Health shocks, job quality, and self-employment in Africa

Monk, Courtenay; Teal, Francis

RECOUP
RESEARCH CONSORTIUM ON
EDUCATIONAL OUTCOMES & POVERTY

RECOUP Working Paper 13

Health Shocks, Job Quality, and Self-Employment in Africa

Courtney Monk and Francis Teal
University of Oxford

January 2008

DFID Department for International Development

UNIVERSITY OF CAMBRIDGE

Health Shocks, Job Quality, and Self-Employment in Africa*

Courtney Monk† and Francis Teal
Centre for the Study of African Economies
Department of Economics
University of Oxford

January 2008

Abstract

This paper examines the relationship between occupation and different types of human capital—skills, education, ability and health. Summary statistics of our panel data from the Ghana Household Worker Survey strongly suggest that height and health vary by occupation. Our regression results show that after controlling for age, gender and movement out of jobs, the self-employed are the most likely to have at least one day of illness (a health shock) in the past year. However, conditional on having at least one day of illness, the self-employed have the lowest expected number of days ill. On the other hand, evidence that large firm workers have longer illnesses than other workers, perhaps reflecting their better employment circumstances that allow more time off when sick, cautions against the use of days of illness as a measure of health that is unbiased by occupational choice. We also investigate labour market outcomes and find that the number of days ill does not affect the labour supply decision. Controlling for entry into the labour force, however, days of illness does have a significant negative impact on earnings, especially if it exceeds 30 days. Our analysis suggests that changes in health and changes in occupation are strongly correlated; understanding the causal links in this relationship should be the focus of future work.

Keywords: Self-employment; Africa; Health
JEL Classification Codes: O12; J24

*The data used in this paper were collected by the Centre for the Study of African Economies (CSAE), Oxford, in collaboration with the Ghana Statistical Office (GSO) in 2006. The survey was funded in part by the Department for International Development (DfID) of the UK as part of its work on assessing the outcomes of education and the Economic and Social Research Council (ESRC) as part of the Global Poverty Research Group. We have been assisted by numerous collaborators in enabling us to collect this data. We are also greatly indebted to Moses Awoonor-Williams, Geeta Kingdon and Andrew Zeitlin for their assistance in the design and implementation of the survey. We would also like to thank participants at the 2007 IZA/World Bank Conference on Employment and Development as well as participants at a Cambridge seminar, who provided several helpful comments on an earlier draft.This paper forms part of the Research Consortium on Educational Outcomes and Poverty (RECOUP). Neither DfID nor any of the partner institutions are responsible for any of the views expressed here.

†Corresponding author: courtney.monk@economics.ox.ac.uk; Economics Department, Manor Road Building, Manor Road, Oxford OX1 3UQ, United Kingdom; Tel: +44 (0)1865 271077.

1 Introduction

The large increases in urban self-employment and rising numbers of microenterprises in Africa raise an interesting issue: while the number of opportunities for employment may be increasing, it is not clear how job quality in this type of employment compares with that of large sector firms. The limited literature on the expansion of self-employment disagrees on this issue. Looking at movements into and out of self-employment in Mexico, Fajnzylber, Maloney, and Rojas (2006) conclude that self-employment is just as desirable as other labour market alternatives. In contrast, Sandefur, Serneels, and Teal (2006) use labour market data for the urban sector in Ghana, Tanzania and Ethiopia to show that incomes for the self-employed and those in small firms are substantially less than those in large firms and the public sector (see also Söderbam, Teal, and Wambugu (2005)). Thus earnings differentials imply that the quality of self-employment and jobs in microenterprises might be substantially inferior compared to employment in large firms.

Yet these findings ignore the effects of health on occupational choice and earnings and the reverse effects of occupation and income on health. Indeed, there is evidence in the literature that health shocks result in a loss in income due to reductions in labor supply and/or productivity (Gertler & Gruber, 2002), and that healthier individuals earn more. In other words, our analysis will be incomplete if we ignore health.

We therefore explore the relationships between employment status, health outcomes and wages in Africa. We are particularly interested in the self-employed and use two waves from new panel data from the Ghana Household Worker Survey to investigate job quality and health for these individuals. The key to understanding the nature of this relationship is observing the individual at different points in time, so as to observe movements between jobs, changes in income and shocks to health. Our panel data set, which is collected at the individual level and contains detailed health and income information, is ideally suited to consider this problem. We also incorporate data on skills and health knowledge, to see how the various aspects of human capital (health, education, skills and ability) come into play. The temporal aspect of the panel gives us the opportunity to investigate whether the movement out of certain jobs impacts health. Whether or not these movements are caused by or are the cause of health shocks is an equally interesting element of health outcomes and has obvious implications for assessing job quality. Dynamic extensions of this kind, as well as exploring the role of lagged health on present health, are particularly insightful.

In this paper we use data from an urban sample in Ghana to look directly at the effect of past illness and current job choice on current income, as well as the impact of past income and past job status on current health. We begin by using the data to ask if job type varies along a range of measures of health outcomes that can be observed in the data. As measuring health is a difficult undertaking, we consider two measures for which we have consistent measurements across the survey years: height and inactive days due to illness or injury.

In our in-depth analysis, we are aware that identifying the effects we are interested in is difficult and confounded by simultaneity and endogeneity issues. First, while health as an outcome is a major dimension of job quality, there is a concern that health may affect employment status by determining participation in the labour force (see Currie and Madrian (1999) for a review). In addition, it is also plausible for health to influence preferences for job type, so that unhealthy individuals favour jobs in which lower productivity is required. This may be more of a factor among the poorest individuals. Indeed, nutrition-based efficiency wage models imply that the sickest individuals will be left out of the formal sector. Theory and empirical evidence both suggest that nutrition, through its impact on overall health, strength, and endurance, is an important determinant of worker productivity (Strauss, 1986; Deolalikar, 1988; Thomas & Strauss, 1997). There are also potential feedback effects because productivity increases income, which in turn can be reinvested in health by buying more inputs or higher-quality inputs (Strauss & Thomas, 1998).

We attempt to establish causality and address these issues of simultaneity by combining the individual data with data at the household level. More importantly, observing past changes in health and income allows us to control for past unobservables and avoid the problem of simultaneous feedback effects. We show that past health shocks decrease current income, and that skills matter both for health and income. The evidence shows that the self-employed may be worse off in terms of health outcomes, as they have a higher likelihood of falling ill. Job movements, especially those out of self-employment and large firms, do play a role in the health story, although the direction of causality is as of yet unclear. Conversely, income has a complex relationship with health, in that people who earn more are more likely to report an illness, but also experience shorter illnesses conditional on having any days ill. This supports the idea that those with better jobs and higher incomes can afford to be ill.

This paper proceeds in five sections. Section 2 offers a brief review of the literature. Section 3 sets out our theoretical framework and empirical strategy, and Section 4 offers relevant details about the data. Results are presented in Section 5, and Section 6 concludes with ideas for further research.

2 Empirical Evidence

Dealing with the complicated interdependent relationships between health, wages, and labour supply is a challenge (see Strauss and Thomas (1998) for a very comprehensive review). In general, the consensus is that better health raises labour participation and wages, and that health shocks lower income and consumption.

2.1 Developed countries

Labour economists who look at health issues have been most interested in the effects of health on wages. Research on data from developed countries is more common that research on developing countries. Due to the rarity of panel data, most have focused primarily on the causal pathways between current health and current income and the endogeneity problems that arise with this approach (Lee, 1982; Ettner, 1996). Relatively little research has been able to exploit intertemporal changes in health and income, although a limited number of studies have contributed to this area.

Haveman, Wolfe, Kreider, and Stone (1994) use a 3-equation structural model and panel data on American white men. The three equations describe health, hours worked (labour supply), and wages. Health here is measured on a zero to three scale indicating the degree to which a medical condition limits one's ability to work. The authors conclude that prior (lagged) poor health lowers wages and work time.

In an approach similar to our own, Smith (2003) investigates the impact on past health changes and changes in socio-economic indicators, like income, on today's health. He also looks at the impact of health changes (shocks) on current income and labour supply. Unfortunately, he is only about to look at household income, rather than individual income, but finds that health shocks have a sharp negative effect on income, operating mainly through labour supply and not through medical expenses.

2.2 Developing Countries

Much of the work on developing countries focuses either on the simultaneous relationship between health and earnings (Thomas & Strauss, 1997; Schultz, 2003; Schultz & Tansel, 1997; Savedoff & Schultz, 2000), or the effect of health shocks on consumption (Wagstaff, 2006; Dercon & Hoddinott, 2003; Pitt & Rosenzweig, 1986). The existing literature that uses panel data to examine health shocks and income shocks has been somewhat limited. In addition, as income is most often measured at the household level, determining the true effects of health on individual income has been difficult. We believe that our ability to measure income at the individual level will give us a much better picture of the true effect of a health shock on labour outcomes.

Gertler and Gruber (2002) use a panel data set from Indonesia, which allows them to look at the effect of major health shocks on changes (first differences) in labour supply, income and medical expenses, and the resulting ability of households to smooth consumption. Their health measure is the individual's ability to perform activities of daily living (ADLs). Their results show that larger health shocks result in larger income losses, but they do not model a recursive health equation in addition their labour outcomes equation.

Lindelow and Wagstaff (2005) use a similar first differences approach with Chinese data and find that illness results in large declines in labour supply and income. However, the income data was measured only at the household level and then calculated on a per capita basis.

3 Theoretical Framework

The theoretical model developed by Grossman (1972) views health as an endogenously determined capital stock. Grossman's premise is that health capital differs from other forms of human capital because the stock of health "determines the total amount of time [one] can spend producing money earnings and commodities" (Grossman, 1972, p. 224). The model predicts that health should be positively correlated with wages, for better health will increase productivity and thus wages, and higher wages will increase the demand for health and medical care.

Let us think, as Grossman does, of health outcomes as a function of several inputs. An adult individual produces health H, which is determined by various factors:

$$H_t = H(H_{t-1}, I_{t-1}, w_{t-1}, B_t, D_{(t,h)}, U, e_{(t,h)}) \tag{1}$$

where I_{t-1} is a vector of health inputs, w_{t-1} is prior wage (in logs), B is a vector of family background characteristics, and $D_{(t,h)}$ is a vector of community level variables affecting health. The underlying time-invariant "healthiness" of the individual, U, is unobservable but known by the individual, while $e_{(t,h)}$ is the component that is both unknown and unobservable to all, e.g. measurement error.

Meanwhile, there exists a relationship that relates the log of real wages w in period t to several inputs (Mincer, 1974):

$$w_t = w(w_{t-1}, H_{t-1}, E_t, B_t, D_{(t,w)}, A, e_{(t,w)}) \tag{2}$$

where E_t is education and $D_{t,w}$ are community level infrastructure variables. Unobservable time-invariant qualities known to the individual like ability are included in A; $e_{(t,w)}$ captures measurement error and random fluctuations in w_t.

3.1 Empirical Model

From the framework above, we develop two approaches. First, we look at health and income in a setting without a lag of the dependent variable, for which the model to be estimated can be thought of as:

$$\begin{aligned} H_{i,t} &= \alpha_0 + \alpha_1 X_{i,t} + \alpha_2 J_{i,t-1} + u_{i,1t} && (3) \\ w_{i,t} &= \beta_0 + \beta_1 H_{i,t-1} + \beta_2 X_{i,t} + \beta_3 J_{i,t} + u_{i,2t} && (4) \end{aligned}$$

where X_t is a vector of time-varying and time-invariant individual characteristics, and J_t and J_{t-1} are vectors of job characteristics. We measure H_t as the inactive days in the past year caused by illness or injury.

We extend this model to incorporate dynamics of the dependent variable, creating structural equations that relate health and income in the following way:

$$\begin{aligned} H_{i,t} &= \alpha_0 + \alpha_1 H_{i,t-1} + \alpha_2 w_{i,t-1} + \alpha_3 X_{i,t} + \alpha_4 J_{i,t-1} + \eta_i + u_{i,1t} && (5) \\ w_{i,t} &= \beta_0 + \beta_1 H_{i,t-1} + \beta_2 w_{i,t-1} + \beta_3 X_{i,t} + \beta_4 J_{i,t} + \eta_i + u_{i,2t} && (6) \end{aligned}$$

Because we are dealing with panel data, this kind of model brings up some serious endogeneity issues, as $H_{i,t-1}$ and $w_{i,t-1}$ will be correlated with the individual effects η_i. We are able to deal with this for the wage equation, by instrumenting $w_{i,t-1}$ with $w_{i,t-2}$, available from the 2004 wave of the survey. However, we have not yet found a suitable instrument for $H_{i,t-1}$; nevertheless, by comparing the static and dynamic results, it is strong evidence that the bias resulting from this source of endogeneity in the health equation may not be a big concern.

We also use a selection equation to account for labour participation, as the wage in Equations 4 and 6 is only observed if one enters the labour force. This stage is identified with three variables: number of children, a dummy for being married or not, and a dummy indicating one's status as the household head.

Our identification strategy is dependent on the ability of height to capture one's history of health and nutrition inputs, so that we can successfully use height as a control for long-term health that may or may not impact the short-term measure of inactive days.[1] The main advantage of this specification is that it allows us to see the effect of a prior health shock on today's income, while controlling for other confounding factors, and without worrying about the endogeneity introduced by the contemporaneous feedback of current income onto current health.[2]

4 Data

The data are from the 2005 and 2006 waves of the Ghana Household Worker Survey (GHWS) conducted by the Centre for African Studies (CSAE). The survey is a representative sample of adult workers in urban areas in Ghana (Accra, Kumasi, Takoradi, and Cape Coast). The number of individuals in this 2005-2006 panel is 1,073, from which 200 were dropped from our sample due to missing values, because they were students, or because they were younger than 15 or older than 60, leaving us with 873 individuals. Table 4 presents simple summary statistics for this sample.

Great care has been taken in the design of the GHWS to measure incomes for workers in both the formal and non-formal sectors. In particular, the information collected on the self-employed is very detailed and, in contrast to many other surveys, allows for much more accuracy when determining earnings. This unique aspect of the GHWS means that we have individual level earnings information for employees in all types of employment.

4.1 Skills

The 2006 wave of the GHWS includes data from four skills tests and a health knowledge test. The skills tests include a Raven's test, an English test, a mathematics test, and a reading test.

[1] There is some evidence in very recent literature regarding the endogeneity of height with respect to earnings (Schultz, 2002); however, we assume here that height is predetermined, as is common practice.

[2] The endogeneity of the education variable in the earnings equation, due to omitted variables like ability and school quality, has been well discussed in the labour economics literature. We do not instrument for education in this paper, as it is not the focus of the analysis; however, previous work on the same data has concluded that instrumenting education does not alter its effects on earnings (Sandefur et al., 2006).

Both the mathematics and reading tests have two sections (elementary and more advanced), which are weighted accordingly in the scoring. The health knowledge test is comprehensive, covering such topics as sources, signs and home treatments for common ailments like diarrhea and heat stroke, as well as a few specific items like polio, malaria, and nutrition.

Although these tests allow for admittedly crude assessments of one's true skills, they are *much* more informative than simple subjective questions regarding literacy and numeracy. The Raven's test is interesting in itself, as it tries to ascertain one's natural intelligence, irrespective of schooling. Having access to such rich data about dimensions of human capital is rare, rendering our results unique and new.

Looking at some summary statistics on skills in Table 1, there are clearly strong differences in worker characteristics across occupations. This suggests that there is a strong selection mechanism of highly skilled workers into good jobs, although some of this is undoubtedly driven by age and gender. In addition, we cannot rule out the reverse possibility that some of these skills are acquired or improved in some occupational settings and not in others. What is most impressive in this table, however, is that those who do not earn any income are actually "more skilled" than the average person in the sample in almost everything except health knowledge. In fact, they are more skilled on average than those in the work force as well. This finding supports the claim that high-skilled workers who are unemployed do not enter self-employment during their unemployment spell but prefer to keep searching for a good formal sector job.

4.2 The Health Variables

The health measures that we will focus on are height and the number of days in the past year that one has been unable to do one's normal activity. These variables both suffer from measurement error. For height, however, only 1.6% of the measurements seemed to be obviously mis-measured and were treated as missing, and we were able to use data from the 2004 wave of the panel to fill in some missing height values. Furthermore, in the sample that we use for the results in this paper, characteristics such as gender, age and years of education do not have predictive power over whether a height measurement is grossly mis-measured or whether it is missing. Therefore, we do not expect the measurement error in the height variable to cause significant endogeneity issues.

In the case of days of illness reported, measurement error is probably present, as is the possibility that more educated and wealthier people may be more likely to report illness (Schultz & Tansel, 1997), and labour supply decisions may be endogenous to this variable as people try to justify leaving employment due to health reasons. Another potential problem with our days of illness variable is that the 2005 survey and 2006 survey were not identical in the way in which these questions were asked. In 2005, respondents were asked for the number of days ill in the past three months, whereas in 2006, respondents' frame of reference was a year. Therefore, to match the 2006 year-long framework, the 2005 days ill variable has been adjusted.[3] In addition, the 2006 survey asked for the number episodes of illness experienced during the past year (up to a maximum of three episodes), and the corresponding number of days ill suffered during each episode. From this information, we construct an aggregated number of days ill across these three episodes. Information about days ill is unavailable for any additional episodes, although only about 1% of the sample reports having more than three episodes, so we do not think this has resulted in a downward bias in the days ill measure.

Some summary statistics regarding occupation and health status are helpful. For the whole sample, 32% reported at least one day of illness in the last year. Table 2 breaks down days of illness in the past year by wave of the survey, gender, and occupation. It is not altogether clear yet from this table what the relationship between occupation and illness is. Women tend to suffer more disabled days than men. Workers in small firms report less illness than others, perhaps because their job stability is relatively precarious. The unemployed and self-employed

[3]It is possible to simply multiply the 2005 measure by four. This makes the strong assumption that health over the past three months is representative of health over the past year. Instead, we took information about health over the last year provided in 2005, in which a person noted his/her frequency of illness, which determined the factor by which the 2005 days of illness measure was multiplied. For example, someone with 20 days of illness in the last three months, but who also reported being ill "almost all of the time" was then deemed to have 80 days of illness in the past year. In any case, this adjustment, as compared with simply multiplying all of the 2005 measurements by four, did not change the results or our conclusions in any significant way. It simply allowed for a smoother empirical distribution and hopefully better estimation.

tend to have more days of illness than the average working individual.

It is also quite interesting to look at the distribution of height by occupation, shown in Figure 1. This graph shows clearly the stark differences in mean across occupations, with employees of large firms and those in the public sector being much taller than others.

5 Results

5.1 Effects for Health

To examine the health-occupation relationship, we first consider the simple model without the complications caused by including a lagged measure of health. Focusing, for the moment, on days of illness as the dependent variable: we have several possible empirical strategies to deal with this type of variable since it is has a high proportion of zeros, a very long tail, and is a count. We chose to focus on a probit model, in which the dependent variable takes the value of one if an individual reported one or more days of illness, but we also offer OLS estimates and a count regression conditional on at least one day of illness.[4]

Estimates from the probit model are found in Table 6 (Table 3 offers variable definitions). There are several things to note from this table, although we can start with the basics. Males are less likely to suffer illness, but this effect is not significant. Age on its own has a positive impact on the probability of having an illness, and its relationship is convex, so age has increasing marginal effects. Surprisingly, education has nothing to do with illness here. Height is also insignificant, suggesting that these short-term illnesses observed in this limited time-frame are pure shocks and have little to do with underlying "healthiness".

In addition, the NHIS variable, a community-level measurement of the proportion of people reporting that they have registered to receive one of the new National Health Insurance Scheme (NHIS) cards, is worth noting. It is significant, and controlling for all other variables at their means, a one percent increase in one's community's take-up rate decreases one's probability of

[4] We also estimated a tobit model, which confirmed the results reported here. Those results are available upon request.

falling ill by just under one percent.

In columns 1 to 3, we see that previous occupational choice does not seem to have a significant impact on becoming sick, except for those in big firms, who have a higher likelihood of falling ill. (The reference group here are those who earned no income last year.) This is most likely a result of the fact that workers in big firms 'can afford to be ill.' Moreover, the relationship between lagged occupation and the actual number of days ill is rather large, as can be seen in the OLS estimates in Table 7 and in the NegBin results in Table 8. These two tables coincide with what the summary statistics suggested, that those who are unemployed in the past year experience many more days of illness on average than others.

Finding that large firm workers may have more days of illness is consistent with the fact that employment circumstances for workers in big firms tend to be much better. On the other hand, it does suggest that days of illness may not be the best measure of health with which to investigate occupational effects. While we are trying to assess job quality, it may be job quality itself which is driving some of the variation in days of illness; in the future, we will need to look at more health measures in order to know how strong this phenomenon is.

After controlling for movement out of employment types, we are able to identify a much stronger effect of self-employment (refer to columns 4-7, Table 6). Now, the self-employed have the highest probability of falling ill, relative to all other categories. However, those who get out of self-employment do just as well as the others; the null hypothesis of whether the coefficients on "self-empl. last year" and "left self-empl." sum to zero is accepted, so we have isolated a positive relationship between self-employment and the probability of getting sick. whereas leaving a large firm is associated with a big rise. We cannot be sure of the direction of causality here—people may leave jobs because of health reasons, or health may alter job choice decisions—so interpretation is difficult. Nonetheless, it is a very interesting finding that the pattern of job choice is key to understanding how occupation and health interact.

The skills results are a little more difficult to interpret. Columns 5 and 6 show the addition of the skills variables. Innate ability, indicated by the Raven score, is linked to a higher probability of illness: scoring ten percent higher on the Raven's test raises the probability by roughly five percent; the result for reading ability is similar. In contrast, English skills have a downward effect on the probability of illness.

We also show the results for including health knowledge and then instrumenting it with parental education and number of children in columns 6 and 7, respectively.[5] It is obvious from these results that health knowledge is indeed highly endogenous to health, i.e. the more unhealthy one is, the more knowledge one acquires. Once instrumented, health knowledge and probability of illness are negatively related, which makes more sense intuitively—presumably the more one knows about health, the more one can prevent illness.

Moving on from the binary restriction imposed by the probit, we turn briefly to a zero-truncated negative binomial model of days ill (Table 8). The dependent variable is number of days ill, for individuals with at least one day. The estimation precision may suffer from the low number of observations, but the regression is nonetheless an interesting addition to the information provided by the probit. Being male reduces the expected count by 34%. Those without earned income in the last year have much more days of illness than the self-employed and those in small firms. In fact, the self-employed have the lowest number of days of illness. This could be because they are actually healthier and thus have more minor illnesses. But combining this with our evidence from the probit, the results support the alternative argument that the self-employed's work circumstances are inferior to those of market sector workers: not only do the get sick with the highest probability, but they then stay sick for the shortest amount of time, presumably because they have to return to work.

Employment in a big firm is associated with a more than doubling of inactive days, which is again some evidence that those whose job security is less tenuous may enjoy more time for convalescence. Similar to the probit model, job movement seems to be important, although

[5] We use a control function approach here, regressing health knowledge on the exogenous regressors; the results of this first stage can be found in Table 5.

more so for those moving out of large firms and those moving into the labour force or out of self-employment.

The skills estimates are all significant. For those who are sick, it seems that intelligence and numeracy help you to get well, whereas English knowledge and literacy are linked to longer spells of illness. It is unclear at this point what exactly this means.

5.2 Effects for Earnings

Let us now look at the selection equation for earned income (Table 9), the first stage of our earnings analysis. The variables that identify this selection are number of children, a dummy for married, and a dummy for household head.

Gender is not a significant determinant of earning an income. That men and women have similar employment rates may be a bit surprising, but this is probably because women are more likely to be self-employed, and our binary earned income variable does not pick up on this nuance. Some may also question why the Raven score variable is negative, and the rest of the skills and education are insignificant. One would expect those who do not earn income to be on average, less educated and less able. The answer may have something to do with signalling and worker quality. It is likely that high quality workers stay out of the labour force, waiting for a good job to come along, and do not seek self-employment in the interim because they fear it would send a negative signal to potential (good) employers. Therefore, many of those without income may in fact be more qualified than not.

As for the health factors, neither height nor the health shock (days ill last year) are important. Although a past illness enters negatively, therefore decreasing the probability of having earned income, it is statistically insignificant.[6]

Now, we can turn to the earnings equations in Table 10. The table shows the results of regressing the log of monthly wages on several explanatory variables. Columns 1 through 4 are estimates

[6] A possible extension, that is beyond the scope of this paper, is to investigate further the fact that height does clearly have something to do with occupation through a multinomial logit selection equation.

without controlling for selection, and Columns 5-8 include an inverse Mills ratio ('heckman') from the probit for earned income (the full model from Column 4, Table 9). The first column offers a standard specification, in which hours, gender, age, education, and occupation all affect earnings significantly. Here, it is clear that men earn more than women, and that age, which may also pick up experience, raises earnings but at a decreasing rate. Employees in small firms earn significantly less than the self-employed, and they both earn much less than others in large firms and the public sector. When skills are added, the overall picture remains the same. And although the skills variables take away some of the education effect, they certainly do not wipe it out, which means that the skills are picking up characteristics that are not measured with years of schooling.

A previous illness, or health shock, has a negative impact on wages, but height does not play any role. This suggests that short-term shocks may affect income but long-term health does not. Moreover, if health is controlling for the unobservables associated with health, then this is a very strong result indeed. It must be acknowledged that outliers have a strong effect, and this effect loses its statistical significance if we take out those people who experienced more than 14 days of illness last year. Even so, it is clear that for some people, who have serious illnesses, these do have a downward impact on earnings.

After controlling for selection, the health shock is still significant (column 7), but we lose our ability to identify this effect precisely when we put the whole model together by adding in skills (column 8). Ability, measured by the Raven's score, enters strongly here and increases earnings.

5.3 Dynamic Effects for Health

Let us now consider a dynamic setting, in which lagged health is included as a determinant of current health, and lagged earnings is included in the earnings equation. This autoregressive model brings up new econometric issues, which have yet to be fully addressed by the authors. However, we include them here because they are an integral part of understanding the intertemporal nature of the health-income relationship.

The estimates for the dynamic probit model are presented in Table 11. It is easy to verify that the point estimates, compared with Table 6, are nearly identical, which may be good evidence that the lagged health variable does not introduce too much bias, although we cannot rule out this possibility. As the results are so similar, we will limit our discussion here to the lagged days ill variable, which is positive and significant. An 10-day increase in illness suffered in the previous year translates into a 7% greater chance of falling ill this year.

Again, it is possible that this effect is being driven by outliers, and if we aggressively restrict the sample to people without long illness spells, the effect's statistical significance diminishes. Using a dummy to indicate an illness in the previous year is also statistically insignificant for the whole sample. If we instead use a dummy that indicates a long sickness (over 1 month) in the previous year, the point estimates of the other variables are essentially unchanged, while the dummy for a big health shock is statistically significant and shows that having a severe illness last year increases the chances of falling ill this year by over 50%. This is important evidence of the persistence of illness, especially for those who experience a somewhat severe health shock.

5.3.1 Earnings or Occupation?

This occupational effect on illness that we have uncovered by including dummies for job movement may have more to do with income variations across occupation than the occupation itself. In order to investigate this question, we need to limit the sample size to those for whom we have observed earned income. These results are presented in Table 12.

If we narrow our sample accordingly and control for the level of lagged income, it is true that the lagged occupation effects lose their statistical significance in the probit model for having any illness—income, for which the coefficient is significant, overrides the occupational differences. However, the dummy for leaving self-employment is still negative and weakly significant, although we have trouble identifying it when the skills are added to the model. The effect of the income level is positive, so that higher income is associated with a higher probability of any sickness. Again, this may be because those with higher incomes can afford to be ill. When we run a count regression but control for the level of income, the coefficient on income is negative but insignificant. Nevertheless, even if some of the occupational effect on health is working through earnings, it does not take away from the fact that workers in different occupations experience different health outcomes.

5.4 Dynamic Effects for Earnings

To examine the dynamic relationship in the earnings equation, we must limit the sample to those people for whom we observe earned income in both 2005 and 2006. Estimates are offered in Table 14 (refer to Table 13 for the relevant comparison of the same sample but without the log of last year's wage).

It is immediately clear that the previous year's log wage dominates almost all of the other explanatory variables. Earnings is clearly a persistent process, and last year's earnings is a strong, significant determinant of earnings today. Column 9-IV is an instrumental variables model which instruments the lag of log wage with an individual's log wage from 2004 to try to purge some of the bias inherent to the dynamic linear panel model.[7] This raises the estimate, but does not change the conclusions. Interestingly, last year's days of illness variable is still significant, despite the power of lagged wages. This is a key result and implies that the effect of a health shock is potentially very strong. Given our earned income probit findings, we conclude that health shocks impact directly on earnings rather than indirectly through labour participation.

6 Conclusion: Policy Implications

This paper attempts to assess the impact of various dimensions of human capital on earnings as well as the relationships between some of these dimensions and health. Skills and ability

[7] For some individuals in the sample this was a predicted value from the regression of 2004 log wage on the 2004 levels of age, education, gender, occupation, and job tenure.

definitely play a role in short-term health, but they are not as important for earnings. Education is not a significant player, nor is health knowledge. Looking closer at the education-health relationship is a potential improvement on our present analysis. In particular, determining whether health operates indirectly through education is an interesting question but is beyond the scope of this paper.

We also investigate the possibility of a virtuous or vicious cycle by which good (poor) health leads to improved (lower) incomes which leads to improved (poorer) health. We present evidence that occupational status, job movements and health are linked. Health shocks do not affect the choice between working and not working, and that illness, at least for some who experience serious health shocks, is persistent over time. Health shocks do have a negative effect on earnings, especially for those who have recently been severely ill.

Overall, the question of whether higher incomes in large firms translate into better health is highly relevant for policy, as we decide whether promotion of microenterprises is truly a pro-poor strategy along non-income dimensions of welfare. We find some evidence that self-employment may cause bad health, but we do not reach the same conclusion for workers in small firms. Furthermore, our results suggest that job movement and health are intimately related. This relationship should be explored more in future research.

Further research into the destination as well as the source of occupational change would be extremely helpful. Using our data, it would be possible to link occupational changes to changes in income, from which we can get a better understanding of why people move jobs, assuming that income jumps indicate a choice to move jobs, and income declines indicate being forced out of a job. This is left as a subject for future work.

Figure 1: **Height by Occupation**

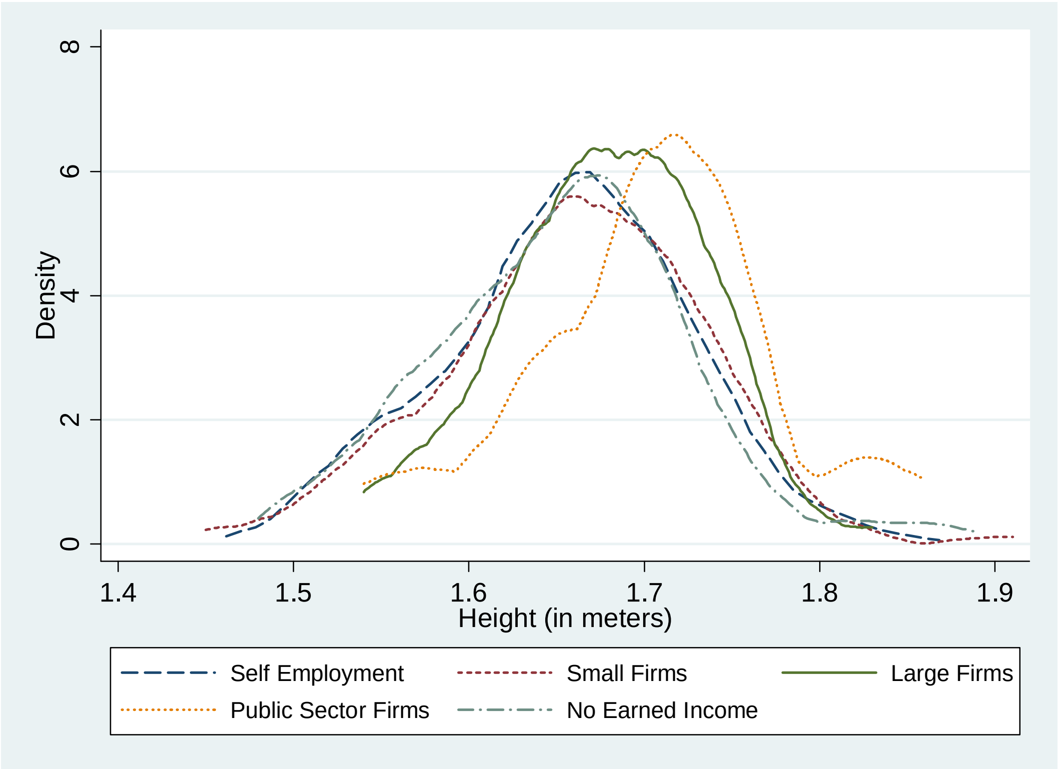

Table 1: **Average test scores, by test type and occupation**

	Self-Employed	**Small Firms**	**Large Firms**	**Public Sector**	**No Earned Income**	**Total**
Raven's (%)	10.2	12.1	16.4	12.0	14.2	12.2
	(15.3)	(17.3)	(19.1)	(17.8)	(17.5)	(16.9)
Mathematics (%)	22.2	29.3	45.7	55.3	31.2	29.9
	(23.6)	(27.3)	(27.7)	(33.3)	(28.7)	(27.9)
English (%)	42.8	56.8	79.6	87.1	55.8	54.7
	(44.5)	(42.5)	(33.1)	(30.7)	(44.7)	(44.5)
Reading (%)	13.7	19.5	38.0	49.6	23.3	21.6
	(26.2)	(30.1)	(36.9)	(37.2)	(32.8)	(32.0)
Health knowledge (%)	38.4	37.6	41.9	55.8	35.7	39.0
	(16.9)	(15.6)	(17.8)	(18.1)	(16.8)	(17.3)

Values reported are sample means; standard deviations are in parentheses.

Table 2: **Days inactive due to illness or injury in the past year**

	Self-Employed	**Small Firms**	**Large Firms**	**Public Sector**	**No Earned Income**	**Total**
2006 Wave						
Proportion with at least						
1 day of illness	0.36	0.25	0.32	0.37	0.25	0.32
	(0.48)	(0.44)	(0.47)	(0.49)	(0.43)	(0.47)
Days of illness	2.98	0.95	2.85	1.84	7.21	3.46
	(14.57)	(1.98)	(13.29)	(3.69)	(44.96)	(23.04)
Days of illness if >0	8.24	3.75	8.90	5.00	29.16	10.90
	(23.38)	(2.24)	(22.48)	(4.66)	(87.58)	(39.94)
2005 Wave						
Proportion with at least						
1 day of illness	0.37	0.25	0.30	0.39	0.27	0.32
	(0.48)	(0.44)	(0.46)	(0.50)	(0.45)	(0.47)
Days of illness	3.48	1.72	3.03	1.61	4.01	3.19
	(16.00)	(6.38)	(16.38)	(2.66)	(21.81)	(16.10)
Days of illness if >0	9.52	6.78	10.21	4.07	14.83	9.90
	(25.38)	(11.36)	(29.08)	(2.81)	(40.30)	(27.22)

Values reported are sample means; standard deviations are in parentheses.

Table 3: **Variable Definitions**

Variable	**Description**
Male	=1 if male.
Age	age in years.
Education	years of education, not including vocational training.
Height	height in meters.
NHIS	community-level measure of the percent of individuals who have registered for an NHIS card (from 0 to 1).
Empl. small firm	=1 if employed in firm with $\leq$10 employees.
Empl. big firm	=1 if employed in firm with more than 10 employees.
Empl. public	=1 if employed in the public sector.
Raven score	correct answers as a fraction of the total on the Raven's test.
Math score	correct answers as a fraction of the total on the mathematics test.
English score	correct answers as a fraction of the total on the English test.
Reading score	correct answers as a fraction of the total on the reading test.
Health knowledge	correct answers as a fraction of the total on the health knowledge test.
ln(hours)	natural log of hours of work per week.
Earned Income	=1 if non-zero wages.
Children	individual's number of children.
Married	=1 if married.
HH head	=1 if head of household.
ln(wage)	natural log of real monthly earnings in Ghanaian cedis for all wage earners (both the self-employed and those in the market sector).

Table 4: **Summary Statistics for Regression Sample**

Male	0.43
	(0.50)
Age (years)	33.85
	(10.46)
Education (years)	8.08
	(4.17)
Height (m)	1.66
	(0.075)
Children	1.29
	(1.69)
Married	0.45
	(.50)
Self-employed	0.47
	(0.50)
Employed in small firm	0.14
	(0.35)
Employed in big firm	0.15
	(0.35)
Employed in public sector	0.04
	(0.20)
No earned income	0.20
	(0.40)
N	873

Values reported are sample means; standard deviations are in parentheses.

Table 5: **Health Knowledge Control Function**

Dependent Variable: Health knowledge score ($\in [0, 1]$)	
Mother's education (years)	.001
	(.001)
Father's education (years)	.003
	(.001)**
Children	.018
	(.003)***
Raven score	-.083
	(.031)***
English score	.082
	(.017)***
Math score	.076
	(.033)**
Reading score	.082
	(.024)***
Cape Coast	-.058
	(.027)**
Accra	-.064
	(.019)***
Kumasi	-.037
	(.019)**
Const.	.314
	(.020)***
Obs.	869
R^2	.246

All standard errors are robust.

Table 6: **Probit Model: Impact of Occupation on Illness**

Dependent Variable: Binary, takes value of one if any days of illness reported, zero otherwise	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	-.111	-.113	-.134	-.095	-.090	-.041	-.046
	(.098)	(.098)	(.104)	(.105)	(.107)	(.109)	(.109)
Age	.016	-.046	-.048	-.063	-.070	-.075	-.068
	(.005)***	(.032)	(.032)	(.033)*	(.033)**	(.033)**	(.033)**
Age^2		.083	.086	.103	.113	.116	.110
		(.042)**	(.042)**	(.043)**	(.043)***	(.043)***	(.043)**
Education (years)	.0007	.004	.003	.007	.005	-.003	-.003
	(.012)	(.012)	(.012)	(.012)	(.014)	(.014)	(.014)
Self-empl. last year	.024	.071	.071	.288	.320	.303	.314
	(.116)	(.118)	(.117)	(.152)*	(.153)**	(.154)**	(.155)**
Empl. small firm last year	.018	.028	.027	.138	.159	.132	.131
	(.161)	(.161)	(.160)	(.219)	(.219)	(.221)	(.222)
Empl. big firm last year	.266	.289	.292	.219	.215	.204	.202
	(.173)	(.174)*	(.174)*	(.222)	(.223)	(.223)	(.222)
Empl. public last year	-.103	-.116	-.119	.254	.252	.180	.165
	(.229)	(.232)	(.233)	(.289)	(.294)	(.297)	(.297)
NHIS	-.923	-.870	-.884	-.952	-.828	-.892	-.848
	(.401)**	(.405)**	(.405)**	(.409)**	(.415)**	(.419)**	(.419)**
Left self-empl.				-.576	-.560	-.511	-.535
				(.208)***	(.210)***	(.211)**	(.211)**
Left small firm				-.004	.020	.034	.024
				(.266)	(.266)	(.268)	(.269)
Left big firm				.553	.572	.568	.621
				(.312)*	(.313)*	(.311)*	(.317)*
Left public				-.642	-.676	-.636	-.594
				(.448)	(.460)	(.463)	(.463)
Left no income/unempl.				.234	.284	.287	.290
				(.179)	(.180)	(.181)	(.182)
Height (m)			.408	.346	.170	.031	.113
			(.689)	(.686)	(.692)	(.701)	(.704)
Raven score					.605	.708	.537
					(.288)**	(.292)**	(.316)*
English score					-.364	-.415	-.233
					(.153)**	(.155)***	(.203)
Math score					.220	.165	.317
					(.284)	(.285)	(.304)
Reading score					.336	.257	.461
					(.223)	(.224)	(.268)*
Health knowledge						.858	-1.265
						(.310)***	(1.490)
control							2.173
							(1.500)
Obs.	873	873	873	873	873	873	869
Pseudo R^2	.038	.041	.042	.055	.067	.073	.075
Log likelihood	-524.918	-522.903	-522.711	-515.25	-509.005	-505.468	-502.9

All standard errors are robust. 'Control' refers to a control function and represents the residuals of a first-stage regression of health knowledge on all exogenous regressors plus parental education and number of children. All regressions also include a control for city of residence. N changes slightly in column 7 due to missing values of the instruments (parental education).

Table 7: **OLS Model: Impact of Occupation on Days of Illness**

Dependent Variable: Days of illness	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	-3.426	-3.410	-2.886	-2.784	-2.544	-2.829	-3.035
	(1.549)**	(1.546)**	(1.414)**	(1.373)**	(1.314)*	(1.469)*	(1.543)**
Age	.308	-.351	.318	.325	.316	.328	.367
	(.142)**	(.592)	(.146)**	(.151)**	(.145)**	(.152)**	(.168)**
Age^2		.889					
		(.919)					
Education (years)	-.005	.029	.012	.023	.119	.163	.140
	(.125)	(.111)	(.126)	(.131)	(.185)	(.201)	(.198)
Self-empl. last year	-6.210	-5.743	-6.169	-9.633	-9.659	-9.532	-9.480
	(3.533)*	(3.344)*	(3.519)*	(5.755)*	(5.736)*	(5.655)*	(5.658)*
Empl. small firm last year	-3.736	-3.656	-3.706	-7.743	-7.472	-7.267	-7.244
	(2.171)*	(2.148)*	(2.162)*	(4.324)*	(4.205)*	(4.070)*	(4.070)*
Empl. big firm last year	-1.217	-1.018	-1.273	-4.537	-4.247	-4.183	-4.240
	(3.122)	(3.074)	(3.145)	(5.258)	(5.211)	(5.174)	(5.197)
Empl. public last year	-7.418	-7.538	-7.383	-10.757	-10.672	-10.242	-10.737
	(3.729)**	(3.794)**	(3.712)**	(6.093)*	(6.096)*	(5.762)*	(5.950)*
NHIS	-3.585	-3.073	-3.261	-2.342	-1.128	-.851	-.227
	(3.939)	(3.753)	(4.112)	(4.084)	(4.270)	(4.526)	(4.355)
Left self-empl.				-.109	-.117	-.417	-.610
				(1.284)	(1.279)	(1.283)	(1.243)
Left small firm				1.207	1.101	.968	.795
				(1.122)	(1.221)	(1.226)	(1.211)
Left big firm				-.960	-1.227	-1.154	-1.182
				(4.025)	(4.065)	(4.092)	(4.136)
Left public				-.555	-.585	-.870	-.319
				(1.944)	(1.915)	(1.835)	(1.917)
Left no income/unempl.				-7.235	-7.462	-7.439	-7.431
				(4.663)	(4.691)	(4.671)	(4.695)
Height (m)			-10.274	-9.668	-9.414	-8.654	-8.246
			(10.724)	(10.512)	(10.876)	(10.080)	(10.156)
Raven score					-3.047	-3.560	-6.163
					(2.785)	(2.797)	(4.047)
English score					-.253	.051	3.000
					(1.976)	(2.017)	(2.106)
Math score					-8.057	-7.740	-5.318
					(3.204)**	(3.000)***	(3.030)*
Reading score					5.830	6.210	9.278
					(3.235)*	(3.600)*	(3.802)**
Health knowledge						-4.745	-36.786
						(6.085)	(20.769)*
control							33.112
							(21.771)
Obs.	873	873	873	873	873	873	869
R^2	.033	.034	.033	.04	.046	.047	.049

All standard errors are robust. ‘Control’ refers to a control function and represents the residuals of a first-stage regression of health knowledge on all exogenous regressors plus parental education and number of children. All regressions also include a control for city of residence. N changes slightly in column 7 due to missing values of the instruments (parental education).

Table 8: **Zero-Truncated NegBin Model: Impact of Occupation on Days of Illness**

Dependent Variable: Days of illness if ≥1	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	-.405	-.385	-.408	-.453	-.409	-.428	-.416
	(.232)*	(.234)	(.243)*	(.236)*	(.221)*	(.221)*	(.221)*
Age	.034	.072	.034	.039	.030	.031	.032
	(.010)***	(.075)	(.010)***	(.010)***	(.010)***	(.010)***	(.010)***
Age^2		-.050					
		(.100)					
Education (years)	.005	.002	.005	.003	.031	.032	.031
	(.025)	(.026)	(.025)	(.025)	(.027)	(.026)	(.026)
Self-empl. last year	-1.064	-1.102	-1.065	-1.560	-1.400	-1.363	-1.325
	(.263)***	(.274)***	(.264)***	(.331)***	(.299)***	(.300)***	(.305)***
Empl. small firm last year	-1.006	-1.008	-1.006	-1.700	-1.114	-1.131	-1.173
	(.381)***	(.381)***	(.381)***	(.510)***	(.496)**	(.492)**	(.494)**
Empl. big firm last year	-.163	-.198	-.163	-.127	-.056	-.003	-.016
	(.393)	(.399)	(.393)	(.493)	(.456)	(.460)	(.460)
Empl. public last year	-1.001	-1.008	-1.004	-1.375	-1.183	-1.172	-1.162
	(.540)*	(.538)*	(.543)*	(.630)**	(.599)**	(.594)**	(.595)*
NHIS	-.015	-.098	-.021	.135	-.190	-.104	-.027
	(.916)	(.930)	(.930)	(.894)	(.789)	(.797)	(.800)
Left self-empl.				.231	-.021	-.027	-.123
				(.552)	(.524)	(.522)	(.531)
Left small firm				.554	.004	.071	.089
				(.593)	(.563)	(.568)	(.566)
Left big firm				-1.520	-1.474	-1.452	-1.474
				(.581)***	(.548)***	(.548)***	(.547)***
Left public				-.745	-.621	-.583	-.450
				(1.078)	(1.001)	(.998)	(1.012)
Left no income/unempl.				-1.090	-1.138	-1.142	-1.182
				(.406)***	(.377)***	(.374)***	(.376)***
Height (m)			.057	-.201	-.991	-.846	-.837
			(1.517)	(1.501)	(1.302)	(1.312)	(1.316)
Raven score					-1.820	-1.868	-2.085
					(.584)***	(.586)***	(.638)***
English score					.683	.706	.945
					(.313)**	(.313)**	(.418)**
Math score					-2.139	-2.066	-1.845
					(.574)***	(.576)***	(.627)***
Reading score					.878	.923	1.137
					(.452)*	(.454)**	(.515)**
Health knowledge						-.481	-3.179
						(.662)	(3.177)
control							2.725
							(3.140)
Obs.	277	277	277	277	277	277	277
Pseudo R^2	.051	.051	.051	.059	.076	.076	.077
Log likelihood	-788.616	-788.493	-788.615	-781.696	-767.844	-767.582	-767.208

All standard errors are robust. ‘Control’ refers to a control function and represents the residuals of a first-stage regression of health knowledge on all exogenous regressors plus parental education and number of children. All regressions also include a control for city of residence.

Table 9: **Probit for Earned Income**

Dependent Variable: Binary, takes value of one if any earned income reported, zero otherwise

	(1)	(2)	(3)	(4)
Male	.038	.014	.010	.009
	(.112)	(.114)	(.119)	(.119)
Age	.143	.149	.148	.149
	(.037)***	(.037)***	(.038)***	(.038)***
Age^2	-.180	-.189	-.188	-.190
	(.049)***	(.049)***	(.049)***	(.050)***
Education (years)	-.011	-.013	-.013	-.014
	(.013)	(.016)	(.016)	(.016)
Children	.035	.041	.041	.039
	(.042)	(.043)	(.043)	(.043)
Married	.250	.244	.245	.238
	(.132)*	(.133)*	(.133)*	(.133)*
HH head	.472	.482	.482	.489
	(.131)***	(.132)***	(.132)***	(.132)***
Raven score		-.683	-.684	-.675
		(.321)**	(.321)**	(.322)**
English score		.123	.123	.130
		(.172)	(.172)	(.172)
Math score		.086	.085	.071
		(.318)	(.318)	(.318)
Reading score		-.058	-.061	-.051
		(.248)	(.249)	(.250)
Height (m)			.086	.062
			(.756)	(.757)
Days ill last year				-.003
				(.003)
Obs.	873	873	873	873
Pseudo R^2	.104	.11	.11	.111
Log likelihood	-390.732	-388.17	-388.164	-387.782

Table 10: **Earnings Model: Impact of Days of Illness on Earnings**

Dependent Variable: ln(wage)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(hours)	.190	.176	.170	.156	.199	.183	.178	.164
	(.071)***	(.072)**	(.073)**	(.074)**	(.071)***	(.072)**	(.073)**	(.073)**
Male	.309	.255	.300	.251	.283	.239	.255	.218
	(.069)***	(.075)***	(.070)***	(.075)***	(.073)***	(.078)***	(.075)***	(.079)***
Age	.067	.064	.064	.060	.036	.042	.004	.010
	(.023)***	(.024)***	(.023)***	(.024)**	(.038)	(.039)	(.041)	(.042)
Age^2	-.067	-.063	-.063	-.059	-.031	-.038	.007	-.0001
	(.031)**	(.031)**	(.031)**	(.031)*	(.047)	(.048)	(.050)	(.052)
Education (years)	-.034	-.033	-.031	-.031	-.032	-.032	-.027	-.027
	(.024)	(.024)	(.025)	(.025)	(.024)	(.024)	(.025)	(.025)
Education^2	.435	.426	.383	.382	.432	.425	.371	.373
	(.171)**	(.171)**	(.175)**	(.175)**	(.169)**	(.170)**	(.169)**	(.170)**
Empl. small firm	-.268	-.266	-.275	-.273	-.262	-.262	-.266	-.266
	(.079)***	(.079)***	(.079)***	(.079)***	(.079)***	(.079)***	(.079)***	(.079)***
Empl. big firm	.184	.192	.156	.166	.202	.204	.185	.190
	(.089)**	(.089)**	(.089)*	(.089)*	(.089)**	(.090)**	(.089)**	(.090)**
Empl. public	.367	.348	.345	.332	.360	.344	.337	.327
	(.144)**	(.144)**	(.143)**	(.144)**	(.144)**	(.144)**	(.143)**	(.143)**
Raven score			.285	.277			.440	.408
			(.187)	(.187)			(.206)**	(.208)**
English score			-.105	-.105			-.151	-.146
			(.107)	(.107)			(.106)	(.106)
Math score			.227	.220			.215	.213
			(.175)	(.174)			(.176)	(.175)
Reading score			.058	.036			.085	.061
			(.144)	(.145)			(.144)	(.145)
Heckman					-.393	-.273	-.744	-.628
					(.361)	(.374)	(.386)*	(.407)
Days ill last year		-.003		-.003		-.003		-.002
		(.001)**		(.001)**		(.001)*		(.001)
Height (m)		.796		.748		.775		.694
		(.531)		(.532)		(.531)		(.535)
Obs.	699	699	699	699	699	699	699	699
R^2	.184	.189	.19	.195	.185	.19	.195	.198

'Heckman' refers to the inverse Mills ratio using predicted values from a first-stage probit for earned income. All standard errors are robust. All regressions also include a control for city of residence.

Table 11: **Probit Model: Impact of Employment and Earnings on Days of Illness, including Days Ill Last Year**

Dependent Variable: Binary, takes value of one if any days of illness reported, zero otherwise							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Days ill last year	.007	.007	.007	.007	.007	.008	.008
	(.003)**	(.003)**	(.003)**	(.003)**	(.003)**	(.003)**	(.003)**
Male	-.107	-.109	-.133	-.096	-.090	-.042	-.047
	(.098)	(.098)	(.104)	(.106)	(.107)	(.109)	(.110)
Age	.015	-.047	-.049	-.064	-.072	-.076	-.070
	(.005)***	(.032)	(.032)	(.033)*	(.033)**	(.033)**	(.033)**
Age^2		.083	.086	.104	.114	.117	.112
		(.042)**	(.042)**	(.043)**	(.044)***	(.044)***	(.044)**
Education (years)	.0007	.004	.003	.007	.005	-.003	-.003
	(.012)	(.012)	(.012)	(.012)	(.014)	(.014)	(.014)
Self-empl. last year	.033	.080	.080	.313	.344	.328	.338
	(.116)	(.118)	(.118)	(.152)**	(.154)**	(.154)**	(.155)**
Empl. small firm last year	.031	.042	.041	.169	.187	.162	.160
	(.162)	(.161)	(.161)	(.220)	(.220)	(.222)	(.222)
Empl. big firm last year	.255	.278	.282	.232	.225	.216	.214
	(.175)	(.176)	(.176)	(.224)	(.225)	(.225)	(.224)
Empl. public last year	-.085	-.098	-.101	.299	.295	.224	.210
	(.229)	(.233)	(.233)	(.288)	(.294)	(.297)	(.297)
NHIS	-.916	-.863	-.879	-.950	-.826	-.893	-.852
	(.403)**	(.406)**	(.406)**	(.411)**	(.416)**	(.420)**	(.420)**
Height (m)			.448	.393	.222	.086	.165
			(.697)	(.694)	(.700)	(.710)	(.712)
Left self-empl.				-.575	-.560	-.508	-.530
				(.208)***	(.211)***	(.211)**	(.211)**
Left small firm				-.006	.019	.032	.023
				(.266)	(.267)	(.269)	(.270)
Left big firm				.543	.551	.556	.614
				(.318)*	(.317)*	(.316)*	(.323)*
Left public				-.672	-.705	-.666	-.626
				(.451)	(.464)	(.468)	(.467)
Left no income/unempl.				.262	.312	.317	.319
				(.180)	(.181)*	(.182)*	(.182)*
Raven score					.594	.698	.541
					(.288)**	(.292)**	(.316)*
English score					-.377	-.427	-.257
					(.154)**	(.157)***	(.205)
Math score					.253	.197	.334
					(.286)	(.287)	(.305)
Reading score					.319	.237	.426
					(.223)	(.225)	(.268)
Health knowledge						.869	-1.094
						(.311)***	(1.498)
control							2.010
							(1.509)
Obs.	873	873	873	873	873	873	869
Pseudo R^2	.043	.046	.047	.061	.072	.079	.081
Log likelihood	-522.225	-520.219	-519.988	-512.319	-506.006	-502.407	-500.022

All standard errors are robust. 'Control' refers to a control function and represents the residuals of a first-stage regression of health knowledge on all exogenous regressors plus parental education and number of children. All regressions also include a control for city of residence. N changes slightly in column 7 due to missing values of the instruments (parental education).

Table 12: **Probit Model: Impact of Occupation and Wages on Illness for 2-year Income Earners**

Dependent Variable: Binary, takes value of one if any days of illness reported, zero otherwise							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Days ill last year	.013	.013	.014	.014	.014	.014	.013
	(.006)**	(.006)**	(.007)**	(.007)**	(.006)**	(.006)**	(.006)**
ln(wage) last year	.163	.166	.168	.168	.140	.136	.137
	(.071)**	(.072)**	(.072)**	(.072)**	(.073)*	(.072)*	(.072)*
Male	-.318	-.319	-.267	-.267	-.291	-.233	-.230
	(.130)**	(.130)**	(.141)*	(.141)*	(.148)**	(.151)	(.152)
Age	.017	-.0004	-.011	-.011	-.015	-.018	-.016
	(.006)***	(.043)	(.044)	(.044)	(.044)	(.044)	(.044)
Age^2		.022	.034	.034	.041	.043	.040
		(.056)	(.057)	(.057)	(.057)	(.056)	(.057)
Education (years)	.005	.005	.008	.008	-.011	-.016	-.016
	(.014)	(.014)	(.014)	(.014)	(.017)	(.017)	(.017)
Empl. small firm last year	.131	.122	.044	.044	.012	-.002	-.007
	(.177)	(.178)	(.217)	(.217)	(.219)	(.222)	(.222)
Empl. big firm last year	.225	.216	.006	.006	-.109	-.097	-.104
	(.189)	(.192)	(.219)	(.219)	(.226)	(.226)	(.225)
Empl. public last year	.006	-.006	.132	.132	-.007	-.058	-.069
	(.254)	(.256)	(.293)	(.293)	(.309)	(.310)	(.312)
NHIS	-1.484	-1.473	-1.509	-1.509	-1.475	-1.571	-1.555
	(.492)***	(.492)***	(.495)***	(.495)***	(.520)***	(.525)***	(.525)***
Left self-empl.			-.468	-.468	-.468	-.422	-.433
			(.270)*	(.270)*	(.273)*	(.273)	(.272)
Left small firm			.005	.005	-.011	-.0001	-.002
			(.305)	(.305)	(.306)	(.310)	(.311)
Left big firm			.529	.529	.578	.564	.562
			(.377)	(.377)	(.384)	(.383)	(.383)
Left public			-.582	-.582	-.607	-.574	-.561
			(.509)	(.509)	(.543)	(.544)	(.542)
Left no income/unempl.							
Height (m)			.145	.145	-.190	-.322	-.273
			(.926)	(.926)	(.929)	(.946)	(.951)
Raven score					.888	.968	.918
					(.363)**	(.370)***	(.404)**
English score					-.243	-.306	-.267
					(.198)	(.203)	(.259)
Math score					.316	.254	.286
					(.367)	(.368)	(.389)
Reading score					.588	.496	.544
					(.281)**	(.281)*	(.338)
Health knowledge						.891	.388
						(.399)**	(1.806)
control							.526
							(1.829)
Obs.	559	559	559	559	559	559	558
Pseudo R^2	.09	.09	.099	.099	.123	.129	.129
Log likelihood	-327.338	-327.254	-324.016	-324.016	-315.542	-313.188	-312.763

All standard errors are robust. ‘Control’ refers to a control function and represents the residuals of a first-stage regression of health knowledge on all exogenous regressors plus parental education and number of children. All regressions also include a control for city of residence. N changes slightly in column 7 due to missing values of the instruments (parental education).

Table 13: **Earnings Model: Impact of Illness on ln(wage) for 2-year Income Earners, excluding lag of ln(wage)**

Dependent Variable: ln(wage)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(hours)	.181	.154	.174	.149	.185	.160	.176	.154
	(.091)**	(.093)*	(.090)*	(.093)	(.091)**	(.093)*	(.091)*	(.093)*
Male	.348	.338	.305	.302	.332	.300	.298	.276
	(.081)***	(.082)***	(.087)***	(.087)***	(.087)***	(.089)***	(.091)***	(.093)***
Age	.075	.071	.073	.069	.058	.023	.066	.032
	(.027)***	(.027)***	(.027)***	(.027)**	(.045)	(.049)	(.046)	(.050)
Age^2	-.078	-.072	-.075	-.069	-.058	-.016	-.066	-.026
	(.035)**	(.035)**	(.035)**	(.035)**	(.055)	(.060)	(.055)	(.061)
Education (years)	-.055	-.052	-.053	-.051	-.053	-.048	-.053	-.048
	(.026)**	(.027)*	(.026)**	(.027)*	(.026)**	(.027)*	(.026)**	(.027)*
Education^2	.527	.472	.517	.467	.523	.457	.516	.457
	(.181)***	(.186)**	(.181)***	(.187)**	(.179)***	(.180)**	(.181)***	(.183)**
Empl. small firm	-.213	-.234	-.216	-.235	-.207	-.222	-.213	-.226
	(.092)**	(.091)**	(.093)**	(.092)**	(.092)**	(.092)**	(.093)**	(.093)**
Empl. big firm	.263	.227	.262	.229	.273	.249	.267	.246
	(.101)***	(.101)**	(.102)***	(.102)**	(.102)***	(.103)**	(.103)***	(.103)**
Empl. public	.404	.362	.386	.349	.403	.359	.386	.348
	(.170)**	(.169)**	(.170)**	(.169)**	(.170)**	(.168)**	(.170)**	(.169)**
Raven score		.228		.223		.362		.327
		(.204)		(.203)		(.231)		(.234)
English score		-.162		-.154		-.195		-.181
		(.122)		(.122)		(.122)		(.123)
Math score		.458		.426		.438		.415
		(.196)**		(.197)**		(.199)**		(.199)**
Reading score		-.043		-.041		-.018		-.023
		(.162)		(.164)		(.164)		(.165)
heckman					-.226	-.596	-.093	-.457
					(.423)	(.469)	(.430)	(.486)
Daysill last year			-.004	-.004			-.004	-.003
			(.001)***	(.001)***			(.001)***	(.001)**
Height (m)			.582	.506			.577	.472
			(.585)	(.584)			(.586)	(.587)
Obs.	559	559	559	559	559	559	559	559
R^2	.189	.199	.194	.204	.189	.202	.194	.205

'Heckman' refers to the inverse Mills ratio using predicted values from a first-stage probit for earned income. All standard errors are robust. All regressions also include a control for city of residence.

Table 14: **Earnings Model: Impact of Illness on ln(wage) for 2-year Income Earners, including lag of ln(wage)**

Dependent Variable: ln(wage)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9-IV)
ln(hours)	.089 (.072)	.078 (.073)	.091 (.073)	.083 (.074)	.089 (.073)	.078 (.074)	.087 (.074)	.080 (.075)	.067 (.077)
ln(wage) last year	.688 (.035)***	.684 (.035)***	.697 (.035)***	.693 (.035)***	.688 (.035)***	.685 (.036)***	.698 (.035)***	.696 (.035)***	.816 (.087)***
Male	.122 (.062)**	.112 (.063)*	.115 (.066)*	.107 (.067)	.124 (.064)*	.114 (.065)*	.129 (.069)*	.121 (.069)*	.094 (.072)
Age	.024 (.020)	.023 (.020)	.023 (.020)	.022 (.020)	.026 (.034)	.024 (.039)	.039 (.033)	.043 (.037)	.045 (.037)
Age^2	-.032 (.025)	-.030 (.025)	-.030 (.025)	-.029 (.025)	-.034 (.041)	-.032 (.046)	-.050 (.040)	-.053 (.044)	-.058 (.044)
Education (years)	-.034 (.019)*	-.037 (.019)*	-.031 (.019)*	-.034 (.019)*	-.034 (.019)*	-.037 (.020)*	-.032 (.019)*	-.036 (.020)*	-.034 (.020)*
$Education^2$	.281 (.113)**	.270 (.117)**	.264 (.111)**	.253 (.116)**	.282 (.113)**	.271 (.116)**	.268 (.112)**	.258 (.118)**	.223 (.120)*
Empl. small firm	-.102 (.073)	-.113 (.073)	-.105 (.074)	-.114 (.074)	-.103 (.073)	-.113 (.073)	-.111 (.074)	-.118 (.074)	-.100 (.077)
Empl. big firm	.074 (.069)	.059 (.069)	.067 (.069)	.052 (.069)	.073 (.069)	.058 (.070)	.056 (.068)	.041 (.069)	.006 (.075)
Empl. public	.085 (.102)	.061 (.104)	.071 (.102)	.048 (.105)	.086 (.102)	.061 (.104)	.071 (.102)	.048 (.105)	-.004 (.108)
Raven score		.013 (.139)		.016 (.137)		.009 (.159)		-.043 (.158)	-.107 (.161)
English score		-.030 (.094)		-.016 (.088)		-.029 (.099)		-.0005 (.090)	.031 (.092)
Math score		.307 (.157)*		.254 (.152)*		.308 (.157)*		.260 (.153)*	.233 (.157)
Reading score		-.108 (.127)		-.079 (.127)		-.109 (.129)		-.089 (.129)	-.100 (.132)
heckman					.027 (.336)	.019 (.393)	.219 (.313)	.259 (.364)	.383 (.378)
Daysill last year			-.007 (.001)***	-.007 (.001)***			-.007 (.001)***	-.007 (.001)***	-.007 (.001)***
Height (m)			-.161 (.445)	-.188 (.444)			-.150 (.444)	-.171 (.444)	-.282 (.453)
Obs.	559	559	559	559	559	559	559	559	559
R^2	.533	.536	.543	.545	.533	.536	.544	.546	.536

'Heckman' refers to the inverse Mills ratio using predicted values from a first-stage probit for earned income. All standard errors are robust. All regressions also include a control for city of residence.

References

Currie, J., & Madrian, B. (1999). Health, health insurance, and the labour market. In O. Ashenfelter & D. Card (Eds.), *The handbook of labor economics, vol. 3* (p. 3309-3414). Amsterdam: Elsevier Science.

Deolalikar, A. (1988). Nutrition and labor productivity in agriculture: Estimates for rural south India. *The Review of Economics and Statistics*, *70*, 406-413.

Dercon, S., & Hoddinott, J. (2003). Health, shocks, and poverty persistence. In S. Dercon (Ed.), *Insurance against poverty.* Oxford: Oxford University Press.

Ettner, S. (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, *15*, 67-85.

Fajnzylber, P., Maloney, W., & Rojas, G. (2006). Microenterprise dynamics in developing countries: How similar are they to those in the industrialized world? Evidence from Mexico. *World Bank Economic Review*, *20*, 389-419.

Gertler, P., & Gruber, J. (2002). Insuring consumption against illness. *American Economic Review*, *92*, 51-70.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*, 223-255.

Haveman, R., Wolfe, B., Kreider, B., & Stone, M. (1994). Market work, wages, and men's health. *Journal of Health Economics*, *13*, 163-182.

Lee, L.-F. (1982). Health and wage: A simultaneous equation model with multiple discrete indicators. *International Economic Review*, *23*, 199-221.

Lindelow, M., & Wagstaff, A. (2005). *Health shocks in China: Are the poor and uninsured less protected?* (World Bank Policy Research Working Paper No. 3740)

Mincer, J. (1974). *Schooling, experience and earnings.* New York, NY: National Bureau of Economic Research.

Pitt, M., & Rosenzweig, M. (1986). Agricultural prices, food consumption, and the health and productivity of Indonesian farmers. In I. Singh, L. Squire, & J. Strauss (Eds.), *Agricultural household models: Extensions, applications, and policy* (p. 153-182). Baltimore: Johns Hopkins University Press.

Sandefur, J., Serneels, P., & Teal, F. (2006). *African poverty through the lens of labour economics: Earnings and mobility in three countries.* (CSAE Working Paper, November)

Savedoff, W., & Schultz, T. (2000). Earnings and the elusive dividends of health. In W. Savedoff & T. Schultz (Eds.), *Wealth from health.* Washington, DC: Inter-American Development Bank. (Chapter 1)

Schultz, T. (2002). Wage gains associated with height as a form of health human capital. *American Economic Review*, *92*, 349-353.

Schultz, T. (2003). Wage rentals for reproducible human capital: Evidence from Ghana and the Ivory Coast. *Economics Of Human Biology*, *1*, 331-366.

Schultz, T., & Tansel, A. (1997). Wage and labor supply effects of illness in Côte d'Ivoire and Ghana: instrumental variable estimates for days disabled. *Journal of Development Economics*, *53*, 251-286.

Smith, J. (2003). *Consequences and predictors of new health events.* (NBER Working Paper No. 10063)

Söderbam, M., Teal, F., & Wambugu, A. (2005). Unobserved heterogeneity and the relation between earnings and firm size: Evidence from two developing countries. *Economic Letters*, *87*, 153-159.

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

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

Thomas, D., & Strauss, J. (1997). Health and wages: Evidence on men and women in urban Brazil. *Journal of Econometrics*, *77*, 159-185.

Wagstaff, A. (2006). The economic consequences of health shocks: evidence from Vietnam. *Journal of Health Economics*. (forthcoming, doi:10.1016/j.jhealeco.2006.07.001.)

For publication enquiries, please contact:

Communication Officer
RECOUP
Faculty of Education
184 Hills Road
Cambridge CB2 8PQ
United Kingdom

tel: (01223) 507231 e-mail: recoup@hermes.cam.ac.uk

www.educ.cam.ac.uk/RECOUP/index.html

ISSN 1754–2464