COMMISSION ON MACROECONOMICS AND HEALTH
A WHO commission examining the interrelations among investments in health, economic growth and poverty reduction

CMH Working Paper Series

Paper No. WG1 : 7

Health, Nutrition, and Economic Prosperity: A Microeconomic Perspective

Author:

Duncan Thomas

Date: May 2001 (revised)

Health, nutrition and economic prosperity: A microeconomic perspective

Duncan Thomas
RAND and UCLA

April, 2001

Prepared for Working Group 1, World Health Organization Commission on Macroeconomics and Health. This paper draws on joint work with Elizabeth Frankenberg, James P. Smith, John Strauss and Jeffrey Yau. Financial support from the National Institutes of Health and National Science Foundation is gratefully acknowledged.

1. Introduction

There is a strong positive association between health and economic prosperity. This is observed in both micro and macro data. It has been shown to be true for a wide array of health indicators and for many different dimensions of economic prosperity. However, isolating a causal effect of health on economic prosperity has proved to be very controversial.

It seems reasonable to suppose that causality runs in both directions. Many studies have documented that higher income individuals invest more in human capital, including health. As income grows, individuals invest in higher quality diets, improved sanitation practices and better health care. Moreover, there is some evidence that investments in health are particularly susceptible to income innovations. Some of this evidence is discussed in Section 2 which sets the stage for the remainder of the paper.

Thinking of health as one dimension of human capital, it is likely that improvements in health will, like education, yield returns in the labor force. If a healthier worker is less susceptible to disease, is more alert and has more energy, then he or she will probably be a more productive worker and command higher hourly earnings. Establishing this link is not straightforward. It is, however, very important. If health does affect economic prosperity then evaluations of health interventions that do not take this into account will tend to understate their economic benefits.

Section 3 discusses a key stumbling block in this literature: measurement of health. It is incontrovertible that health is multi-facetted and spans many domains of an individual's life. In contrast with education which is often reasonably well summarized for a particular individual in a single statistic (highest grade attained), no similar summary statistic for health status has been shown to perform as well at the individual level. Moreover, health is a stock that evolves over the life course and many health problems that are associated with early life experiences -- including intra-uterine growth -- only emerge much later in life. We argue that the time is ripe for investment in long term longitudinal studies that provide the information infrastructure necessary to conduct scientific studies that seek to isolate the causal effects of different domains of health on a variety of behaviors. Good measurement of health will form one of the pillars of that information infrastructure.

We turn in Section 4 to the micro-based evidence on the impact of health on prosperity. Three complementary approaches to evaluating how health affects economic and social behaviors are described: experimental studies that involve some form of randomized treatment-control design; quasi-experiments that attempt to construct a randomization scheme by exploiting variation in the social, economic, political or health environment; and observational studies. We argue that each of these approaches has distinct advantages and disadvantages. Greater integration of experimental and non-experimental methods are likely to reap large rewards.

There is a solid body of evidence indicating that some dimensions of nutritional status have a causal effect on work capacity, productivity and, in all likelihood, economic and social prosperity in low income settings. While there are good reasons to suppose that other dimensions of health and nutrition also affect labor outcomes, the scientific evidence demonstrating these links is not as well established. With rapid changes in the technology of health measurement rendering it feasible to conduct physical assessments for many health domains in a field setting, in combination with the dramatic improvement in the quality of socio-economic data from low income settings over the last few decades and the emergence of long-term longitudinal studies, these are exciting times. The scientific community can reasonably expect that there will be many important breakthroughs in our understanding of the causal links between health and economic prosperity over the next few years.

2. Effects of economic prosperity on health

A good deal of evidence indicates that income has a causal impact on health. Evidence at the household level indicates that increases in economic resources are invested in improved diet, better sanitation and health practices, increased use of health services and, possibly, more effective use of these services. (See Behrman and Deolalikar, 1988; and Strauss and Thomas, 1995, for discussions of the evidence at the micro level. Ruger, Jamison and Bloom, 2001, summarize the macro literature. See also the 1993 World Development Report and 1999 World Health Report.)

An important body of evidence demonstrates that, over the long run, there is a positive association between economic prosperity and health of a population as measured by morbidity, mortality or nutritional status. (See, for example, Preston, 1975; Fogel, 1992, 1994, 1999; Steckel, 1995.) Information on the attained stature of adults in the United States and Europe has been combined in creative ways with information on indicators of economic status by Fogel and his collaborators to yield important insights into the links between nutrition and economic prosperity.

A similar strategy using data from developing countries today has the potential to provide further insights into the linkages between health and economic growth. The upper panel of Figure 1 displays the relationship between year of birth and attained height of male and female adults who were measured in the 1997 wave of the Indonesia Family Life Survey (IFLS). There was substantial and essentially uninterrupted growth in attained adult height during the generation prior to the 1955 birth cohort. On average, Indonesian males and females grew by about 1.5 cms in each decade during this period. Subsequent cohorts have not fared as well. Among males born between 1955 and 1962, there was very little improvement in attained height. Although the 1963 through 1966 cohorts did experience substantial improvements in height, by 1967, growth levelled off again. Among females, growth in height was arrested slightly later than for males (in the 1960 cohort) and there is no difference in attained height of all women born during the 1960s and the first half of the 1970s.

The substantial increases in height enjoyed by the cohorts born prior to 1955 suggest considerable improvements in the diet, health practices of parents and the health and sanitary environments these people experienced during early childhood. The economic and political crisis of the 1960s, and the turmoil associated with it, apparently took its toll on the longer-term health of children born during that era of Indonesia's history.

A closer examination of the evidence suggests that the link between economic performance and stature of adults in Indonesia is not straightforward. The relationship between adult stature and year of birth is reproduced in the lower panel of Figure 1 and the relationship is overlaid with estimates of real GNP per capita in the year of birth for the period after the Second World War to the mid-1970s. While there are no reliable estimates of GDP prior to 1945 for what is now Indonesia, we do know that the late 1930s and early 1940s was a period of tremendous disruption; there is no evidence that height of Indonesians was impacted.

During the decade following the end of World War II, Indonesia enjoyed substantial increases in real GDP per capita. In 1954, inflation shot up and although it was quickly brought under control, a short recession ensued with positive economic growth returning in the late 1950s. Hyperinflation emerged in the late 1950s and there was an extended and very deep recession from 1960 though 1966 at which point Indonesian on a path of rapid economic growth for 30 years.

Bearing in mind that adult stature is largely determined during the first few years of life, including the fetal period, the effects of economic downturns during those years in a person's life will plausibly be reflected in attained height of an adult. If the impact of a downturn in the macro-economy differs across households or if there is scope for households to shift resources across time, then the link between economic disruptions and the stature of an average adult may be quite complex. Apparently, it is.

The recession around 1955 appears to have been accompanied by a very substantial decline in the nutritional status of the 1955-1958 birth cohort of males; in contrast, there was very little disruption in the growth of females in this cohort. As the economy expanded during the late 1950s and early 1960s, nutritional status of both males and females improved. But, with the onset of the recession that began around 1962, improvements in the nutritional status of females halted: there are no differences in the attained stature of females born between 1960 and 1975. Males, however, appear to have been protected from the effects of this economic downturn, at least initially. Improvements in their nutritional status continued until the 1967 cohort and then, as with females, halted abruptly. Digging a little further, we find that the recession of the mid 1950s took its greatest toll on rural males and it is they who benefitted most from the economic growth of the late 1950s. The recession of the 1960s affected urban and rural dwellers and it affected children across the entire distribution of socio-economic status.

The fact that changes in the macro-economy did not have the same impact on males and females may reflect gender-specific differences in resilience to reduced health inputs and/or differences in behavioral responses of families to economic fluctuations. Evidence from the most recent crisis in Indonesia suggests that the latter is at least part of the story.

After three decades of sustained economic growth, Indonesia is in the throes of a major economic crisis. In late 1997, the Indonesian rupiah came under pressure and declined from Rp 2,500 per US$ to Rp 4,000. In early 1998, the rupiah collapsed. In the space of a few days, it fell to Rp 15,000 per US$. Inflation spiralled to 80% and GDP declined by 12-15% in 1998. Figure 2 provides some preliminary evidence on the impact of the crisis.

The figures are based on data from the 1997 and 1998 waves of IFLS. Height for age and weight for height z-scores for children under 9 are arrayed against age (measured at the time of each survey) in the left hand panel. There is very little evidence that height of children has been deleteriously impacted by the crisis and, apart from very young children, weight given height is little changed. While reduced weight, conditional on height, of the youngest children is not significant it may be substantively important and it may presage reduced linear growth. (There is no evidence of differences by gender in these data.)

In contrast with the evidence for children, weight has declined substantially and significantly among adults of all ages. The decline is larger for women, relative to men, and it is larger for older adults relative to younger adults. For example, as shown in Table 1, among women age 40 through 65, body mass index (BMI) which is weight divided by height squared has declined by 0.42 kg/m^2 ($p<0.001$) and there has been a 25% increase in the fraction of these women whose BMI is below 18 kg/m^2 ($p<0.001$). These declines likely reflect a combination of change in diet and increases in energy expenditures (since the proportion of people working in the labor market or in a family business has increased since the onset of the crisis). One interpretation of the evidence is that, in the immediate term, adults have sought to protect the nutritional status of their children and grand-children.

Have other dimensions of children's health and well-being been protected from the crisis? The answer is a resounding no. Table 1b summarizes additional results from IFLS. Children's use of health care, particularly public sector care, has declined substantially. The decline is concentrated in preventive services for young children as evidence by a 60% decline in use of the *posyandu*, the community health post where these services are typically provided. There has been some substitution to other sources of care although this substitution is not complete since relative to 1997, in 1998 22% few children under 3 received Vitamin A, a micro-nutrient that helps ward off illness.

It is possible that short term declines in the weight of adults and reduced use of preventive health care among children have modest impacts on well-being in the longer term. If however, adult BMI remains low and if children do not receive the health care they need, then there may be serious health consequences of the crisis. Clearly the medium and longer term effects of the crisis can only be revealed as new data become available. Nonetheless, the evidence from IFLS suggests that households have adopted behaviors that seek to protect children from the deleterious consequences of the crisis. More generally, it is likely that household and family members will respond to changes in the economic and social environment around them when making decisions about investments in health -- and human capital -- as well as decisions about the allocation of time. (See, for example, Pitt, Rosenzweig and Hassan, 1979.) This should be kept in mind when evaluating the evidence on the effects of individual health on economic prosperity.

Prior to discussing some of that evidence, in the next sub-section, we examine the issue of measurement of health status in the context of understanding the links between health and economic status.

3. Effects of health on economic prosperity: Measurement

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. Health status is also hard to measure. 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 studies of the links between health and economic status are to be interpreted in a meaningful way.

Three issues arise in the context of linking health and labor outcomes. First, if different dimensions of health affect time allocation and labor productivity differently, it is important to measure and model those differences. For example, physical strength may be an important attribute in the determination of wages of a manual labor but not among those with more sedentary occupations. A physical injury might have a devastating impact on the earnings capacity of the former but a more muted impact on the earnings of the latter. Moreover, individuals are likely to choose their time allocation and possibly investments in other dimensions of human capital in response to their health status. A less robust person, for example, may choose to acquire skills that are rewarded in the labor force independent of the person's strength. There have, to date, been rather few studies that simultaneously relate multiple dimensions of health to labor outcomes. To a large extent this is a reflection of the availability of data. There is likely to be a substantial return to investments in this regard. It is useful to contrast the literature on education and labor outcomes. Education is also multi-dimensional and includes amount of time spent in school, nature of the curriculum, quality of schooling at each stage, extent of learning in school, post-schooling training and skill acquisition. However, years of schooling has proved to be a very good overall summary of educational attainment and has been the work horse of the literature linking education to labor market success. No such summary indicator has been widely accepted in the literature on the health of individuals.[1]

[1]Life expectancy has proved to be a valuable tool in the arsenal of macro-economists (although see Behrman and Rosenzweig, 1994, for a discussion of the quality and interpretation of those data). At the individual level, future years of life may not adequately capture all domains of current health that affect economic prosperity. First, and most immediately, there are very few surveys that contain such data, particularly in low income settings. Such surveys would need to either interview respondents during their working lives and subsequently identify the date of death either through repeat interviews or linking with mortality records. The former would require very long panels (which are not abundant) and the latter would require good mortality records (which are even less abundant in low income countries). Second, there is limited evidence on the association between longevity and health limitations during one's working life. The link may be quite tight in very poor societies in which people work until they die but rather looser in societies where people live for many years beyond retirement. In that case, the information about domains of health

The fact that health is a stock which reflects both genotype and phenotype influences further complicates measurement. Many genotype influences are difficult to observe and may only be revealed later in life. Moreover, health at a point in time combines the cumulative effects of phenotype factors including an individual's behavior through the life course as well as the health and socio-economic environments to which the individual has been exposed. Capturing all of these influences is extremely difficult. Indeed, recent evidence suggests that exposure to nutritional distress during key periods of intra-uterine growth results in health problems that emerge only in middle or later life such as coronary heart disease, strokes, diabetes and hypertension (Barker, 1997). Barker has shown, for example, if there are inadequate nutrients provided during the fetal period in which arteries are developed, an individual is more susceptible to hardened arteries and vascular disease during adulthood. As a second example, even among normal birthweight babies, short length relative to head size is associated with less developed livers which, in turns, is associated with elevated levels of cholesterol. Evidence from the Dutch famine of 1944-45 yields similar conclusions regarding the role of fetal exposure to nutrition insults. Those people who were exposed to the famine during the last trimester of fetal growth are more likely to suffer from diabetes than similar children born before or after the famine.

The British cohort studies, and similar studies in other European countries, have demonstrated that the evolution of health over the life course is very complex (Wadsworth and Kuh, 1997). For example, low birthweight, shortness at birth or thinness at birth are all indicative of reduced intra-uterine growth and have been shown to be associated with increased prevalence of coronary heart disease and biological risk factors associated with that disease (such as hypertension, non-insulin dependent diabetes, abnormalities in lipid metabolism and blood coagulation). (Wadsworth and Kuh, 1997.) Following a cohort of children from birth, Eriksson et al (1999) show that it is those children who were thin at birth but heavy around age 11 who are at greatest risk of death from coronary heart disease. In fact, they report that a high ponderal index (birth weight/length^3) provides protection against this disease.

Clearly, a snapshot recording an individual's health at a particular point in time will reveal an incomplete picture of the person's health status. If current labor market outcomes are correlated with

---

which affect economic prosperity that is contained in future years of life may vary systematically with the level of income of an individual, community or country.

prior health behaviors which are not observed, interpretation of the relationship between concurrent health and labor outcomes becomes very complicated.

The third, related, measurement issue revolves around the fact that individuals may vary in their propensity to report health problems and that propensity may be systematically related to labor market outcomes. The issue has been raised in the context of disability insurance and early retirement in the United States where it has been argued some individuals report an illness or disability in order to become eligible for health-related social benefits (Parsons, 1982; Bound, 1991). It is likely that the issue is of more general concern.

A large literature discusses the validity and limitations of different health indicators. Murray and Chen (1992) provide a good review. Some of the more 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.

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. The authors treat the physician assessment as truth and conclude 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). Moreover, even in a clinic, many health problems involve eliciting information from the respondent (such as pain) in which case physician reports are themselves, contaminated by differences in the propensities of respondents to report ailments. It is reasonable to argue that clinical measures are not perfect and that in some instances, one's perception of one's health is a legitimate factor in decisions about time allocation and labor market outcomes.

Reported and measured anthropometry

A key complexity that arises in isolating the role of differences in the propensity of individuals to report health problems is that "true" health status is seldom known. To side-step that complexity and focus attention on the role of reporting differences, we examine indicators of health (or, more precisely, nutritional) status which can be measured without any ambiguity: height and weight. We exploit the fact that in the National Health and Nutrition Examination Survey (NHANES), respondents provide self-reports and are measured by a trained anthropometrist. Figure 3 displays the difference between measured and self-reported height and weight for the same individuals age 20 through 90 in the third round of NHANES (conducted between 1988 and 1994).[2] 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 1 cm 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. This 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 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 2 reports the correlates associated with the difference between reported anthropometry and measured anthropometry. The age splines essentially replicate the shapes in Figure 3. 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 overstate 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

---

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

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 health 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 incorporate 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.

BMI is often used as a summary indicator of current nutritional status and reflects, in part, net energy intake. The gap between reported and measured BMI is presented in the third panel of Table 2. Since women tend to under-report weight, BMI tends to be understated with the magnitude of the gap being approximately constant until around age 50 when the gap rises (because under-reporting of weight diminishes while over-reporting of height emerges). Women around age 60 tend to under-report BMI by 1 kg/m^2. Reported and measured BMI of males are very similar for men under age 40. Among older men, the effect of over-stating height results in BMI being under-stated and, for those around age 70, the average understatement is 0.5 kg/m^2. Better educated men tend to understate BMI more (reflecting the combined effect of understating weight and overstating height more as education rises). Among women, education is not related to mis-reporting BMI (as the understatement of both height and weight cancel each other out). Perceptions of ideal weight also significantly affect errors in reported BMI. Respondents who consider themselves to be over-weight tend to report their BMI is 0.6 kg/m^2 less than is measured and those who consider themselves under-weight underestimate BMI by almost 0.5 kg/m^2. BMI based on self-reported height and weight clearly contains measurement error that is not trivial and varies systematically with the characteristics of the respondent in a complex way.

In these analyses, we have sought to determine whether there are systematic differences in reporting of health status, relative to "true" health across the socio-economic distribution which might contaminate analyses of the links between health and labor outcomes. The implications of these results are twofold. First, there are differences in reporting across socio-economic status, as measured by education, which is, in turn, related to labor outcomes. Second, there is evidence of regression towards the mean in self-reported height and weight. If this is also true in other indicators of health status, then the variance of the health of a population will tend to be understated. If deviations from the norm are indicative of poor health, self-reports will likely understate the incidence of poor health. Moreover, whereas in models of the effect of health on labor outcomes, random measurement error in health will result in downward biased estimates of its effect, the "reporting differences" uncovered here are likely to substantially complicate interpretation of relationships between self-reported health and labor outcomes.

Reported health, measured health and respondent characteristics

We turn next to health indicators for which "truth" is not known. 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. The most obvious examples 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.[3] Respondents may incorporate knowledge of family history (and thus genetic health endowment) into their answer.[4] 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, 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, physicians reports will measure "true" health with error. It is, therefore, not clear what to make of comparisons between self-reports and physician reports in this case.

To provide an illustration, we draw on data from the second wave of the Indonesian Family Life Survey (IFLS2).[5] In the survey, respondents were asked a long battery of health questions that

---

[3]Self-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, Feskins, and Kromhout, 1993).

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

[5]The 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: 94.5% 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,

included 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. 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).

This evaluation of a respondent's general health by the healthworker and the respondent's own assessment of general health provides an obvious contrast. While the mean healthworker score declines as self-reports move from good to average and then to poor, there is a good deal of variation in the healthworker scores within each self-reported GHS category. This is to be expected. It is consistent with the notion that health status is complex and not adequately summarized by one trichotomous indicator. It also suggests there is a good deal of heterogeneity within each self-reported category. However, there is also substantial overlap in the marginal distributions of healthworker scores across the GHS categories which may be 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 physician (or healthworker) error from respondent "reporting differences" 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 characteristics and self-reported health with the correlations between the same

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.

characteristics and measured assessments affords an opportunity to provide some insights into 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.

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}:

$$\theta^S_{ij} = \theta^*_i + \varepsilon_i + \varepsilon_j + \varepsilon_{ij} \qquad [1]$$

where θ^* is multi-dimensional and hence vector-valued. 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.[6]

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.

---

[6]It 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.

Allowing underlying dimensions of health and the components associated with reporting effects to depend on socio-economic characteristics, x, yields:

$$\theta^S_{ij} = \theta^*_{ij}(x) + \varepsilon_i(x) + \varepsilon_j + \varepsilon_{ij}(x) \qquad [2]$$

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.[7]

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 a puff test) and that participation is related to the characteristic. Thus

$$\theta^M_{ij} = \theta^*_{ij}(x) + \varepsilon_j + \varepsilon_{ij} \qquad [5]$$

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.

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

Table 3 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 two 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 and the education of the respondent (which combines SES and information about health). The models also control height, age and location of residence of the respondent. (See Thomas and Frankenberg, 2001a, for details.)

Panel A of the table focuses on physical assessments conducted by the health worker. The first two reflect nutritional status of the respondent. Body mass index (BMI), weight/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, the presence of worms, 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.[8]

Lung capacity is higher among better educated respondents who also tend to perform better on the timed sit to stand test which captures muscular-skeletal problems and primarily reflects lower body motor functioning. (Each respondent was asked to stand from a sitting position and then return to the seated position five times. The total time taken was recorded in seconds.) The fifth physical measure is blood pressure; in the regressions, we estimate the probability a respondent's measured blood pressure is high ($\geq$160, $\geq$100 Hg mm). For both men and women, blood pressure 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

[8]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.

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, as well as with education.

Of these measures, only BMI and hemoglobin satisfy the condition that measurement error 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, attitude and home all of which are likely to be linked to SES.

With all of this in mind, there is one key point to be drawn from panel A of the table. 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.

Panel B of the table 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.

The striking result in the regressions in panel B of the table 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, the possibility remains that a respondent's notion of "good" health will be influenced by his or her own behaviors and prior experience and may vary with socio-economic status, interaction with the health system and so on.

To illustrate, consider differences in reported health of 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 (1983) and others have written extensively on this topic and suggest that perceptions of health (ε_i) differ systematically 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 panel B 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.

As noted above, 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. 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. While public provision of disability insurance and unemployment insurance is uncommon in developing countries, there may well be a tendency to report oneself as in poor health if one is not working.

Moreover, in low income settings, it is not obvious that "limited activity" is well-defined. Health problems that limit the activity of a laborer may be 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.

To summarize, the evidence in panel B of the table 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.

Panel C of the table 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 panel C of the table. 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 may 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 that 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).

Further analyses indicate that physical assessments are significant predictors of GHS (and, in particular, seem to be better predictors of whether a respondent reports him or herself as being in poor health). The physical assessments are also able to predict a good deal of the variation in reported ADLs. GHS and ADLs clearly contain important information about the health of a respondent -- and it is likely that they contain information that will be very difficult to reveal in physical measures. A central issue for the interpretation of analyses of the links between health and labor outcomes revolves around whether self reports of GHS and ADLs are also affected by "reporting" differences that are correlated with those labor outcomes. With these issues in mind, we turn next to a discussion of estimation and interpretation of labor outcome functions.

4. Effect of health on economic prosperity: Estimation and interpretation

It is useful to distinguish two classes of studies within the empirical literature on the links between health and economic prosperity at the micro level. Experimental studies involve some form of randomization of treatments and controls and proceed to examine the impact of an intervention on economic outcomes. These studies can provide direct evidence on the causal effect of health on the outcomes studies. Observational studies, in contrast, are typically based on survey data and estimated in conjunction with a model of the behaviors of individuals and households that seeks to provide a plausible argument for interpreting the evidence in a causal framework. We begin with a discussion of experimental studies and then turn to evidence from observational studies.

Experimental studies of nutrition and labor outcomes

The link between nutrition and productivity arguably provides the best documented evidence in the literature on inter-relationships between health and economic prosperity. Moreover, there is evidence that, along with genotype and environmental influences, diet plays a role in the etiology of many chronic diseases. Nutrition provides a good starting point for an assessment of the evidence.

In recent years substantial strides have been made in our understanding the links between nutrition and health in low income settings. In earlier decades, it was generally believed that poor nutrition arose primarily because of inadequate energy or protein and effort was focussed on increasing energy intakes among the poor. As it became apparent that protein-energy malnutrition was only one element in improving the nutritional status of people in low income settings, the focus has shifted towards better understanding the influence of micro-nutrients such as iron, iodine, zinc, calcium and several key vitamins on health and nutrition. This work suggests that labor outcomes are probably influenced by both macro- and micro-nutrients.

Experimental designs are well-suited to isolate the impact of specific nutrients on labor outcomes. Several such studies have demonstrated that there is good reason to believe there is a causal effect of iron deficiency on reduced work capacity. Haas and Brownlie (2001) provide an excellent review.

Iron plays an essential role in oxidative energy production. Iron deficient anemia (IDA) -- that is low levels of hemoglobin (Hb) in combination with abnormal levels of other iron indicators -- is associated with, *inter alia*, greater susceptibility to disease, fatigue and reduced child development. In severe cases, it is associated with elevated infant and maternal mortality. Iron deficiency affects physical activity through primarily two pathways. First, as hemoglobin levels decline, the maximum amount of oxygen that the body can use (aerobic capacity) declines. Second, as iron stores are depleted, the amount of oxygen available to muscles declines, reducing endurance, and the heart works harder to produce the same amount of activity.

Rigorous studies of both animals and humans have demonstrated a causal relationship between iron deficiency and reduced maximum aerobic capacity (VO_2max). For example, studies indicate that experimentally induced anemia results in about a 30% decline in VO_2max whereas iron supplementation for around 12 weeks produces about a 25% increase in $V0_2$max (Celsing et al., 1986; Li et al., 1994; Woodson et al. 1978). There is also evidence that IDA is associated with reduced endurance at below maximal work rates. In contrast, iron deficient (but non-anemic) individuals -- individuals with normal Hb but depleted iron stores as evidenced by, for example, low transferrin saturation or elevated levels of transferrin receptor (TfR) -- may also suffer from fatigue, but there is little evidence that for these individuals iron status has any effect on VO_2max or endurance. (Scrimshaw, 1991; Haas and Brownlie, 2001).

Demonstrating that maximal capacity and endurance are impeded by iron deficiency does not necessarily inform us about the economic consequences of iron deficiency in daily life. Those consequences may be more closely aligned with energy efficiency which is the amount of physiological energy required to perform a given task and is usually assessed by indirect calorimetry. Laboratory studies indicate that iron deficiency impairs energetic efficiency (Zhu and Haas, 1998; Li et al. 1994). Li and collaborators, for example, conducted a randomized treatment-control study of Chinese female cotton mill workers. After 12 weeks of iron supplementation, they found a 5% increase in gross and net energetic efficiencies among the treatments relative to the controls. They also observed a significant reduction in heart rates and a 17% increase in the production efficiency of the women. There was no increase in work output (which was constrained by the technology of the mill) but time spent on leisure activities increased among the treated as did energy expenditure in these activities. A similar finding is reported by Edgerton et al. (1979): iron supplementation of Sri

Lankan female tea plantation workers is associated with increased voluntary activity. These results are important because they suggest iron deficiency affects how time is allocated by an individual.

Whereas causal effects of iron status on aerobic capacity and endurance have been extremely well-documented in both animals and humans, the link with economic success is less well established. Obviously experimental studies that examine the relationship between iron deficiency and work output or productivity are limited to a field setting. The strongest evidence is provided by a longitudinal study of nearly 400 male rubber tree tappers and weeders in Indonesia (Basta *et al.*, 1979). Baseline health measures indicated that 45% of the study population was anemic (Hb<13g/dl). It is likely that hookworm explains a good part of the high levels of anemia (with examination of stools of a 10% sub-sample indicating that worms were present in 88% of the men). Among the anemic, baseline productivity (measured by kilograms of latex collected by tappers per day and the area of trenches dug by weeders) was about 20% lower than the productivity of non-anemic workers. In the experiment, workers were randomly assigned to one of two groups (irrespective of their anemia status). The treatments were given a daily iron supplement (100 mg ferrous sulphate) for 60 days; the controls were given a placebo. Workers received an incentive payment to take the pills as scheduled. At the end of the period, blood hemoglobin, aerobic capacity (measured by the Harvard step test) and output of those who were initially anemic, and received the treatment, increased to nearly the levels of the non-anemic workers (whose biological indicators did not change). Among those in the control group who were anemic, productivity and blood hemoglobin levels also rose, although the increase was substantially smaller than among those in the treatment group. (This is attributed by Basta *et al.* to the effect of the incentive payment, a claim which is corroborated by a comparison of dietary intakes before and after the experiment: they point out that, during the experiment, anemics in the control group spent more on leafy greens as well as other foods which provide higher amounts of iron.)

The results suggest that the output of workers who are IDA can be raised by around 20% through supplementation. This is a very large effect. An issue that is not addressed in the Basta et al. study is the possibility of selective attrition. Although the study involved a sample of 156 tappers, the final analytical sample included only 77 of those workers. If those who attrited from the study were not random with regard to productivity (and productivity change), then the advantages of the experimental design will be dissipated.

In fact, two other studies suggest the effects of iron supplementation may be considerably smaller. Edgerton et al. (1979) report supplementation had small effects on the amount of tea picked by female pickers in Sri Lanka and Li et al. (1994) found small increases in output of female cotton mill workers in China although, as noted above, energy efficiency among these workers did increase significantly and technology impeded productivity improvements at the cotton mills.

While there may be controversy about the precise magnitude of the effect of inadequate iron in the diet on productivity and energy efficiency, there is general agreement that iron deficiency does reduce productivity of adults and cognitive achievement of children. (See Pollitt, 1997 and 2001 for reviews of the literature on child development.) Results from less specific food supplementation interventions are not as clear. Imminck and Viteri (1981a, b) report that sugar cane cutters in Guatemala who received calorie supplements were no more productive than controls. Randomization in this study was at the village level and it may be that changes in productivity between villages during the study confound the estimates. In contrast, calorie supplementation had a small but significant positive impact on the amount of road dug by road construction workers in Kenya where the 47 study subjects were randomized at the individual level (Wolgemuth *et al.*, 1982).

Experimental studies of the effects of nutrient supplementation have several key advantages over non-experimental evaluations. First, it is possible to isolate specific nutrients and examine their impact on outcomes in isolation or in combination with other factors (which may or may not be under the control of the experimental design). It is plausible, for example, that the calorie supplementation of Kenyan road diggers resulted in elevated levels of key micro-nutrients and they are the driving force behind elevated productivity. Among the Indonesian rubber tappers, there was no evidence that energy intake had an independent effect on productivity over and above iron absorption. It is also possible in experiments to examine the effects of supplementation as well as the effects of experimentally reducing nutrients which can be very useful in identifying likely biological mechanisms. Second, and perhaps the greatest advantage of treatment-control designs in the context of the links between health and labor market outcomes is that they provide an opportunity to isolate causal mechanisms.

Implementation of these kinds of focussed studies is not without potential pitfalls. First, if treatments benefit from the intervention and controls do not, there are reasons to expect attrition to be selective. Failure to take that into account can unravel the key advantages of an experimental evaluation. Second, interventions have tended to be targeted at specific individuals. If individuals share the benefits of the intervention with other household and family members, then the effects of the intervention on the subjects will be under-estimated because of these spillover effects. This might arise, for example, if road workers who received a calorie supplement ate less food at home, to make room for other household members. This raises the more general issue noted above that there may be behavioral responses to health interventions and that responses of households may be quite complex. Third, it may be difficult to detect the effects of supplementation in some types of work or institutional settings. As an example, the scope for Chinese cotton mill workers to increase output is limited by the technology of production and the speed with which their co-workers operate. The immediate effects of supplementation on work output among these workers are likely to be muted. However, workers who receive supplementation might reap the benefits of elevated productivity by moving to other tasks, other factories or other jobs in which case short-term studies will tend to miss the benefits of supplementation. Of course, if subjects who move (or change jobs) are not followed then these effects will be missed altogether. The Chinese study also suggests that focussing on productivity effects may miss an important part of the link between health and prosperity. In addition to changes in hours worked and type of work, it is possible that healthier workers will allocate more time to non-work activities including leisure and home production. This will likely result in increased levels of well-being (and possibly even indirect benefits on productivity through elevated levels of functioning). Enhanced productivity at home may also provide benefits for the next generation if, say, parents invest more in their children.

Examining the effects of experimental health interventions on labor market outcomes of subjects, their psycho-social well-being, time allocation and productivity in non-market work are feasible. They call, however, for a modified approach to evaluation that is broader in scope (and includes measurement of a wide array of health, nutrition, social and economic indicators of both subjects and members of their households) as well as longer-run in time frame (to capture the immediate and longer-term effects of interventions). Integrating the benefits of experimental designs with broad purpose repeated observation social surveys is likely to be profitable.

Experimental studies of other dimensions of health and labor outcomes

This point is underscored when we contrast two experimental studies that focus on schistosomiasis. Caused by parasitic worms that live in slow-moving water, repeated exposures to schistosomiasis will usually result in fevers, aches and, often, fatigue. In one study, sugar case workers in Tanzania were randomly assigned to two groups, one of which received chemotherapy. Prior to the treatment, those workers who showed signs of exposure to schistosomiasis cut less cane per day. After treatment, their output increased although not to the same level as those who had no signs of schistosomiasis. (Fenwick and Figenschou, 1972). A slightly earlier study of sugar workers in the Cameroons found no effect of chemotherapy on output of exposed workers (Gateff *et al.*, 1971). In both studies, schistosomiasis was treated effectively. The reason for the difference in results regarding productivity is not at all clear. It may be because the sample size in the Camerooni study is too small to detect effects, the time frame of the study may have been too short or it may be that other, unmeasured factors (such as other dimensions of health) may constrain their productivity.

That said, experiments and quasi-experiments indicate that several domains of health have a causal impact on economic prosperity. A recent experiment in Britain randomly assigned men with back pain to an exercise program or usual primary care management. After a year, the treatments reported less back pain and fewer days of missed work relative to the controls. (Moffett et al. 1999).

Changes in the price of health care have also served as a useful tool for assessing the impact of health on labor outcomes. The RAND Health Insurance Experiment randomly assigned subjects to different combinations of deductibles and co-payments (with the most generous program providing free care). Those who received free care used more health care although the impact on health outcomes was muted except for the poorest and sickest. (Newhouse, 1993.) It turns out that females who received free care increased their labor force participation rate relative to other females; a similar finding emerged for males who had not completed high school.

A similar experiment in Indonesia involved changes in the prices of health services. In the experiment, user fees at public health centers were raised in randomly selected "treatment" districts while prices were held constant (in real terms) in neighboring "control" districts. A baseline household survey was conducted prior to the intervention and the same households were re-surveyed two years later. Health care utilization declined in those areas where prices were increased. In addition, labor force participation declined in the treatment areas, relative to controls, with effects being particularly large (and significant) for men and women at the bottom of the education distribution, those whom we would expect to be the most vulnerable. The most plausible interpretation of both the HIE and Indonesian results is that the average treatment effects on labor supply indicate a causal role of improved health on the allocation of time to the labor market (Dow *et al.* 2001).

Results from Canada support this conclusion. During the 1960s and early 1970s, Canada introduced national health insurance. Exploiting the fact that the introduction of the system was phased across provinces and occupations, Gruber and Hanratty (1995) find that employment and wages increased as workers were covered by national health insurance. The authors conclude that labor demand rose because workers were more productive either because there was increased job mobility and therefore better matching of skills or because they were healthier as a result of being covered by health insurance.

While the introduction of national health insurance was not designed as an experiment, Gruber and Hanratty take advantage of the fact that some people were covered by the system earlier than others. The plausibility of their results rests crucially on the extent to which this "natural experiment" approximates random assignment. (The authors provide a compelling argument in favor of this interpretation.) It is feasible to design health interventions to provide a similar "natural experimental" in order to evaluate the effect of the intervention -- on health status and on other outcomes including economic prosperity. It is unfortunate that there have been relatively few such designs. Programs that seek to reduce, control or eradicate malaria, tuberculosis, onchocerciasis and HIV would seem to be especially fertile ground.

Observational studies of nutrition and labor outcomes

Hands down, the best documented fact in observational studies is that there is a significant return to height in the labor market. As noted above, seminal work by Fogel (1994) documented secular increases in height that parallel economic growth in the historical literature; similar patterns have been documented for many of today's low income countries. At the micro level, many studies have demonstrated that height has a positive impact on hourly earnings. (See Strauss and Thomas, 1998, for a review of the evidence.) While the empirical result is very robust, its interpretation is complex. Taller people are probably stronger -- an attribute that is likely to be more highly rewarded in a low income setting. However, height is much more than a just proxy for strength. Part of height is genotype and, therefore, reflects family background. Moreover, since height is largely determined in early childhood, it reflects investments made by parents when the worker was a young child, including not only investments in nutrition but also broader health and human capital investments.

Counter to that intuition, Behrman and Rosenzweig (2001) find that cross-section estimates of the effect of height on wages is under-estimated; they find that controlling endowments by contrasting monozygotic twins, the effect of height on wages increases three-fold.

In contrast with height, BMI varies through adolescence and adulthood and thus may capture both longer run and shorter run dimensions of nutritional status and health. Clearly BMI is related to energy intake, net of output; it has also been shown to be related to maximum oxygen uptake during physical work (VO_{2max}), which in turn, is related to aerobic capacity and endurance, independent of energy intake (Spurr, 1983, 1988; Martorell and Arroyave, 1988). Whether this is an important pathway through which health may influence productivity is not obvious since many jobs do not require sustained physical effort. Treadmill tests suggest that excess (fat) weight affects the efficiency at which energy is transferred to work output (Cureton, 1992).

Evidence on labor market outcomes is mixed. In higher income settings, high BMI (or obesity) is thought to reduce wages (Averett and Korenman, 1996). Controlling endowments, however, Behrman and Rosenzweig (2001) find that this association disappears and they argue that it likely reflects reverse causality.

In most low income countries, obesity is relatively uncommon (although African women provide an exception). The focus of the literature outside of the developed world has been on the link between low levels of BMI and labor outcomes. Haddad and Bouis (1991), for example, find that BMI has no effect on earnings of rural workers in the Philippines; using the same sample, Foster and Rosenzweig (1993, 1994) report that BMI does affect the wage of workers who earn a time-wage rather than piece-rate. They argue that health is difficult to observe and employers use BMI as a marker for health. In urban Brazil, Thomas and Strauss (1997) find that BMI affects the hourly earnings of both employees and the self-employed, suggesting BMI is more than just a health marker. They argue that BMI is probably correlated with strength since its effect is largest among the least educated who are more likely to do manual labor. Glick and Sahn (1998) report that BMI is associated with elevated wages among self-employed males and females in urban Guinea as well as males who work in the market sector. See also Croppenstedt and Muller (2000) for evidence from Ethiopia. While there is no empirical evidence relating BMI to labor force participation or hours, BMI (or weight for height) has been shown to affect the proportion of working time that is spent on very physically demanding activities by men (Pitt, Rosenzweig and Hassan, 1990; Bhargava; 1997;

Fafchamps and Quisumbing, 1999). It is surely the case that BMI at a point in time partly reflects previous health and human capital investments and so a correlation between BMI and productivity may be capturing the influence of those prior investments. Dynamics between BMI and productivity have not been examined. They are complicated because BMI has both stock and flow dimensions and thus reflects both contemporaneous changes in prices and incomes but as well as prior influences. In addition, there may be complex lags in how changes in BMI are translated into changes in aerobic capacity and endurance. The fact that weight can be drawn down in times of need to convert energy further complicates the dynamics.

Nutrient intakes

Some of the earliest studies of the links between health and productivity focussed on calorie intakes. Treating calorie intake as a choice that is influenced by food prices, Strauss (1986) and Sahn and Alderman (1988) report farm output and wages of males, respectively, rise with calorie intake. However, that result has been called into question by studies that control unobserved heterogeneity in nutrient intake and productivity with an individual-specific fixed effect and find no such link (Deolalikar, 1988; Haddad and Bouis, 1991) although Foster and Rosenzweig (1993, 1994) do find that piece-rate work -- an arguably good measure of productivity -- is affected by calorie intake after controlling individual fixed effects.

Measurement of nutrient intakes is difficult; see, for example, the discussion in Strauss and Thomas (1995). All these studies have used some form of self-reported intake. Food prepared for consumption and left overs were weighed daily for a week in a Brazilian survey of 50,000 households: using those data Thomas and Strauss (1997) report that per capita calorie and protein intakes have a significant impact on wages on both the self-employed and employees. Moreover, they find that wages increase with improvements in quality of diet, measured by the fraction of calories consumed from protein.

That evidence is consistent with results from the INCAP study of school children in Guatemala in which treatments were given a high calorie while controls were not. The treatments were healthier, performed better in school and had greater work capacity ($V0_2$max) in early adulthood (Haas et al., 2000). There is, however, some debate about the interpretation of the intervention which was also rich in several micro-nutrients which have been shown to be related to cognition and work capacity, (Pollitt, 1997).

Knaul (2000) examines the link between wages of adult Mexican women and age at menarche, an alternative indicator of early childhood nutrition. Since age at menarche in this sample was collected retrospectively, she notes there is ample evidence suggesting it is likely to be measured with error and so adopts an instrumental variables approach for estimation. As instruments, she uses indicators of health infrastructure in the community measured at the time of the survey making the assumption, as she notes, that the present distribution of services and resources is correlated with the distribution that prevailed at the time and in the place where the woman grew up. She finds that earlier age at menarche (which is indicative of improved nutritional status) results in higher wages. When she allows the effect of age at menarche to be non-linear she finds that the positive association between wages and early onset of puberty is limited to those women who report age at menarche at 14 or less (which is slightly above the mean); for women who report later age at menarche, wages decline as nutrition improves. These are provocative results. Her findings, however, are not replicated in IFLS. In those data, women were asked age at menarche twice; assuming reporting error is random, the average of the two reports should reduce the influence of this noise. Semi-parametric estimates based on those data indicate no relationship between (average) reported age at menarche and wages.

Observational studies of other dimensions of health and labor outcomes

Relatively few socio-economic surveys contain an extensive battery of physical assessments apart from anthropometric measures. Several surveys contain self-reported health indicators such as GHS, ADLs, days of limited activity and days in bed as well as specific morbidities or symptoms. A small number also contain information about emotional health (see Frank, 2001) and even fewer provide reports of health of parents and siblings although they may be very useful proxies for genotype influences on labor outcomes.

As discussed in the previous section, these indicators contain information about a respondent's perception of his or her health and will reflect important domains of health that are not captured by the physical assessments. Indeed, GHS has been shown to be a powerful predictor of future morbidity and mortality, after controlling a series of other observable health status indicators. (see Idler and Benyamini, 1997, for a review). However, as noted above, self-reports do raise an additional level of complexity in analyses of labor outcomes. (See, for example, Bound, 1991; Mathiowetz and Laird, 1994; Dow et al, 1997.)

ADLs have been used extensively in the study of health status and health care utilization in the United States (e.g., Manning et al, 1982; Manton and Woodbury, 1992) and more recently in developing countries (Strauss, et al., 1993). Several studies in the United States have documented a correlation between ADLs and labor force participation of older men; Stern (1989); Blau, Gilleskie and Slusher (1997); Bound et al (1999). Evidence in LDCs is more limited (but see Lavy, Palumbo and Stern, 1995; Schoenbaum, 1995; Swaminathan and Lillard, 2000) although ADLS have been shown to be strongly correlated with socio-economic status (Strauss et al, 1993). If ADLs reflect physical functioning, they should affect productivity in more physically demanding jobs and also affect type or sector of work.

Studies that have examined the relationship between the number of days ill (or days in bed) and labor outcomes tend to find that hourly earnings and days ill are negatively related (Schultz and Tansel, 1997; Ribero and Nunez, 2000). Murrugarra and Valdivia (2000) use data from urban Peru and treat reported days ill as jointly determined with wages which are measured as income over the last 12 months divided by hours worked over that period. Household sanitary infrastructure and costs of reaching local health centers are assumed to be instruments and appear to perform well (according to over-identification tests). They report that an extra day ill (in the 4 weeks prior to the survey) results in a 1% decline in hourly earnings of male workers in the wage sector and a 3% decline among those who are self-employed. The decline for females is around 2%. They proceed to measure these effects across the distribution of wages and find that among male wage workers, days ill are penalized only at the bottom of the wage distribution. Among the self-employed, the effects are more uniformly distributed.

Putting aside questions regarding the legitimacy of the instruments, the result may reflect differences in labor supply and the nature of employment contracts across sectors and across the distribution of earnings. Specifically, if hours worked in the last year is measured as "normal hours" and does not take into account time spent ill during the prior 12 months, then "wages" given by earnings divided by hours will combine both productivity effects and labor supply effects. For example, a self-employed person who was ill for 1 day would have supplied less labor and, therefore, earned less income (assuming income is related to labor supply). We would expect to see the relationship reported by Murrugarra and Valdivia. If low wage workers are paid by piece or on a time basis, and do not receive wages when they do not work, the same pattern will emerge for those workers. If higher wage workers also receive benefits and so are paid when ill, there should be no link between reported days ill and income over the previous year (as long the number of days ill did not exceed that allowed by the contract). It is, apparently, difficult to disentangle the relationships between health, labor supply, wages and form of contract.

Dynamics between health and economic prosperity

There has been very little study of the dynamic links between health and economic prosperity. This is an important gap in the scientific literature for several reasons. First, health is a stock that evolves over time, and prior health behaviors -- and health shocks -- likely influence current economic status. Virtually nothing is known about the speed with which the effects of health transitions at the individual level are transmitted to the labor market in low income settings. Does a period of poor health (or a negative health shock) put a worker on a permanently lower wage trajectory or do the negative consequences of ill-health dissipate as health subsequently improves? The extent of catch up likely depends on the nature of the health problem, the structure of the labor market and characteristics of the worker including age, education and the extent to which the individual has a buffer of resources on which to draw in bad times.

A second advantage of examining the dynamic inter-play between health and economic prosperity is that it is likely to help pin down some of the mechanisms through which the two are correlated. Apart from the econometric advantages associated with analyses based on repeated observations of the same person over time (see, for example, Deolalikar, 1988; Haddad and Bouis, 1991; Foster and Rosenzweig, 1995), it is also feasible to exploit "natural experiments" that arise from unanticipated variation in the life of a respondent. Smith (1999) provides a clever example.

Using longitudinal survey data of older adults in the United States, he attempts to isolate unanticipated changes in health between rounds of the surveys and measure the impact of those "shocks" on wealth. Under the assumption that the timing of the onset of chronic condition and their severity is largely unanticipated (controlling such factors as smoking, weight, exercise and health of siblings and parents), he finds that mild shocks result in a $3,600 reduction in wealth. Severe shocks, however, bring about a $17,000 or 7% reduction in household wealth. A small part of this reduction can be attributed to out-of-pocket medical expenses. Smith concludes that reduced labor supply because of the health shock is the key factor affecting wealth: he estimates that a severe health shock results in a reduction of 4 hours of work per week and a 15% decline in the probability a 50-60 year old adult is in the labor force.

To illustrate some of the difficulties that observational studies confront, we provide some examples drawn from the first two waves of IFLS, which were conducted in 1993 and 1997. The surveys contain multiple health indicators including anthropometry, physical assessments and self-reports. (See Frankenberg and Thomas, 2001b, for a fuller discussion of the results reported below.) more details.) Both height and weight of all respondents were measured in 1993 and 1997. In 1997, a nurse or recently qualified doctor traveled with the interviewing team and visited each household to record various measures of physical health for each household member. Each healthworker received special training in taking the measurements. Hemoglobin levels were measured using a pin-prick. While few prime age males would be considered moderately anemic (Hb<120g/L), the rate among older males (age >40) rises to over 20%. To detect cardio-pulmonary problems, blood pressure and lung capacity were measured. Lung capacity also reflects strength (and is therefore related to height). At the end of the physical assessment, the healthworker evaluated each respondent's health status on a 9-point scale (1=poor, 9=excellent) and recorded comments about the individual's health.[9]

Respondents in all waves of IFLS are also individually interviewed by a trained enumerator who asks detailed questions ranging over their entire life histories. These include a battery of questions about health status and use of health care. We examine two classes of health status indicators: self-reported general health status (GHS) and difficulties performing basic and intermediate tasks associated with daily living (ADLs). About 12% of men reported themselves as being in "good" health, and about 10% in "poor" health. While these fractions remained approximately constant 1993 and 1997, a very large fraction of respondents transitioned into (or out of) good (or poor) health during this time. About 8% of men reported having difficulty carrying a heavy load and 16% reported having difficulty walking one kilometer.

---

[9]Measures of weight were taken using Seca UNICEF scales, and recumbent length or standing height was measured with Shorr measuring boards. Both instruments have been used in survey work in other countries and are suitable for fieldwork given their portability, durability, and accuracy. The floor-model scales have a digital readout and are accurate to the nearest 0.1 kg. Children who were unable to stand on their own were held by a parent and weighed (after the scale had been adjusted to zero with just the parent alone on the scale). Standing height was measured for adults and children over age 2, and recumbent length was measured for younger children. Blood pressure and pulse were measured with an Omron digital measuring device. Hemoglobin was assessed using the hemocue method. Three measurements of lung capacity were recorded using Personal Best peak flow meters.

We focus on the correlations between these health indicators and three labor outcomes measured in 1997: labor force participation (estimated as a probit), the logarithm of hours worked (during the previous month) and average hourly earnings (in the previous month). To highlight some of the advantages of thinking about the dynamics underlying the links between health and economic prosperity, the repeated observation nature of the survey is exploited and health indicators from 1993 and 1997 are included in the regressions. (The regressions include controls for age and education of the respondent, province and sector of residence, measures of community infrastructure that might be related to labor demand and an overall price index. Estimates of variance-covariance matrices take account of correlations in unobservables that arise because of the cluster design of the surveys.)

We begin with anthropometry. Taller men earn significantly higher wages. More robust men, as indicated by greater weight, also earn higher wages and they tend to work longer hours. Since the regressions control BMI in 1993, BMI in 1997 can be interpreted as weight gain between 1993 and 1997: weight gains are associated with elevated wages and greater work effort. (There is some evidence that the association between BMI and wages is not linear: it is flat among those men with BMI < 18.5. Since we are unable to detect this non-linearity in the regressions presented below, we adopt a linear specification throughout.)

Males who report themselves as currently being in poor general health are less likely to be working and they work few hours. Since GHS in 1993 is controlled, it appears to be transitions into poor health that are associated with reduced work. This result parallels Smith's evidence on wealth decumulation in the United States. Having difficulty walking a kilometer and carrying a heavy load are also associated with reduced labor force participation and, in the case of carrying heavy loads, the effects appear to be persistent. However, conditional on being employed, those who have difficulty walking a kilometer earn higher wages. It is hard to imagine this correlation is not dominated by unobserved heterogeneity: it seems very likely that it reflects systematic reporting differences that are related to socio-economic status as discussed in Section 3 above.

The final panel of the table displays the association between labor outcomes and physical assessments (which were conducted in 1997 and later waves of IFLS). Controlling BMI and self-assessed health, elevated blood pressure is associated with reduced labor force participation (at a 10% size of test). Hemoglobin levels are associated with increased wages and there is a suggestion that males in households that use iodized salt earn higher wages.[10] At the end of the physical assessment in IFLS2, the healthworker evaluated each respondent's health status on a 9-point scale (1=poor, 9=excellent). Controlling all other health indicators, the healthworker evaluation of the respondent's health status has no predictive power on labor outcomes.

One approach to assessing whether these correlations reflect any causal effects of health on labor outcomes is to re-estimate the models, replacing 1997 labor indicators with 1993 indicators. If, conditional on current health (in 1993), future health (in 1997) is correlated with labor outcomes in 1993 then we can conclude that unobserved heterogeneity likely influences both health and labor outcomes. Results from those models (which are not reported) indicate that BMI in 1997 is correlated with wages in 1993 (controlling BMI in 1993) suggesting that weight gain in 1997 may be the result of higher income and not vice-versa. Similar results emerge for poor GHS in 1997 and whether the household uses iodized salt. Reverse causality is apparently a serious concern in these models.

There are two standard approaches to dealing with unobserved heterogeneity. We begin with instrumental variables estimates. In both waves of the survey, detailed data were collected about respondents' communities and about public and private facilities available for health care, the price of services and the quality of services. These characteristics serve as instruments along with relative prices of ten foods (which are also collected at the community level in IFLS). Plausible arguments can be made that the instruments are invalid (because of program placement or selective migration, for example); our feeling is that those concerns are most troubling for health infrastructure characteristics whereas *relative* prices of foods do not belong in wage functions (of those not working in the food production sector). Labor supply functions are considerably harder to identify. With this concern in mind, the first stage regressions does a reasonably good job of predicting the health indicators in the sense that F tests for joint significance of the identifying instruments are significant. However, these test statistics are not large (usually in the range of 2 through 12) which is further cause for pause. (Overidentification tests indicate that the instruments are not correlated with unobservables in the labor outcome functions.) For the sake of comparability with the next set of estimates, we have dropped the physical assessments conducted only in 1997. The results are presented in Table 5 and indicate there

[10] In an effort to determine whether household members were likely to be iodine-deficient, the iodine content of salt in the household was tested; about half the population consumes salt that has not been iodized.

is rather little evidence in support of a causal effect of health on labor outcomes except that taller men earn more and those men who have difficulty carrying a heavy load earn lower wages.

If the source of unobserved heterogeneity is a time-invariant respondent-specific factor, its influence in the regressions may be removed by including a person-specific fixed effect. In essence, the regressions amount to examining the association between changes in labor outcomes (participation, hours or wages) and changes in health status. Of course, all individual characteristics that do not change through time are swept out of the model. This applies to observed characteristics (such as height and inherited health status) as well as to characteristics that are not observed (such as the propensity to report oneself as being in poor health). To the extent that "reporting propensities" can be treated as linear and additive effects, they are swept out of these models.

Fixed effects estimates are reported in Table 6a. The results are strikingly different from those based on instrumental variable estimation. Low BMI, poor GHS, difficulty walking a kilometer and carrying a heavy load are all associated with reduced labor force participation. Poor GHS is associated with lower hours of work among those working. BMI is the only health indicator that is significantly related to wages.

The differences between the instrumental variable and fixed effects estimates may be because the instruments are invalid (or weak or both) or because the assumptions of the fixed effects models are wrong. We can directly test one of those assumptions. Recasting the model in terms of changes in labor outcomes and changes in health status, the model assumes that prior health has no direct effect of changes in labor outcomes. This imposes assumptions about the persistence of health innovations that are easily tested by including prior health in a model of the change in labor outcomes. Results are presented at the foot of the bottom panel of Table 6b. In the case of labor force participation, it is clearly rejected. The assumption is not rejected in the case of hourly earnings.

Intuitively, the fixed effects model assumes that changes in BMI cause changes in wages. We noted above, however, that 1997 BMI is related to 1993 wages, suggesting that causality might run in the opposite direction. This insight calls for combining fixed effects with instrumental variables estimation. Natural instruments are changes in relative food prices. Results are presented at the bottom of Table 7 for wages. (We do not provide results for participation since the assumptions underlying the fixed effects model were rejected and because changes in relative prices are not likely to be good instruments.) Although the instruments appear to be valid (in the sense that they are not correlated with unobservables in the wage function), they do not predict changes in BMI very well. (The F statistic is significant but fairly small.) The estimated effect of BMI on wages is very large and significant. The magnitude of the effect is, arguable, implausibly large and suggests that the variation in prices in these data is not sufficient to pin down the causal effect of BMI on wages.

This summary of results from IFLS provides some insights into the types of difficulties that are encountered in observational studies. The example highlights the importance of taking into account unobserved heterogeneity in modelling the link between health and labor outcomes -- and also highlights that there are likely to be complex dynamics underlying these linkages.

5. Conclusions

A substantial body of literature indicates that nutritional status affects labor outcomes, particularly productivity. While the exact mechanisms underlying these relationships are not entirely clear, this literature is distinguished by the co-existence of carefully conducted experimental studies and observational studies that have documented sizeable effects of nutrition on productivity indicators.

The most convincing evidence in this domain has documented there is a causal impact of iron deficient anemia on work capacity, energy efficiency and productivity. The tight link between results that have emerged from many of the animal and human experimental studies and explanations that are based on the underlying biological mechanisms that link iron absorption with cell functioning has proved to be a powerful force in support of the conclusion that iron deficiency reduces work capacity. Yet important gaps remain. The link between iron deficiency and higher hourly earnings has been less well-established. It is plausible that as iron deficiency is alleviated resulting in an increase in an individual's work capacity, hourly earnings will increase, hours of work will increase and income will rise. It is also plausible, however, that part of these benefits will be taken in greater leisure or investment in home production activities such as child rearing or community activity. The relative importance of these pathways has not been explored.

Much less is known about the link between other micro-nutrients and economic prosperity although biological evidence suggests an important role for several vitamins and minerals as well, possibly, for macro-nutrients. Both experimental and observational studies that focus on these mechanisms are likely to be extremely valuable.

A plausible argument can be -- and has been -- made that other dimensions of health likely affect economic prosperity. Relative to the links between nutrition and labor outcomes, these links have not been as well established in the scientific literature. In part, this is a reflection of limited data on non-nutrient physical health indicators, particularly in observational studies. In part, it is a reflection of the difficulties associated with using self-reported health status. Creative design of experimental studies that might shed new light on the meaning of different health indicators is likely to have a considerable pay-off. For example, it might be of value to assess whether GHS or psycho-social well-being responds to a health intervention -- be it a change in the price of health care, quality of care or nutrition supplementation -- and link those results to other behaviors (such as use of health care) or health indicators (such as iron deficiency).

Remarkably little is known about the dynamics linking innovations in health and innovations in economic prosperity. This is a serious gap in our knowledge and reflects the paucity of long term longitudinal studies that contain detailed information on health as well as on economic status. Very few surveys have systematically collected physical assessments of health and so most studies are forced to rely on self-reported health indicators. This substantially complicates the analysis and interpretation of the evidence. With the rapid changes in technology of measuring many health indicators, it would seem that surveys might profitably invest in adopting some of those technologies.

There is also very limited evidence on how individuals cushion themselves from bearing the full brunt of the burden associated with changes in health, be they anticipated or unanticipated. In developed countries, health insurance and social services likely play an important role. In less developed countries, where insurance is limited and social services are few, the household, family and possibly even the employer may play a central role. Incorporating these aspects of behavior substantially complicates analyses of the links between health and economic prosperity. It will likely also substantially enrich our understanding of those linkages.

Better integrating studies based on experimental designs with the approaches adopted in observational studies will yield substantial returns. On the one hand, observational studies will benefit from following the tradition of experimental studies which pay greater attention to measurement of health. Expanding the horizon of experimental studies to look beyond the immediate outcomes of interest, incorporating socio-economic outcomes as well as the measurement and analysis of behavioral responses to the treatment and following respondents for extended periods of time is likely to be very profitable.

References

Averett, S. and S. Korenman. (1996). "The economic reality of the beauty myth," *Journal of Human Resources*, 31(2):304-30.

Basta, S. K. Soekirman, and N. Scrimshaw. (1979). "Iron deficiency anemia and productivity of adult males in Indonesia," *American Journal of Clinical Nutrition*, 32:916-25.

Barker, D. (1997). "Maternal nutrition, fetal nutrition and diseases in later life," *Nutrition*, 13(9):807-13.

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.

Behrman, J., and A.. Deolalikar. (1988). "Health and nutrition," in H. Chenery and T.N. Srinivasan (eds.) *Handbook of Development Economics*, Vol. 1, Amsterdam: North Holland.

Behrman, J. and M. Rosenzweig. (2001). "The returns to increasing body weight," mimeo, University of Pennsylvania.

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.

Blau, D., D. Gilleskie, and C. Slusher. (1997). "The effects of health on employment transitions of older men." Presented at the 1997 Population of America Meetings.

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.

Bound, J. (1991). "Self-reported objective measures of health in retirement models," *Journal of Human Resources*, 26:106-38.

Bound, J., M. Schoenbaum, T. Stinebrickner, and T. Waidmann. (1999). "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, 6(2):179-202.

Celsing, F., E. Blomstrand, B. Werner, P. Pihlstedt, and B. Ekblom. (1986). "Effects of iron deficiency on endurance and muscle enzyme activity in man," *Med. Sci. Sports Exerc.*, 18:156-161.

Cleveland, W. S. (1980). "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, 74(368):829-36.

Croppenstedt, A., C. Muller. (2000). "The impact of farmers' health and nutritional status on their productivity and efficiency: Evidence from Ethiopia", *Economic Development and Cultural Change,* 48.3:475-502.

Cureton, K. (1992). "Effects of experimental alterations in excess weight on physiological response to exercise and physical performance," in *Body composition and physical performance: Applications for the Military Services*, National Academy of Science Press.

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, P. J. van der Maas, and 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.

Deolalikar, A. (1988). "Nutrition and labor productivity in agriculture: estimates for rural south India," *Review of Economics and Statistics*, 70:406-13.

Dow, W., P. Gertler, R. Schoeni, J. Strauss and D. Thomas. (2001). "Health Care Prices, Health and Labor Outcomes: Experimental Evidence," Labor and Population Program Working Paper Series 97-01, RAND, Santa Monica.

Edgerton, V., G. Gardner, Y. Ohira, K. Gunawardena, and B. Senewiratne. (1979). "Iron-deficiency anemia and its effect on worker productivity and activity patterns," *British Medical Journal*, 2: 1546-9.

Eriksson, J. G., T. Forsen, J. Tuomilehto, P. D. Winter, C. Osmond, D. J. P. Barker, (1999). "Catch-up growth in childhood and death from coronary heart disease: longitudinal study" *British Medical Journal*, Feb 13, 318:427-31.

Fafchamps, M. and A. Quisumbing. (1999). "Human capital, productivity and labor allocation in rural Pakistan", *Journal of Human Resources*, 34.2:369-406.

Fenwick, A., and Figenschou, B. M. (1972). "The effect of Schistosoma Mansoni on the productivity of cane cutters on a sugar estate in Tanzania," *Bulletin of the World Health Organization*, 47(5): 567-72.

Fogel, R. (1992). "Second thoughts on the European escape from hunger: Famines, chronic malnutrition and mortality rates," in S. R. Osmani (ed.), *Nutrition and Poverty*, World Institute for Development Economics Research Studies in Development Economics, New York: Clarendon Press, pp. 243-86.

Fogel, R. (1994). "Economic growth, population theory and physiology: The bearing of long-term processes on the making of economic policy," *American Economic Review* 84(3):369-95.

Fogel, R. (1999). "Catching up with the economy", *American Economic Review* 89(1):1-21.

Foster, A. and M. Rosenzweig. (1993). "Information flows and discrimination in labor markets in rural areas in developing countries," in L. Summers and S. Shah (eds.), *Proceedings of the World Bank Annual Conference on Development Economics, 1992: Supplement to The World Bank Economic Review and The World Bank Research Observer*. Washington, DC: World Bank, 1993, pp. 173-203.

Foster, A. and M. Rosenzweig. (1994). "A test for moral hazard in the labor market: Contractual arrangements, effort, and health," *Review of Economics and Statistics*, 76(2):213-27.

Francis-Williams, J., and P. Davies. (1974). "Very low birth weight and later intelligence," *Developmental Medicine and Child Neurology*, 16:709-28.

Frank, R. (2001) "Mental health and labor market performance," mimeo, Harvard University.

Glick, P., D. E. Sahn. (1998). "Health and productivity in a heterogeneous urban labour market", *Applied Economics*, 30.2:203-216.

Gateff, C. et al. (1971). "Influence de la Bilharziose Vesicale sur la Rentabilite Economique d'une Population Adult Jeune du Cameroon," *Annales de la Societe Belge de Medicin Tropical*, 51(3):309-24.

Gruber, J., and M. Hanratty. (1995). “The labor-market effects of introducing National Health Insurance: Evidence from Canada,” *Journal of Business and Economic Statistics* 13(2):163-73.

Haas, J. E., and T. Brownlie IV. (2001). “Iron deficiency and reduced work capacity: A critical review of the research to determine a causal relationship,” *J. Nutr.* 131:676S-690S.

Haas, J. E., T. Brownlie, and Y. Zhu. (2000). “Critical review of the evidence that iron deficiency anemia causes reduced work capacity.” Mimeo. Cornell University.

Haddad, L., and H. Bouis. (1991). “The impact of nutritional status on agricultural productivity: wage evidence from the Philippines,” *Oxford Bulletin of Economics and Statistics*, 53:45-68.

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):1103-10.

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.

Imminck, Maarten and Viteri, Fernando. (1981a). "Energy intake and productivity of Guatemalan sugarcane cutters: An empirical test of the Efficiency Wage Hypothesis--Part 1," *J. of Dev. Econ.*, 9(2):251-71.

Imminck, Maarten and Viteri, Fernando. (1981b). "Energy intake and productivity of Guatemalan sugarcane cutters: An empirical test of the Efficiency Wage Hypothesis--Part 2," J. *of Dev. Econ.*, 9(2):273-87.

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*, 34:105-21.

Knaul, F. M. (2000). “Health, nutrition, and wages: Age at menarche and earnings in Mexico,” in W. D. Savedoff and T. P. Schultz (eds.), *Wealth from Health: Linking social Investments to Earnings in Latin America*, Washington, DC: Inter-American Development Bank, pp. 35-70.

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):930-42.

Krueger, D. E. (1957). "Measurement of prevalence of chronic disease by household interviews and clinical evaluations," *American Journal of Public Health*, 47:953-60.

Lavy, V., M. Palumbo, and S. Stern. (1995). "Health care in Jamaica: Quality, outcomes, and labor supply," Living Standards Measurement Study Working Paper No. 116, World Bank.

Li, R., X. Chen, H. Yan, P. Deurenberg, L. Garby, and J. Hautvast. (1994). “Functional consequences of iron supplementation in iron-deficient female cotton workers in Beijing, China,” *American Journal of Clinical Nutrition*, 59:908-13.

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.

Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and S. Marquis. (1987). "Health insurance and the demand for medical care: Evidence from a randomized experiment," *American Economic Review*, 77:251-71.

Manton, K. and M. Woodbury. (1992). "Grade of membership analysisin the epidemiology of aging." in R. Wallace and F. Woolson, (eds.) *The Epidemiologic Study of the Elderly*, pp. 333-357. New York: Oxford University Press.

Martorell, R., and G. Arroyave. (1988). "Malnutrition, work output and energy needs," in K. Collins and F. Roberts (eds.), *Capacity for Work in the Tropics*, Cambridge: Cambridge University Press.

Mathiowetz, N., and T. Laird. (1994). "Getting better? Change or error in the measurement of functional limitations," *Journal of Economic and Social Measurement*, 20:37-62.

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):1100-05.

Moffett, J. K., D. Torgerson, S. Bell-Syer, D. Jackson, H. Llewlyn-Phillips, A. Farrin, J. Barber. (1999). "Randomised controlled trial of exercise for low back pain: Clinical outcomes, costs, and preferences", *British Medical Journal*, July 31, 319:279-283

Murray, C. and L. Chen. (1992). "Understanding morbidity change," *Population and Development Review*, 18(3):481-503.

Murrugarra, E., and M. Valdivia. (2000). "The returns to health for Peruvian urban adults by gender, age, and across the wage distribution," in W. D. Savedoff and T. P. Schultz (eds.), *Wealth from Health: Linking social Investments to Earnings in Latin America*, Washington, DC: Inter-American Development Bank, pp. 151-88..

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. (1993). *Free for all? Lessons from the RAND Health Insurance Experiment*, Cambridge: Harvard University Press.

Pijls, L. T. J., J. M. Feskins, 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):840-8.

Pitt, M., M. Rosenzweig, and M. N. Hassan. (1990). "Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries," *American Economic Review*, 80:1139-56.

Pollitt, E. (1997). "Iron deficiency and educational deficiency," *Nutritional Reviews*, 55(4):133-40.

Pollitt, E. (2001). "The developmental and probabilistic nature of the functional consequences of iron-deficiency anemia in children, *J. Nutr.* 131:669S-675S.

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.

Ribero, R., and J. Nuñez. (2000). “Adult morbidity, height, and earnings in Colombia,” in W. D. Savedoff and T. P. Schultz (eds.), *Wealth from Health: Linking social Investments to Earnings in Latin America*, Washington, DC: Inter-American Development Bank, pp. 111-50.

Ruger, J. P., D. T. Jamison and D. E. Bloom. (2001). "Health and the economy," in M. H. Merson, R. E. Black, and A. J. Mills (eds.), *International Public Health*, Gaithersburg: Aspen, pp. 617-66..

Schoenbaum, Michael. (1995). "The Health Status and Labor Force Behavior of the Elderly in Taiwan," Comparative Study of the Elderly in Asia, Research Report No. 95-32, Population Studies Center, University of Michigan.

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):291-6.

Scrimshaw, N. (1991). “Iron deficiency,” *Scientific American*, October, 46-51.

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.

Smith, J. P. (1999). "Healthy bodies and thick wallets: The dual relation between health and economic status," *Journal of Economic Perspectives*, 13(2):145-66.

Spurr, G. B. (1983). “Nutritional status and physical work capacity,” *Yearbook of Physical Anthropology*, 26:1-35.

Spurr, G .B. (1988). "Body Size, Physical Work Capacity and Productivity in Hard Work: Is Bigger Better?," in J. C. Waterlow (ed.), *Linear Growth Retardation in Less Developed Countries*, Nestle Nutrition Workshop Series Vol. 14, New York: Raven Press, pp. 215-44.

Steckel, R. (1995). "Stature and the Standard of Living," *J. of Econ. Lit.*, 33(4):1903-40.

Stern, S. 1989. “Measuring the effect of disability on labor force participation,” *Journal of Human Resources*, 24:361-95.

Strauss, J. (1986). “Does better nutrition raise farm productivity?” *Journal of Political Economy*, 94: 297-320.

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.

Strauss, J., and D. Thomas. (1995). “Human resources: Empirical modeling of household and family decisions,” in T. N. Srinivasan and J. Behrman (eds.), *Handbook of Development Economics*, Vol. 3, North Holland.

Strauss, J., and D. Thomas. (1998). "Health, nutrition and economic development," *Journal of Economic Literature*, 36:766-817.

Swaminathan, S., and L. Lillard. (2000). “Health and labor market outcomes: Evidence from Indonesia,” mimeo, University of Michigan.

Thomas, D., and J. Strauss. (1997). "Health and wages: Evidence on men and women in urban Brazil," *Journal of Econometrics*, 77:159-85.

Thomas, D. and E. Frankenberg. (2001a). "The measurement and interpretation of health in social surveys", forthcoming in A. Lopez, C. Murray and C. Mathers, (eds.) *Summary measures of population health*, World Health Organization.

Thomas, D. and E. Frankenberg. (2001b). "The dynamics of health, nutrition and economic prosperity", mimeo.

Wadsworth, M. E. J. and D. J. L. Kuh. (1997). "Childhood influences on adult health: A review of recent work from the British 1946 National Birth Cohort Study, the MRC National Survey of Health and Development," *Pediatric and Perinatal Epidemiology*, 11:2-20.

Waldron, I. (1983). "Sex differences in illness incidence, prognosis and mortality: Issues and evidence," *Social Science and Medicine*, 17(16):1107-23.

Wolgemuth, J. et al. (1982). "Worker productivity and the nutritional status of Kenyan Road construction laborers," *Amer. J. of Clinical Nutrition*, 36(7):68-78.

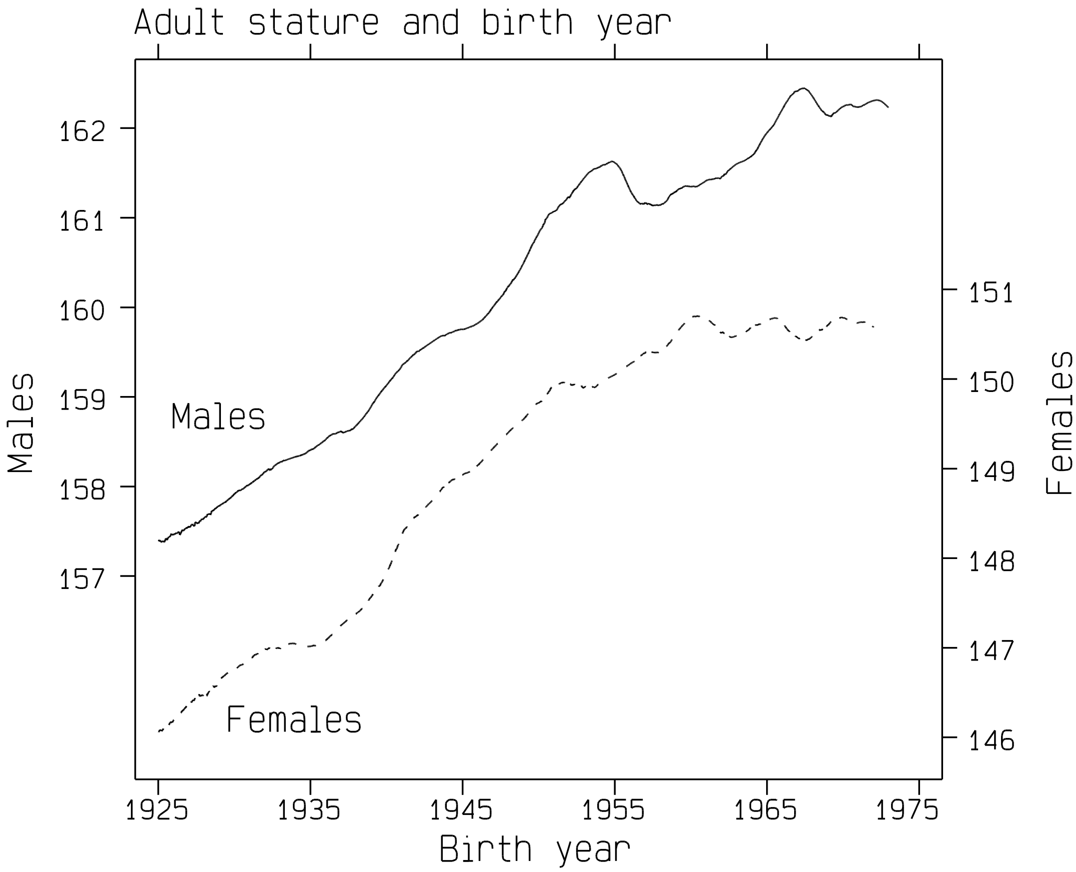
Woodson, R. D., R. E. Wills, and C. Lenfant. (1978). "Effect of acute and established anemia on O_2 transport at rest, submaximal and maximal work, *J. Appl. Physiol.* 44:36-43.

World Bank. (1993). *World Development Report 1993*, Oxford: Oxford University Press.

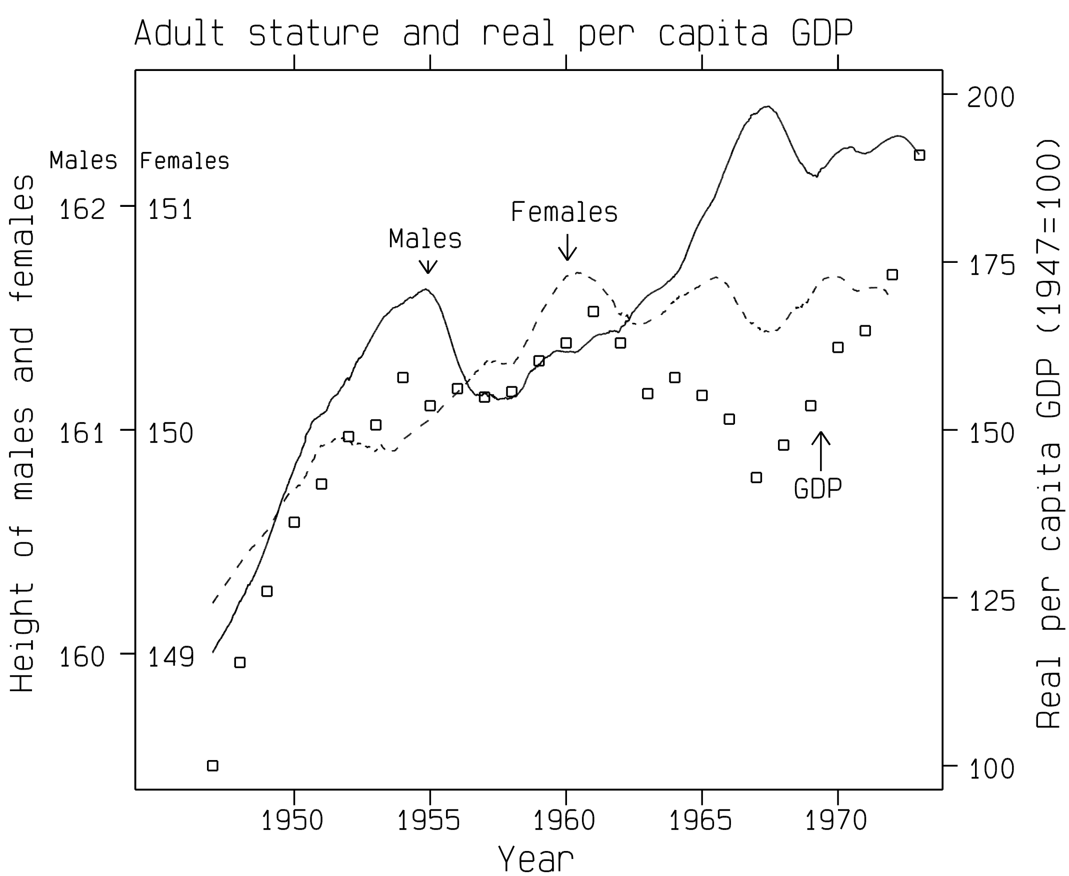
Zhu, Y. I., and J. D. Haas. (1998). "Altered metabolic response of iron-depleted nonanemic women during a 15-km time trial," *J. Appl. Physiol*. 84:1768-75.

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



Adult stature and birth year
162
161
160
159
158
157
Males
Males
Females
151
150
149
148
147
146
Females
1925
1935
1945
1955
1965
1975
Birth year



Adult stature and real per capita GDP
Males
Females
162
151
161
150
160
149
Height of males and females
Males
Females
GDP
200
175
150
125
100
Real per capita GDP (1947=100)
1950
1955
1960
1965
1970
Year

Figure 2: Height, weight and BMI in a crisis

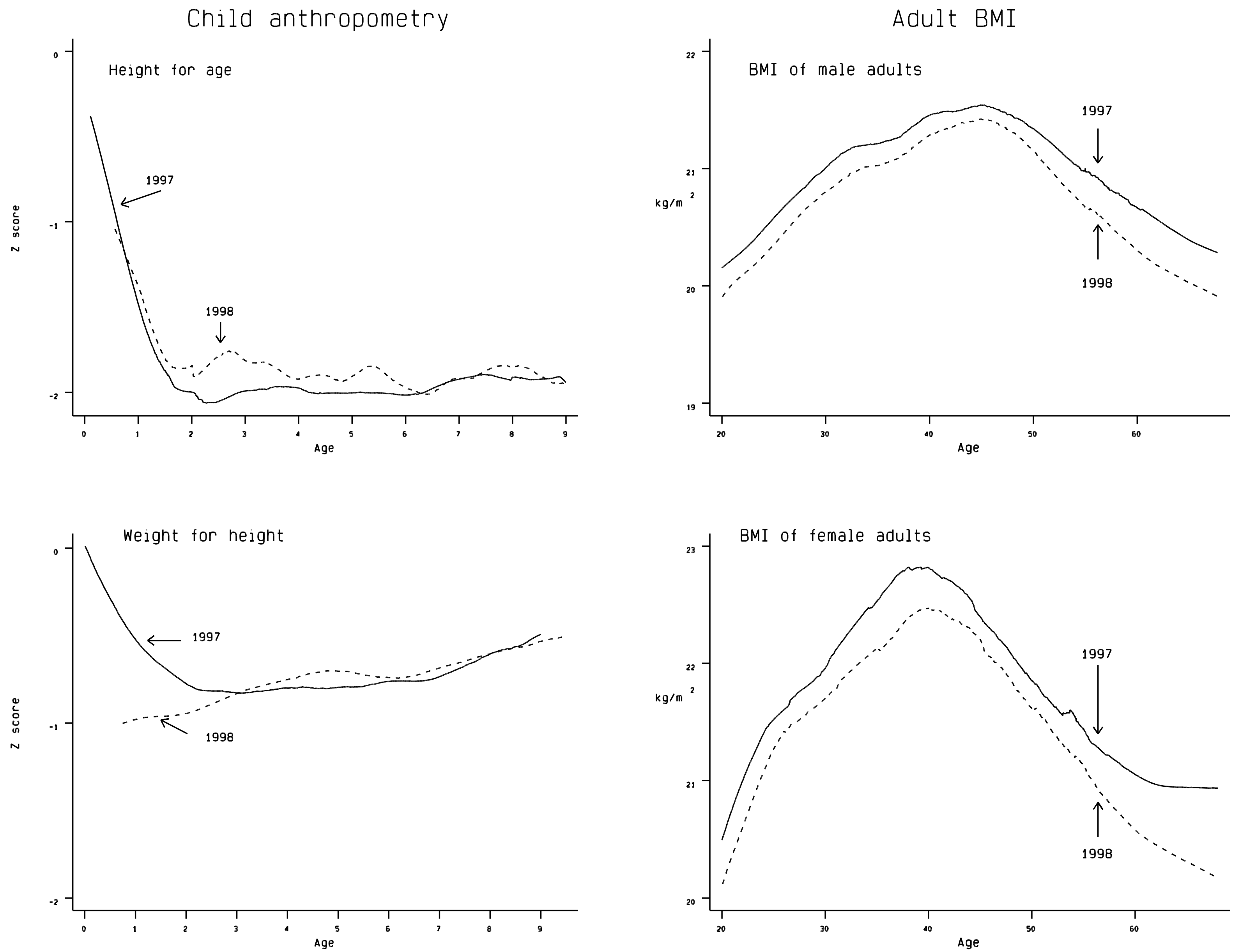


Figure 3: Height and weight of adults

Differences between self-reports and measurements by age

(NHANES3)

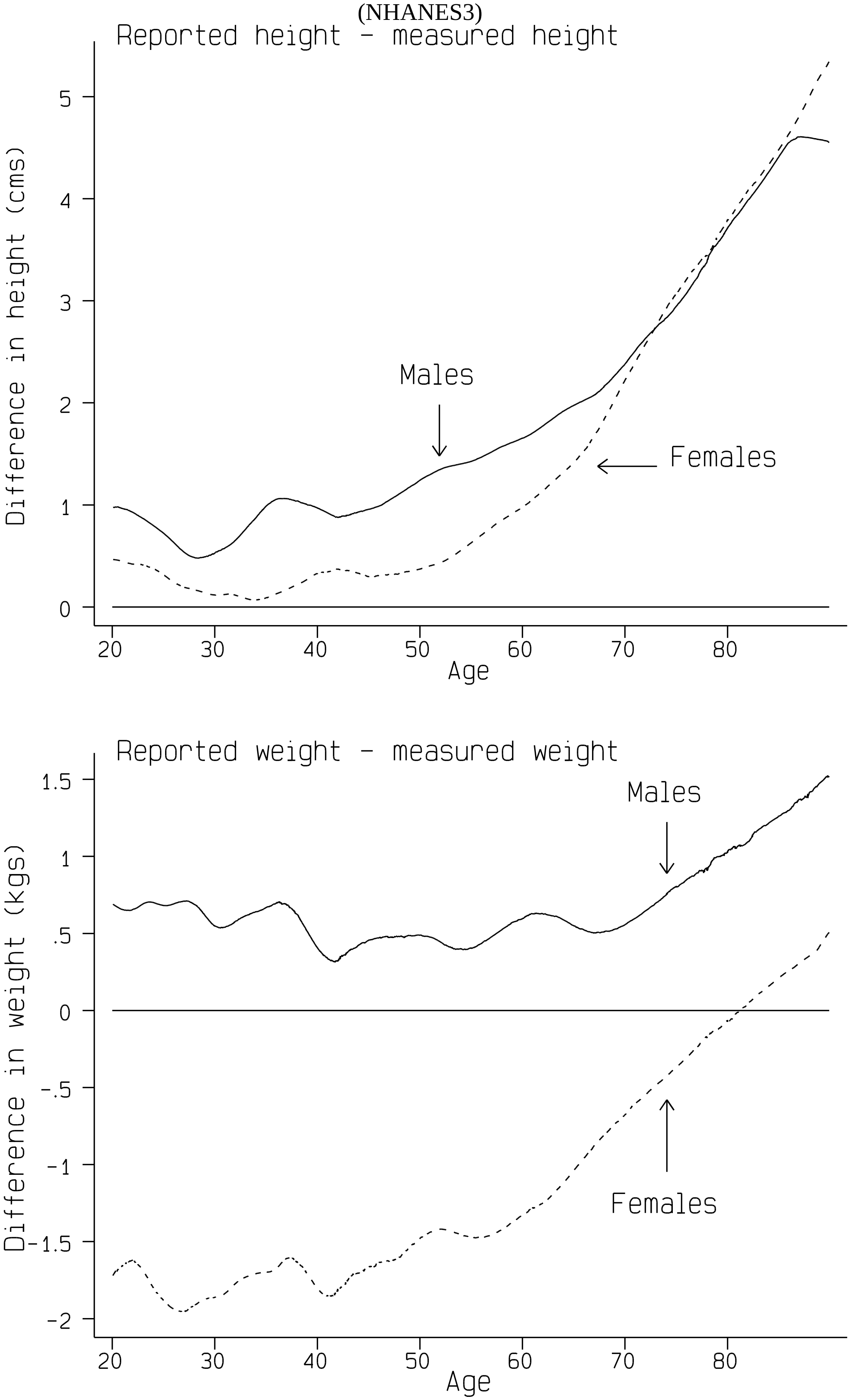


Table 1a: Changes in nutritional status of Indonesians Between 1997 and 1998

Child height-for-age and weight-for-height z scores

Age	Height-for-age	Weight-for-height
6 mths-2 years	0.02	-0.09
	(0.08)	(0.08)
3-4 years	0.03	0.13
	(0.06)	(0.06)
5-9 years	0.02	0.05
	(0.04)	(0.03)

Adult BMI

Age	Males	Females
22-39 years	-0.14	-0.26
	(0.05)	(0.05)
40-68 years	-0.29	-0.42
	(0.05)	(0.06)

Source: IFLS2/2+; Change is 1998 measure-1997 measure. (standard errors in parentheses).

Table 1b: Changes in use of health services by Indonesians %change between 1997 and 1998

Use of health care by children age < 15	-38%
Use of *posyandu* by children < 5	-60%
Children age < 3 received Vitamin A	-22%
Use of health care by adults	-9%

Source: IFLS2/2+. Change is 1998-1997 expressed as % of 1997.
All changes are significant (p<0.001).

Table 2: 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 years	0.002	-0.009	0.024	0.027	0.009	0.011
(spline)	[0.17]	[0.91]	[2.1]	[1.96]	[1.97]	[1.71]
Age 40-60	0.035	0.025	-0.011	0.01	-0.016	-0.005
	[3.56]	[2.39]	[1.24]	[0.75]	[3.69]	[0.8]
Age 60-80	0.113	0.152	0.003	0.032	-0.03	-0.033
	[9.85]	[12.08]	[1.86]	[3.2]	[6.24]	[5.56]
Year of education	0.050	-0.104	-0.053	-0.072	-0.035	0.013
	[2.75]	[4.12]	[2.95]	[3.32]	[4.18]	[0.94]
ℓn(per capita HH income)	-0.085	-0.109	-0.057	-0.02	0.002	0.031
	[1.28]	[1.56]	[0.73]	[0.23]	[0.06]	[0.7]
(1) if self-report						
as overweight	0.12	0.047	-1.541	-1.43	-0.627	-0.645
	[1.14]	[0.39]	[12.07]	[10.43]	[11.74]	[9.35]
as underweight	0.048	0.017	1.253	1.069	0.471	0.447
	[0.24]	[0.05]	[6.19]	[5.47]	[5.77]	[4.25]
Intercept	0.92	2.63	1.413	-0.737	0.246	-1.102
	[1.38]	[4.21]	[1.95]	[0.93]	[0.82]	[2.84]
F(all covariates)	44.51	59.13	40.58	41.13	50.21	27.43
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
R^2	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 3: Health and respondent characteristics

Indonesia (IFLS2)

A: Physical health asssessments	BMI		Hemoglobin		Lung capacity		Sit->stand		Blood pressure		HW evaln	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ℓn(PCE) below median (spline)	0.170	0.216	0.042	0.057	0.021	-0.004	-0.002	-0.006	0.013	0.002	0.200	0.184
	[6.62]	[7.47]	[1.48]	[2.10]	[0.76]	[0.19]	[0.05]	[0.18]	[0.56]	[0.07]	[4.97]	[5.00]
-- above median (spline)	0.107	0.117	0.055	-0.014	0.041	0.014	0.025	0.018	0.02	0.074	0.066	0.084
	[3.41]	[3.64]	[1.74]	[0.65]	[1.58]	[0.85]	[1.02]	[0.62]	[0.61]	[1.96]	[1.73]	[2.30]
Education 0-5 yrs (spline)	0.023	0.072	0.037	0.020	0.032	0.011	-0.019	-0.010	0.021	0.017	0.019	0.028
	[3.31]	[8.3]	[4.17]	[2.65]	[4.19]	[2.07]	[1.96]	[0.97]	[2.7]	[1.95]	[1.7]	[2.97]
-- >=6 yrs (spline)	0.034	-0.029	0.020	-0.001	0.030	0.014	-0.018	-0.020	0.002	-0.012	0.028	0.029
	[7.74]	[5.4]	[4.62]	[0.15]	[7.10]	[4.16]	[4.49]	[4.17]	[0.47]	[2.54]	[4.96]	[4.87]

B: Self-reported health	Walk 5 kms with difficulty		Carry heavy load with difficulty		GHS good		GHS poor		# days limited activity		# bed days	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln(PCE) below median (spline)	-0.008	0.042	-0.012	-0.030	0.028	0.000	-0.050	0.034	-0.012	0.079	-0.004	0.018
	[0.35]	[1.39]	[0.53]	[1.05]	[0.84]	[0.00]	[1.58]	[1.30]	[0.37]	[3.2]	[0.12]	[0.72]
-- above median (spline)	-0.017	0.017	0.007	0.078	0.007	0.013	0.002	-0.031	-0.047	-0.04	-0.031	-0.047
	[0.96]	[0.56]	[0.37]	[2.30]	[0.22]	[0.49]	[0.07]	[1.25]	[1.54]	[1.62]	[0.82]	[2.54]
Education 0-5 yrs (spline)	-0.008	0.001	-0.019	-0.008	0.001	0.005	-0.015	0.010	0.004	-0.019	0.000	-0.018
	[1.03]	[0.07]	[2.32]	[0.95]	[0.07]	[0.77]	[1.55]	[1.25]	[0.42]	[2.59]	[0.02]	[2.54]
-- >=6 yrs (spline)	-0.004	-0.005	0.004	-0.002	-0.002	0.003	-0.010	-0.019	-0.006	-0.003	-0.001	-0.001
	[1.20]	[1.03]	[1.33]	[0.41]	[0.44]	[0.52]	[2.36]	[4.01]	[1.3]	[0.73]	[0.17]	[0.26]

C: Self-reported morbidities	Cough		Breathing difficulty		Joint pain		Fever		Diarrhoa		Nausea	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln(PCE) below median (spline)	-0.006	0.080	-0.035	0.030	-0.022	0.037	-0.020	0.046	-0.008	0.034	0.051	0.114
	[0.19]	[3.03]	[1.09]	[1.20]	[0.74]	[1.32]	[0.63]	[1.78]	[0.24]	[1.36]	[2.03]	[4.29]
-- above median (spline)	-0.001	0.036	0.010	-0.013	-0.017	0.026	-0.034	-0.001	0.017	0.053	0.007	-0.029
	[0.03]	[1.21]	[0.27]	[0.46]	[0.58]	[1.00]	[1.30]	[0.02]	[0.55]	[1.48]	[0.24]	[0.95]
Education 0-5 yrs (spline)	-0.020	-0.020	-0.040	-0.019	0.006	-0.003	-0.011	-0.025	-0.006	-0.001	-0.005	0.010
	[2.06]	[2.60]	[3.42]	[2.65]	[0.6]	[0.33]	[1.06]	[2.97]	[0.77]	[0.16]	[0.63]	[1.30]
-- >=6 yrs (spline)	-0.013	-0.006	-0.012	-0.015	-0.012	-0.006	-0.016	-0.014	0.002	-0.009	-0.006	-0.012
	[2.85]	[1.20]	[3.00]	[3.65]	[2.73]	[1.19]	[3.56]	[2.72]	[0.45]	[1.83]	[1.48]	[2.26]

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 4: Labor market outcomes and health indicators
OLS estimates for Indonesian males (IFLS)

	Labor force participation	ℓn hours per month	ℓn hourly earnings
Nutrition			
ℓn Height	-0.157	-0.001	1.658
	[1.22]	[0.00]	[3.58]
ℓn BMI (in 1997)	-0.005	0.347	0.800
	[0.09]	[2.43]	[4.60]
ℓn BMI (in 1993)	0.065	-0.030	0.187
	[1.38]	[0.20]	[1.02]
General health status			
(1) good (in 1997)	-0.026	0.018	0.047
	[1.89]	[0.36]	[0.88]
(1) good (in 1993)	0.010	0.025	0.058
	[1.01]	[0.69]	[1.33]
(1) poor (in 1997)	-0.037	-0.162	0.072
	[2.44]	[3.37]	[1.16]
(1) poor (in 1993)	-0.022	-0.002	-0.007
	[1.20]	[0.03]	[0.10]
Activities of daily living			
(1) walk 1km-hard (in 1997)	-0.055	-0.057	0.169
	[2.88]	[1.32]	[2.53]
(1) walk 1km-hard (in 1993)	0.024	0.030	-0.050
	[0.96]	[0.49]	[0.51]
(1) heavy load-hard (in 1997)	-0.133	0.052	-0.090
	[5.03]	[0.93]	[0.97]
(1) heavy load-hard (in 1993)	-0.108	-0.068	-0.047
	[3.14]	[0.69]	[0.32]
Hemoglobin spline≤10g/dl	-0.019	0.046	0.048
	[0.83]	[1.04]	[0.63]
spline>10g/dl	-0.002	-0.010	0.027
	[0.86]	[1.12]	[2.47]
(1) HH iodized salt	-0.020	-0.010	0.074
	[1.80]	[0.30]	[1.70]
ℓn Lung capacity	0.024	0.046	0.008
	[1.37]	[0.94]	[0.11]
(1) Blood press (mild)	-0.027	-0.041	-0.039
	[1.76]	[0.98]	[0.69]
(1) Blood press (mod)	-0.053	-0.067	0.142
	[1.77]	[0.85]	[1.34]
Healthworker evaluation (1-9)	0.009	-0.003	-0.008
	[1.62]	[0.15]	[0.35]
R^2	0.44	0.10	0.31
F(all covariates)	44.58	9.16	29.41
	[0.00]	[0.00]	[0.00]
χ^2(All health)	9.45	2.45	6.73
	[0.00]	[0.03]	[0.00]
χ^2(97 health)	9.79	2.58	4.9
	[0.00]	[0.03]	[0.00]

Notes: Source IFLS1 (93) and IFLS2 (97). Asymptotic t statistics in parentheses robust to arbitrary heterosckedasticity and permit correlations within clusters.

Table 5: Labor market outcomes and health indicators
Instrumental variables estimates for Indonesian males (IFLS)

	Labor force participation	ℓn hours per month	ℓn hourly earnings
ℓn Height	-0.309	0.361	1.550
	[0.79]	[0.41]	[2.08]
ℓn BMI (in 1997)	0.379	1.823	3.810
	[0.42]	[0.74]	[1.34]
ℓn BMI (in 1993)	-0.080	-5.394	-1.713
	[0.07]	[1.37]	[0.40]
(1) GHS - good (in 1997)	-0.324	2.046	-0.513
	[1.00]	[1.62]	[0.34]
(1) GHS - poor (in 1997)	0.199	-1.017	-0.119
	[0.66]	[1.27]	[0.15]
(1) Walk 1km-hard (in 1997)	-0.502	0.133	0.702
	[1.10]	[0.11]	[0.71]
(1) Heavy load-hard (in 1997)	-0.750	0.877	-2.626
	[1.48]	[0.55]	[1.61]
(1) GHS - good (in 1993)	-0.176	0.627	-0.173
	[0.79]	[1.00]	[0.25]
(1) GHS - poor (in 1993)	0.531	-2.252	1.55
	[1.27]	[1.38]	[0.83]
(1) Walk 1km-hard (in 1993)	-0.159	1.158	3.203
	[0.25]	[0.69]	[1.68]
(1) Heavy load-hard (in 1993)	0.004	-0.575	-5.093
	[0.01]	[0.28]	[2.18]
F(all covariates)	27.09	3.87	19.49
	[0.00]	[0.00]	[0.00]

Table 6a: Labor market outcomes and health indicators
Fixed effects estimates for Indonesian males (IFLS)

	Labor force participation	ℓn hours per month	ℓn hourly earnings
ℓn BMI	0.135	0.223	0.59
	[3.89]	[1.68]	[3.48]
(1) GHS - good	-0.002	0.054	0.011
	[0.18]	[1.56]	[0.25]
(1) GHS - bad	-0.031	-0.109	0.06
	[2.66]	[2.65]	[1.1]
(1) Walk 1km-hard	-0.056	-0.068	0.06
	[4.18]	[1.46]	[0.97]
(1) Heavy load-hard	-0.182	-0.032	-0.053
	[10.72]	[0.51]	[0.6]
R^2	0.38	0.06	0.09

Table 6b. Fixed effects and instrumental variables fixed effects estimates
Indonesian males (IFLS)

	ℓn(hourly earnings)	LFP
Fixed effects		
ℓn BMI	0.59	0.14
	[3.5]	[3.98]
F(all)	20.80	105.17
	[0.00]	[0.00]
F(Fixed effect)	1.74	1.09
	[0.00]	[0.00]
Significance of level of health in 1993		
F(levels of health)	0.70	12.20
	[0.62]	[0.00]
Instrumental variables fixed effects		
ℓn BMI	5.68	.
	[2.0]	
GMM Overid test		
F(prices)	0.67	.
	[0.88]	