel of the figure, men tend to overstate their height -- by around 1cm until age 50 when the overstatement increases with age. Apparently, as men shrink with age, they do not update their height. Women also overstate their height but the extent of overstatement is small (and not significant) until they reach age 50.

The difference between reported and measured weight is presented in the lower panel. Weight does vary through the day and between days which will affect the variance of the deviation; it should not, however, affect the first moment of the difference which we would expect to be zero. It is not. Whereas, on average, men tend to overstate their weight by nearly two-thirds of a kilogram, women tend to understate weight by nearly 1.5 kgs. (The standard errors are 0.07 kg in each case and so the differences

⁶The figures relate the gap between reported and measured anthropometry with respondent age. The non-parametric locally weighted smoothed scatterplots (Cleveland, 1980) use a tricube weighting function and 20% bandwidth.

are significant.) Weight under-reporting is greatest among younger women, declines with age and, among women in their 80s, reported weight exceeds measured weight.

In principle, if it is known that men tend to report themselves as taller and heavier than they are and that women report themselves as lighter than they are, it would be possible to adjust reported height and weight (taking account of differences by age group) and not incur the costs of conducting anthropometric measures. That conclusion would be premature.

Table 1 reports the correlates associated with the difference between reported anthropometry and measured anthropometry ($\theta_{ij}^S - \theta_{ij}^M$). The age splines essentially replicate the shapes in Figure 1. In household surveys, it is often argued that better educated respondents provide more accurate answers; in the health measurement literature this might be attributed to the better educated having a more complete set of information. However, we see that better educated men tend to over-report height *more* than those who are less educated. Better educated women, on the other hand, tend to report heights that are closer to the truth. For weight, the reverse is true. Better educated men tend to report weights that are closer to the truth whereas better educated women tend to understate their weight more than those with less schooling. Apparently the link between education and accuracy of reporting is not simple.

To probe more deeply, the regressions also include controls for whether the respondent feels he or she is underweight or overweight. They are powerful predictors of the difference between reported and measured weight: those who feel they are overweight tend to understate their weight whereas those who feel they need to put on weight tend to overstate their weight. The same result emerges if we replace these controls with indicators for respondents who are trying to gain or lose weight. (Hope springs eternal?)

Apparently, even in the case of indicators that are easily verified, such as height and weight, the gap between perceptions and reality is large, significant and depends on a host of factors. Putting aside the fact that older respondents, whose height declines with age, do not appear to update their perception of their height, one is tempted in this example to conclude that self reports reflect not only θ^* but also the respondent's own perception of some ideal health status. Men want to be taller and as education increases their ideal height increases; women want to be thinner and as their education increases their ideal weight decreases. And those whose weight deviates from their ideal want to move towards that ideal.

Using the same source of data but examining child height, Strauss and Thomas (1996) observe that the gap between maternal reports and measurements tend to be smaller among higher income and better

educated mothers and the gaps are smaller among older children. They suggest this likely reflects differences in the frequency that children are measured and would arise if mothers of higher socioeconomic status (SES) measure their children (or visit health centers that measure their children) more frequently than women of lower SES. The declining gap across the age distribution is consistent with this interpretation since the rate of growth in height declines with age.

Taken together, these results suggest mapping the association between self reports of health and "true" health will be difficult -- even in cases in which there is no argument about the meaning of "truth". Things become even more complicated in the more general case in which the terms ξ in [2] are not reasonably assumed to be zero.

3. Reported health, measured health and respondent characteristics

The majority of studies that cross-validate reported health with measures of health have contrasted specific symptoms or diagnoses reported by a respondent in an interview with clinical assessments. Many of the health indicators recorded in health interviews are, however, very difficult to cross-validate because there is no obvious clinical counterpart. This is, perhaps, least controversial in the case of general health indicators which are often collected in interview surveys; indeed, many multi-purpose surveys focus primarily on "global" or "summary" health status indicators precisely because the number of items required to "fully" characterize health status is large.

The most commonly collected item is self-reported *general health status* (GHS). One of the most extensively documented relationships to emerge in the literature on health status is that self-reported GHS is a significant predictor of subsequent mortality. (See McCallum, Shadbolt and Wang, 1994; Idler and Benyamini, 1997, provide a recent review.)

It is not obvious what accounts for the strong relationship between self-rated health and subsequent mortality. It is likely that self-reported health is very inclusive with respect to both the range and severity of conditions that it reflects.⁷ Respondents may incorporate knowledge of family history (and thus genetic

⁷Self-reported health may pick up diseases that have begun to affect respondents' health but that have not yet been diagnosed. However, one of the studies that found significant impacts of GHS on subsequent mortality found no effects of GHS on the onset of chronic diseases (Pijls, Feskens, and Kromhout, 1993).

health endowment) into their answer.⁸ Possibly self-rated health is associated with practices and resources that affect subsequent health. In-depth interviews suggest that the presence or absence of particular health problems or of positive and negative health behaviors shapes the responses of the majority of answers to questions on general health status. Moreover, this work suggests that factors which lead to a rating of good health differ from those associated with a rating of poor health (Krause and Jay, 1994). Respondents in "good" health reported choosing this rating on the basis of comparisons with other people far more frequently than did respondents who rated themselves in fair or poor health. All of the respondents in poor health cited health problems or difficulty with physical functioning (not health comparisons or health behaviors) as the reason for their choice of rating (Krause and Jay, 1994). Several studies confirm these findings and the results indicate that it would be incorrect to treat "good" and "poor" health symmetrically (Smith, Shelley, and Dennerstein, 1994).

While one can certainly ask a physician to provide an assessment of an individual's overall health, θ^M , unless the physician has a complete individual and family history, a good working relationship with the respondent, and knows a considerable amount about the respondent's life style, θ^M will measure "true" health, θ^* , with error. Given the model outlined above, it is not clear what to make of comparisons between θ^S and θ^M since we know nothing about ε and ξ .

To provide an illustration, we draw on data from the second wave of the Indonesian Family Life Survey (IFLS2).⁹ In the survey, respondents were asked a long battery of health questions that included

⁸This is not the entire explanation. Even after controlling family history, several studies demonstrate that GHS continues to be a significant predictor of mortality (Pijls, Feskens, and Kromhout, 1993; Borawski, Kinney, and Kahana, 1996; Deeg et al., 1989).

⁹The IFLS is an on-going longitudinal survey of individuals, households, families and communities in Indonesia. The first wave was conducted in the last half of 1993 and included a sample of 7,224 households in 321 communities in 13 of the archipelago's provinces; the sample is representative of 83% of the population of Indonesia. The second wave sought to re-interview IFLS1 respondents four years later in 1997: over 94% of the IFLS1 households were located and interviews were completed with more than 30,000 respondents. Enumerators interviewed all household members (apart from young children who were interviewed by proxy) to obtain comprehensive information on consumption, income and wealth of the family, labor market histories, education histories, migration histories, marriage histories and fertility histories of each household member. While the survey is fundamentally multi-purpose, it does contain a good deal of detail on health status and health behaviors. Each respondent completed a battery of questions about their own health, use of health care, health insurance, smoking and the health of non co-resident family members. In addition, a trained healthworker visited each household and conducted a series of physical assessments in the home. IFLS2 was coming out of the field as Indonesia fell into the biggest economic and financial crisis in three decades and so we returned to the field and conducted a re-survey of 25% of the enumeration areas included in IFLS. That survey, IFLS2+, was completed in late 1998 and the next wave, IFLS3, was completed during the last half of 2000. The surveys provide a unique opportunity to assess the impact of the crisis on the health and well-being of the Indonesian population.

their current general health status, difficulties with a series of Activities of Daily Living (ADLs), questions on specific morbidities (including probes to obtain details) and extensive questions about use of health care. In addition, a trained health worker (a nurse or recently qualified doctor) visited the household separately and conducted a series of physical assessments. These measures were selected to represent different dimensions of health that are of epidemiological relevance to Indonesia (and most developing countries) while also being feasible to field in a household survey setting. For each adult respondent, the health worker measured height, weight, hemoglobin status, lung capacity, blood pressure, and the speed with which the respondent was able to stand up five times from a sitting position.¹⁰ Additionally, at the end of the assessment and after a brief discussion with the respondent, the healthworker rated the person's health relative to someone of average health of the same age and sex using a scale of 1 (for very poor) through 9 (for excellent).

Figure 2 presents the joint distribution of self-reported GHS and healthworker rating. It is comforting that the mean healthworker score declines as self-reports move from good to average and then to poor. As noted above, it has often been observed that the distinction between "good" and "average" is less clear-cut than the difference between "average" and "poor." This is reflected in the scores assigned by the healthworkers. The marginal distributions of the healthworker scores within each GHS category demonstrate that there is a good deal of heterogeneity within each category. There has to be: it is hard to imagine that a construct as complex as health status will be adequately summarized by one trichotomous indicator. The extent of overlap in the marginal distributions across the GHS categories is, however, very substantial and is, perhaps, cause for pause. While it is likely that a complete physical examination, the results of a battery of tests and a family history would reduce the variance in the healthworker scores, the central issue will likely remain: it is not possible to separately identify ϵ and ξ when comparisons are made between self-reports and measures of health when neither can legitimately be construed as revealing some intrinsic level of health.

We turn, therefore, to an alternative strategy in order to better understand the nature of information contained in self-reported and measured health indicators. Contrasting the correlations between respondent

¹⁰Height is measured with Shorr wooden measuring boards, weight with SECA scales specially designed for fielding anthropometry surveys conducted by UNICEF. Blood hemoglobin levels are measured using the Hemocue portable photometer. Lung capacity is measured with Personal Best peak flow meters; each respondent performs the test three times and the top score is used in the analyses below. Blood pressure is measured using an Omran self-inflating meter with a digital readout.

characteristics and self-reported health with the correlations between the same characteristics and measured assessments affords an opportunity to provide a characterization of how the meaning of health varies with those characteristics. For concreteness, consider education as an example. The correlation between it and self-reported health will reflect at least two components: the influence of education on underlying intrinsic health and also the relationship between education and the respondent's perception of his or her own health. In the context of the additive model above, both the underlying dimensions of health and the components associated with reporting effects may depend on socio-economic characteristics, x :

$$\theta_{ij}^S = \theta_{ij}^*(x) + \varepsilon_i(x) + \varepsilon_j + \varepsilon_{ij}(x) \quad [4]$$

Education will be associated with an individual's reporting propensity, ε_i , if better educated people are better informed about their health and are, therefore, more likely to identify a health problem (conditional on a particular level of θ^*). If what is construed as "normal" (or "difficult") varies with education, then that too will be captured in $\varepsilon_i(x)$.

The relationship between "reporting effects" and education is likely to differ depending on the particular self-reported indicators and so ε_{ij} is allowed to be a function of x . For example, a person with little or no education may not know about health ailments that afflict him or her and so may report GHS as very good; a better educated person, however, with the same level of θ^* but more information about it may perceive himself or herself to not be in very good health. In contrast, whether a person has difficulty breathing -- a relatively obvious problem -- may be less prone to "reporting effects" of this form. In general, the more "objective" the indicator -- or the more objectively defined the reference category -- and the more obvious the health problem, the less likely that "reporting effects" will vary with education or any other element of the vector of socio-economic characteristics, x .¹¹

In contrast to self-reports, correlations between socio-economic characteristics and physical assessments will not be affected by individual characteristics unless the assessment involves participation of the respondent (such as performing the puff test) and that participation is related to the characteristic. Thus

$$\theta_{ij}^M = \theta_{ij}^*(x) + \varepsilon_j + \varepsilon_{ij} \quad [5]$$

¹¹See, for example, Mackenbach et al. (1996) who report greater under-reporting of chronic conditions by those with less education in the Netherlands.

and so differences in the effects of x on elements of θ^S and θ^M that are closely related to the same underlying health may be interpreted as being informative about $(\partial \varepsilon_i / \partial x + \partial \varepsilon_{ij} / \partial x)$, that is the "reporting effects" in self-reports. These comparisons will be contaminated if the health indicators differ in their utility as measures of θ^* which argues for basing inferences on multiple health status indicators.

Table 2 presents results of the correlations between a series of health indicators and respondent characteristics for adults age 20 through 80 in IFLS2. In addition to gender, we focus on four respondent characteristics: *per capita* expenditure (PCE) levels in the household which is a measure of longer-run household resources and can be thought of as an indicator of SES; the education and height of the respondent, both of which reflect human capital investments earlier in life and, finally, age. The models include splines in the covariates to permit non-linearities in their effects.

Table 2A focuses on physical assessments conducted by the health worker. The first two reflect nutritional status of the respondent. Body mass index (BMI), $\text{weight}/\text{height}^2$, is thought to be correlated with physical capacity and VO_2max , and extremes of BMI have been shown to be related to elevated morbidity and mortality. For both males and females, BMI tends to increase as household resources increase, particularly among the poorest, and as education increases, at least for those with less than 6 years of schooling. (Among better educated women, BMI declines with education after controlling PCE.)

Low levels of hemoglobin indicate iron-deficiency which has been shown to be linked to susceptibility to disease, fatigue and lower levels of productivity. Hemoglobin levels likely reflect the combination of a diet that is higher in animal proteins (a primary source of iron) and greater absorption capacity (which is reduced by disease insults, loss of blood and also by diets that are high in rice). As with BMI, better educated men have higher levels of hemoglobin; the same is true of women in the lower half of the SES spectrum. Greater attained height is associated with elevated hemoglobin counts suggesting that the effects of early childhood nutrition experiences may reach into adulthood; an alternative interpretation that height is capturing the influence of SES is not very appealing given the latter's relatively modest effect in the regression. Hemoglobin levels of men tend to decline with age; this may reflect the impact of better nutrition among more recent cohorts or the cumulative effect of disease insults among older cohorts. This pattern among women is complicated by the fact that reproductive age women are prone to elevated levels of anemia.

Stature and lung size are positively correlated and so height is a very good predictor of lung capacity. Lung capacity is also higher among better educated and younger males and females. Although urban dwellers tend to be taller, and have higher BMI, there is no difference in lung capacity between urban and rural residents which presumably reflects lower air quality in many urban areas in Indonesia.

The time it takes for a respondent to stand from a sitting position is intended to capture muscular-skeletal problems and primarily reflects lower body motor functioning. Each respondent was asked to repeat the test five times and the total time taken was recorded in seconds. Like lung capacity, better educated and younger respondents perform better on this test.

The fifth physical measure is blood pressure; in the regressions, we estimate the probability a respondent's measured blood pressure is high (≥ 160 , ≥ 100). For both men and women, blood pressure is elevated among urban residents and tends to rise with age until around 60. Blood pressure also increases with education until completion of primary school; among better educated males, the probability of high blood pressure is constant and declines with education among women. Since elevated blood pressure tends to be correlated with weight (and thus rise with SES) and with stress (and thus possibly rise with wages), it is plausible that the relatively modest relationships between SES and blood pressure reflect the effect of treatment among the better off and the better educated.

Finally, the health worker's evaluation of each respondent's general health status is positively associated with PCE, especially among the poorest, with education and with height. There is little association with age -- which is as it should be since the healthworker was instructed to compare the respondent's health with the average for a person of the same age.

Of these measures, only BMI and hemoglobin satisfy the condition that measurement error, ξ , is unlikely to be related to respondent characteristics. For the others, things are a bit more complicated. Lung capacity and sit-to-stand times require interaction between the healthworker and respondent. In the field, not all respondents approached these tests the same: older, higher SES women were sometimes more tentative than other respondents. In our judgement, the impact of these differences is relatively modest. Health interventions that moderate blood pressure are more likely among higher SES respondents, and so we are cautious in our interpretation of those results. While the healthworkers were instructed to only evaluate the health of the respondent, we cannot rule out the possibility that the healthworker was influenced by the respondent's demeanor and attitude which is likely to be linked to SES.

With this in mind, there are several general observations that might be drawn from Table 2A. First, measured health tends to be positively associated with SES -- PCE or education -- particularly in the case of those indicators that do not involve respondent interaction. Second, except in those cases where there are reasons to expect otherwise, the measured indicators suggest that health status declines with age.

Table 2B presents comparable regressions drawing on some of the self-reports of health status collected in IFLS2. Many health interview surveys include questions about ADLs, the respondent's perception of his or her ability to perform physical activities such as walking or lifting heavy items, or activities that are necessary in daily life such as bathing or dressing, Instrumental Activities of Daily Living, (IADLs). We examine two of the questions included in IFLS2 and report, in the first columns of the table, the characteristics associated with a higher probability that a respondent reports having difficulty walking 5 kms (columns 1 and 2) and difficulty carrying a heavy load (columns 3 and 4).

It has been argued that questions about ADLs are easy for respondents to understand as they ask about activities that are well-defined and capture (or are good proxies for) important dimensions of functional health status. Theoretical work has related ADLs to the underlying health dimensions they are intended to tap (Nagi, 1965; Kopec, 1995; Johnson and Wolinsky, 1993). While there is controversy in the literature regarding the success with which ADLs do (and should) match objective measures of functioning (Daltroy et al., 1995; Hoeymans et al., 1996), if predictive power of future health problems is a metric against which to evaluate health indicators, ADLS appear to perform fairly well. Studies have found that scales of physical functioning based on ADLs are significant predictors of subsequent mortality, net of covariates such as age and even self-reported GHS (Reuben, Siu, and Kimpau, 1992; Scott et al., 1997). Nevertheless, they are not without problems. The notions "difficulty" and "heavy" are not explained in the survey and depend on the respondent's own perception. For some respondents, the ADLs are outside of the range of their own experience. And, for many of the ADLs that are commonly included in surveys, few prime age respondents report difficulties completing the activity.¹² This is reflected in the regressions in Table 2B which indicate that the incidence of difficulties increases substantially with age although the age at which difficulties emerge is lower than that typically observed in higher income countries (see Strauss et al., 1993, for a detailed discussion).

¹²This suggests there would be a substantial pay-off to the development of ADL-type questions that discriminate among prime age adults.

The striking result in the regressions in Table 2B is the absence of a negative correlation between SES and difficulty with either ADL (except in the case of men with little education and carrying a heavy load). In fact, among women in the top half of the distribution of PCE, a higher level of PCE is associated with *more* difficulty carrying a heavy load. The same is not true of walking 5 kms nor is it true of men; one is tempted to conclude that the act of carrying a heavy load is not something higher income women view as part of their daily activities, and reporting difficulties is more a reflection of perception of the appropriateness of the activity than physical constraints.

Columns 5 through 8 report the relationship between GHS and respondent characteristics. The characteristics associated with respondents who report themselves as being in good health are in the first two columns; the characteristics associated with being in poor health are in the second pair. Two issues cloud interpretation of these responses. First, it is unclear what reference the respondent uses as the benchmark: is it the average person in the survey (country), in the community or someone in the respondent's peer group? While this can, in principle, be addressed by explicitly specifying the reference group in the interview,¹³ it is likely that a respondent's notion of "good" health will be influenced by his or her own experience and will thus vary with socio-economic status, interaction with the health system and so on.

Results for physical assessments and ADLs indicate that health status tends to decline with age; the age profile for GHS is substantially more muted, especially in the case of respondents who report themselves as being in good health. If all respondents use the average in the country as the reference, we would expect a decline in health status with age that is comparable with the declines observed for physical assessments and ADLs; if the respondents use someone of the same age as the reference, there would be no systematic relationship with age. The estimated age profile suggests that no single reference is used by all respondents and that the reference health amounts to some combination of the national average and the respondent's peer group. This complicates drawing inferences about θ^* with these sorts of data.

Moreover, even among those who report themselves in poor health, there are marked differences between males and females. A significant incidence of poor health emerges much earlier among females (during their thirties) than males (during their fifties) -- a pattern also observed for carrying a heavy load. Waldron (1982) and others have written extensively on this topic and suggest that perceptions of health

¹³See, for example, the Matlab Health and Social Survey in Bangladesh and pilot studies conducted by WHO.

(ϵ_i) differ systematically between men and women. Contrasting the age profiles in Table 2B, it appears that these differences vary with the specific health measure (i.e., in context of the model above, ϵ_{ij} differs between men and women).

The issue of the reference health emerges again when we consider the relationship between GHS and SES. If the reference is the national average, we would expect GHS to improve with SES; if the reference is one's peer group (which is typically comprised of people with similar levels of SES), there is no reason to expect an association. We find no association between SES and the probability that a respondent reports himself or herself as being in good health. There is, however, a negative association between the probability of reporting poor health and education, but only among those who have completed primary school. If these patterns reflect differences in the reference health, they suggest a complex interaction among different measures of SES, levels of SES, age and gender.

The lack of a clearly defined reference health status is not the only problem with interpreting GHS. Rather, the observed patterns probably also reflect differences in perceptions of what is "good" or "poor" health as well as differences in the information respondents have about their health.

Dow et al. (1999) present some evidence supporting the second contention. Drawing on data from two social experiments, one conducted in the United States and one conducted in Indonesia, they show that as the price of health care is increased, there is a decline in use of care. However, GHS tends to improve where prices are raised and, conversely, GHS worsens where prices are lowered. If one were to interpret that evidence at face value, one would conclude that *raising* the price of health care *improves* health. In fact, when physical assessments of health are examined, the evidence suggests that θ^* improved when prices were reduced and respondents saw health professionals more often. Dow et al. conclude that it is not seeing the doctor that makes one sick, but seeing the doctor does increase one's information set and thereby influences one's perception of one's health. Since both experiments involved randomization of the treatment, the analyses are not contaminated by the potential for reverse causality -- that those who see the doctor more often are more likely to report themselves as ill. See Newhouse et al., (1998) for a fuller description of the U.S. experiment.

The last two health indicators in Table 2B are the number of days the respondent reports having had to limit his or her normal activity in the previous four weeks and the number of days spent in bed during that time. Many studies have argued that because these indicators are conceptually very clear and

do not demand judgements of the respondent about the meaning of "poor" or "difficulty," they are likely to be rather good measures of health. We are more sanguine.

It is well known in developed countries that the incidence of reporting limited activity rises substantially when people are applying for or are receiving public assistance that is linked to disability. This is less likely to be a problem in most developing countries where such assistance is not common. It has also been suggested that people who are unemployed (or who have retired) are more likely to report limited activity as a rationalization for their not working.

In low income settings, however, it is not obvious that "limited activity" is well-defined. Moreover, it is quite likely that health problems that limit the activity of a laborer are quite different -- and possibly more serious -- than those that limit the activity of a sedentary worker. The meaning of "limited activity" is likely to vary across SES. It is also likely to vary with employment status: a self-employed person whose income is directly tied to his or her being actively at work is less likely to have limited normal daily activities (failed to go to work) than someone whose income includes such benefits as sick pay (and so may miss a day of work without incurring an income penalty). The meaning of "limited activity" is likely, once again, to vary systematically with SES.

The regression results support this view. In contrast with the physical assessments, for males, there is no association between SES and either the number of days of limited activity or the number of bed days. Among females, the number of days of limited activity declines with education but *increases* with PCE (among women in the lower half of the PCE distribution) suggesting that among these women, as income increases, they are more likely to be sick. We suspect that is not the case and that, rather, they are more likely to report their activities are limited by health problems.

As Sadana (2000) shows, it is possible to construct questions about specific activities which are informative about health. As she notes, these kinds of questions need to be tailored to the particular context and sub-population of interest which limits their value in national and cross-national studies.

To summarize, the evidence in Table 2B suggests that it is difficult to interpret self-reports about the kinds of general indicators of health status that are commonly collected in health interviews. The more specific the question, there may be less ambiguity in the interpretation of the question by the respondent, but, unfortunately, that does not necessarily translate into a more readily interpreted answer by the analyst.

Table 2C turns to specific questions about morbidities. As discussed above, it is these sorts of questions that have been the basis for the majority of the cross-validations of self-reports and clinical assessments and several studies have described their strengths and weaknesses. We will not, therefore, dwell on them.

In IFLS2, questions were asked about whether the respondent had experienced a particular morbidity, reading down a list of 30 items. If so, additional questions were asked about the nature of the morbidity in some cases -- such as the type of coughing -- to elicit more information. Six morbidities are included in Table 2C. In general, the better educated are less likely to report the presence of a morbidity; there is little association with PCE.

The contrast between coughing and breathing difficulties is informative. The regressions suggest higher PCE women are more likely to report they suffer from coughing problems. They are not, however, more likely to report having breathing difficulties and, recall, they do not tend to have lower levels of lung capacity. One suspects that the positive link between SES and reported coughing reflects, in part, differences in what is considered to be a "cough" across the SES distribution, a difference that does not carry through to the more general term, "breathing difficulty." A similar argument might be made for the incidence of reported "nausea" which rises with PCE for both men and women (among those below median PCE). Why does the incidence of coughing differ across the SES distribution? It is possible that poorer women are not aware of coughing whereas higher income women are; while we cannot rule that out, it seems more likely to us that poorer women consider a particular amount of coughing to be normal whereas higher income women report that same amount of coughing as an illness. (See Bhatia and Cleland, 1995.) It is clear that the precise nature of questions about morbidities makes a difference. Sindelar and Thomas (1991) argue that evidence from Peru suggests relative to diagnoses (which often involve interaction with the health system) questions about symptoms are less prone to differences in interpretation across the SES distribution. See also Zurayk et al. (1995) and Murray and Chen (1992).

4. Self reports and physical assessments

If self-reports are contaminated by ϵ , can they be used to say anything about θ^* ? To address this question, Tables 3 and 4 present the relationships between self-reports and physical assessments. Substituting [5] in [4] we estimate:

$$\theta_{ij}^S = \beta' \theta_{ij}^M + \varepsilon_i(x) + \varepsilon_j + \varepsilon_{ij}(x) + \varepsilon_j + \varepsilon_{ij} + u_{ij} \quad [6]$$

to provide an empirical assessment of the links between self-reports of health, θ^S and physical measures, θ^M . If they are unrelated, then β will be zero indicating that the self-report conveys no information about any of the physical assessments.

We begin with GHS in Table 3.¹⁴ The physical assessments do not do a good job of discriminating between respondents in average health and those in good health. Even the healthworker evaluation is unable to discriminate between these states. (In fact, those who take longer to stand from a sitting position are more likely to report themselves in better health!) In contrast, the assessments, θ^M , do much better in isolating those who report themselves as being in poor health; this is true for lung capacity, blood pressure, hemoglobin levels and BMI. It is also the case that the healthworker evaluation has predictive power that is independent of these assessments. (This does not reflect reverse causality since, recall, the healthworker does not know how the respondent answered the questions about self-assessed health.) There can be little argument that poor GHS does contain information about θ^* .

There is, however, cause for caution. In the model [6], SES will predict θ^S if there are unmeasured aspects of health that are correlated with SES. In that case, given the positive correlation between SES and physical assessments, we would expect $\partial\theta^S/\partial x$ to be positive. If, conditional on measured health, there are differences in perceptions of health across the SES distribution, then those will enter through ε in [6]. Those effects are likely to be negative as higher SES is likely associated with a greater propensity to report poor health. Thus, if $\partial\theta^S/\partial x$ is negative, then the reporting effects must dominate the role of unobservables. For the case of females who report themselves as being in poor health, this turns out to be the case: among those below median PCE, as PCE increases, women are more likely to report themselves as being in poor health. A similar pattern emerges among women with 5 years or less of schooling, although the reverse is true among the better educated. We conclude that for GHS the links between θ^S , θ^M and ε are complex and complicate interpretation of self-reported GHS.

Table 4A focuses attention on the two ADLs. The first regression in each pair includes the physical assessments. They all tend to predict difficulties with walking and carrying a heavy load suggesting that the ADLs are also doing a good job of capturing some dimensions of these assessments.

¹⁴The regressions include controls for age (spline) and location of residence. They are suppressed from the tables.

(There are, however, some instances in which the relationships are not as one might expect -- higher lung capacity is associated with greater difficulty carrying a heavy load among men.)

The second regression in each pair adds self-reported GHS to the list of covariates. We have two motivations. First, we wish to assess the extent to which GHS is able to soak up the predictive power of the physical assessments. Given that all the assessments are continuous indicators and GHS is simply a trichotomous variable, the fact that significant coefficients decline by between 10 and 30 per cent indicates GHS does contain information about θ^* .

Our second motivation for including GHS in the regression is to examine the extent to which it is correlated with other self-reported health measures after controlling measured health. The correlation is high and, for the case of poor health, significant. There are two possible interpretations. Poor GHS is reflecting unmeasured aspects of health which are also reflected in the ADLs. Or, we are simply picking up the effect of individual idiosyncratic reporting effects, ε_i , which are common to both GHS and the ADLs. It is not possible to distinguish between these hypotheses with these data.

Table 4B presents results for morbidities and activity limitations; we report only those specifications that include all physical assessments and GHS. Respondents with low lung capacity (and women with low BMI) are more prone to coughing and breathing difficulty; fever likely emanates from a number of sources and is associated with most of the physical assessments; neither joint pain nor nausea are well-predicted with the physical assessments included in IFLS2; BMI appears to predict the number of days of limited activity. For all of these self-reports, GHS is significantly correlated with the self-reports. In the case of days of limited activity, the inclusion of GHS increases the regression R^2 by a factor of four. We suspect that correlated reporting effects, ε_i , is the driving force behind that observation.

The regressions have been re-estimated including PCE and education of the respondent to determine whether reporting differences that are correlated with SES persist after controlling measured health and GHS. The results are reported for a subset of the self-reported health indicators in Table 5; all regressions also include the full set of covariates listed in Table 4.

Controlling measured health and GHS, better educated men are more likely to report having difficulty carrying a heavy load. As PCE increases, women in households with PCE above the median are more likely to report having difficulty carrying a heavy load. It would seem that this ADL is not immune

to differences in perceptions of what is meant by the question across the SES distribution for both men and women.

Women in households with PCE below the median are more likely to report having a cough and suffering from nausea as PCE increases; the same is true for the probability a male reports suffering from nausea. Increases in PCE in these households is also associated with a greater probability a woman reports health has limited her activities.

The fact that SES is not correlated (or is negatively correlated) with the other self-reported health indicators does not mean that those indicators are not contaminated by reporting effects. Rather, the evidence in Table 5 indicates that the propensity to report health problems increases with SES and this propensity dominates the positive relationship between SES and improved health evident in Figure 2. This provides a powerful test of the hypothesis that interpretation of self-reported health indicators is far from straightforward.

5. Conclusions

Using data from NHANES3 and IFLS2, we have examined the links between self-reported health status, physical assessments of health and SES. The empirical results highlight the complexity of the interactions among these indicators -- without even venturing into the even more complex arena of mental health, pain and health difficulties that are very difficult to diagnose.

There can be little doubt that self-reports of health obtained in health interview surveys contain important information about the health status of respondents. There is, however, evidence that the self-reports also contain information about the respondents' own characteristics -- their education, their standard of living, their interaction with the health system and their beliefs about what is "good health." Disentangling these pieces is not straightforward (see, also, Sadana, et al, 2001).

There are likely to be very substantial pay-offs to investments in collecting and disseminating more experimental and quasi-experimental evidence on the links between self-reports, physician assessments and respondent characteristics. Studies of the introduction of disability income for workers whose health prevented them from working has yielded important insights into the way self-reports have responded to changes in economic incentives (Parsons, 1991). There are a large number of other potential quasi-experiments -- health interventions, introduction of new or improved services, information campaigns,

changes in prices of care or changes in health insurance schemes. By exploiting the variation induced by such (quasi-)exogenous changes in the environment that people live in, and by contrasting the impact of the changes on perceptions of health and health assessments, it should be possible to shed some light on the mechanisms that underlie the evolution of self-reported health status.

Now is a good time to conduct these sorts of studies. The rate of technological advances in the measurement of health status is nothing short of stunning, and it is becoming increasingly feasible to conduct tests for many common ailments in a field setting as part of a large-scale household survey that also contains a battery of self-reported health indicators, information about respondent behaviors (such as use of health care, employment, daily activities) and good measures of socio-economic status. If these data are collected longitudinally, it will be possible to exploit the variation that arises as social, economic and demographic environments change either because of shifts in policy or because of external forces. There also seems to be considerable scope for learning about the meaning of self-reports through increased experimentation in questions included in interview surveys that contain large samples of respondents who span a wide spectrum of SES and behaviors.

The collection of survey data that integrates self-reports of health, clinical or physical assessments and includes substantial heterogeneity in respondent backgrounds is an important first step. As those data are placed in the public domain and extensively analyzed, so they will surely serve as powerful tools for scientists to make more headway in understanding the meaning of health status indicators collected in social surveys and, thereby, reveal information about the health of a population that can be interpreted with confidence.

References

- Belcher, D. W., A. K. Neumann, F. K. Wurapa and I. M. Lourie. (1976). "Comparison of morbidity interviews with a health examination survey in rural Africa," *American Journal of Tropical Medicine and Hygiene*, 25.:751-8.
- Bhatia, J. C. and J. Cleland. (1995). Self-reported symptoms of gynecological morbidity and their treatment in South India," *Studies in Family Planning*, 26.4:203-16.
- Borawski E.A., J. M. Kinney and E. Kahana. (1996). "The meaning of older adults' health appraisals: congruence with health status and determinant of mortality," *J Gerontol B Psychol Sci Soc Sci.*, 51.3:S157-70.
- Butler, J.S., R. Burkhauser, J. Mitchell and T. Pincus. (1987). "Measurement error in self-reported health variables," *Review of Economics and Statistics*, 644-650.
- Cleveland, W. S. (1980). "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, 74.368:829-36.
- Daltroy, L. H., C. B. Phillips, H. M. Eaton, M. G. Larson, A. J. Partridge, M. Logigian, and M. H. Liang. (1995). "Objectively measuring physical ability in elderly persons: The physical capacity evaluation," *American Journal of Public Health*, 85.4:558-799.
- Deeg D.J., R. J. van Zonneveld RJ, P. J. van der Maas, J. D. Habbema. (1989). "Medical and social predictors of longevity in the elderly: total predictive value and interdependence," *Social Science and Medicine*, 29.11:1271-80.
- Dow, W., P. Gertler, R. Schoeni, J. Strauss and D. Thomas. (1999). "Health prices, health outcomes and labor outcomes: Experimental evidence," RAND Working Paper.
- Fisher G., G. Pappas, M. Limb. (1996) "Prospects, problems, and prerequisites for national health examination surveys in developing countries," *Social Science and Medicine*, 42.12:1639-50.
- Freij, L. and S. Wall. (1977), *Exploring child health and its ecology: The Kirkos Study in Addis Ababa -- An evaluation of procedures in the measurement of acute morbidity and a search for causal structure*, Umea: University of Umea.
- Hoeymans, N., E. J. M. Feskens, G. A. M. van den Bos, and D. Kromhout. (1996). "Measuring functional status: Cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990-1993)," *J. Clin. Epidemiol.*, 49.10: 1103-1110.
- Iburg, K. M. (2001). "Self-reported and observed measures of health status," Prepared for *Global conferences on summary measures of population health*, Geneva: World Health Organization.
- Idler, E. L., and Y. Benyamini. (1997). "Self-rated health and mortality: A review of twenty-seven community studies," *Journal of Health and Social Behavior*, 38:21-37.
- Jette, A. M. (1994). "How measurement techniques influence estimates of disability in older populations," *Soc. Sci. Med.* 38.7: 937-942.
- Johnson, R. J., and F. D. Wolinsky. (1993). "The structure of health status among older adults: Disease, disability, functional limitation, and perceived health," *Journal of Health and Social Behavior*, 34: 105-121.

- Kooiker, S. E. (1995). "Exploring the iceberg of morbidity: A comparison of different survey methods for assessing the occurrence of everyday illness," *Soc. Sci. Med.*, 41.3: 317-332.
- Kopec, J. A. (1995) "Concepts of disability: the Activity Space Model," *Social Science and Medicine*, 40.5:649-56.
- Krause, N. M., and G. M. Jay. (1994). "What do global self-rated health items measure?" *Medical Care*, 32.9: 930-942.
- Kroeger, A. (1985). "Response errors and other problems of health interview surveys in developing countries," *World Health Statistics Quarterly*, 38:16-33.
- Krueger, D. E. (1957). "Measurement of prevalence of chronic disease by household interviews and clinical evaluations," *American Journal of Public Health*, 47:953-60.
- Mackenbach, J., C. Looman and J. van der Meer. (1996). "Differences in the misreporting of chronic conditions, by level of education: The effect on inequalities in prevalence rates," *American Journal of Public Health*, 8.5:706-11.
- Maddox, G. L., and D. O. Clark. (1992). "Trajectories of functional impairment in later life," *Journal of Health and Social Behavior*, 33:114-125.
- Maddox, G. L., D. O. Clark, and K. Steinhauser. (1994). "Dynamics of functional impairment in late adulthood," *Soc. Sci. Med.*, 38.7: 925-936.
- McCallum, J., B. Shadbolt, and D. L. Wang. (1994). "Self-rated health and survival: A 7-year follow-up of Australian elderly," *American Journal of Public Health*, 84.7:1100-1105.
- Murray, C. and L. Chen. (1992). "Understanding morbidity change," *Population and Development Review*, 18.3:481-503.
- Nagi, S. Z. (1965). "Congruency in medical and self-assessment of disability." *Indian Medical Surgery*, 38.3:27-36.
- National Center for Health Statistics (NCHS). (1965). "Health interview responses compared with medical records," *Vital and Health Statistics*, 2.7.
- Newhouse, J. et al. (1998). *Free for all*, Cambridge: Cambridge University Press.
- Pappas G., P. J. Gergen, M. Carroll. (1990). "Hypertension prevalence and the status of awareness, treatment, and control in the Hispanic Health and Nutrition Examination Survey (HHANES), 1982-84," *American Journal of Public Health*, 80.12:1431-6.
- Parsons, D. (1991). "Self screening in targetted public transfer programs," *Journal of Political Economy*, 99.4:859-76.
- Picavet, H. S. J., and G. A. M. van den Bos. (1996). "Comparing survey data on functional disability: The impact of some methodological differences," *Journal of Epidemiology and Community Health*, 50:86-93.
- Pijls, L. T. J., J. M. Feskens, and D. Kromhout. (1993). "Self-rated health, mortality, and chronic diseases in elderly men: The Zutphen study, 1985-1990," *American Journal of Epidemiology*, 138.10:840-848.

- Reuben, D. B., A. L. Siu, and S. Kimpau. (1992). "The predictive validity of self-report and performance-based measures of function and health," *Journal of Gerontology, Medical Sciences*, 47.4:M106-M110.
- Ross, D. A. and J. P. Vaughan. (1986). "Health interview surveys in developing countries: A methodological review," *Studies in Family Planning*, 17.2: 78-94.
- Sadana, R. (2000). "Measuring reproductive health: Review of community-based approaches to assessing morbidity," *Bulletin of the World Health Organization*, 78.5:640-54.
- Sadana, R. (2001). "Self report versus observed health status: Evaluation of cross-population comparability," Prepared for *Global conferences on summary measures of population health*, Geneva: World Health Organization.
- Sadana, R., C. Mathers, A. Lopez, C. Murray and K. Iburg. (2000). Comparative analyses of more than 50 household surveys on health status, WHO/GPE Discussion Paper 15, Geneva.
- Scott, W. K., C. A. Macera, C. B. Cornman, and P. A. Sharpe. (1997). "Functional health status as a predictor of mortality in men and women over 65," *J. Clin. Epidemiol.*, 50.3: 291-296.
- Segovia, J., R. F. Bartlett, and A. C. Edwards. (1989). "An empirical analysis of the dimensions of health status measures," *Soc. Sci. Med.*, 29.6: 761-768.
- Shusterman, D., J. Z. Kaplan, and C. Canabarro. (1993). "Immediate health effects of an urban wildfire," *West J. Med.*, 158:133-138.
- Sindelar, J. and D. Thomas. (1991). "Measurement of child health: Maternal response bias," mimeo, Yale University.
- Smith A.M., J. M. Shelley, L. Dennerstein. (1994). "Self-rated health: biological continuum or social discontinuity?," *Social Science and Medicine*, 39.1:77-83.
- Stewart, A. and J. Ware. (1992). *Measuring functioning and well-being: The Medical Outcomes Study*, Durham: Duke University Press.
- Strauss, J. and D. Thomas. (1996). "Measurement and mismeasurement of social indicators," *American Economic Review*, 86.2:30-34.
- Strauss, J., P. Gertler, O. Rahman, and K. Fox. (1993). "Gender and life-cycle differentials in the patterns and determinants of adult health," *Journal of Human Resources*, 28.4:791-837.
- Waldron I. (1983). "Sex differences in illness incidence, prognosis and mortality: issues and evidence," *Social Science and Medicine*, 17.16:1107-23.
- Zurayk, H. et al. (1993). "Concepts and measures of morbidity," *Health Transitions Review*, 3.1:17-39.
- Zurayk, H. et al. (1993). "A community study of gynecological and related morbidities in rural Egypt," *Studies in Family Planning*, 24.3: 175-86.
- Zurayk, H. et al. (1995). "Comparing women's reports with medical diagnoses of reproductive morbidity conditions in rural Egypt," *Studies in Family Planning*, 26.1:14-21.

Figure 1: Height and weight of adults
Differences between self-reports and measurements by age
(NHANES3)

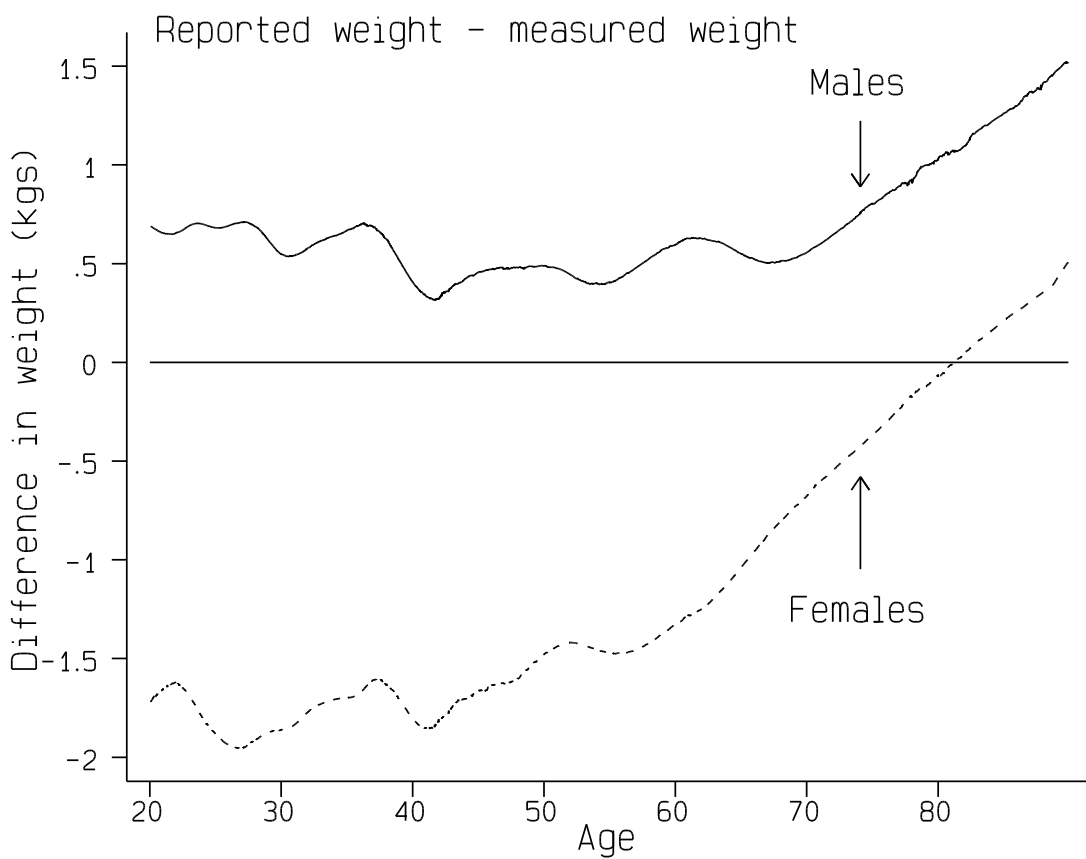
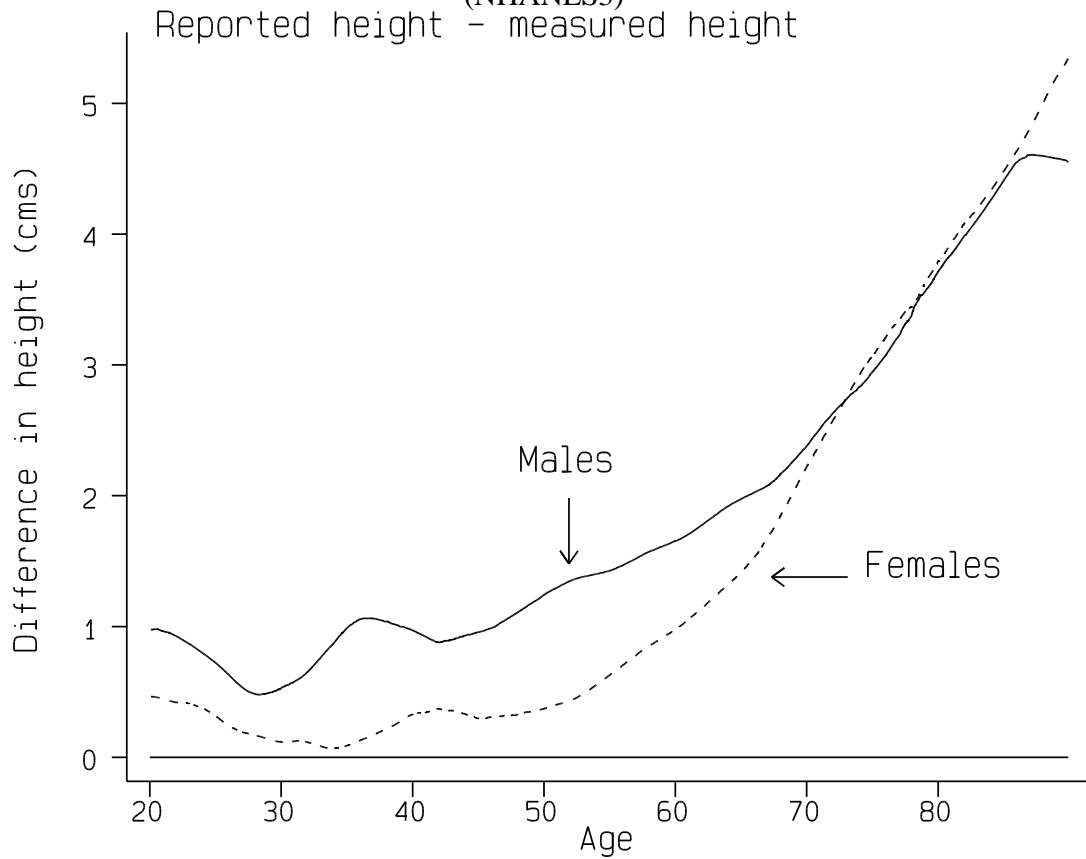
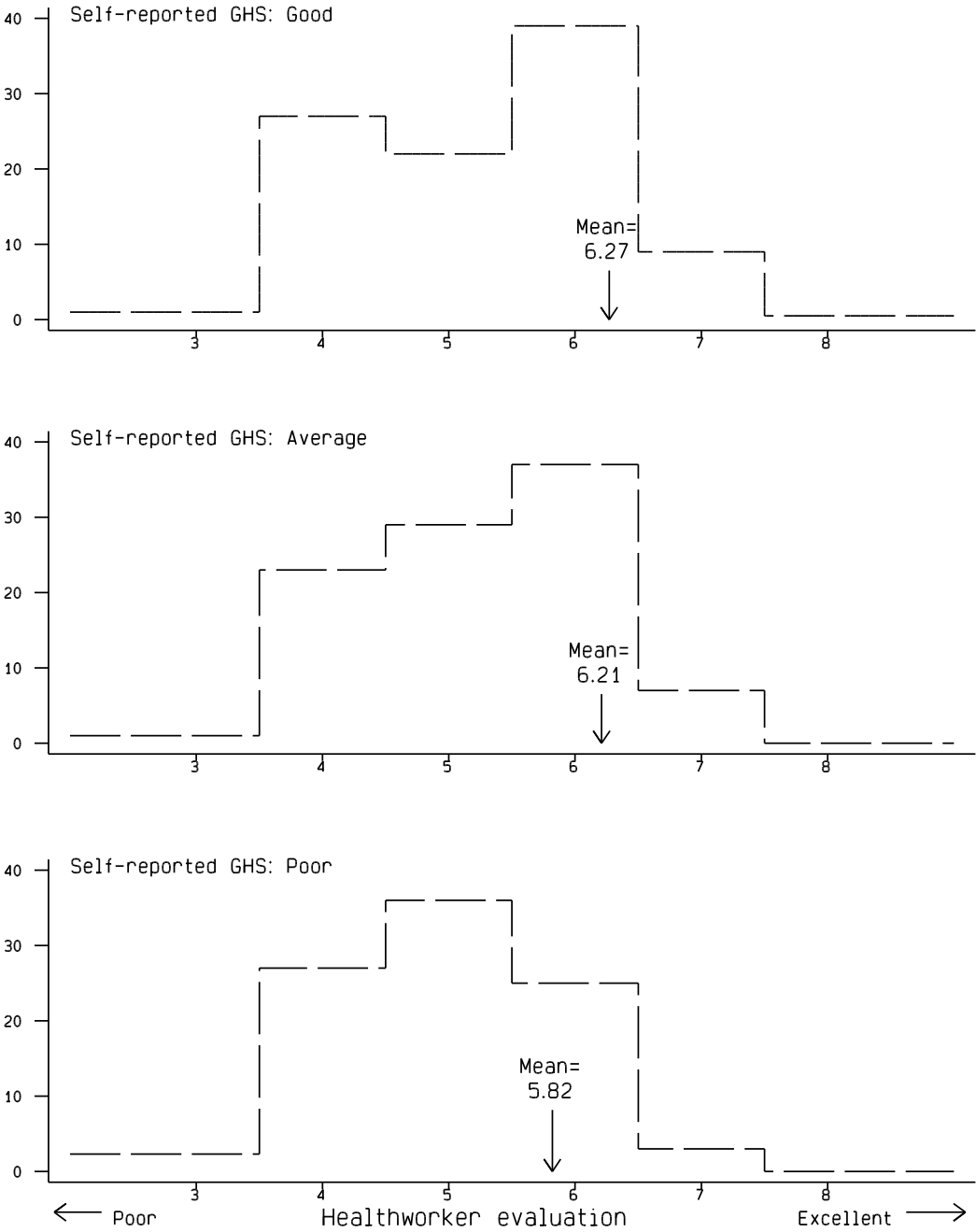


Figure 2: General health status
 Joint distribution of self-reported GHS and healthworker evaluation



Source: IFLS2

Table 1: Difference between reported anthropometry and measured anthropometry
United States NHANES 3

	Reported height- measured height		Reported weight- measured weight		BMI(based on reports)- BMI(based on measurements)	
	Male	Female	Male	Female	Male	Female
Age 20-40 (spline)	0.002 [0.17]	-0.009 [0.91]	0.024 [2.1]	0.027 [1.96]	0.009 [1.97]	0.011 [1.71]
Age 40-60	0.035 [3.56]	0.025 [2.39]	-0.011 [1.24]	0.01 [0.75]	-0.016 [3.69]	-0.005 [0.8]
Age 60-80	0.113 [9.85]	0.152 [12.08]	0.003 [1.86]	0.032 [3.2]	-0.03 [6.24]	-0.033 [5.56]
Yrs education	0.050 [2.75]	-0.104 [4.12]	-0.053 [2.95]	-0.072 [3.32]	-0.035 [4.18]	0.013 [0.94]
ln(per capita income)	-0.085 [1.28]	-0.109 [1.56]	-0.057 [0.73]	-0.02 [0.23]	0.002 [0.06]	0.031 [0.7]
(1) if self-report as overweight	0.12 [1.14]	0.047 [0.39]	-1.541 [12.07]	-1.43 [10.43]	-0.627 [11.74]	-0.645 [9.35]
as underweight	0.048 [0.24]	0.017 [0.05]	1.253 [6.19]	1.069 [5.47]	0.471 [5.77]	0.447 [4.25]
Intercept	0.92 [1.38]	2.63 [4.21]	1.413 [1.95]	-0.737 [0.93]	0.246 [0.82]	-1.102 [2.84]
F(all covariates)	44.51 [0.00]	59.13 [0.00]	40.58 [0.00]	41.13 [0.00]	50.21 [0.00]	27.43 [0.00]
R2	0.096	0.149	0.064	0.058	0.085	0.037

Notes: Sample is 3,870 males and 3,682 females age 20 through 90 interviewed and measured in NHANES3 (1988-1994). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method.

Table 2A: Physical health assessments and respondent characteristics

IFLS2

	BMI		Hemoglobin		Lung capacity		Sit->stand		Blood pressure		HW evaln	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)	Male (7)	Female (8)	Male (9)	Female (10)	Male (11)	Female (12)
In(PCE) below median (spline)	0.17 [6.62]	0.216 [7.47]	0.042 [1.48]	0.057 [2.10]	0.021 [0.76]	-0.004 [0.19]	-0.002 [0.05]	-0.006 [0.18]	0.013 [0.56]	0.002 [0.07]	0.2 [4.97]	0.184 [5]
-- above median (spline)	0.107 [3.41]	0.117 [3.64]	0.055 [1.74]	-0.014 [0.65]	0.041 [1.58]	0.014 [0.85]	0.025 [1.02]	0.018 [0.62]	0.02 [0.61]	0.074 [1.96]	0.066 [1.73]	0.084 [2.3]
Education 0-5 yrs (spline)	0.023 [3.31]	0.072 [8.3]	0.037 [4.17]	0.02 [2.65]	0.032 [4.19]	0.011 [2.07]	-0.019 [1.96]	-0.01 [0.97]	0.021 [2.7]	0.017 [1.95]	0.019 [1.7]	0.028 [2.97]
-- >=6 yrs (spline)	0.034 [7.74]	-0.029 [5.4]	0.02 [4.62]	-0.001 [0.15]	0.03 [7.1]	0.014 [4.16]	-0.018 [4.49]	-0.02 [4.17]	0.002 [0.47]	-0.012 [2.54]	0.028 [4.96]	0.029 [4.87]
Height (cms)	-0.002 [0.95]	-0.003 [1.06]	0.006 [2.83]	0.003 [1.82]	0.025 [11.6]	0.016 [11.27]	0.001 [0.46]	-0.001 [0.35]	0.001 [0.48]	-0.004 [1.48]	0.014 [5.53]	0.01 [4.03]
Age 20-30 (spline)	0.04 [8.42]	0.047 [9.42]	0.000 [0.04]	-0.007 [1.57]	-0.004 [0.86]	0.011 [3.42]	0.016 [3.98]	0.014 [3.5]	0.002 [0.83]	0.008 [3]	-0.006 [1.11]	0.002 [0.38]
Age 30-40 (spline)	0.022 [4.73]	0.032 [5.9]	-0.016 [2.82]	-0.004 [0.97]	-0.011 [2.41]	-0.006 [1.94]	0.016 [4.1]	0.015 [3.69]	0.011 [3.29]	0.01 [2.54]	-0.006 [1.23]	-0.001 [0.27]
Age 40-50 (spline)	-0.01 [2.05]	-0.018 [3.35]	-0.019 [3.33]	-0.001 [0.24]	-0.029 [6.05]	-0.022 [6.93]	0.02 [4.06]	0.022 [4.21]	0.012 [2.16]	0.022 [3.17]	-0.016 [3.16]	-0.011 [2.29]
Age 50-60 (spline)	-0.023 [4.04]	-0.031 [4.2]	-0.02 [3.07]	0.005 [0.85]	-0.049 [8.78]	-0.026 [6.29]	0.037 [5.75]	0.061 [7.41]	0.018 [1.96]	0.031 [2.79]	-0.002 [0.28]	-0.006 [0.88]
Age 60-70 (spline)	-0.009 [1.17]	-0.001 [0.05]	-0.012 [1.48]	-0.014 [1.92]	-0.034 [4.42]	-0.018 [3.46]	0.04 [4.06]	0.041 [3.17]	0.006 [0.39]	-0.007 [0.39]	-0.012 [1.63]	-0.007 [0.88]
Age 70-80 (spline)	-0.05 [4.27]	-0.053 [2.81]	-0.021 [1.32]	0.002 [0.12]	-0.037 [2.83]	0.000 [0.03]	0.051 [2.53]	0.051 [1.91]	0.029 [1.09]	0.057 [1.4]	-0.023 [1.76]	0.013 [0.92]
(1) urban	0.116 [4.24]	0.269 [6.75]	0.045 [1.03]	-0.006 [0.16]	0.067 [1.63]	0.035 [1.09]	-0.015 [0.27]	-0.026 [0.43]	0.072 [3.05]	0.068 [2.28]	-0.063 [0.88]	-0.028 [0.43]
F(all covariates)	30.34 [0.00]	37.03 [0.00]	31.95 [0.00]	8.59 [0.00]	107.43 [0.00]	75.17 [0.00]	50.27 [0.00]	54.38 [0.00]	6.86 [0.00]	11.02 [0.00]	57.13 [0.00]	79.14 [0.00]
R2	0.134	0.116	0.124	0.033	0.341	0.263	0.291	0.297	0.036	0.054	0.306	0.277

Table 2B: Self-reported health and respondent characteristics

IFLS2

	Walk 5 kms with difficulty		Carry heavy load with difficulty		GHS good		GHS poor		# days limited activity		# bed days	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)	Male (7)	Female (8)	Male (9)	Female (10)	Male (11)	Female (12)
ln(PCE) below median (spline)	-0.008 [0.35]	0.042 [1.39]	-0.012 [0.53]	-0.03 [1.05]	0.028 [0.84]	0.000 [0.00]	-0.05 [1.58]	0.034 [1.3]	-0.012 [0.37]	0.079 [3.2]	-0.004 [0.12]	0.018 [0.72]
-- above median (spline)	-0.017 [0.96]	0.017 [0.56]	0.007 [0.37]	0.078 [2.3]	0.007 [0.22]	0.013 [0.49]	0.002 [0.07]	-0.031 [1.25]	-0.047 [1.54]	-0.04 [1.62]	-0.031 [0.82]	-0.047 [2.54]
Education 0-5 yrs (spline)	-0.008 [1.03]	0.001 [0.07]	-0.019 [2.32]	-0.008 [0.95]	0.001 [0.07]	0.005 [0.77]	-0.015 [1.55]	0.01 [1.25]	0.004 [0.42]	-0.019 [2.59]	0.000 [0.02]	-0.018 [2.54]
-- >=6 yrs (spline)	-0.004 [1.2]	-0.005 [1.03]	0.004 [1.33]	-0.002 [0.41]	-0.002 [0.44]	0.003 [0.52]	-0.01 [2.36]	-0.019 [4.01]	-0.006 [1.3]	-0.003 [0.73]	-0.001 [0.17]	-0.001 [0.26]
Height (cms)	-0.002 [1.14]	-0.005 [2.26]	-0.003 [1.73]	-0.008 [3.6]	0.002 [0.89]	0.001 [0.3]	-0.001 [0.18]	-0.002 [0.71]	0.001 [0.19]	-0.003 [1.16]	-0.001 [0.31]	-0.001 [0.2]
Age 20-30 (spline)	-0.005 [1.92]	0.000 [0.04]	-0.001 [0.48]	-0.013 [2.79]	-0.016 [2.13]	0.009 [1.52]	0.003 [0.55]	0 [0.06]	0.003 [0.5]	-0.007 [1.57]	-0.001 [0.26]	-0.01 [2.3]
Age 30-40 (spline)	0.009 [3.04]	0.003 [0.68]	0.004 [1.73]	0.011 [2.38]	-0.005 [0.87]	-0.006 [1.2]	0.001 [0.15]	0.012 [2.78]	-0.007 [1.55]	0.007 [1.79]	-0.004 [0.82]	0.009 [2.23]
Age 40-50 (spline)	0.006 [1.52]	0.036 [6.03]	0.003 [0.93]	0.024 [4.02]	-0.002 [0.37]	-0.01 [1.91]	0.01 [1.94]	0.015 [2.7]	0.012 [1.87]	0.001 [0.24]	0.011 [1.83]	-0.005 [1.44]
Age 50-60 (spline)	0.051 [7.88]	0.048 [5.87]	0.042 [5.84]	0.061 [6.77]	-0.019 [3.15]	-0.011 [2.25]	0.029 [3.92]	0.012 [1.52]	0.021 [2.44]	0.005 [0.59]	-0.009 [1.06]	0.015 [2.1]
Age 60-70 (spline)	0.04 [4.12]	0.029 [2.98]	0.052 [4.7]	0.044 [3.47]	-0.002 [0.39]	-0.005 [1.04]	0.026 [2.2]	0.013 [1.11]	0.000 [0.01]	0.013 [1.17]	0.025 [1.8]	-0.005 [0.44]
Age 70-80 (spline)	0.065 [4.1]	0.038 [2.35]	0.073 [3.45]	0.07 [3.05]	0.000 [0.01]	0.005 [0.56]	-0.013 [0.59]	0.06 [2.11]	0.004 [0.16]	0.043 [1.74]	-0.008 [0.32]	0.052 [1.87]
(1) urban	0.019 [0.92]	0.068 [2.02]	0.022 [1.04]	0.103 [2.94]	0.024 [0.61]	0.056 [1.81]	-0.036 [1.22]	0.046 [1.41]	0.002 [0.05]	0.027 [0.9]	-0.008 [0.23]	-0.002 [0.07]
F(all covariates)	35.74 [0.00]	66.77 [0.00]	20.35 [0.00]	57.54 [0.00]	14.46 [0.00]	19.06 [0.00]	12.99 [0.00]	12.92 [0.00]	4.06 [0.00]	8.23 [0.00]	3.11 [0.00]	2.75 [0.00]
R ²	0.218	0.158	0.201	0.186	0.09	0.06	0.067	0.052	0.018	0.024	0.008	0.014

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method allowing within-cluster correlations.

Table 2C: Self-reported morbidities and respondent characteristics
IFLS2

	Cough		Breathing difficulty		Joint pain		Fever		Diarrhoea		Nausea	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)	Male (7)	Female (8)	Male (9)	Female (10)	Male (11)	Female (12)
ln(PCE) below median (spline)	-0.006 [0.19]	0.08 [3.03]	-0.035 [1.09]	0.03 [1.2]	-0.022 [0.74]	0.037 [1.32]	-0.02 [0.63]	0.046 [1.78]	-0.008 [0.24]	0.034 [1.36]	0.051 [2.03]	0.114 [4.29]
-- above median (spline)	-0.001 [0.03]	0.036 [1.21]	0.01 [0.27]	-0.013 [0.46]	-0.017 [0.58]	0.026 [1.00]	-0.034 [1.3]	-0.001 [0.02]	0.017 [0.55]	0.053 [1.48]	0.007 [0.24]	-0.029 [0.95]
Education 0-5 yrs (spline)	-0.02 [2.06]	-0.02 [2.6]	-0.04 [3.42]	-0.019 [2.65]	0.006 [0.6]	-0.003 [0.33]	-0.011 [1.06]	-0.025 [2.97]	-0.006 [0.77]	-0.001 [0.16]	-0.005 [0.63]	0.01 [1.3]
-- >=6 yrs (spline)	-0.013 [2.85]	-0.006 [1.2]	-0.012 [3.00]	-0.015 [3.65]	-0.012 [2.73]	-0.006 [1.19]	-0.016 [3.56]	-0.014 [2.72]	0.002 [0.45]	-0.009 [1.83]	-0.006 [1.48]	-0.012 [2.26]
Height (cms)	-0.003 [1.21]	-0.004 [1.62]	-0.004 [1.83]	0.003 [1.42]	-0.001 [0.36]	-3.72 [0.00]	0.001 [0.26]	-0.004 [1.74]	-0.001 [0.38]	0.001 [0.38]	-0.001 [0.39]	-0.002 [0.6]
Age 20-30 (spline)	0.002 [0.28]	-0.002 [0.4]	0.002 [0.32]	-0.004 [0.92]	0.014 [3.25]	0.009 [2.11]	-0.006 [1.03]	-0.002 [0.44]	0.006 [0.9]	0.01 [1.91]	-0.007 [1.25]	0.004 [0.6]
Age 30-40 (spline)	-0.009 [1.69]	0.005 [1.07]	0.002 [0.42]	0.007 [1.76]	0.006 [1.22]	0.027 [6.09]	-0.005 [0.97]	-0.003 [0.62]	-0.008 [1.61]	-0.005 [1.00]	0.000 [0.08]	-0.007 [1.26]
Age 40-50 (spline)	0.013 [2.17]	0.009 [1.7]	0.000 [0.02]	0.008 [1.42]	0.02 [3.43]	0.032 [5.46]	-0.004 [0.76]	0.005 [0.89]	0.008 [1.53]	0.001 [0.17]	-0.008 [1.62]	-0.002 [0.37]
Age 50-60 (spline)	0.012 [1.58]	0.000 [0.04]	0.024 [3.25]	0.013 [1.76]	0.018 [2.46]	0.023 [3.02]	0.008 [1.17]	0.003 [0.45]	-0.011 [1.73]	0.005 [0.65]	0.000 [0.04]	-0.003 [0.48]
Age 60-70 (spline)	0.017 [1.88]	0.013 [1.29]	0.034 [2.76]	-0.008 [0.69]	0.032 [3.21]	0.01 [1.01]	-0.008 [0.92]	-0.014 [1.37]	-0.003 [0.43]	-0.009 [0.99]	-0.006 [0.83]	-0.012 [1.48]
Age 70-80 (spline)	0.018 [1.16]	0.011 [0.55]	-0.027 [1.2]	0.028 [1.18]	-0.028 [1.57]	0.035 [1.87]	0.019 [1.18]	0.007 [0.34]	0.01 [0.67]	-0.011 [0.81]	0.002 [0.16]	0.011 [0.66]
(1) urban	0.005 [0.14]	0.028 [0.89]	-0.007 [0.23]	0.037 [1.34]	-0.052 [1.74]	0.011 [0.35]	0.001 [0.02]	0.057 [1.69]	0.029 [0.92]	0.01 [0.32]	-0.017 [0.69]	-0.002 [0.08]
F(all covariates)	10.78 [0.00]	7.67 [0.00]	12.21 [0.00]	10.56 [0.00]	25.48 [0.00]	41.91 [0.00]	6.97 [0.00]	8.81 [0.00]	4.75 [0.00]	3.04 [0.00]	6.84 [0.00]	11.93 [0.00]*
R2	0.042	0.027	0.054	0.037	0.085	0.136	0.031	0.032	0.012	0.013	0.028	0.039

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method allowing within-cluster correlations.

Table 3: General health status, physical health and SES

	GHS good				GHS poor			
	Male		Female		Male		Female	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
ln(PCE) below median (spline)		0.588 [0.63]		-0.268 [0.34]		-0.96 [0.93]		1.702 [2.02]
-- above median (spline)		-0.061 [0.07]		0.265 [0.36]		0.435 [0.5]		-0.74 [0.95]
Education 0-5 yrs (spline)		-0.008 [0.03]		0.094 [0.47]		-0.176 [0.55]		0.478 [1.91]
-- >=6 yrs (spline)		-0.095 [0.7]		0.083 [0.57]		-0.197 [1.45]		-0.571 [3.72]
Lung capacity <median (spline)	-0.006 [0.68]	-0.008 [0.83]	0.003 [0.4]	0.002 [0.2]	-0.098 [4.99]	-0.096 [4.75]	-0.021 [2.05]	-0.017 [1.71]
-- >median (spline)	0.011 [2.12]	0.012 [2.19]	0.009 [0.58]	0.01 [0.62]	0.000 [0.03]	0 [0.03]	-0.018 [1.4]	-0.015 [1.16]
Sit->stand time <median (spline)	0.903 [1.69]	0.938 [1.72]	1.268 [2.47]	1.294 [2.48]	0.559 [1.29]	0.554 [1.25]	0.763 [1.44]	0.818 [1.54]
-- >median (spline)	0.102 [0.55]	0.119 [0.61]	-0.165 [0.99]	-0.136 [0.8]	0.454 [1.4]	0.494 [1.46]	0.376 [1.47]	0.305 [1.19]
BP diastolic <150 (spline)	0.078 [2.09]	0.079 [2.09]	-0.051 [1.49]	-0.051 [1.44]	-0.069 [1.53]	-0.068 [1.47]	-0.062 [1.61]	-0.066 [1.68]
-- >150 (spline)	-0.08 [2.09]	-0.078 [2.01]	-0.024 [0.91]	-0.022 [0.85]	0.163 [2.52]	0.166 [2.54]	0.115 [2.41]	0.112 [2.32]
BP systolic <90 (spline)	-0.025 [0.46]	-0.035 [0.63]	0.057 [1.17]	0.051 [1.04]	0.14 [2.51]	0.143 [2.51]	-0.077 [1.26]	-0.079 [1.27]
-- >90 (spline)	0.074 [0.74]	0.077 [0.75]	-0.102 [1.48]	-0.108 [1.55]	-0.054 [0.41]	-0.074 [0.55]	0.175 [1.52]	0.183 [1.58]
Hemoglobin <12 (spline)	-0.632 [1.07]	-0.552 [0.92]	0.296 [0.94]	0.312 [0.98]	-1.435 [1.58]	-1.199 [1.31]	-1.618 [2.95]	-1.606 [2.89]
-- >12 (spline)	0.634 [2.26]	0.725 [2.52]	0.65 [1.68]	0.701 [1.8]	-0.241 [0.88]	-0.209 [0.76]	0.724 [1.63]	0.762 [1.7]
BMI <18 (spline)	1.079 [1.63]	1.012 [1.48]	0.484 [1.09]	0.501 [1.11]	-4.233 [3.29]	-4.428 [3.38]	-3.363 [3.63]	-3.67 [3.85]
-- 18-28 (spline)	0.038 [0.24]	0.055 [0.32]	0.166 [1.27]	0.156 [1.15]	-0.411 [2.48]	-0.359 [2.16]	-0.137 [0.99]	-0.182 [1.3]
-- >28 (spline)	-0.391 [0.42]	-0.347 [0.36]	0.668 [1.28]	0.673 [1.28]	0.41 [0.45]	0.584 [0.62]	-0.227 [0.48]	-0.286 [0.61]
Height (cms)	0.022 [0.35]	0.018 [0.28]	0.005 [0.09]	-0.006 [0.11]	0.012 [0.16]	0.024 [0.33]	-0.004 [0.06]	0.001 [0.01]
HW evaluation	0.286 [0.51]	0.286 [0.5]	0.096 [0.23]	0.088 [0.21]	-1.601 [3.58]	-1.429 [3.17]	-1.58 [3.2]	-1.551 [3.07]
Age 20-30 (spline)	-0.415 [1.96]	-0.46 [2.14]	0.195 [1.24]	0.214 [1.33]	0.085 [0.55]	0.092 [0.6]	0.14 [1.01]	0.09 [0.65]
Age 30-40 (spline)	-0.127 [0.79]	-0.142 [0.88]	-0.196 [1.39]	-0.187 [1.26]	0.009 [0.07]	-0.029 [0.21]	0.417 [2.94]	0.38 [2.66]
Age 40-50 (spline)	-0.036 [0.26]	-0.021 [0.15]	-0.213 [1.49]	-0.216 [1.49]	0.185 [1.1]	0.165 [0.97]	0.298 [1.68]	0.337 [1.88]
Age 50-60 (spline)	-0.486 [2.89]	-0.496 [2.9]	-0.232 [1.72]	-0.234 [1.67]	0.514 [2.24]	0.563 [2.39]	0.124 [0.49]	0.144 [0.55]
Age 60-70 (spline)	-0.043 [0.26]	-0.058 [0.35]	-0.129 [0.88]	-0.118 [0.8]	0.535 [1.46]	0.431 [1.17]	0.211 [0.57]	0.281 [0.76]
Age 70-80 (spline)	0.054 [0.26]	0.057 [0.27]	0.217 [0.91]	0.21 [0.87]	-1.217 [1.73]	-1.157 [1.63]	1.775 [1.98]	1.694 [1.89]
(1) urban	0.638 [0.64]	0.658 [0.62]	1.544 [1.86]	1.471 [1.65]	-2.008 [2.18]	-1.325 [1.36]	1.414 [1.43]	1.81 [1.7]
F(all covariates)	12.12 [0.00]	10.95 [0.00]	14.27 [0.00]	13.41 [0.00]	12.87 [0.00]	11.28 [0.00]	10.43 [0.00]	10.83 [0.00]
R2	0.092	0.094	0.062	0.064	0.092	0.093	0.067	0.069

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method allowing within-cluster correlations.

Table 4A: Self-reported ADLs, physical health and GHS

	Walk 5kms with difficulty				Carry heavy load with difficulty			
	Male		Female		Male		Female	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Lung capacity <median (spline)	-0.080 [4.11]	-0.057 [2.94]	-0.016 [1.24]	-0.013 [0.98]	-0.051 [2.82]	-0.033 [1.74]	-0.016 [1.52]	-0.013 [1.25]
-- >median (spline)	0.005 [1.02]	0.005 [1.00]	0.003 [0.16]	0.006 [0.31]	0.009 [2.44]	0.008 [2.39]	0.013 [0.92]	0.015 [1.12]
Sit->stand time <median (spline)	0.6 [1.58]	0.466 [1.29]	1.077 [1.49]	0.945 [1.31]	0.56 [1.78]	0.452 [1.54]	0.345 [0.51]	0.209 [0.32]
-- >median (spline)	1.369 [4.05]	1.263 [4.07]	0.971 [3.46]	0.907 [3.28]	0.839 [2.44]	0.753 [2.35]	1.239 [4.52]	1.189 [4.45]
BP diastolic <150 (spline)	-0.065 [1.48]	-0.049 [1.2]	-0.037 [0.76]	-0.026 [0.54]	-0.052 [1.52]	-0.039 [1.19]	-0.017 [0.39]	-0.007 [0.15]
-- >150 (spline)	0.17 [2.6]	0.132 [2.12]	0.129 [2.35]	0.109 [2.01]	0.176 [3.03]	0.145 [2.66]	0.119 [2.35]	0.103 [2.08]
BP systolic <90 (spline)	0.13 [2.2]	0.097 [1.71]	-0.006 [0.07]	0.008 [0.1]	0.067 [1.54]	0.041 [0.94]	-0.029 [0.43]	-0.019 [0.29]
-- >90 (spline)	-0.042 [0.31]	-0.03 [0.23]	0.1 [0.77]	0.071 [0.55]	-0.075 [0.67]	-0.065 [0.59]	0.065 [0.55]	0.042 [0.36]
Hemoglobin <12 (spline)	-1.217 [1.56]	-0.879 [1.25]	0.236 [0.36]	0.51 [0.79]	-1.037 [1.38]	-0.763 [1.09]	-0.905 [1.76]	-0.68 [1.37]
-- >12 (spline)	0.077 [0.3]	0.131 [0.51]	-1.175 [1.94]	-1.299 [2.15]	0.375 [1.85]	0.419 [2.06]	-0.795 [1.65]	-0.913 [1.91]
BMI <18 (spline)	-3.896 [3.64]	-2.909 [2.81]	-3.602 [3.61]	-3.033 [3.03]	-4.037 [3.74]	-3.24 [3.28]	-2.546 [2.39]	-2.075 [1.96]
-- 18-28 (spline)	-0.063 [0.37]	0.034 [0.21]	0.024 [0.12]	0.047 [0.24]	-0.137 [1.00]	-0.059 [0.46]	-0.32 [2.00]	-0.304 [1.92]
-- >28 (spline)	1.067 [0.92]	0.972 [0.87]	2.02 [2.73]	2.057 [2.79]	0.34 [0.49]	0.263 [0.38]	2.169 [3.22]	2.187 [3.22]
Height (cms)	-0.094 [1.51]	-0.097 [1.61]	-0.177 [1.96]	-0.176 [1.96]	-0.107 [1.96]	-0.109 [2.04]	-0.232 [3.15]	-0.231 [3.17]
HW evaluation	-0.166 [0.42]	0.207 [0.53]	0.953 [1.57]	1.22 [2.02]	0.003 [0.01]	0.305 [0.9]	0.512 [0.86]	0.736 [1.25]
(1) self report GHS=poor		0.234 [12.45]		0.170 [9.77]		0.190 [11.86]		0.143 [10.01]
(1) self report GHS=good		0.004 [0.38]		0.002 [0.08]		0.003 [0.46]		0.021 [1.27]
Age 20-30 (spline)	-0.254 [2.2]	-0.272 [2.42]	-0.068 [0.32]	-0.092 [0.44]	-0.032 [0.44]	-0.047 [0.62]	-0.439 [2.72]	-0.463 [2.87]
Age 30-40 (spline)	0.357 [2.85]	0.356 [2.97]	0.127 [0.61]	0.057 [0.27]	0.139 [1.79]	0.138 [1.84]	0.325 [2.02]	0.27 [1.68]
Age 40-50 (spline)	0.086 [0.54]	0.043 [0.28]	1.388 [5.52]	1.338 [5.32]	0.014 [0.13]	-0.021 [0.18]	0.711 [3.49]	0.673 [3.33]
Age 50-60 (spline)	1.818 [6.82]	1.7 [6.6]	1.734 [4.96]	1.714 [4.98]	1.192 [5]	1.096 [4.82]	1.783 [5.74]	1.77 [5.85]
Age 60-70 (spline)	1.398 [3.49]	1.273 [3.33]	1.035 [2.51]	0.999 [2.42]	1.51 [3.99]	1.409 [3.91]	1.188 [2.76]	1.161 [2.75]
Age 70-80 (spline)	2.085 [3.13]	2.369 [3.72]	1.403 [2.09]	1.102 [1.66]	1.984 [2.81]	2.214 [3.14]	2.21 [2.88]	1.951 [2.68]
(1) urban	0.183 [0.22]	0.651 [0.8]	2.862 [2.07]	2.619 [1.92]	0.557 [0.81]	0.935 [1.44]	3.447 [2.97]	3.212 [2.78]
F(all covariates)	29.33 [0.00]	36.28 [0.00]	54.16 [0.00]	61.73 [0.00]	15.53 [0.00]	19.15 [0.00]	46.47 [0.00]	50.47 [0.00]
R2	0.238	0.281	0.166	0.179	0.218	0.264	0.194	0.207

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method allowing within-cluster correlations.

Table 4B: Self reported morbidities, physical health and GHS

	Cough		Breathing difficulty		Joint pain		Fever		Nausea		# days ltd activity	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Lung capacity <median (spline)	-0.071 [3.07]	-0.035 [2.36]	-0.137 [6.89]	-0.048 [5.40]	0.016 [0.80]	0.002 [0.15]	-0.024 [1.26]	-0.020 [1.54]	-0.013 [1.01]	-0.007 [0.75]	-0.088 [0.40]	-0.012 [0.12]
-- >median (spline)	-0.021 [2.53]	-0.046 [2.38]	-0.009 [1.91]	-0.008 [0.79]	-0.014 [1.95]	0.001 [0.06]	0.002 [0.31]	0.013 [0.78]	0.001 [0.22]	-0.012 [0.73]	0.005 [0.08]	-0.142 [1.34]
Sit->stand time<median (spline)	1.781 [2.29]	0.201 [0.25]	0.109 [0.25]	0.147 [0.34]	-0.523 [0.81]	-0.750 [1.01]	1.967 [3.22]	2.210 [3.02]	-0.135 [0.33]	0.500 [0.41]	-1.804 [0.80]	4.129 [0.80]
-- >median (spline)	-0.923 [2.35]	-0.148 [0.48]	-0.053 [0.16]	-0.152 [0.83]	0.573 [1.48]	0.060 [0.19]	-0.241 [0.76]	0.012 [0.05]	-0.299 [0.12]	-0.289 [1.51]	2.008 [0.60]	3.516 [1.30]
BP diastolic <150 (spline)	-0.088 [1.34]	-0.051 [0.89]	-0.054 [1.41]	-0.046 [1.52]	0.015 [0.23]	-0.016 [0.33]	-0.115 [1.86]	-0.073 [1.36]	-0.038 [0.91]	-0.080 [1.93]	-0.872 [1.69]	-0.115 [0.29]
-- >150 (spline)	-0.087 [1.02]	0.050 [0.79]	0.052 [0.98]	0.006 [0.14]	0.029 [0.41]	0.025 [0.39]	0.076 [1.20]	-0.034 [0.71]	0.064 [1.59]	-0.038 [1.03]	0.262 [0.35]	-0.131 [0.26]
BP systolic <90 (spline)	0.039 [0.42]	0.070 [0.86]	0.073 [1.30]	0.052 [1.15]	-0.066 [0.71]	-0.052 [0.73]	0.091 [1.10]	0.074 [0.97]	-0.001 [0.01]	0.050 [0.79]	0.443 [0.64]	-0.880 [1.35]
-- >90 (spline)	0.294 [1.53]	-0.161 [1.10]	-0.002 [0.02]	0.116 [1.15]	0.117 [0.71]	0.074 [0.56]	0.026 [0.18]	-0.011 [0.09]	-0.127 [1.42]	0.084 [0.85]	1.051 [0.66]	1.644 [1.32]
Hemoglobin <12 (spline)	-1.031 [0.98]	0.602 [0.97]	-0.335 [0.52]	-0.398 [1.03]	-1.464 [1.56]	-0.614 [1.07]	-0.559 [0.61]	-0.158 [0.27]	-0.135 [0.22]	0.266 [0.61]	-8.544 [0.84]	6.741 [1.53]
-- >12 (spline)	-0.511 [1.23]	-1.223 [1.87]	-0.189 [0.76]	-0.131 [0.40]	-0.217 [0.61]	0.289 [0.56]	-0.521 [1.35]	-0.972 [1.68]	0.569 [2.01]	-0.651 [1.50]	-2.654 [0.96]	-8.846 [2.30]
BMI <18 (spline)	0.122 [0.08]	-4.451 [4.31]	-0.123 [0.12]	-3.784 [3.77]	2.095 [1.59]	-0.099 [0.10]	-2.089 [1.68]	-2.725 [2.59]	0.676 [0.80]	-0.507 [0.56]	-14.193 [0.95]	-4.309 [0.39]
-- 18-28 (spline)	-0.127 [0.48]	0.249 [1.14]	-0.100 [0.69]	-0.122 [1.02]	0.039 [0.16]	0.434 [2.35]	-0.294 [1.25]	-0.461 [2.43]	0.093 [0.60]	0.030 [0.20]	-2.936 [1.55]	-0.752 [0.44]
-- >28 (spline)	1.200 [0.75]	0.523 [0.66]	-0.312 [0.37]	0.626 [1.46]	1.537 [1.24]	0.152 [0.22]	1.034 [0.82]	0.701 [1.06]	1.665 [1.50]	0.328 [0.62]	25.737 [1.75]	4.985 [0.81]
Height (cms)	-0.08 [0.77]	-0.095 [0.90]	-0.077 [1.20]	0.095 [1.74]	0.008 [0.10]	0.028 [0.30]	0.003 [0.04]	-0.163 [1.74]	-0.053 [0.78]	-0.016 [0.22]	0.241 [0.29]	-0.463 [0.66]
HW evaluation	-0.568 [0.74]	-0.755 [1.10]	-0.020 [0.05]	0.053 [0.14]	-0.207 [0.32]	-0.46 [0.73]	-1.179 [1.85]	-0.918 [1.55]	0.085 [0.21]	-0.311 [0.60]	-5.115 [0.95]	-8.205 [1.59]
(1) self report GHS=poor	0.214 [10.78]	0.157 [8.63]	0.198 [10.84]	0.121 [9.03]	0.220 [11.92]	0.182 [10.75]	0.215 [10.73]	0.197 [10.80]	0.126 [8.01]	0.118 [7.48]	3.679 [12.93]	2.787 [14.19]
(1) self report GHS=good	-0.098 [4.73]	-0.096 [4.91]	-0.020 [2.18]	-0.119 [1.22]	-0.016 [0.92]	-0.343 [2.02]	-0.049 [2.92]	-0.061 [3.77]	-0.032 [3.30]	-0.048 [4.14]	-4.271 [4.32]	-0.484 [5.66]
F(all covariates)	16.31 [0.00]	12.12 [0.00]	13.32 [0.00]	11.95 [0.00]	23.54 [0.00]	38.67 [0.00]	10.08 [0.00]	14.87 [0.00]	8.61 [0.00]	11.43 [0.00]	8.14 [0.00]	12.13 [0.00]
R2	0.070	0.047	0.121	0.072	0.110	0.155	0.062	0.061	0.050	0.052	0.118	0.101

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on Huber-type variance-covariance estimation based on influence function method allowing within-cluster correlations.

Table 5: Self-reported health, physical health and SES

	Difficulty carrying a heavy load		Cough		Difficulty breathing		Nausea		# Days of limited activity	
	Male (1)	Female (2)	Male (1)	Female (2)	Male (1)	Female (2)	Male (1)	Female (2)	Male (1)	Female (2)
ln(PCE) below median (spline)	-0.02 [0.03]	-1.054 [1.09]	0.234 [0.15]	3.783 [2.93]	-0.633 [0.71]	0.848 [1.19]	1.711 [2.22]	3.534 [4.08]	4.941 [0.45]	26.606 [3.09]
-- above median (spline)	0.27 [0.43]	2.764 [2.46]	0.036 [0.02]	1.981 [1.45]	0.361 [0.39]	-0.175 [0.23]	0.1 [0.11]	-0.761 [0.8]	-14.789 [1.52]	-10.99 [1.31]
Education 0-5 yrs (spline)	-0.46 [1.79]	-0.23 [0.81]	-0.674 [1.48]	-0.833 [2.35]	-0.791 [2.47]	-0.452 [2.32]	-0.096 [0.43]	0.34 [1.35]	3.889 [1.4]	-6.906 [2.7]
-- >=6 yrs (spline)	0.188 [2.06]	0.025 [0.14]	-0.498 [2.2]	-0.138 [0.55]	-0.254 [2.21]	-0.368 [3.14]	-0.196 [1.51]	-0.322 [1.93]	-0.808 [0.48]	0.887 [0.59]
F(all covariates)	17.72 [0.00]	47.33 [0.00]	14.6 [0.00]	11.33 [0.00]	12.33 [0.00]	10.83 [0.00]	7.35 [0.00]	10.62 [0.00]	7.41 [0.00]	11.1 [0.00]
R2	0.266	0.21	0.072	0.049	0.122	0.074	0.05	0.055	0.122	0.104

Notes: Sample is 6,435 males and 7,658 females age 20 through 80 interviewed and measured in IFLS2 (1997). t statistics in parentheses based on influence function method allowing within-cluster correlations.