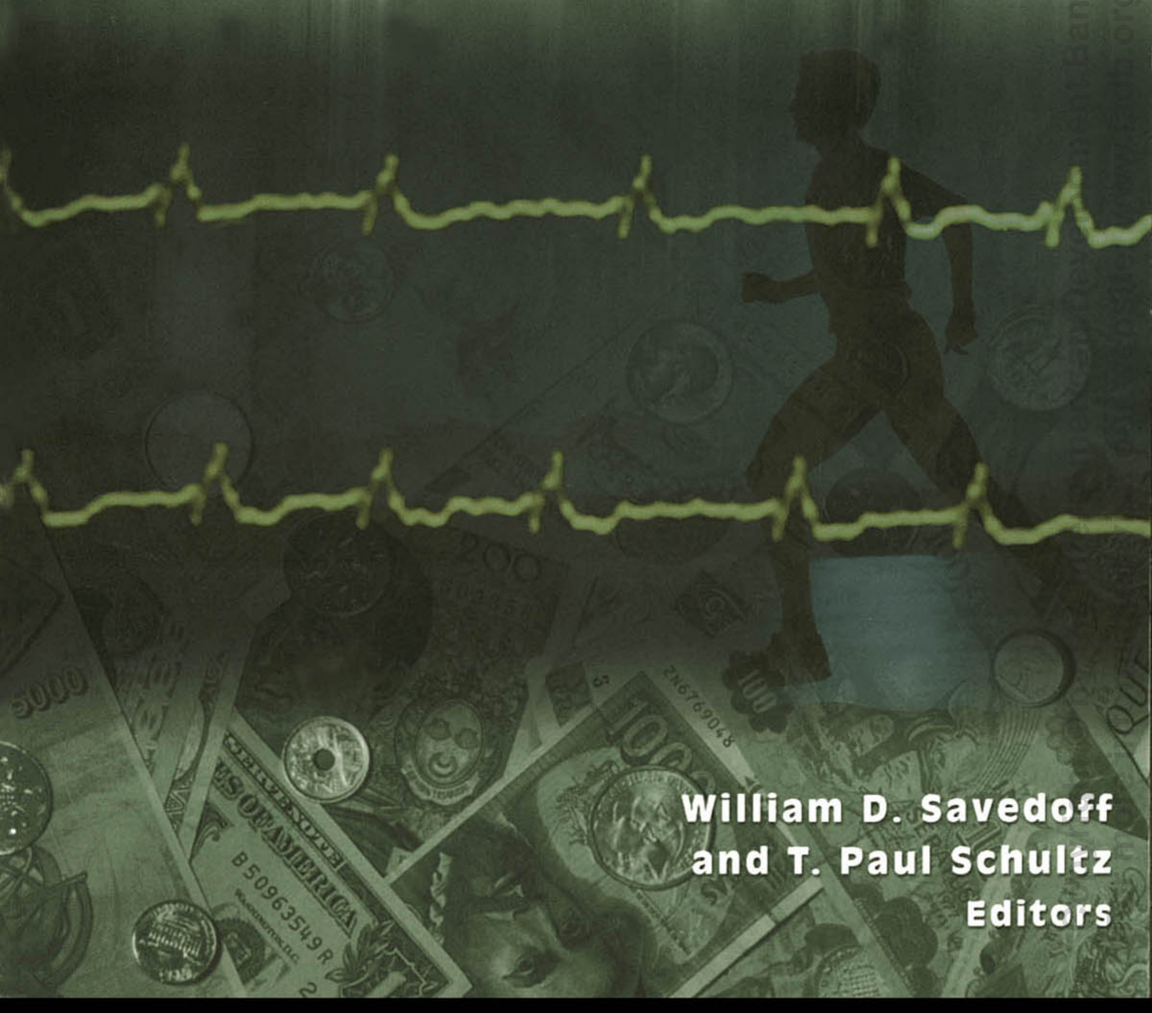
Inter-American Development Bank

Wealth from Health

Linking Social Investments to Earnings in Latin America



William D. Savedoff
and T. Paul Schultz
Editors

LATIN AMERICAN RESEARCH NETWORK

This page intentionally left blank

Wealth from Health

Linking Social Investments to Earnings in Latin America

William D. Savedoff
and T. Paul Schultz
Editors

INTER-AMERICAN DEVELOPMENT BANK
Washington, D.C.
2000

Latin American Research Network
Inter-American Development Bank

The Inter-American Development Bank created the Latin American Research Network in 1991 in order to strengthen policy formulation and contribute to the development policy agenda in Latin America.

Through a competitive bidding process, the Network provides grant funding to leading Latin American research centers to conduct studies on economic and social issues selected by the Bank in consultation with the region's development community.

Most of the studies are comparative, which allows the Bank to build its knowledge base and draw on lessons from experiences in macroeconomic and financial policy, modernization of the state, regulation, poverty and income distribution, social services, and employment.

Individual country studies are available as working papers and are also available in PDF format on the internet at http://www.iadb.org.res.41.htm.

1300 New York Avenue, N.W.
Washington, D.C. 20577

Cover: Images © 2000 PhotoDisc, Inc.

To order this book, contact:
IDB Bookstore
Tel: 1-877-PUBS IDB/(202) 623-1753
Fax: (202) 623-1709
E-mail: idb-books@iadb.org
www.iadb.org/pub

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the official positions of the Inter-American Development Bank.

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Wealth from health : linking social investment to earnings in Latin America / William D. Savedoff and T. Paul Schultz, editors.

p. cm.
Includes bibliographical references.
ISBN:1886938849

1. Health—Effect of Income on. 2. Medical economics—Latin America. 3. Latin America—Medical policy. 4. Latin America—Social conditions—Health aspects. 5. Labor productivity—Latin America—Health aspects. I. Savedoff, William D. II. T. Paul Schultz. III. Inter-American Development Bank.
306.461 W357—dc21

CONTENTS

ACKNOWLEDGMENTS

The studies in this book were financed by the Latin American Research Network of the Inter-American Development Bank. The initial impulse for this research project emerged from a meeting at which George Alleyne, Director of the Pan-American Health Organization, Enrique Iglesias, President of the IDB, and Nancy Birdsall, Executive Vice President of the IDB, agreed that the two institutions should pursue joint research on the economic impact of health investments. As a result of that meeting, the studies found in this book were commissioned by the IDB, while PAHO contracted parallel studies on the impact of health at the macroeconomic level. In addition to the hard work put into the case studies by their authors, the project would not have been possible without collaboration and suggestions from many friends and colleagues. In particular, Rubén Suárez was instrumental in the initiation of this research, while Duncan Thomas and John Strauss provided encouragement and valuable comments at different times. Others who provided input to specific chapters include Ricardo Aparicio, Rodolfo Cermeño, David Cutler, Julio Frenk, Silvestre Frenk, and Ana Langer. Research assistance was provided for various chapters by María Angélica Albino, Ana Mylena Aguilar, Héctor Arreola, Maite Guijarro, Erik López J., Martha Miranda, María del Carmen Franco Juárez, Pablo Sanabria, Tami Aritomi, and César Calvo. We offer special thanks to Elena Zuñiga, Víctor Guerrero, Daniel Hernández, CONAPO, INEGI, and all the other people and institutions that gave us access to the data that was critical to this study. Additional financial support for specific studies was provided by the Centro de Investigación in Mexico, the Economic Growth Center at Yale University, and the Consorcio de Investigación Económica in Peru. Norelis Betancourt and Raquel Gómez took care of the many administrative, logistical, and communication needs involved in conducting research in different countries, while John Smith, Rita Funaro, and Ron Weber were responsible for the final editing and publication of the working papers and the manuscript for the book that you now hold in your hands.

PREFACE

Latin America has been embroiled in debates about the appropriate nature of social policy for decades. For many years, the need for public involvement in financing or providing social services was justified for social reasons—fairness, equity, human rights, or the advancement of culture. At the same time, the view that social services could be seen as investments in human capital was gaining currency in academic circles. In particular, research demonstrating the strong association between education and earnings provided backing for the current emphasis on assuring that every child has access to a good quality school. And the analysis of education and its economic impact have revolutionized how we think about inequality, poverty, and social mobility.

Health is clearly important to human activity, and some would argue that it is even more fundamental than education. Learning itself is an activity that is enhanced or inhibited by personal health. Yet research into the economic impact on health has been relatively unknown until now. Pioneering studies by Frank T. Falkner, James M. Tanner, Richard H. Steckel, and Robert W. Fogel have laid a foundation for recognizing the long-term impact of health conditions on economic growth, but the difficulties of such research cannot be underestimated. Unlike education, investments in health occur throughout an individual's lifetime. Furthermore, an individual's health is not primarily associated with one activity (like schooling) but with numerous factors—environmental, behavioral, and genetic.

This Latin American Research Network book continues a series that takes advantage of the large and growing capacity for policy relevant research in our region. It combines new data available in Latin America with the latest methodological advances to untangle the complex relationships between health and income with the goal of enriching our understanding of how public policy can improve health and human well-being. Uncovering evidence of the economic impacts of health and its determinants will move us forward in the continuing debates over social policy, and build a greater appreciation for the important role health plays in the socioeconomic development of the region.

Ricardo Hausmann
Chief Economist
Inter-American Development Bank

This page intentionally left blank

CHAPTER 1

Earnings and the Elusive Dividends of Health

William D. Savedoff and T. Paul Schultz[1]

This chapter looks at the relationship between health and income. After discussing the general context of health improvements in Latin America during the last few decades, the study elaborates on the interrelationships between the physical and social determinants of health; the complexities that arise in attributing earnings differentials to variations in health status; and the difficulties of accurately measuring health status. It presents a methodology for estimating the impact of health on earnings that addresses problems of measurement error and endogeneity. It then summarizes the main findings of the studies that appear in subsequent chapters. These studies show that health status does have a significant, although modest, impact on earnings in four Latin American countries. Furthermore, environmental conditions (such as housing and sanitation) appear to have significant impacts on health status, compared to health services and public health facilities, which show little influence. The universally strong relationship between education and earnings is only modestly reduced by the inclusion of health status despite a general expectation that estimated returns to education were, in part, capturing the frequently unmeasured effects of health. By analyzing these relationships together—health determinants and the impact of health on earnings—we can assess the magnitude and importance of the human capital component of health status, validate and compare a range of health indicators, and identify promising areas for public policy to invest in health improvements.

[1] Dr. William D. Savedoff is a Senior Economist with the Social Programs and Sustainable Development Department of the Inter-American Development Bank. Dr. T. Paul Schultz is Malcolm K. Brachman Professor of Economics and Demography at Yale University.

Introduction

Common sense tells us that healthier people are more productive and that wealthier people can obtain goods and services and shape their environments in ways that make them healthier. However the magnitude of these causal relationships is largely unknown. Quantifying how health affects productivity and how wealth affects health is a starting point for understanding the role of health as a form of "human capital"—that is, as a characteristic of individuals, like education, that can be increased and improved through investing time and resources. In turn, understanding this aspect of human capital is critical for evaluating the role public policy plays in encouraging households, private firms, and public sector institutions to make efficient investments in health.

Developing this understanding means grappling with several difficult questions. What constitutes an "investment" in health? How is "health" to be measured? How does health affect productivity and, in turn, yield higher earnings? Is health a statistically significant and relatively substantial factor in determining the earnings capacity of people? Or is it largely a "consumption" good for which demand increases as people's incomes rise?

To help answer these questions this book will focus on health investments, health outcomes, and labor productivity. It uses a common framework to analyze these factors in several countries, and shows how improvements in data and analysis might advance our understanding of their interrelationships and lead to improved policy choices. At the core of the book are a series of studies of health indicators, based on household surveys in Latin America. Taken together, they shed light on which indicators of personal health are most informative; how health varies across populations by differences in age, sex, education, and region of residence; and how these health indicators are meaningfully related to wages or labor earnings per hour worked.

This research agenda involves complex quantitative and qualitative issues. On the one hand statistical inference must be used to interpret individual indicators of health that are measured with considerable error and that are possibly associated with other unobserved factors that affect wages, such as biological, psychological, and economic endowments and motivations.

On the other hand, there are fundamental questions about the nature and purposes of health that economists may feel uncomfortable in answer-

ing. Should health be improved only to increase personal market incomes? Surely not. But to the extent that measurements of health can be linked with higher wages, then a statistically valid way will have been found to connect health policy choices to improved health, increased income, and perhaps reduced poverty and economic inequality. Increasing income opportunities is one way to generate sustainable advances in well-being, but other benefits of improved health should also be assessed and included in any global evaluation of policy options.

Good Health Despite Relatively Slow Growth

A population's health is arguably one of the best indicators of its level of "development." This premise is a basic element in efforts to find measures of development other than those based on per capita income—whether it is the UNDP's Human Development Indicator (which includes life expectancy) or the call by Amartya Sen (1998) for measuring human "capabilities" that include health status.

For Latin America, the use of income to measure progress leads to the conclusion that the region has developed relatively slowly. Latin America has lagged behind world performance in income growth during the last few decades—and performed poorly compared to the rapidly growing nations of East Asia. Even when the yardstick shifts to measuring progress through human capital investments in educational attainments, Latin America performs well below the international pattern. In fact, the gap in educational progress between Latin America and other regions appears to be growing.

By contrast, using health as a measure of progress would suggest that Latin America has developed quite fast. In fact, Latin America's health status has improved remarkably during recent decades. The process of increasing life expectancy that took as much as 150 years to accomplish in many European countries has swept much of Latin America in less than 50. Average life expectancy in Latin America was only 58 years in 1962, and stands at just over 70 years in the mid-1990s. Life expectancy increased largely due to reduced mortality rates among children and those over 45. Infant mortality declined from almost 100 per 1,000 live births in the early 1960s to less than 30 in 1995. Currently 11 out of 22 Latin American countries report infant mortality rates below 30 per 1,000 live births. Life expectancy in Latin America is almost 10 percent below the rates enjoyed by the more devel-

oped OECD countries; but after controlling for differences in income, the gap disappears.[2]

It is important to qualify these gains in several ways. Most importantly, these rates are national averages and do not necessarily reflect comparable gains in all population groups—whether by sex, region, or ethnic group. In general, health outcomes are significantly worse in rural areas and appear to be worse among lower income groups as well.[3] A study in southern Brazil showed that among children born with normal birth weight, the mortality rate is five times higher in families with incomes below $50 per month than in families whose incomes are more than $150 per month (Victoria et al., 1992). The infant mortality rate in Peru is almost five times higher in the poorest than for the uppermost quintile. Inequalities have also been documented in Guatemala, where neonatal mortality per 1,000 live births is only 18 in urban areas, compared to 29 in rural areas and 32 among the indigenous population.[4]

Despite these qualifications, great strides have been made and this progress can be attributed to many different factors. Some people credit *public health policy*—particularly the provision of sanitary infrastructure, potable water, and immunization campaigns against infectious diseases such as smallpox. Others point to *rising income levels* that allow individuals and families to raise dietary intake and quality and obtain adequate shelter and clothing. Another line of reasoning emphasizes the expansion of *personal medical services*, particularly publicly financed social security and publicly dispensed medical care targeted at the middle classes and the poor. Finally there are those who focus on other *social and cultural transformations* associated with changes in work, urbanization, status inequality, mobility, education, and the environment.

All of these factors do appear to contribute to improvements in health conditions, but the relationships are not always strong. Much of the world-

[2] These are the findings by Piras and Savedoff (1999), based on a cross-country sample of 151 countries and using control variables that include income; availability of doctors, nurses, and hospital beds; sanitation and potable water; geographic variables; and average caloric intake. A variety of specifications were used, and as long as income was included, the regression would explain most of Latin America's "advantage" in life expectancy.

[3] See *Health Systems Inequalities and Poverty in Latin America and the Caribbean*, PAHO/UNDP/ World Bank Equilac/IHEP Project, forthcoming.

[4] For data on Peru and Guatemala, see *Health Systems Inequalities and Poverty in Latin America and the Caribbean*, PAHO/UNDP/World Bank Equilac/IHEP Project, forthcoming.

wide decline in mortality rates preceded the introduction of public health initiatives—and many of these were inadequately applied in Latin America. Although income is closely related to rising health status, the region has experienced entire decades of economic crisis (like the 1980s) without seeing a reversal in trend, or even a clear slowdown in the decline of mortality rates. By some measures, notably child and infant mortality, the region may not fare quite as well once adjustments are made for its income level (Piras and Savedoff, 1999). Indeed some studies claim that income is itself the key factor in explaining the differences in health conditions among countries—which, in some ways, would reinstate income as a valid cross-country measure of welfare and a proxy for health (Pritchett and Summers, 1996).

Personal medical services have clearly expanded since the 1950s, but they account for little if any of the health improvements detected. Even social trends such as status inequality, mobility, and urbanization would appear to put Latin America at a significant disadvantage compared with the rest of the world; yet the region performs close to the international average.[5] Some have noted that the one exception among these trends is that Latin America has invested relatively more in girls' schooling than have other developing regions. Since women's educational attainment is strongly correlated with reduced child and adult mortality rates, this may be a powerful factor in accounting for the continued health improvement in Latin America despite income growth stagnation and a slowing pace of average educational improvements during the 1980s (Schultz, 1995a).

Understanding Latin America's performance in health status is very difficult because the determinants of health are not well documented. This is partly due to problems in collecting data on health conditions and partly due to lack of agreement about which health indicators are most accurate and useful for analytical purposes. One rarely used way to evaluate health indicators is to test them against other observable effects of health status. The studies conducted for this book take advantage of extensive data on individual earnings to see if it is possible to document relationships between health status (measured imperfectly by several different indicators) and earnings. Because we believe that healthier people are likely to be more produc-

[5] For a recent discussion of Latin America's high levels of inequality, see Inter-American Development Bank (1998).

tive and therefore, on average, better remunerated in their employment, it should be possible to validate or qualify the various yardsticks for health.

Additionally, this approach may offer insight into the indirect benefits of health. In a sense, it tells us how much health is "worth" in the market. Although this does not change our fundamental interest in improving health conditions, regardless of market valuation, it does provide more knowledge about health's impact on other aspects of individual and social life.

Understanding Health as a Form of "Human Capital"

The notion of "human capital" emerged in the economic literature in the 1960s as recognition grew that many improvements in human capacities are due to investments of time and resources. This concept led to the development of theories and testable hypotheses to better understand choices made by individuals and their families regarding education, labor force participation, health, fertility, and many other aspects of behavior. The concept of human capital does not imply that genetic, cultural, ethnic, or social factors do not influence behaviors. Rather it introduces a way of understanding an additional dimension of human capacities by drawing attention to those elements of our capacities that we, as individuals, can influence through our choices and behavior. The key aspects of human capital are that the investments involved affect human capacities over a reasonably long time frame and that people make choices regarding these investments in relation to the expected future returns of those "investments," among other things.

In the specific case of health, two questions arise. First, what is the effect of a change in health status on the productive capabilities of an individual? Second, what individual, family, and community characteristics effect changes in health status? Answering the latter question, in particular, requires focusing primarily on characteristics that are malleable. This focus has two sources: the need to know which policies can improve health, and awareness that it is impossible to properly measure the impact of health on earnings without recognizing that individuals make choices about health investments based on their expectations of how health affects their life opportunities. With answers to these two questions, it becomes possible to assess the resource costs of modifying those conditions that will improve health, and to calculate the internal rate of return on those outlays as a human capital investment.

To evaluate the returns to health human capital involves many of the same problems that have occupied economists for some time in estimating the returns to schooling. There are, however, a few added complications. First, there is no immediately obvious metric for individual health status. By contrast, there is some agreement that "years of schooling completed by a worker" is a reasonable first approximation for the physical units of education, although the equivalence may be further refined to include various qualitative dimensions of the education being measured (Becker, 1964; Mincer, 1974). There is no comparable consensus about how one can measure the stock of health as human capital. For an aggregate population for which age-specific mortality can be estimated, "life expectancy at birth" (or from some other age) has the appeal of a demographic summary measure of expected survival within the physical limits of life. Such a measure can also be refined by other qualitative aspects of life that people value, for instance the timespan that is disability-free. However, there is no comparable summary measure at the individual level for attributing to a given person an "expected lifetime."

The second complication—measurement error—is shared with schooling, but may be more serious in the case of health. If there is little consensus on how to measure individual health status, and most measures considered involve self-reporting by a survey respondent, it seems likely that the range of error in recording a truly latent variable will be more substantial for health status than for completed years of schooling.

The third complication is the difficulty of assigning weights to the inherited, environmental, and behavioral factors that affect health outcomes. Health human capital (*H*) can be considered as an outcome of an individual's genetic endowment (*Hg*), and his or her behaviorally and environmentally determined accumulation of health (*Hb*). Difficulties arise because the two sources of variation are not only determined by different factors, they may also exert different effects on labor productivity. The productive benefit of the socially accumulated *Hb* is usually the more relevant variable in evaluating policy interventions to improve health.[6] If the stock of educational capital were measured not by school years completed but by achievement tests,

[6] We set aside here the question of technologies that may be available to modify genetic potential. Nor do we consider public policies that promote eugenics, or selective reproduction—most notoriously in nationalistic programs that seek so-called "racial purity."

then the same dichotomy between genetic potential and acquired abilities would arise and make it necessary to decompose the effects. The fact that years of education has become a commonly accepted measure of human capital is due to the regularly observed correlation it has with earnings, although educators continue to grapple with finding measurements to differentiate the effectiveness of various forms of pedagogy in terms of their "learning outcomes."

In sum, health presents serious problems for measurement at the individual level through surveys. It presents more serious problems for analysis because of measurement error. Finally, for the purposes of both policy and analysis, it needs to be decomposed explicitly into two functionally different components of health capital: a genetic endowment and a socially acquired set of assets and liabilities.

What insights can be gleaned from the literature on health economics? First, it has been emphasized that the health heterogeneity of individuals can bias direct estimation of health production functions (h) that seek to characterize the technological relationship between health inputs (I) and health outcomes (H) and residual variation (e_1) (Rosenzweig and Schultz, 1983). It is reasonable to expect that individuals, their families, and perhaps their medical advisors will know more about the severity of illness or the frailty of the individual (g) than does the social analyst trying to account for health outcomes. The health production function can thus be described, where g and e_1 are unobserved:

$$H = h\left(I, g, e_1\right). \qquad [1.1]$$

Consequently, the demand for medical care and other health-related inputs (I) may be modified through a function (d) that has the following elements: the private knowledge of the individual, family, and medical system (g); other demand-limiting factors (X), such as health input prices, household income, etc.; and another error, or residual, variation (e_2).

$$I = d\left(X, g, e_2\right). \qquad [1.2]$$

For example, individuals who are ill will seek out medical care, introducing a negative partial correlation in function (h) between demand for curative health inputs (I) and good health (H), rather than the positive cor-

relation that is expected because people who receive preventive care are less likely to be ill. If the unobserved part of an individual's initial health heterogeneity is subsumed in the error in the health production function, this error is likely to be correlated with the observed use of health inputs, imparting an omitted variable (or simultaneous equation) bias to the health production function when the latter is estimated by single-equation methods such as Ordinary Least Squares (OLS).

A solution to this health heterogeneity problem is to treat the health inputs as endogenous or behaviorally controlled, and employ instrumental variables (X), such as the local variation in prices or access to health inputs, as natural instrumental variables to predict the health input demands. One can then estimate without bias the health production function by two-stage methods (Rosenzweig and Schultz, 1983). Assuming that the input prices and access are not correlated with individual health heterogeneity, and that the price and access variables explain a statistically significant share of the variation in input demand, these instrumental variable estimates of the health production technology are consistent and have desirable properties (Bound et al., 1995).

In trying to evaluate how health human capital affects wages or labor productivity, an analogous problem arises. Assume that outcomes of health human capital can be decomposed into two components: Hb, which is explained by the technological effect of socially controlled inputs responding to exogenous constraints (X); and a remainder (Hg) that subsumes genetic heterogeneity, differences in preferences, other unexplained factors, and specification and stochastic errors from both the production and input demand equations (Schultz, 1996):

$$H = Hb\left(X\right) + Hg\left(g, e_1, e_2\right). \qquad [1.3]$$

Only the first component can be viewed as a form of reproducible human capital, derived from predicting the health outcome from the fitted reduced form for health that embodies both the health production function and the health input demand equations. If the goal is to evaluate how this overall health human capital affects productivity, it is necessary to predict from the reduced-form health function, and use only this predicted health component in the wage equation as a human capital variable that is uncorrelated with health heterogeneity. Following the health production

function literature, the appropriate instruments for predicting the human capital component in the health outcome would include variables such as the local price of health inputs and access to health care for individuals, and the presence of community institutions/investments that affect exposure to disease and adequacy of treatment when ill. It is also arguable that household or family wealth will increase the individual's receipt of health inputs and care, and thus contribute to health human capital (Schultz and Tansel, 1997).

Each of the seven studies in this book extracts from available data sources certain instruments on which to base prediction of the behaviorally controlled variation in health human capital. Just as social scientists perennially struggle to decompose variation in achievement due to nature (genetics) and nurture (human capital), this book offers no entirely satisfactory method for resolving the riddle. In all seven studies, some genetic variation in health will be correlated with household income, price, and community variables and thus be embedded in the predicted health outcome. Thus the instrumental variable estimates of the effect of health capital on wages will contain some unavoidable portion of the genetic health effect, in addition to the socially induced variation in health related to the individual and community instrumental variables.[7]

Labor productivity, approximated by the hourly wage rate, is then fitted to variation in individual human capital stocks, where health human capital (*Hb*) is only the variation in health status that is accounted for by the instrumental variable, and hence uncorrelated with e_1 and e_2, which is desirable because these production and demand errors are likely to be correlated with e_3 due to omitted variables in the wage function and other sources of simultaneous equation bias:

[7] Hausman (1978) specification tests could be performed to determine whether the health human capital variable appears to be exogenous (or endogenous) in the wage function, which is analogous to testing whether the human capital and residual health components have statistically the same coefficient in the wage equation. Endogeneity of the health human capital variable in the wage equation would be confirmed if the behavioral and genetic/residual components of the health human capital variable received significantly different coefficients in the wage equation. These tests are likely to differ not only as the choice of instruments varies but also as the indicators of health are changed. In some health measures, such as height, the genetic component appears to account for most of the variable's variation, whereas self-reported categorical or disability measures of health are more readily explained by individual/family/community instrumental variables.

$$W = w\left(Hb\left(X\right), E, Z, e_3\right). \quad [1.4]$$

In this equation, W is the logarithm of the hourly wage rate; w is the wage function; E is education; Z represents other observed factors affecting the wage that are not behaviorally determined; and e_3 is the error in the wage equation.

This instrumental variable approach to estimating the effect of health human capital on the wage function has the additional benefit of correcting for the measurement error in the health indicators, a problem that is potentially more serious in the case of health than education.[8] The empirical specification of this model for estimating the effects of health human capital on productivity involve the choice of the health indicator variable (H), the specification of the instrumental variables (X) that account for the health indicator but are not otherwise related to the wage rate or other household choice variables, and the additional control variables (Z) included in the wage equation. Some studies are based on surveys that provide information on community characteristics that are candidates for inclusion in X or Z. Most household surveys ask respondents their municipality or place of residence. Thus sample clusters can be aggregated from individual responses to other questions so as to draw conclusions about specific community characteristics such as the average distance from a household to the nearest clinic for all observations in the sample, whether they have reported using those services or not. In other instances, government records on the climate, geography, and infrastructure of each sample cluster community can be merged with results from the individual survey respondents who reside there to provide additional contextual information that may be relevant for local health and labor markets. This estimation approach presumes that place of residence is not related to average health heterogeneity, to individual preferences that affect a person's health, to migration decisions, to the location of health and other service infrastructure, and other relevant individual or community choices and actions (Rosenzweig and Wolpin, 1986; Schultz, 1988).

The sample of wage earners is also interpreted as being representative of all people who could make investments in health, education, etc. In

[8] However if the measurement error in health is not of the classical form (that is, uncorrelated with other explanatory variables in the wage function), then the estimation problem may not be completely resolved by this instrumental variable estimation strategy.

reality the subsample of wage earners may be selected for characteristics that are related to the unexplained variation in health and wages. Where possible, the studies assess whether standard methods for dealing with sample selection bias shed any light on this potential problem, and it fortunately appears in these cases not to be a serious source of bias. To implement these methods, however, it is generally necessary to specify some exclusion restriction that implies certain variables significantly affect who is selected into the sample of wage earners, but this selection variable should not affect the wages they are offered. This identifying variable, which is excluded from the wage function, presumably affects only the reservation wage or the value of time in activities other than wage employment. It is also generally assumed that the sample selection equation has a specific functional form, such as the standard normally distributed error that can be modeled as a probit and estimated jointly with the wage equation by maximum likelihood methods.

Study Findings

The main finding of these studies on health and productivity is that healthier people receive higher wages. Though the effect of health on wages varies in magnitude depending on the health measure used in the analysis, it is generally significant when estimated by instrumental variable methods, even after controlling for selected individual and community characteristics and for some job attributes as well. The studies provide evidence that public health services and community conditions, as well as private health inputs and reduced exposure to disease, are associated positively with the health of adults and also with greater individual income-generating capacity. It demonstrates that, as elusive as it may appear, we do have measures that can represent the human capital dimension of health.

Distinguishing Ways to Measure Health

What does "good health" mean? The term is decidedly relative—depending on the era in which the question is asked and the society that frames the question. Sometimes the definition is so nebulous it is little more than a dimly defined perception of potential fitness. Often it seems best to fit health into that singular category of abstract concepts that are impossible to mea-

sure but about which one can say, "I know it when I see it." Unfortunately our poor understanding of what determines health and its effects is, in large part, a direct consequence of difficulties in measuring both. As previously discussed, no consensus singles out one variable that best aggregates the diverse dimensions of health in a feasible indicator at the individual level. There is nothing that could function for health in the same way "completed years of schooling" serves as a proxy for a range of educational outcomes in the fields of labor and education studies.

Measures of health status can be broadly divided into four categories: relative assessment of health status, self-reported morbidity, functional limitations, and health and nutritional outcomes. These indicators can be derived from surveys in which respondents are asked to rank themselves relative to some ideal sense of "good health" or relative to others in their community, or asked about specific symptoms, the number of days ill in some reference period, functional limitations, height, or weight. Observed measures range from the official registration of mortality and formal epidemiological surveillance data, to surveys in which individuals are tested for vital signs, symptoms, functional limitations, or size (height and body mass).[9] Tables 1.1 and 1.2 provide an overview of the case studies that follow and summarize the health indicators that they use.

In studies that use self-assessment indicators, respondents are asked to compare their health to others in society or in their age group. This measure has two main problems. First, there is no clear benchmark of similar health for others or for the respondent's age group. Thus the respondent may rely on knowledge of the health status of other persons in his/her neighborhood or socioeconomic class, rather than apply a common standard embracing the entire survey population. Second, any self-reported health status could reflect experiential conditioning and perceptions that are potentially correlated with socioeconomic behaviors and outcomes, rather than quantify only the individual's actual condition. Nonetheless these general indicators of health status have been shown to be significantly related to subsequent morbidity and mortality of the individual, and have been used in evaluating the National Health Experiment in the United States (Manning et al., 1982).

[9] For discussion of alternative measures see Murray et al. (1992) regarding adult mortality in developing countries.

Table 1.1 Summary of Case Studies

Chapter	Author	Country	Year	Survey	Sample Coverage	Health Variable(s)
2	Knaul	Mexico	1995	ENPF	Women aged 15–54	Age at menarche
3	Parker	Mexico	1994	ENMV	Ages 60–79	Disabled, disabled days, self-reported relative health status, functional limitations (ADLs)
4	Ribero and Nuñez	Colombia	1993	CASEN	Ages 18–70	Disabled, days disabled
	Ribero and Nuñez	Colombia	1991	ENH	Urban area ages 18–60	Adult height
5	Murrugarra and Valdivia	Peru	1994	PLSMS	Urban area ages 16–60	Ill and days ill
6	Cortez	Peru	1995	ENAHO	Ages 18–70	Ill and days ill
7	Espinosa and Hernández	Nicaragua	1993	ENV	Ages 18–65	Ill and days ill

Note: For more details, see the specific study in subsequent chapters.

The second group of indicators is based on asking respondents about illnesses or disability in some retrospective period (which can range from 14 days to 180 days, depending on the survey). These questions sometimes simply ask whether or not the respondent was ill or injured. Sometimes the survey probes further, asking about the number of days ill or whether the person was sufficiently ill to be disabled (i.e., unable to engage in his/her regular activity) during the reference period. Questions of this nature are included in many labor force and household general surveys, and have not been subjected to much analysis as an indicator of acute or chronic health status, perhaps because they are viewed as subjective. It has been hypothesized that reports of days lost to disability among wage earners may be a more reliable indicator of health because the illness must be sufficiently severe to forego wage-earning activities, setting a threshold from which to infer the severity of the illness (Schultz and Tansel, 1997).

Table 1.2 Summary of the Main Health Status Variables Used in the Case Studies

Country (Year)	Variable	Definition
Mexico (1995)	Age at menarche	Age (year and month) of first menstruation
Mexico (1994)	ADL	The sum (from 0 to 4) of being able to carry out basic activities (walking up stairs, walking more than 30 m, carrying a heavy object, and doing light domestic tasks)
Mexico (1994)	Disability	A binary variable distinguishing people who reported being sick, injured, or hospitalized in the previous 180 days
Mexico (1994)	Self-reported health status	Ordinal indicators (on scale of 1 to 5) in response to questions regarding the respondent's own health, and own health compared to others of a similar age
Colombia (1991)	Height	Adult height measured in centimeters
Colombia (1993)	Disability	Individuals who due to illness or injury were unable to work at some time during the month before the survey
Peru (1994)	Days ill	Number of days of reported illness or injury (including those not necessarily severe enough to miss a workday) during the four weeks prior to the survey
Peru (1995)	Days ill	Number of days of reported illness (including those not necessarily severe enough to miss a workday) during the 15 days prior to the survey
Nicaragua (1993)	Days ill	Number of days of reported illness or injury (including those not necessarily severe enough to miss a workday) in the month prior to the survey.

Note: For more details, see the specific study in subsequent chapters.

This kind of self-reported statement regarding whether or not the respondent was ill in recent weeks was the most common indicator used in the case studies. This includes the living-standard measurement surveys in Peru for 1994 (Murrugarra and Valdivia) and Nicaragua for 1993 (Espinosa and Hernández), the 1993 CASEN survey in Colombia (Ribero and Nuñez), and the 1995 National Household Survey (ENAHO) in Peru (Cortez). In Mexico (Parker), using the 1994 National Mexican Aging Survey, these questions were also asked, but only the dichotomous part of the question was used and the recall period was significantly longer (180 days). By contrast,

the ENAHO survey in Peru uses a recall period of only two weeks. The Colombian study focused on days disabled (i.e., that the individual was unable to go to work). In Ghana and Côte D'Ivoire, similar studies have used comparable indicators of days ill or days disabled (Schultz, 1996; Schultz and Tansel, 1997). As discussed in detail below, the number of people who respond that they were unable to work due to illness is generally one-half to one-third as many as those who report having been ill.

This indicator is not perfect. It is subject to recall error, and the definition of what it means to be ill may vary culturally and according to personal disposition. But it is no accident that this indicator is favored by studies investigating the relationship between health and income. The evidence necessary to link health and income is difficult to find because health surveys, rich in health indicators, rarely collect data on income or wealth; while labor and household surveys rarely collect more than a few health-related indicators. Self-reported days of illness are fairly easy to incorporate in the labor and household surveys, and so these have become the *de facto* choice of health indicator for these kinds of studies.[10]

The number of days ill appears to be a reasonable measure of poor health. It varies systematically with age, gender, income, and education. Individuals who were older, less educated, and poorer generally reported more incidents and days of illness in these studies (see Tables 1.3, 1.4, and 1.5). In general, more women reported being ill in the four weeks prior to the survey than did men; but of those who reported illnesses, women tended to report fewer days of illness than did men. Most of the studies noted how skewed this measure was—large shares of the population report no illness at all, making it difficult to distinguish degrees of health status among them. For illnesses lasting more than three or four days, there is a "lumping" of answers around discrete numbers of weeks. And finally, there is a truncation at 30 days, which in some cases is quite significant. Cortez deals with these problems by using a transformation of the days-ill variable ($H = 1/(1+D)$) that varies from 1 for those who report no days of illness, and approaches 0

[10] When researchers want to use data sets that are richer in health indicators to look at socioeconomic interrelationships, they are reduced to using imperfect proxies such as occupations (e.g., Davey Smith et al., 1990) or place of residence (Phillimore et al., 1994), or to constructing wealth indicators from other variables using DHS (Bonilla-Chacin and Hammer, 1999; and Montgomery et al., 1997).

Table 1.3 Morbidity Indicators by Sex and Age (percentage ill at any time during reference period)

	Country, period, and indicator			
	Colombia (1993)	Peru (1995)	Peru (1994)	Nicaragua (1993)
Age	Prior month	Prior 15 days	Prior 4 weeks	Prior month
groups	Unable to work	Ill	Ill or accident	Ill or accident
Men				
18–24	5.0	17.7	17.0	12.8
25–34	5.2	20.1	22.0	18.6
35–44	5.9	22.7	26.0	20.9
45–59	7.0	28.4	30.0	28.3
60–70	12.1	37.9	35.0	36.7
All	6.2	23.2	29.0	20.2
Women				
18–24	6.2	21.5	22.0	14.7
25–34	6.1	27.3	26.0	20.6
35–44	7.6	32.3	31.0	27.4
45–59	7.9	38.8	41.0	33.9
60–70	9.8	47.7	42.0	43.1
All	7.0	30.8	32.0	24.1

as *D* approaches infinity.[11] Murrugarra and Valdivia address the problem of truncation by using a two-limit tobit model to extract the predicted latent health variable. Schultz and Tansel (1997) fit a quadratic approximation to the number of days disabled.

The days-ill variable tended to be correlated with other health measures. In general, the higher threshold implied by whether or not the respondent was able to work reduces the number of reported events to between one-half and one-third of those reporting any illness. Murrugarra and

[11] Of course since the question asks only about the last 30 days, the minimum value for *II* is 0.034. The transformation is only an approximation to the problem of truncation bias since the respondent does not indicate when the illness started (it could have been earlier) or whether the illness is continuing at the time of the survey.

Table 1.4 Morbidity Indicators by Sex and Education (percentage ill at any time during reference period)

	Country, period, and indicator		
	Colombia (1993)	Peru (1994)	Nicaragua (1993)
	Prior month	Prior 4 weeks	Prior month
Education	Unable to work	Ill or accident	Ill or accident
Men			
0 years	8.0	37.0	n.a.
1–6 years	6.5	32.0	38.8
7–12 years	6.1	20.0	19.6
13+ years	2.4	20.0	17.7
All	6.2	24.0	33.3
Women			
0 years	9.1	42.0	—
1–6 years	7.6	36.0	43.1
7–12 years	6.2	24.0	22.6
13+ years	6.6	25.0	19.1
All	7.0	30.0	35.8

Valdivia in Peru found that approximately 28 percent of the sample reported being ill, but only 10 percent reported being disabled. Ribero and Nuñez in Colombia found that about 15 percent reported being ill, but only a little more than 6 percent reported being disabled. It is interesting to note that, in this latter case, men reported being ill less than women (13 percent versus 18 percent), but a higher portion of men than women who reported being ill were unable to work (47 percent versus 37 percent). Schultz and Tansel (1997) also find that disability was reported about half as much as illness in two West African countries. Ribero and Nuñez also estimated the effect of both days ill and days disabled on earnings and found that they yield similar results. For elderly Mexicans, Parker calculates that disabled days are significantly correlated with the number of functional limitations, and inversely correlated with questions about the respondent's health status relative to others their own age.

Table 1.5 Morbidity by Quintiles of Individual Hourly Earnings (percentage ill at any time during reference period)

	Country, period, and indicator	
	Colombia (1993)	Peru (1995)
	Prior month	Prior 15 days
Income Quintile	Unable to work	Ill
	Men	
1	4.6	29.0
2	5.2	21.4
3	4.9	20.4
4	5.6	21.4
5	4.7	19.3
	Women	
1	6.4	35.1
2	8.4	29.5
3	7.7	28.0
4	5.6	25.7
5	6.8	28.3

A third group of health indicators is related to functional limitations that the individual experiences in Activities of Daily Living (ADLs). These indicators can be tabulated by observation, or through self-reported answers to specific questions. In contrast to self-reported general health status, or incidences of illness or days disabled, the number of functional limitations appears to be more concrete and less subjective; and the focused form of the questions even permits the interviewer to validate partially the response. It is argued, therefore, that health indicators based on functional limitations in performing daily tasks tend to be less correlated with socioeconomic endowments, conditioning factors, and perceptions, and are thus less likely to be biased by subjective factors (Strauss et al., 1995). ADLs appear to represent an adequate continuum of health states among the elderly, for whom physical limitations on everyday activities are commonplace, and have been shown to replicate clinical health status reliably in high-income populations (Stewart

and Ware, 1992). How well indicators of functional limitations differentiate health status among the nonelderly is more uncertain because of infrequent incidence in most younger populations. Yet ADLs remain promising health status indicators because they can be readily collected in surveys while reducing the subjective bias that dilutes other measures of adult health status.

In the study by Parker, the number of ADLs that can be performed with difficulty or cannot be performed at all was derived from responses to questions related to the respondent's ability to walk upstairs, walk 300 meters, carry a heavy object, and engage in light domestic tasks. Parker validates this measure against self-assessed health measures and demonstrates that it is a significant determinant of earnings after controlling for other factors.

The final category of indicators includes measures of health and nutrition outcomes. Such outcomes as height and menarcheal age result from a complex accumulation of nutritional inputs and health care, offset by exposure to infectious disease, and modified by the burden of work activities that combine to influence human physical growth (Faulkner and Tanner, 1986; Fogel, 1994). Nutrition and a healthy environment appear to be critical in the rate of uterine and early childhood development, while conditions in adolescence may further modify the outcome when physical growth briefly accelerates during puberty. Adult height is an indicator that is commonly used to measure those aspects of childhood nutritional and health status that have lifetime consequences on an individual's health and physical and mental capacities (Fogel, 1994; Thomas and Strauss, 1997). Age at menarche is another indicator of physical growth that is affected by improvements in nutrition and health, but it has never been studied previously as an indicator of health human capital. Although these health outcome indicators are thought to reflect childhood conditions, they have also been linked to acute and chronic health problems among adults and to levels of physical functioning among the elderly (Costa, 1996).

In the chapter on Colombia, Ribero and Nuñez analyze data for height among urban adults and confirm findings in other countries that height is increasing systematically over time and that it is consistently correlated with higher earnings. Furthermore, by analyzing "height" and "disabled days," this chapter suggests that the former is a more robust indicator of health since it is more consistently and strongly associated with earnings.

Knaul has broken new ground by utilizing an unusual measure of health: women's age at menarche, or the onset of first menstruation. His-

torical evidence shows that menarche has been declining steadily in most high-income countries, and studies have related this decline to improved nutrition and health in childhood. It appears that this measure, like height, is a useful marker for the level of health attained in childhood—a "stock" that could possibly affect later risk of morbidity and mortality as well as productivity.

Health Affects Income

Using these varied measures of health status and morbidity, all of the studies find some association between health and wages. This impact varies in magnitude and reliability depending on the health measure used for analysis, the particular sample, and the range of instrumental and control variables that are employed (see Tables 1.6 and 1.7).

Despite the problems of endogeneity and measurement discussed previously, there are some variables that appear to be significantly related to wages, even in a simple regression without using instrumental variable techniques. An individual's height was a statistically significant predictor of earnings in the Colombian study, even after controlling for education and other personal characteristics and even without using instrumental variables. When elderly Mexicans ranked their health relative to others their own age, this measure was also significant in a simple wage equation. The rest of the health variables were not significant at the 10 percent level, with the exception of days ill in the Peruvian ENAHO survey (see Chapter 6). Either health has little impact on productivity, or the endogeneity and measurement problems associated with these health indicators are generally quite serious.

The results of applying a two-stage analysis to these problems demonstrate that health *does* have an impact on earnings after all. For the most part, health indicators that were not significant in single-stage regression were significant after endogeneity and measurement errors were corrected through the use of instrumental variable techniques. Furthermore, the magnitude of impacts estimated for most health indicators increased with the two-stage estimation procedure.

Earnings are lower for people who are predicted to be ill, and are systematically lower for those who are expected to be ill longer. The magnitudes, however, vary widely—between 1 percent and 58 percent. For the most part, an additional day ill was associated with a decline of between 1

Table 1.6 Estimated Impact of Illness on Hourly Earnings (with and without instrumental variables)

	Ill or Disabled		Days Ill		Ill or Disabled		Days Ill	
Country (year)	with	without	with	without	with	without	with	without
	Urban Men				**Rural Men**			
Peru (1995)			–0.93***	–0.09***			–3.46***	–0.09*
			(4.7%)				(14.2%)	
Peru (1994)			–1.21**	n.s.				
			(1.2%)					
Colombia (1993)	–0.28***	n.s.	n.s.	n.s.	–0.41***	n.s.	–32.9***	n.s.
	(28%)				(41%)		(32%)	
Mexico (1994)	–3.29**	n.s.			–3.29***	n.s.		
	(96%)				(96%)			
Nicaragua (1993)			0.16*	n.s.			n.s.	n.s.
			(16%)					
	Urban Women				**Rural Women**			
Peru (1995)			–1.06***	–0.07***			–2.25***	–0.15*
			(3.4%)				(6.2%)	
Peru (1994)			–2.41**	n.s.				
			(2.4%)					
Colombia (1993)	–0.14*	n.s.	n.s.	n.s.	–0.19*	n.s.	–13.5***	n.s.
	(14%)				(19%)		(13%)	
Mexico (1994)	n.s.	n.s.			n.s.	n.s.		
Nicaragua (1993)			n.s.	–0.02*			n.s.	n.s.
				(2%)				

Notes: For definitions of variables, see Table 1.2; Mexico (1994) is a combined rural and urban sample; Nicaragua (1993) had significant coefficients in combined samples, but not when the sample was disaggregated by sex and area; Peru (1995) coefficients are listed as negative here for comparability even though the estimates in Chapter 6 are positive, due to the reciprocal transformation of the dependent variable; (n.s.) means not statistically significant; blanks indicate that the dependent variable or particular sample was not used. Numbers in parentheses represent the percentage effect calculated from the coefficient in the log-linear specification.

* Significant at 10% level.

** Significant at 5% level.

*** Significant at 1% level.

n.s. Not significant at the 5% level.

Table 1.7 Health Impact on Hourly Earnings by Indicator (with and without instrumental variables)

	Men		Women	
Indicator	With	Without	With	Without
Height in Colombia (1991)	7.9*** (7.9%)	0.72***	6.8*** (6.8%)	0.48***
Age at menarche in Mexico (1995)			-0.26*** (26%)	n.s.
ADLs in Mexico (1994)	-0.86*** (58%)	n.s.	n.s.	n.s.

Notes: For definitions of variables, see Table 1.2; (n.s.) means not statistically significant; blanks indicate that the dependent variable or particular sample was not used; numbers in parentheses represent the percentage effect calculated from the coefficient in the log-linear specification.

*** Significant at 1% level.

n.s. Not significant at the 5% level.

percent and 4 percent in hourly earnings. This was found to be the case in both Peruvian studies and in Colombia. The largest impact was found for elderly Mexicans, men over 65 years old, for whom an additional day ill was associated with a 58 percent decline in earnings. Although this may be overestimated, it plausibly suggests that health status may have a stronger effect on productivity and earnings potential as individuals age, a conclusion supported by Murrugara and Valdivia in Peru.

The studies that looked at whether or not an individual reported any illness also showed an impact on productivity. Ribero and Nuñez found that having been sick in the last month decreases the earnings of men by 28 percent and women by 14 percent in urban areas, while the effect was larger in rural areas—41 percent and 19 percent, respectively. Parker provides evidence that being ill reduces the wages of elderly men by at least 40 percent.

For the two physical growth measures of health "stock," the instrumental estimates were also quite significant. Knaul presents evidence that a one-month increase in a woman's age at menarche is associated with a wage decline of 2.3 percent; while Ribero shows that each additional centimeter of height increases urban male earnings by 8 percent and urban female earnings by about 7 percent.

Health and Education: Is There Synergy?

It has long been suspected that the returns associated with education, documented in virtually every household survey, may be overestimated as a result of being positively correlated with other socioeconomic factors that might raise earnings. Among the omitted variables that can lead to bias is health.

Although the studies in this book do confirm that educational returns may be somewhat inflated when there is a failure to control for health status, the magnitude of this effect appears to be modest. An individual's schooling is widely observed to be associated with improved health indicators (Knaul, Parker, and Espinosa and Hernández), and most of the studies find that wage returns to schooling are smaller when health variables are included in the estimated wage function (see Table 1.8). In general, a year of additional completed schooling was associated with 5 percent to 20 percent higher wages when health was ignored, and this estimated private return to schooling declined by less than a tenth when controls were included for health human capital. Thus, although all the studies suggest improvements in health have contributed to the substantial gains in labor productivity in Latin America, controlling for health indicators does not markedly weaken the estimated role in the region of schooling as a stimulant to growth.

In addition to potentially misattributing the measures of returns to education, there is evidence that health and education may interact in important ways—making a joint impact greater than the sum of their separate contributions. For example, specific studies have argued that healthier children benefit more from schooling, are more likely to remain in school, and are more likely to reach higher levels of schooling than those who are less healthy.[12] A few of the studies in this book sought evidence for this complementarity by introducing interaction terms between health and education, but the results were not always significant or conclusive. The strongest results appear in the study by Cortez, in which he found positive and significant effects on wages from an interaction of health and education in his sample of male workers. However in the female sample, the interaction term was not significant. In Knaul, the interaction of the health indicator and education also had a positive impact on wages, although this result was

[12] See for example Moock and Leslie (1986) and Glewwe and Jacoby (1992).

Table 1.8 Returns to Education with and without Controls for Health

Study	Health Variable	No Health Variable	With Health Variable	With Instrumented Health Variable
		Men		
Colombia (1993)	Days disabled	8.5%	8.5%	8.4%
Colombia (1991)	Height	9.8%	9.5%	9.1%
Mexico (1994)	Days ill		11.7%	6.8%
Mexico (1994)	ADLs		11.9%	10.2%
Nicaragua (1993)	Days ill	8.6%	8.6%	8.5%
Peru (1994)	Days ill	8.0%	8.0%	7.5%
Peru (1995)	Days ill	8.3%	8.2%	7.0%
		Women		
Colombia (1993)	Days disabled	10.4%	10.4%	8.8%
Colombia (1991)	Height	9.6%	9.5%	9.0%
Mexico (1994)	Days ill		18.7%	19.6%
Mexico (1994)	ADLs		19.6%	18.6%
Mexico(1995)	Age at menarche		13.5%	13.1%
Nicaragua (1993)	Days ill	7.3%	7.3%	7.0%
Peru (1994)	Days ill	8.6%	8.5%	6.9%
Peru (1995)	Days ill	6.9%	6.8%	6.0%

Notes: Returns to education are estimated at the sample mean, using coefficients reported in the case studies. All include both urban and rural populations except Colombia (1991) and Peru (1994). Where necessary, population-weighted averages were calculated and splined or dichotomous specifications were converted to average annual returns to schooling. The Mexican studies did not report earnings equations that excluded the health variable.

statistically weak. Nonetheless, these studies yielded more support for the positive interactions (complementarity) between health and schooling than did Schultz (1995b) in West Africa.

Subpopulation Comparisons: Are the Differences Real or Imagined?

Most of the studies estimated health effects on earnings among different subpopulations that are expected to have different outcomes for biological,

social, or economic reasons. The evidence shows that the individual's gender, age, schooling, rural/urban residence, and household wealth are frequently related to his/her health status indicators. In most studies, health and wage estimates are disaggregated by sex because of the substantial differences in the level and structure of earnings by gender, the higher mortality rates among men within each age cohort, the marked differences in height, and other likely gender differences in morbidity and health status that might arise for biological or behavioral reasons and then be modified by economic development.

Each time that health impacts are estimated separately for subpopulations, the interpretation of findings is complicated by an underlying question: Are the categories exogenous or endogenous? Sex is perhaps the most exogenous of these divisions, but choosing to enter the labor force is not exogenous, so adjustments have to be made for the fact that women in the labor force differ systematically from the population as a whole. The studies often find significant differences in health impacts for men and women, but it is difficult to generalize the results.

It is reasonable to expect that there would be gender differences in the elasticity of wages with respect to health. Other studies have shown that labor force participation decisions and an individual's potential earnings in the labor market, after controlling for other personal characteristics, lead to different patterns of remuneration between men and women. In most of the studies presented in this book, health appears to have more impact on wages for men than for women (compare Tables 1.9 and 1.10). Male earnings appear to be more sensitive than female earnings to differences in health in Colombia (Ribero and Nuñez), in the Peruvian national survey data (Cortez), and among elderly Mexicans (Parker). Women's earnings were more sensitive to health in the Peruvian study that focused on urban areas using the PLSMS and in the Nicaraguan study. Looking at a wider range of studies does not lead to a clear pattern. The gender differences could be real, but they could also be a consequence of smaller samples for working women than men, increasing the margin for sample selection bias to distort the estimates for women.

Ethnicity may also be exogenously determined, like gender, but place of residence introduces more concerns about endogeneity. People move, and their migrations are often related to selecting places that improve their earnings, and potentially their ability to obtain access to health services. Rural and urban populations differ by health conditions, labor markets, employment opportunities, and welfare. Although disaggregating the samples by

this distinction may be informative, it could also introduce some bias because of the substantial migration of individuals, during a lifetime, from one region to another. In Latin America during recent decades, this has taken the predominant form of rural to urban migration. Expectations are that the rural-to-urban migrants will tend to be better educated and possibly healthier than the rural population they left behind, but possibly have received lower-quality schooling and amassed lower health human capital than the native urban population among whom they settle.

Study comparisons of health or productivity within the rural or urban sectors may thus be affected substantially by the distinctive characteristics of the migrants, interjecting a possible sample selection bias. The health characteristics of the community where migrants currently live may diverge sharply from those of the community in which they grew up and in which many adult health outcomes were shaped. The magnitude of migration suggests that about a fourth of the population may be mismatched (Schultz and Tansel, 1997). Several of the studies in this book tried to address this bias by using variables in the survey regarding whether or not the respondent was an immigrant. But this information was only available in a few cases. Unfortunately, none of the surveys asked respondents about their place of birth, which would provide an even better identification of immigrants.

The mismatch is crucial because of the temporal lag between childhood investments in human health capital and eventual adult outcomes. The current health conditions and policies of the community of residence may not approximate the conditions when individuals were making critical childhood investments in health. In general, rural areas are associated with poorer health, even after controlling for schooling and other individual, family, and community characteristics (Murrugarra, Knaul, and Parker). And in several of the studies, health appears to have a greater impact on earnings in rural than urban areas. Nevertheless the health indicator studies reported here, and most others as well, have not corrected rigorously for these matching problems. To estimate the impact on adult health of childhood health investments requires additional data on migration histories and the past evolution of community health programs.

When differences are considered across occupations, formal versus informal employment, and income levels, the questions of bias due to self-selection and sorting make interpretation even more difficult. The studies that enter this territory can be considered suggestive, and demonstrate the

Table 1.9 Impact of Selected Variables on the Health Status of Urban Men

Variable	Peru[a] (1994)	Mexico (1995)	Mexico[a] (1994)	Peru (1995)	Colombia (1991)	Nicaragua (1993)
Individual Characteristics						
Education	−0.72		0.006*	−0.43***		−0.026
Age or experience	−0.68**		−0.005**	−0.18***	0.43*	−.004
Migrant	−11.54**		−0.24			
Personal Health Services						
Distance to health center	−13.38**					
Food price	−7.27**			−0.98**		
Physicians per capita						
Nurses per capita						0.032
Housing						
Sewerage and potable water			1.25*	1.54	7.0*	0.083
Adequate floor or roof	2.71**		0.06**	0.28*		−0.078

Notes: Returns to education and age are estimated at the sample mean using coefficients reported in the case studies; although other policy variables were used as instruments, they have not been included here for purposes of presentation, and information about variables that were not included can be found in subsequent chapters; Mexico (1994) is based on the analysis of ADLs, and Colombia (1991) on adult height.

a/ For urban and rural sample combined with a dummy variable for residing in urban areas.

* Significant at 10% level.

** Significant at 5% level

*** Significant at 1% level.

need for more-refined models and further analysis.[13] Murrugarra and Valdivia use quintile regressions to demonstrate that earnings are more sensitive to changes in the instrumented health variables for lower-income than for higher-income individuals. It would be important to discover which policy interventions most benefited the health and earnings of low-wage workers,

[13] One study conducted by Vijverberg (1995) explicitly addresses such problems in modeling the selection process between rural and urban areas, and between wage and self-employment categories.

Table 1.10 Impact of Selected Policy Variables on Health Status of Urban Women

Variable	Peru[a] (1994)	Mexico (1995)	Mexico[a] (1994)	Peru (1995)	Colombia (1991)	Nicaragua (1993)
Individual Characteristics						
Education	0.75	.025**	0.10	–0.17**		0.01
Age or experience	–0.24**	–0.049	0.09**	–1.83**	0.83*	0.52
Migrant	1.98		0.21			
Personal Health Services						
Distance to health center	9.65*					
Food price	–13.3**			–0.72***		
Physicians per capita		–0.26				
Nurses per capita						–0.013
Housing						
Sewerage and potable water		0.006***	–0.30	–0.15	0.50	–0.25
Adequate floor or roof	2.86**	0.012***	0.67***	–0.28***		–0.23

Notes: Returns to education and age are estimated at the sample mean using coefficients reported in the case studies; although other policy variables were used as instruments, they have not been included here for purposes of presentation, and information about them can be found in subsequent chapters; Mexico (1994) is based on the analysis of ADLs, and Colombia (1991) on adult height.

a/ An urban and rural sample combined with a dummy variable for residing in urban areas.

* Significant at 10% level.

** Significant at 5% level.

*** Significant at 1% level.

and which policy interventions were on balance more favorable for high-wage workers. Future research could then seek to assess the distributional impact of alternative public health policies and programs.

Instrumental Variables as Policy Guides

All of the studies used an instrumental variable procedure to estimate the impact of health indicators on wages. There were two reasons. First, there is

plausible evidence that errors in measurement of health human capital are substantial, and second, the heterogeneity across workers in their reported health is likely to be correlated with the unexplained variation in their wages. To correct for these, the studies exploited a range of variables that are linked to health, such as housing infrastructure, community sanitation, and potable water. To instrument health, Murrugarra and Valdivia used adequate ceilings as a measure of *housing quality*, while Parker used running water and proper floors, and Ribero and Nuñez used crowding (i.e., persons per room). Such variables are likely to reflect household wealth, and more generally the community's wealth, which stimulates demand for investment in health human capital. *Health policy interventions* were also used as instruments, and equated largely with access to various public programs. Murrugarra and Valdivia used distance to health centers, while Parker used the number of hospital beds, Knaul used the number of physicians, and Ribero and Nuñez used percentage of people enrolled in the social security system. Other aspects of *family wealth* are also used to instrument for the health indicators, such as home ownership (Ribero and Nuñez) and private savings (Parker). Tests of overidentification and other specification tests may help to show that alternative reasonable choices in identifying instrumental variables will lead to similar estimates of the impact from health status indicators on wage productivity.

For guiding public policy, the primary focus centers on community characteristics that can be modified through collective actions. Such characteristics include those that could influence exposure to health risks or disease, including population density, transportation, sanitation, or access to potable water. They might also include local population traits that affect behaviors and generate externalities, such as education. These variables are summarized for urban men and women in Tables 1.9 and 1.10.

Access of individuals to community health and educational services is expected to ameliorate the health consequences of exposure to these risks, and can be reflected in the general availability and price of health services and health-related goods, such as food, nutritional supplements, and medicines. For example, some studies found that the community's health resources (in Knaul, the number of physicians per capita; in Espinosa, the number of nurses; and in Parker, the number of hospital beds) are associated with increases in residents' health status. Findings also linked health indicators to the local prices of basic food staples such as potatoes (Murrugarra and

Valdivia), and to rural community coverage rates under national social security programs that provide health insurance (Ribero and Nuñez). The average distance from community households to a health center and the average waiting time in those centers were used as proxies for the local prices of public health services, but were not found consistently to be significantly related to adult health indicators.

The policy variables that most consistently showed an impact on health were those related to housing characteristics and sanitation. For example at the community level, the household living space (the number of people per room), the proportion of homes with proper floors or ceiling materials, and suitable drainage and sanitation facilities were all associated with increased individual health status in the studies by Knaul, Espinosa and Hernández, and Ribero and Nuñez.

A Research Agenda for Improved Policy

All seven Latin American studies in this book, each based on microanalyses of seven different household surveys, find evidence to support the notion that healthier individuals are more productive and earn more per hour worked. Given the lack of consensus among experts on how to measure adult health and the resulting variation in survey questionnaire designs, the various studies understandably have measured health status or health human capital using widely different criteria. Yet all of them found that using instrumental variable techniques to isolate the impact of the individual, family, and community resources, prices, and environment on the individual's health indicator showed the effect of adult health on productivity to be positive and generally statistically significant. The elasticity of earnings with respect to health at the sample means ranges from 0.09 to more than 0.80, but in general, these estimates indicate that earnings may be quite sensitive to small but consistent improvements in health (see Table 1.11).

The next step is more difficult. These aspects of health are hard to link precisely to the available measures of current community social services that one might associate with investments in public health and development infrastructure. These studies demonstrate that the use of instrumental variable estimation is warranted because it deals reasonably with three problems: measurement error embedded in health indicators, the heterogeneity in individual endowments and compensating behavioral patterns, and the dual

nature of health differences (i.e., a genetic foundation modified at the margins by the accumulation of a stock of human capital).

Yet the measures of policy choices that research teams can use as instrumental variables needs to be improved. Major improvement in knowledge about the relationships between public policy interventions and adult health could occur if data were assembled on the historical time series variation in regional health programs that could be matched to individual migration histories. The production of health is distinctively a process of long gestation. Tracing an individual's access to health services and exposure to disease during the critical early childhood years is critical for testing definitively the hypotheses of social scientists and physicians that public policies are responsible for the dramatic increases in life expectation and health in the twentieth century. Great strides could then be made to fine-tune estimates of how public policy interventions can be best designed to improve future adult health status and thereby enhance adult productivity and welfare.

Two of the measures of adult health—height, and age at menarche—are determined during the individual's development life cycle by physical growth processes and then remain more or less constant. Because these variables are determined by adulthood, say age 20 to 25, the cross-sectional variation in these variables associated with the age of each respondent makes it possible to construct a historical time series. From this, one can infer how these indicators of health have improved within a population over time, assuming there are no systematic errors in reporting associated with age. Both of the studies in this book that use these indicators find strong secular improvements in health among their surveyed populations, comparable to that noted in high-income countries since the Industrial Revolution and in a few observed low-income countries (Strauss and Thomas, 1998; Schultz, 1995b and 1999b). More-frequent inclusion in household surveys of these anthropometric indicators of physical health and early nutritional health status could open the door to comparative analyses of the inequality of health in Latin America. Although the surveys used here did not contain information on adult weight-for-height, commonly known as a body mass index (BMI), this measure is as important as height alone is for future data collection. Collecting data on BMI would allow researchers to go beyond the long-term research that utilizes height data and begin to understand short-term effects, including the impact of current malnutrition and poor health status on earnings.

Table 1.11 Summary of Estimated Percent Change in Wages Associated with Changes in Various Indicators of Adult Health

Health Indicator, Country (source)	Females	Males
Height (1 centimeter)		
Colombia (Ribero and Nuñez, 1999) – urban areas only	7%	8%
Ghana (Schultz, 1996)	*8%*	*6%*
Brazil (Alves et al., 1999)	*6%*	*4%*
Age at Menarche in Years (Elasticity with sign reversed)		
Mexico (Knaul, 1999)	0.47	—
Body Mass Index (Weight kg/Height M²; 1 unit change)		
Ghana (Schultz, 1996)	*7%*	*9%*
Côte d'Ivoire (Schultz, 1996)	*14%*	*9%*
Illness and/or Inability to Work as Usual (Elasticity with sign reversed)		
Colombia (Ribero and Nuñez, 1999) – Days unable to work, pop. wtd. average elasticities for rural and urban samples	0.04	0.07
Colombia (Ribero and Nuñez, 1999) – Unable to work, pop. wtd. average elasticities for rural and urban samples	0.010	0.017
Peru (Murrugarra and Valdivia, 1999) – Days sick, urban only, pop. wtd. average elasticity for wage earners and self-employed	0.07	0.04
Peru (Cortez, 1999) – Reciprocal of days sick plus one with sign reversed, average elasticity for rural and urban areas	0.10	0.20
Mexico Elderly (Parker, 1999) – Number of days sick or injured in last 180 days	n.s.	0.81
Nicaragua (Espinosa and Hernández, 1999) – Days sick	0.16	n.s.
Ghana (Schultz and Tansel, 1997) – Days unable to work	—	*0.11–0.24*
Côte d'Ivoire (Schultz and Tansel, 1997) – Days unable to work	—	*0.09–0.28*
Functional Limitations (Elasticity with sign reversed)		
Mexico (Parker, 1999) – Number of ADLs for working individuals over 60 years of age	n.s.	0.38

Notes: Italics indicate studies not included in this book; all elasticities are reported at the sample mean.

n.s. Not statistically significant at the 5% level.

— Not applicable.

In sum, gathering migration histories for individuals and assembling better databases on health conditions and policy variations by small regions of residence will lead to stronger analyses and to greater reliability in their findings. Then the debate could be confidently joined about which policies have the greatest promise of reducing over time identified inequalities in health in Latin American countries, and thereby strengthen efforts to ameliorate the region's notably high level of earnings inequality.

CHAPTER 2

Health, Nutrition, and Wages: Age at Menarche and Earnings in Mexico

Felicia Marie Knaul[1]

This study uses a human capital framework to evaluate the impact of investments in health and nutrition on labor market productivity in Mexico. The research extends the existing literature by proposing age at menarche as an effective indicator for analyzing the long-term impact of health and nutritional investments during childhood and adolescence on the earnings of women. The findings reported in this chapter suggest that those impacts are considerable. The instrumental, errors-in-variables model suggests that a decline of one year in age at menarche is associated with an increase in wages of 23 to 26 percent. Factors associated with declining age at menarche include urbanization, increased levels of education, and improved living conditions. In particular, variables that measure access to public services and the quality of housing appear significant. The proportion of the community with earth flooring in their homes and the proportion that lack toilet or drainage facilities are particularly strong correlates of cumulative health status as measured by age at menarche. Access to personal health services appears to have little marginal impact on this indicator of childhood health conditions.

[1] Dr. Felicia Marie Knaul is currently with the World Health Organisation in Geneva and the University of Lausanne. She was Director of the Program on Health Economics, Centro de Investigación y Docencia Económicas (CIDE), México, D.F., when this research was undertaken.

Introduction

A nation's potential for economic growth and development is reflected in the health and nutritional status of its population. For the individual, particularly at low levels of income, health may be an essential determinant of productive capacity in the labor market, and hence of the earnings needed to escape poverty.

The relationship among health, nutrition, and income—particularly as expressed in nutrition-based efficiency wages—has been an important element in theories of economic development for some time (Leibenstein, 1957; Rozensweig, 1988; Strauss and Thomas, 1998). Economic history has also been advanced considerably by recent efforts to deepen and extend the analysis of how long-run changes in the health of populations are related to the process of economic development and structural transformation (Fogel, 1994; Steckel, 1995). Knowledge of the link between health and income has been enriched by empirical evidence of how health affects wages and productivity among poorer populations (Thomas and Strauss, 1998). The relationship between labor productivity and health is now being explored in an integrated human capital framework (Schultz, 1996 and 1997; Schultz and Tansel, 1997; Thomas and Strauss, 1997). Models of economic growth have been expanded to include the importance of health as a human capital input (Barro and Sala-I-Martin, 1995). All of these breakthroughs are a product of advances in both economic theory and data quality. The surge now taking place in applied research also reflects growing awareness of the opportunities for, and challenges in, formulating more-effective health policies.

This chapter uses a human capital framework to evaluate the impact of health and nutritional investments during childhood and adolescence on adult labor market productivity in Mexico. The research broadens the existing literature by proposing age at menarche as an effective analytic indicator. Like adult height and body mass index (BMI), which have been widely used as indicators, changes in age at menarche also reflect the secular increase in national economic development (Brundtland and Walløe, 1973; Marshall, 1978; Malcolm, 1978; Wyshak and Frisch, 1982; Wyshak, 1983, Manniche, 1983; Wellens et al., 1990; Hulanicka and Waliszko, 1991; Liestøl and Rosenberg, 1995). Age at menarche has steadily decreased at a rate of approximately three to four months per decade in many countries during

the past 150 years. This decrease reflects a variety of socioeconomic factors, particularly improved childhood nutritional status. Despite its parallels with adult height as an indicator of general health, age at menarche has not been used as an indicator in previous analyses of the impact of investments in health on economic development.

This chapter examines the correlates of age at menarche in the framework of a reduced-form health production function. Particular emphasis is placed on the importance of policy-sensitive health variables as determinants of age at menarche, and hence long-term health among women. Hourly wages are used to measure the impact that investment in health and nutrition early in the life cycle has on future labor market productivity. Age at menarche is presented as a proxy for certain aspects of the health and nutritional components of human capital. The integrated human capital framework that underlies the theoretical model is developed in Schultz (1997), and has been applied in such works as Schultz (1996), Schultz and Tansel (1997), and Thomas and Strauss (1997).

The model uses an instrumental variable approach to address the significant degree of measurement error that is inherent in retrospective information on menarche. The instruments used to identify menarche are based on the availability of personal health services, access to public services, housing quality, average levels of education, and access to educational facilities in the community. A number of variables are included in the wage function to control for variation that is related to genetic and other determinants of menarche. These variables are expected to be uncorrelated with the reproducible component of health human capital.

It is important to note that the measure of the impact from age at menarche on labor market productivity is a lower-bound estimate of the welfare impact from ill health (Schultz and Tansel, 1997). First, age at menarche measures only a few of many dimensions of health. In particular, it is a cumulative measure reflecting investments in early nutritional status and other investments in childhood health. Second, labor market productivity and wages reflect only one aspect of the myriad complications that adult ill health has on personal and family welfare.

To be precise, menarche will be defined here as the onset of the first menstrual cycle. Puberty is a collective term that summarizes a set of morphological and physiological changes that are the result of complex developmental processes in the central nervous and endocrine systems. In women

these include the adolescent growth spurt, the development of secondary reproductive organs and sex characteristics, changes in body composition, and development of the circulatory and respiratory systems leading to increases in strength and endurance. Menarche is a relatively late event in physical development that typically occurs after the adolescent growth spurt and typically follows the peak in growth (Marshall, 1978; Tanner, 1962).

The analysis in this chapter proceeds in several stages. The first section gives a snapshot of how health in Mexico has evolved in recent decades. The second discusses menarche as an indicator of health and nutritional status. The third provides an overview of the data used in the analysis. The fourth summarizes the model and the estimation strategy, following Schultz (1996) and Schultz and Tansel (1997). The fifth provides descriptive statistics, with particular emphasis on the distribution of menarche by cohort, level of education, and hourly wages. Results of the first-stage, reduced-form estimates of the health production function are given in section six. The seventh section presents the instrumental variable estimates of the wage regressions, emphasizing the relationship between age at menarche and wages. Conclusions and policy recommendations comprise the final section.

Achievements in Health and Nutrition in Mexico

Mexico is a particularly interesting case for studying the evolution of age at menarche. While the country is well into the epidemiological transition characteristic of development, its pace and scope have been characterized as "protracted and polarized" (Frenk, Bobadilla et al., 1989) as a result of income inequalities and unequal access to resources such as health services. Mexico thus faces a dual challenge. Mexico has a backlog of "pretransitional" diseases—many of which are infectious, are based on nutritional deficiencies related to infant and maternal mortality, and can be prevented with relatively inexpensive public health interventions. These diseases are juxtaposed with an increasing health burden from chronic, noninfectious illness. It is the first challenge, communicable diseases, that disproportionately affects the poor (Frenk, Bobadilla et al., 1989; Frenk, Lozano et al., 1994).

As in many Latin American countries, the decline in Mexican mortality rates has occurred quickly. Life expectancy almost doubled between the early 1900s and 1950, and now exceeds 70 years. Infant mortality has dropped considerably: from 323 per 1,000 live births in 1910 to nearly 40 per 1,000

during the last decade (Frenk, Bobadilla et al., 1989; Bobadilla et al., 1993). The proportions of deaths related to maternal mortality and malnutrition also have declined substantially (Frenk, González-Block et al., 1994).

While historical data on nutrition is scarce, there is evidence to suggest that malnutrition has been rising in some rural areas and declining in others. Overall, the proportion of rural children between one and five years old whose height is normal for their age increased from 49 percent in 1974 to 52 percent in 1996. Furthermore, both mortality and morbidity from nutritional deficiencies declined according to data from 1990 to 1996 (Dirección General de Estadística e Informática, 1998; Avila-Curiel et al., 1998). Figures for Latin America and the Caribbean are both more accessible and more dramatic. The prevalence of nutritional deficiency dropped from 19 percent of the population in 1969–1971 to 15 percent in 1990–1992, and is projected to reach 7 percent in 2010 (FAO, 1996).

Health indicators for Mexico, while clearly improving, show less progress than they should relative to the country's level of economic development. Reduction in the proportion of deaths from infectious disease has lagged behind many other countries in the region. The figure for Mexico declined from 30 percent to 13 percent between 1960 and 1985; while Argentina, Cuba, Costa Rica, and Chile are now well below 10 percent (Frenk, Lozano et al., 1994). Similarly the ratio of deaths from infectious and parasitic diseases to deaths from noncommunicable diseases is approximately unity in Mexico, but less than 0.5 for several other countries with similar per capita income levels.

There are also substantial differences within and between regions and municipalities in Mexico, reflecting the high degree of inequality in both health status and the distribution of health services. Infant mortality in the poorer southern states declined from approximately 147 per 1,000 live births in the early 1960s to 92 per 1,000 in the 1980s, while the rates for the wealthier northern region were 92 and 28, respectively. The ratio of infant mortality between the southern and northern regions went from 1.6 to 3.3 during the same period (Bobadilla et al., 1993). The differences within states also suggest important inequalities based on rural versus urban residence.

The health situation that has resulted from this prolonged and polarized epidemiological transition places a heavy burden on a relatively extensive but inefficient health system. The Mexican system is dualistic. The poor and uninsured have access to the public health system run by the Secretariat

of Health. In contrast, the insured working population has the right to use the Mexican Social Security Institute, or the Instituto Mexicano del Seguro Social (IMSS), which covers close to half of the labor force and their families. Despite this coverage, many people use private health services and pay for them out-of-pocket. This is an indicator of overlap in, the inefficiencies of, and dissatisfaction with the health system (Zurita et al., 1997; Knaul et al., 1997; and Frenk, González-Block et al., 1994).

Furthermore, the distribution of health services not only parallels existing inequalities in the health status of the population, but it often intensifies them. For example, Frenk, Duran-Arenas, et al. (1995) report approximately 200 people per physician in Mexico City, a figure that exceeds the average in many developed countries. The ratio, however, soars in the poorer states and in rural areas. Oaxaca has an estimated 1,120 inhabitants per doctor, and Chiapas 1,370. In partial recognition of and response to systemic deficiencies, reforms have been initiated at IMSS, and a system-wide process of decentralization is well under way (Frenk, 1997).

Menarche as a Measure of Health and Nutritional Status

Fogel (1994) and Steckel (1995) highlight the secular improvements in mortality and morbidity, and their relationship to a complex set of factors associated with economic development. These factors include improvements in nutritional status, medical technology, health care access, education, public health facilities, and hygiene.

Height and weight have traditionally been used as predictors of morbidity and mortality risk among children. More recently adult height and BMI have been put forward as indicators of the probability of dying or of developing chronic diseases at middle and late ages (Fogel, 1994; Thomas and Strauss, 1997; Schultz, 1996), and as measures of living standards (Steckel, 1995).

Adult height and BMI measure different aspects of nutrition and health. Adult height is considered an indicator of nutritional status during infancy, childhood, and adolescence. BMI is a measure of current nutritional status. This evidence is analyzed and extended by Fogel (1994), who documents the secular increase in average height and BMI in several European countries between the seventeenth and nineteenth centuries. This evidence is used to develop an argument for the importance of physiological factors in economic growth.

The research summarized in this chapter adds another dimension to the existing literature on health's importance as a reproducible form of human capital, using age at menarche as an indicator of results from investments in childhood and adolescent nutrition and health. This parallels the work of other researchers who have used adult height. The logic of associating menarche or adult height with investments in health and nutrition is based on the idea that changes over time in average height or in age at menarche for a fixed population, with no variation in its mix of biological groups, stem from changes in reproducible human capital investments and changes in disease environments (Schultz, 1996; Fogel, 1994; Steckel, 1995). Furthermore, a number of studies have shown the significant impact on labor productivity of investment in health and nutrition as measured by height and BMI in an integrated human capital framework. These include Schultz (1996) for Côte d'Ivoire and Ghana, and Thomas and Strauss (1997) for urban Brazil.

Several authors have pointed to the importance of average age at menarche as an overall comparative indicator of population health, timing of maturation, and nutritional status (Hediger and Stine, 1987; Malcolm, 1978). Furthermore, close linkage among adult height, the timing of the adolescent growth spurt, height for age, and age of menarche has been documented in a variety of countries and settings (Malcolm, 1978). Trussel and Steckel (1978) use data on height velocity for female slaves transported within the United States in the 1800s to predict probable age at menarche. Díaz de Mathman et al. (1968) find that Mexican adolescents who are malnourished are significantly older at menarche than those who are well nourished.[2] Yoneyama et al. (1988) found that age at menarche was an important predictor of adult height in a study of Japanese girls.

Some evidence suggests that age at menarche may be an important complementary indicator to adult height, and possibly a more accurate tracer of early nutritional status in certain cases. Specifically, "catch-up" growth may allow certain individuals to attain a normal height, given the expectations of their genetic group, despite having suffered from malnutrition or poor health during childhood (Floud, 1994). That is, catch-up growth reduces the effectiveness of adult height as a measure of cumulative health status to the extent that it disguises the incidence of malnutrition and ill health during youth

[2] Based on a sample of indigenous youth in Mexico City.

that is expected to affect later productivity. A delay in menarche, on the other hand, may be a more dependable tracer of malnutrition and ill health during childhood and adolescence because it is a one-time event that occurs during puberty. For example Eveleth (1978), citing the study by Dreizen et al. (1967), suggests that girls suffering from chronic malnutrition in a poor, rural area of the United States were delayed in age at menarche and skeletal maturation as compared to a control group. The completed height of these two groups was not significantly different, although the malnourished group was shorter than the control group during the period of adolescent growth. Laska-Meirzejewska (1970) found that age at first menstruation was more sensitive to external conditions related to socioeconomic status and family well-being than were height or weight among a sample of Polish girls. Further, Liestøl and Rosenberg (1995) suggest that menarcheal age, possibly related to changes in weight, may be more sensitive than height is to regional differences in poverty among schoolchildren in Oslo.

Both age at menarche and adult height have demonstrated secular improvements. These improvements are likely to be closely related to increased nutritional standards (Trussel and Steckel, 1978). Marshall (1978) evaluates a group of studies of age at menarche and concludes that, despite differences in data quality, they are remarkably consistent in demonstrating an average decline of three to four months per decade over the past 100 years. The secular decline is also evident over the past 100 to 150 years in a variety of developed countries based on aggregate trends (Wyshak and Frisch, 1982). The estimated rate of decline is between two and three months per decade. Brundtland and Walløe (1973) cite evidence from North America, Japan, and Europe to show that girls have been maturing four to five months faster per decade during the past 50 years. More-recent studies have confirmed this tendency for well-nourished women in the United States born since 1920 (Wyshak, 1983), Denmark since the 1940s (Manniche, 1983), Flemish women in the nineteenth century (Wellens et al., 1990), Poland since approximately 1950 (Hulanicka and Waliszko, 1991), and Norway among schoolchildren since the 1920s (Brundtland and Walløe, 1973; Liestøl and Rosenberg, 1995). These studies further suggest that the trend is coming to a halt, coincident with a threshold age at menarche, among some well-nourished groups of high economic status in developed countries (Brundtland and Walløe, 1973).

The determinants of age at menarche can be divided into genetic and environmental factors, in which the latter are widely thought to reflect nu-

tritional differences. The literature on adolescent growth widely concurs in establishing the link between malnutrition in infancy and childhood, later age at menarche, and a slowdown in growth (Díaz de Mathman et al., 1968; Landa Rico et al., 1968; Ramos Galván et al., 1969; Marshall, 1978; Eveleth, 1978; Frisch and Revelle, 1970; Maclure, 1991; Liestøl, 1982; Trussell and Steckel, 1978). Environmental factors such as socioeconomic status, urban residence, number of siblings, birth order, racial differences, climate, altitude, physical activity, psychological stress, season of year, and presence of a related male in the family have all been put forward, with the first two being the most consistently associated with menarche (Eveleth, 1978; Marshall, 1978; Malcolm, 1978; Moisan et al., 1990; Weir et al., 1971; Komlos, 1989; Ulijaszek et al., 1991; Bojlén and Weis, 1971; Valenzuela et al., 1991; Delgado and Hurtado, 1990; Cumming, 1990; Treloar and Martin, 1990; Graber et al., 1995; Bielicki et al., 1986). Racial differences also figure prominently in many of these studies. These partially reflect variation in socioeconomic and climatic factors but may also have an important genetic component.

Heredity-related or genetic factors may dominate among well-nourished populations (Stark et al., 1989) and appear to be more important among later cohorts (Treloar and Martin, 1990). This is supported both by important population differences and by studies comparing twins to other siblings. These studies show much larger differences in age at menarche between siblings who are not twins (Eveleth, 1978; Marshall, 1978).

In summary, literature from a variety of countries demonstrates a secular decline in the age at menarche throughout the world. This research suggests that, while a variety of environmental and genetic factors may also come into play, menarche occurs earlier among females who were healthier and better nourished as girls and adolescents. For the purposes of the research presented below, the crucial hypothesis developed here is that age at menarche is a plausible proxy for measuring part of the differences in adult labor market productivity among women that result from investments in nutrition and health during childhood and adolescence.

Data Sources

The main data source for this study is the National Family Planning Survey (NFPS), or Encuesta Nacional de Planificación Familiar (ENPF), undertaken by the Consejo Nacional de Población (CONAPO) in 1995. The NFPS

includes an individual, a household, and a community questionnaire. The individual survey is directed toward the target population of women aged 15 to 54. This part of the NFPS was answered directly by each woman and includes detailed fertility and marital histories, as well as socioeconomic characteristics and work activity. The household questionnaire considers socioeconomic characteristics, family structure, work activities, and the condition of the dwelling. The community survey was carried out in *localidades* (primarily those with less than 5,000 inhabitants[3]) and was directed at a community leader. This part of the NFPS includes information about basic community characteristics, including information on access to and use of health and educational facilities.

The sampling frame of the NFPS is designed to overrepresent the poorest, most rural states. In particular, 9 states account for 90 percent of the sample. The information for the other 23 and more populous states is given by the remaining observations, which amount to approximately 1,000 cases in the overall sample of women and 300 female wage earners. The survey includes expansion factors that are designed to restore the balance among states and to provide appropriate estimates for the country as a whole. Still, given the low number of observations for the 23 undersampled states, the analysis of age at menarche using the expansion factors proved to be somewhat unstable. As a result, the information in this chapter is based on the unexpanded figures, which implies that the estimates do not necessarily reflect the distributions within the population as a whole.[4] In order to account for the important geographic differences in socioeconomic status and the availability of health services, the regressions include either a dummy variable for rural versus urban residence (rural = 1) and another for the overrepresented states (Chiapas, Guanajuato, Guerrero, Hidalgo, México, Michoacán, Oaxaca, Puebla, and Veracruz = 1), or a full set of state dummies.

It is important to highlight that women who were interviewed in the individual survey portion of the NFPS self-reported all variables.[5] This clearly

[3] While the questionnaire and manuals report that the community segment was exclusively directed at *localidades* with a maximum of 2,500 inhabitants, the data analysis of responses by community leaders shows a large proportion contain between 2,500 and 5,000 inhabitants.

[4] The descriptive figures using expansion factors are available from the author upon request.

[5] The fact that only self-reported information is accepted in the survey generates a sampling problem. Approximately 10 percent of women aged 18 to 54 identified in the household sur-

improved the quality of the data in the sense that the self-reported responses are likely to be more correct. In particular, the labor market variables were reported as part of the household survey, often by a proxy respondent, and then solicited from individual women as self-respondents. Comparing the two responses disclosed substantial differences. For greater reliability, this research study uses the self-reported information.[6]

There are two severe restrictions inherent in the data the NFPS was designed to collect. First, the only measure of adult health is menarche, making it impossible to be more encompassing about the impact of different aspects of health status on wages. Second, the data includes no information on place of birth or history of migration.

To complement information from the community segment of the NFPS, this research uses three other sources of information about municipalities. The first is the *Indicadores Socioeconómicos e Indice de Marginación Municipal* disseminated as the Sistema del Indice de Marginación Municipal (SIMM) and developed as an index of marginality applicable to all municipalities in Mexico. The SIMM was generated by CONAPO (1993) in conjunction with the National Water Commision (Comisión Nacional del Agua) and based on the results of the 1990 general population census undertaken by the Instituto Nacional de Estadística, Geografia e Informática (INEGI). The SIMM includes the percentage of people in each municipality who are illiterate adults, did not complete primary school, live in dwellings that are overcrowded, have earthen floors, lack electricity, lack toilets and drainage facilities, lack running water, live in *localidades* with less than 5,000 inhabitants, or are workers earning less than twice the minimum salary per month.

---

vey are excluded from the individual data that includes age at menarche. The primary reason for exclusion was because they could not be located for an interview. This introduces a particular form of selection bias into the research because the women identified by the household survey but not found for the individual survey are more likely to be younger, working, and better educated and to earn a higher income (see Knaul, 1999, Appendix 1, Table 1). This bias cannot be explicitly handled using the econometric techniques applied in this paper given that age at menarche, as the key variable, is missing for these women. Since women with these characteristics tend to be younger at menarche according to the available sample, the results of the impact of menarche on productivity may be biased downward.

[6] It would be interesting to conduct a more careful analysis of these differences in the future as a means of bounding the possible error in household surveys, in which proxy respondents are common.

Another source of information is a database—jointly developed by researchers at the Colegio de México, CONAPO, and Johns Hopkins University—using the records of the Secretariat of Health and IMSS (Wong et al., 1997). This database also includes information on private-sector health services and personnel compiled in the Economic Census undertaken by INEGI. The information on altitude comes from the Sistema de Información Municipal en Bases de Datos (SIMBAD) compiled by INEGI (1995) and is based on cartographic data. All three data sets were merged with the NFPS at the level of the municipality, with information available for both urban and rural areas.

Modeling Health Production and Labor Productivity

Schultz (1996, 1997) models household demand for human capital as a derived demand for the uses of various human capital stocks. Summarizing this work, household demand for input (j) for individual (i) is given as:

$$I_{ij} = a_j Y_i + \beta_j X_i + \mu_{ij} \qquad j = H \text{ or } M, E, R, \text{ and } B \qquad [2.1]$$

where a critical distinction is made between the Y and the X variables. The Y variables affect the demand for human capital through their impact on wage structures. Because the Y variables affect wages, they have an impact on the incentives to invest in human capital. The Y variables may also have a direct impact on the demand for human capital for other reasons. By contrast, the X variables affect the demand for human capital without affecting wage opportunities. The error term is given by μ.

The inputs in an integrated framework include adult height (H) or age at menarche (M) as indicators of early investments in nutrition and health (see Fogel, 1994; and discussion above), education (E) (Becker, 1993; Mincer, 1974; Griliches, 1977), migration from region of birth (R) (Schultz, 1982), and body mass index (B) as an indicator of adult nutritional status and current health (Fogel, 1994; Thomas and Strauss, 1997). In this chapter, and given the data limitations, only age at menarche and education are considered.

The individual girl and her family maximize a single-period utility function that includes health (h^*), proxied by menarche (m^*); the nonhealth-related consumption bundle (C); and annual time allocated to nonwage activities (H_2):

$$U = U(h^*, C, H_2) \quad [2.2]$$

Equation [2.2] is maximized subject to the budget, time, and health production constraints:

$$RI = HI \cdot P_1 + CP_2 = W \cdot H_1 + V, \quad [2.3]$$

$$T = H_1 + H_2 \quad [2.4]$$

where *RI* is market income, *P*s are market prices, *W* is the wage rate, and *V* is annual household income from nonlabor assets. Total available time (*T*) is divided into wage work (H_1) and nonwage activities.

Cumulative health status is produced over the individual's lifetime and begins with parents' and personal investment in nutrition, disease-preventing interventions and practices, and in health-conserving behaviors. These health inputs (*HI*) and the individual's heterogeneous endowments (*G*) unaffected by family or personal behavior combine to determine the individual's cumulative health status (h^*), proxied by age at menarche (m^*):

$$h^* = f(m^*) \quad [2.5]$$

where,

$$m^* = m^*(HI, G, \varepsilon). \quad [2.6]$$

In equation [2.6], ε is the error term in the health function. The estimates of the determinants of age at menarche are used as the first stage of the estimation of the wage function. This chapter includes only reduced-form estimates of the health production function equations [2.5] and [2.6].

Expanding on the semilogarithmic framework of Mincer (1974), the hourly wage of the individual is a function of her cumulative health status as proxied by age at menarche, acquired skills related to education, experience as a quadratic function of aging, the vector of exogenous variables (*Y*) that are included additively, and other unobserved forms of human capital transfers and genetic endowments:

$$W_i = \sum_{j=i}^{n} \left(d_j I_{ij}\right) + tY_i + \phi_i. \qquad [2.7]$$

The econometric strategy is based on an errors-in-variables model identified using instrumental variables. This parallels Schultz (1996), Schultz and Tansel (1997), and Thomas and Strauss (1997). The two-stage instrumental variables approach is designed to correct for the downward bias of the estimated effect of health on wages due to the errors in measuring age at menarche. Reported age at menarche may diverge from true age at menarche by a measurement error (e):

$$m_i = m_i^* + e_i, \qquad [2.8]$$

where e is assumed to be a random variable that is uncorrelated with the other determinants of health or modeled aspects of behavior. Note that it is the correlation between ϕ and e that gives rise to bias due to heterogeneity or simultaneity in estimating the wage function. The correlation between the error in the wage function and unobserved health heterogeneity leads to simultaneous equation bias if the observed health inputs are related to the unobserved health heterogeneity. To correct for this problem it is necessary to include, in the health demand function, variables (such as prices or access to health services) that affect health input demand yet are not correlated with the health heterogeneity. These variables generate a series of exclusion restrictions that permit identification of the unbiased wage function.

There is ample evidence to support the hypothesis that age at menarche is measured with considerable error, particularly using the type of retrospective data available for this research. The literature on measuring age at menarche highlights the issue of recall error in this type of data. Of the existing means of determining age at menarche, the cross-sectional retrospective method is considered inferior to longitudinal (repeated questioning of adolescents) or status quo (proportion of adolescents who have menstruated by a given age) methods (Marshall, 1978; Brundtland and Walløe, 1973). Several studies have measured the recall error by comparing the results from these different methods. Most notably for this study, Cravioto et al. (1987) found that among adolescents from rural Mexico the correlation coefficient between age at menarche from longitudinal data and from recall data collected four years after menarche was only 0.61. Similarly only 70 percent of the adolescents could recall their age at menarche within one

year of the actual date. In a study of Swedish teens, Bergsten-Brucefors (1976) found that four years after menarche, only 63 percent of respondents could recall age at menarche within three months of the correct date. Hediger and Stine (1987) discuss studies showing that recall capacity falls off rapidly four to five years after the event and then stabilizes. They highlight the finding by Bean et al. (1979) that in a group of U.S. women, an average of 34 years after the event, approximately 90 percent were able to recall their age at menarche within one year. In their own work, Hediger and Stine (1987) find that about half their sample group of U.S. adolescents had low recall ability while the other half remembered relatively accurately for several years after the event. The authors suggest that recall probability is not as closely related to the length of elapsed time as other studies imply.

Recall bias is likely to be associated with three other types of error in the data used in this study. First, because age at menarche is reported in completed years, there will be a consistent downward bias in the mean age. Second, the many conflicting feelings associated with adolescence and therefore with menstruation may induce young women to provide inaccurate information, particularly if menarche was very early, very late, or especially traumatic. Another source of error may be due to women not clearly identifying the onset of menstruation. Finally, another source of error may be related to "telescoping." That is, women may tend to report a later or an earlier age at menarche as they get older or as the event becomes more distant in time.[7]

As mentioned above, in order to control for the bias due to simultaneity and measurement error, the econometric analysis in this chapter is based on an instrumental variables approach. The instrument set consists of community health infrastructure and water and sanitation conditions, as well as the level of education in the community. These variables are assumed to affect the demand for health human capital inputs, and to be uncorrelated with unobserved health heterogeneity or measurement error, thus identifying the wage equation. The instruments are selected using the results of the literature review, the descriptive analysis, and the regressions of the production function of health measured by age at menarche. The specification of the instrumental variables approach to the errors-in-variables is evaluated using Hausman tests (Hausman, 1978; Greene, 1997). In addition, the ro-

[7] For a more complete discussion of these issues, see Knaul (1999).

bustness of the instrumental variable estimate is explored by varying the instrument set.

The other human capital inputs that can be measured using the NFPS are education and post-schooling, potential years of experience. Education is analyzed using a linear specification in years. Experience is formulated in the wage equations, in the traditional Mincerian fashion, as age less years of education less 6, and included as a quadratic (Mincer, 1974).

Another potential source of bias in analyzing the impact of health results from what Schultz (1996) calls "aggregation." This bias arises when inputs that have different productive effects on wages are combined in a single indicator. In a cross-section, the fraction of the variance in age at menarche that can be explained by environmental factors may have either a smaller or a larger impact on productivity than the fraction that is largely unaccounted for and based, in part, on genetic variability. Aggregating the two sources of variation in the single measure of age at menarche may provide misleading results about the impact on productivity of changes in variables that affect only the environmental aspects of menarche.

This form of bias is partially offset by including, in both the wage and human capital demand functions, a series of variables that are related to the genetic component of age at menarche and to environmental factors. These controls are included in order to avoid relying on intergroup genetic variation to identify the wage effects of the reproducible component of health human capital (Schultz, 1996). The independent exogenous control variables are selected based on the results of the literature review and the findings for the determinants of age at menarche discussed below. Individual characteristics include age and the place of residence. Using the information from the rural *localidad* survey included in the NFPS, the proportion of the population that does not speak Spanish (this information is not available for the individual), and distance in kilometers to the nearest and most-frequented market are also used. As mentioned above, this information is only available for small rural communities. For the larger municipalities and cities, these two variables are coded as 0. The controls also include altitude in meters above sea level for each municipality, and a dummy for rural residence and another for the poorer states that are overrepresented in the sample.

The sample is restricted to include women whose declared age of menarche fell sometime between the ages of 10 and 17 years old, based on the fact that the proportion who dated it earlier or later is very low. This also

makes it possible to exclude from the production function and wage equations women aged 17 and younger. The age restriction is useful for the later extension to the wage function since it guarantees a completed profile of menarche for the women aged 18 and over at the time of the survey. Excluding these youths from the wage equation is also supported by the fact that many are still in school and not earning a wage. This restriction reduces the sample by only 1 percent.

Just under 30 percent of the adult women in the sample work and earn a positive wage, suggesting the need to identify, and correct for, sample selection bias using full information, maximum-likelihood estimates of the two-stage technique originally developed by Heckman (1979). Unfortunately the available data do not include sufficient information on exogenous determinants of labor force participation to identify the selection equation. The only exogenous measures of wealth included in the survey are the physical characteristics of the home, and the sample correction term is repeatedly insignificant when identified based on these variables. For this reason, the analysis does not include a correction for sample selection. Based on the selection-corrected regressions that were undertaken, this omission is unlikely to bias the findings regarding the impact of age at menarche on productivity and wages.[8] The results of the corrected sample selection model can be found in the appendix of Knaul (1999).

It is important to mention two general points regarding the regression analysis. First, all of the standard errors are calculated using White's robust heteroskedasticity-consistent estimator (White, 1980; Greene, 1997). Further, all of the instrumental and exogenous control variables that suffer from

[8] The instrumental variable wage equations were also calculated using a selection equation for labor force participation. The selection term was identified by including a series of arguably endogenous variables in the participation equations. In particular, a dummy variable for marriage, as well as a series of indirect indicators of wealth (measured by physical characteristics of the home and access to services) were included in the probit regression. The results of including the sample selection term derived from this analysis have very little impact on the magnitude, sign, or level of significance of the menarche variable in the wage equations. The analysis was again repeated using only the measures of the physical characteristics of the home. Using this estimation strategy, the sample selection term is repeatedly insignificant, suggesting that the identifying variables are too weak to permit a precise estimation of the characteristics of exclusion from the labor force. Schultz and Tansel (1997), using data from the Côte d'Ivoire and Ghana, find that the selectivity correction term is insignificant in predicting the impact of disabled days on productivity.

Table 2.1 Means and Standard Deviations for NFPS Variables by Sample Set and Wage Earners

	All		Wage Earners	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Ln hourly wages			1.48	1.05
Menarche	13.15	1.34	13.13	1.37
Ln menarche	2.57	0.10	2.57	0.11
Menarche^2	174.63	35.71	174.19	36.44
(Menarche) x (years of education)	75.82	53.80	91.73	60.43
Community-level policy variables				
Public Health Services (popn. shares)				
Earthen floors (%)	33.96	24.14	30.74	22.90
Without toilet or drainage facilities (%)	34.65	24.05	31.88	24.37
Overcrowded housing (%)	65.29	10.59	63.67	10.67
Without running water (%)	30.76	23.42	27.99	22.29
Personal Health Services				
Distance (km) to health center[a]	2.43	6.58	1.31	4.45
- Health center distance missing[b]	0.14	0.34	0.08	0.28
Physicians per capita (*100)	0.001	0.001	0.001	0.001
Presence of community health worker	0.68	0.47	0.74	0.44
Educational Capital				
Incomplete primary education (%)	48.88	17.66	45.75	17.92
Illiterate (%)	20.76	13.81	19.20	13.45
Distance to secondary school (km)	12.62	17.86	8.72	15.81
- School distance missing[b]	0.36	0.48	0.25	0.44
No secondary school	0.45	0.50	0.32	0.47
Other human capital variables				
Education (years)	5.81	4.12	7.05	4.63
Experience (years)	18.98	11.97	18.74	11.87
Experience^2	503.65	529.99	492.07	516.06
Controls for ethnicity and residence				
Non-Spanish-speaking (%)	4.20	15.00	2.30	10.40
- Non-Spanish share missing[b]	0.03	0.16	0.02	0.15
Altitude (1,000 meters)	1.41	1.45	1.40	1.45

Table 2.1 (continued)

Variable	All Mean	All Std. Dev.	Wage Earners Mean	Wage Earners Std. Dev.
- Altitude missing values[b]	0.02	0.13	0.01	0.11
Rural	1.59	0.49	0.45	0.50
Oversampled states[c]	0.92	0.27	0.92	0.27
Distance to market (km)	10.02	15.10	7.65	13.97
- Market distance missing[b]	0.02	0.14	0.01	0.09
Age	31.78	9.85	32.78	9.58
Observations	10,839		3,158	

Note: The sample includes women, aged 18 to 54, whose reported age at menarche was between 10 and 17.

a/ For nearest nonprivate health center; urban areas were counted as 0 km. See text for additional explanation.

b/ Applies only to rural areas since the variable is assumed to be 0 for urban areas.

c/ Oversampled states are: Chiapas, Estado de México, Guanajuato, Guerrero, Hidalgo, Michoacán, Oaxaca, Puebla, and Veracruz.

a small number of missing values are recoded with the median value. A dummy value for each variable is added to signal that the observations originally had a missing value. This guarantees comparability across regressions since the number of observations remains constant.

The wage regression uses hourly wages as the dependent variable. The adjustment to hourly wages is done by converting hours worked, when reported by day or month, to hours worked per week, using the days worked during the past week. Similarly when labor earnings are reported by a period other than the week, they are first adjusted to weeks and then divided by hours worked per week.[9]

The means and standard deviations of all variables are reported in Table 2.1 for the full sample of 18- to 54-year-old women, as well as for those with positive wages and for the sample of women living with their

[9] It is important to note that the information on hours refers to the principal job, while labor earnings refers to all jobs. There is no way to adjust for this difference since the survey does not mention the total number of jobs. Still, the proportion of women in INEGI's quarterly National Urban Employment Surveys who report a second job is very low.

mothers. The figures are presented without expansion factors, given the sampling features of the survey discussed above.

Patterns in the Age at Menarche in Mexico

The mean age in the sample of women in the NFPS who experienced menarche between ages 10 and 17 is 13.1 years with a standard deviation of 1.3 (Figure 2.1). The age distribution is concentrated at 12, 13, and 14 years. Moreover, only 1.2 percent of the 521 youth aged 15 at the time of the sample said that they had not yet menstruated. The average age of 13.1 years coincides relatively closely with figures collected for the 1960s and 1970s for certain European countries, although in several other developed countries, including the United States, average ages approach 12.5 years (Marshall, 1978). By comparison, Díaz de Mathman et al. (1968) reported an average age of 12 years (confidence interval of +/– 13 months) and 13.4 years (confidence interval of +/– 10 months) among well-nourished and poorly nourished young women from Mexico City, respectively. The estimated overall average was 12.8 years (confidence interval of +/– 16 months). Using the status quo method, Jacobo and Malacara (1985) found an age at menarche of 12.8 years (confidence interval of +/– 1.3 years) in a population of urban Mexican adolescents with no significant difference based on socioeconomic status.

There is small but consistent negative correlation between menarche and age cohort (Table 2.1). The time trend in age at menarche across birth cohorts is measured with more precision following Schultz (1996), by regressing menarche on age and controlling for the proportion of the population in a community that does not speak Spanish, the only available information on ethnic background. The OLS linear trend suggests a rate of decline of slightly less than one month per decade in Mexico ($b = .011$, $t = 8.54$). The finding of a long-run decline is consistent with other studies summarized previously that show a secular decline during the past 100 to 150 years in a variety of countries (Wyshak and Frisch, 1982; Marshall, 1978). Still, the rate of decline is one-quarter to one-half the reported fall in developed countries.[10]

[10] A large part of this divergence is likely due to differences in data collection strategies or may reflect the slower pace of improvements discussed in the earlier section overviewing the evolution of health status in Mexico.

Increases in education are also associated with a slightly more pronounced decline in age at menarche (Table 2.2). Women with no formal education report an average age of menarche of 13.3 years, compared to 13.2, 13.0, and 12.8 years for women with at least some primary, secondary, or higher education, respectively. This partly explains the cohort effect since education levels have increased substantially in Mexico during past decades. Still, the inverse relationship between menarche and education is also evident within cohorts.

Rural residence is associated with being older at menarche—13.2 versus 13.0 years in the urban areas. This is a lower-bound estimate since it is based on current residence. It is likely that half the sample of women living in the urban areas at the time of the survey were rural residents during infancy and childhood or at the time of menarche. The higher rural figure is related to a variety of factors, including the greater prevalence of malnutrition, the scarcity of health services, as well as selective migration, poverty, and educational achievement.

Table 2.2 Age at Menarche by Cohort, Level of Education, and Rural–Urban Residence

Variable	Mean	Mode	Std. Dev.	Skewness	Kurtosis	Obs.
Age Cohort						
18–24	13.03	13	1.32	0.30	2.88	3,258
25–34	13.10	13	1.35	0.19	2.92	3,537
35–44	13.23	13	1.34	0.15	2.79	2,556
45–54	13.35	14	1.36	–0.07	2.66	1,546
Years of Education						
0	13.27	13	1.31	0.03	2.79	1,574
1–6	13.23	13	1.31	0.16	2.82	5,574
7–12	12.99	13	1.38	0.30	2.89	3,127
13+	12.83	12	1.41	0.35	2.80	616
Residence						
Rural	13.00	13	1.37	0.21	2.81	4,483
Urban	13.24	13	1.73	0.18	2.80	6,416

Menarche shows a weak but steadily declining pattern with respect to the distribution of hourly labor income (Figure 2.1, Table 2.3). Women in the lowest wage quartile have a reported mean age of 13.3 years, compared to 13.0 in the highest decile. It is interesting to note that the average age at menarche is virtually identical among labor force participants and those who do not work.

In summary, the tabulations in cross-sectional data show that menarche tends to be inversely related to age, education, and wages and to be lower in urban areas. While the trends tend to be small, the patterns are consistent both overall and within cohorts. The findings coincide with the expected link to malnutrition, and to the socioeconomic determinants of age at menarche cited previously.

The Determinants of Age at Menarche and the Instrument Set

A reduced-form health production function is estimated in this section to evaluate individual, community, and regional determinants of age at menarche. The variables include both exogenous controls, and variables that are excluded in the second stage and used as instruments in the wage equation.

The instrument set is comprised of 11 variables related to the accessibility of public and health services, the quality of housing, and the level and

Figure 2.1 Distribution at Menarche by Hourly Wage Quartiles (Women 18–54)

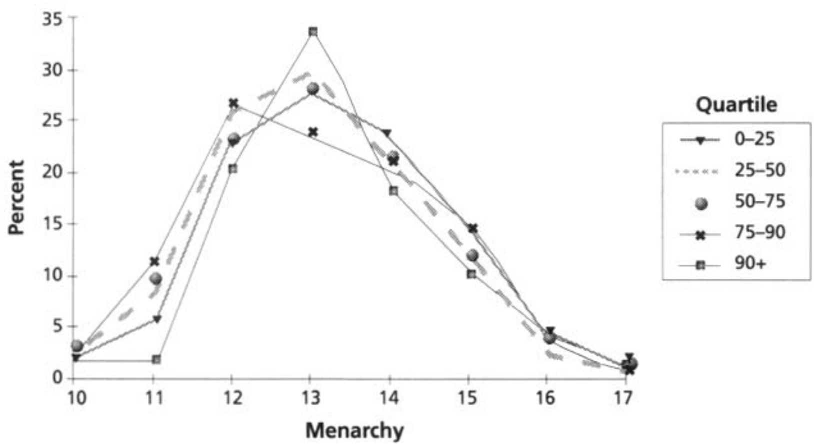


Table 2.3 Age at Menarche by Hourly Wage

	Wage Percentiles:				
Statistic	**0–25**	**26–50**	**51–75**	**76–90**	**91+**
Mean	13.25	13.08	13.18	13.02	12.99
Std. Dev.	1.30	1.36	1.42	1.39	1.43
Skewness	0.17	0.11	0.14	0.30	0.23
Kurtosis	2.68	2.79	2.74	2.85	2.68
Observations	770	835	762	476	313

availability of educational resources. This assumes that the present distribution of services and resources is correlated with the distribution that prevailed at the time when, and in the place where, the woman grew up and experienced menarche.

The first four instruments are indicators of the lack of basic services in the community and the presence of poor-quality housing. These factors should be associated with an older age at menarche since they indicate higher health risks and poverty. The variables included in the instrument set are the proportion of individuals in each municipality whose homes have an earthen floor; who do not have indoor drainage or sanitary facilities; who live in overcrowded conditions; and who lack access to running water. This information comes from the Sistema del Indice de Marginación Municipal (SIMM).

Increasing availability of and access to health services should be negatively correlated with age at menarche. Three instruments are used as indicators of the availability or accessibility of personal health services. The first variable is distance from the *localidad* to the nearest health center in kilometers, using data from the NFPS. The sign is expected to be positive. This variable is somewhat difficult to interpret given the multiplicity of health service providers discussed in the overview of Mexican health status at the beginning of this chapter. Services may be provided through the Secretariat of Health; the social security system; state and other public welfare systems; pharmacies; and to a lesser degree in the rural areas, private clinics and individual physicians and practitioners. The survey includes information on all of the public and social security clinics, although no information is given on

distance to private clinics. The variable therefore is designed to measure the distance to the nearest public or social security clinic. While the uninsured are theoretically unable to use social security clinics, many of the units in rural areas form part of the IMSS Solidaridad system that is open to the general public and targeted toward the poor and uninsured.

The other two instrumental variables for the accessibility of personal health services are related to the presence of trained practitioners. The number of physicians per capita is constructed by combining the information from the community survey of the NFPS with the information on municipalities. Thus physicians per capita is calculated by using the NFPS data for all individuals living in *localidades*, and by using information at the level of the municipality for residents of larger conglomerates.[11] This variable enters into the second set of instruments, along with the variables for earthen floors, drainage, and distance to clinics. A dummy variable indicating the presence of a community health worker at the level of the *localidad* is derived from the NFPS data. This variable is expected to be particularly important in the smaller and poorer communities. Both the physician and the community health worker are expected to be negatively related to age at menarche.

Education is considered an important input in enhancing the ability of the individual and the family to make more efficient use of health technology (World Bank, 1993). Given the externalities that apply to public health services, education is likely to have an impact at both the individual and community levels. For this reason, measures of the level of education and of access to community educational services constitute another group of instruments. The average level of education is measured as the proportion of the population with some primary schooling and the proportion over age 15 who report that they are illiterate using data from the SIMM. Presuming that the present allocation of educational services is related to earlier patterns, access to education is measured by the distance in kilometers to the nearest secondary school and by a dummy if the *localidad* has no secondary school. These data come from the community model of the NFPS. Each of these measures is expected to be associated with being older at menarche.

[11] Physicians include only those individuals who have completed their medical training and been licensed. Students undertaking their social service are not counted.

The empirical results suggest that, as hypothesized, the public service and housing variables are significant in the regression of age at menarche (Table 2.4). The largest and most significant effect is associated with the presence of sanitary facilities and the proportion of the population living in dwellings with an earthen floor. The marginal impact of overcrowding and running water is insignificant. Of particular interest are the findings that the personal health service variables have almost no significant impact. Only the variable for physicians is significant at the 10 percent level in rural areas. The sign, as hypothesized, is negative. The variables measuring the average level of education in the community are also generally insignificant. One surprising result is the negative, significant coefficient on illiteracy in urban areas. This could be related to multicollinearity. The measure of distance to the nearest secondary school is positive and significant in the rural equation. The rest of the measures are generally insignificant.

Based on the results of the F-tests reported at the bottom of the table, the instrument set is jointly significant for the regression on the full sample, on the rural subsample, and on the urban subsample. Further, the public service and housing variables are also jointly significant in each of the regressions. The group of personal health service variables is much weaker and is significant at the 11 percent level only for the full sample. The educational capital variables are also weak, although they are jointly significant at the 4 percent level for the full sample.

The coefficients on the control variables support the descriptive results presented in the first section. Education and especially age are also important determinants of menarche. Coincident with the descriptive results, age at menarche decreases with rising levels of education and occurs earlier among younger cohorts. It is somewhat challenging to interpret the role of the education variable since menarche and secondary school are likely to be coincident. It is probable that this variable is measuring issues related to the educational capital of the family in which the women grew up. Further, the fact that all of the education variables, both at the level of the woman and the community, are only significant in the urban areas may be related to differential migration. It is probable that more-educated women are more likely to move from rural to urban areas.

The age variable is presented using a quadratic specification in order to approximate more closely the specification used later, in the wage equation. While each term in the quadratic specification is insignificant, the Wald

Table 2.4 Determinants of Age at Menarche by Residence

Variables and Statistics	All		Urban		Rural	
Community-level policy variables (instrumental variables)						
Public Services and Quality of Housing						
Earthen floors (percent*100)	0.473	*3.61*	1.233	*4.04*	0.297	*2.02*
No toilet or drainage facilities (percent*100)	0.404	*3.50*	0.630	*2.42*	0.397	*2.98*
Overcrowded housing (percent*100)	0.150	*0.57*	0.343	*0.79*	–0.023	*0.07*
Without running water (percent*100)	0.028	*0.31*	–0.326	*1.58*	0.074	*0.68*
Personal Health Services						
Distance to health center (km)[a]	0.091	*0.35*			0.125	*0.48*
- Health center distance missing[b]	–0.097	*2.15*			–0.087	*1.90*
Physicians per capita (*100)	–0.288	*1.61*	–0.264	*0.63*	–0.369	*1.76*
Presence of community health worker	0.037	*1.03*			0.029	*0.77*
Educational Capital						
Incomplete primary education (%)	–0.167	*0.71*	–0.399	*0.90*	–0.343	*1.11*
Illiterate (%)	–0.284	*1.08*	–1.428	*3.04*	0.316	*0.97*
Distance to secondary school (km)	0.148	*1.44*			0.195	*1.85*
- School distance missing[b]	–0.049	*1.07*			–0.046	*1.00*
No secondary school	–0.096	*1.56*			–0.083	*1.35*
Other human capital variables						
Education (years*100)	–0.122	*3.18*	–0.247	*4.56*	0.012	*0.21*
Age (years*100)	–0.147	*0.16*	0.172	*0.12*	–0.159	*0.13*
Age^2 (years*100)	0.014	*1.04*	0.005	*0.24*	0.018	*1.06*
Controls for ethnicity and residence						
Non-Spanish-speaking (%)	–0.038	*0.37*			–0.071	*0.65*
- Non-Spanish share missing[b]	–0.447	*4.24*			–0.464	*4.30*
Altitude (1,000 meters)	0.009	*0.92*	–0.015	*0.55*	0.008	*0.72*
- Altitude missing values	–0.096	*0.91*			–0.142	*1.30*
Rural	0.250	*4.35*				
Oversampled states[c]	0.126	*2.48*	0.137	*1.95*	0.129	*1.68*
Distance to market (km*100)	–0.372	*2.97*			–0.318	*2.39*
- Market distance missing[b]	0.292	*2.50*			0.323	*2.76*
Constant	12.659	*60.93*	12.796	*37.41*	12.903	*47.33*

Table 2.4 (continued)

Variables and Statistics	All	Urban	Rural
F-statistic	15.160	11.870	7.530
R^2	0.030	0.030	0.030
N	10,831	4,459	6,372
Wald test of joint significance[d]			
All community-level policy variables	9.260 (0.00, 13)	7.800 (0.00, 7)	5.770 (0.00, 13)
Public service and quality-of-housing variables	9.770 (0.00, 4)	6.460 (0.00, 4)	4.890 (0.00, 4)
Personal health service variables	1.910 (0.12, 4)	0.400 (0.53, 1)	1.680 (0.15, 4)
Educational capital variables	2.390 (0.04, 5)	6.810 (0.00, 2)	1.710 (0.13, 5)
Age and age²	16.630 (0.00, 2)	2.710 (0.07, 2)	17.960 (0.00, 2)

Notes: The dependent variable is age at menarche. The sample is women aged 18 to 54 who reported age at menarche between ages 10 and 17. Coefficients reported from OLS regressions with absolute t-values in italics.

a/ For nearest nonprivate health center; urban areas were counted as 0 km. See text for additional explanation.

b/ Applies only to rural areas since the variable is assumed to be 0 for urban areas.

c/ Oversampled states: Chiapas, Estado de México, Guanajuato, Guerrero, Hidalgo, Michoacán, Oaxaca, Puebla, and Veracruz.

d/ Wald test of joint significance, F-statistic with probability at 10 percent significance and degrees of freedom in parenthesis.

tests of joint significance show that age is an important determinant of menarche. The linear specification (not presented in the table) suggests that an increase of one year in age is associated with an increase of one month in the age at menarche for the complete sample. The effect is larger and more significant in the rural areas, reaching approximately 1.5 months per year. In the urban areas, the effect is only one-half month.

The regression results for the control variables for ethnicity and residence also reinforce the descriptive results. Age at menarche occurs at an older age in rural areas and in the oversampled, poorer states. The variable for distance to the nearest market is negative and significant, again suggesting that more-urbanized areas have younger ages of menarche. Furthermore, ease of contact with other populations and with a flow of goods and services may reflect both higher incomes and better access to health care services. Finally, the term for the missing values on the variable for the proportion of

the population that does not speak Spanish is negative and significant. This is likely to reflect the fact that the information is missing for a few larger municipalities that are registered in the NFPS as *localidades* yet are registered with much larger populations in other sources.

The findings of this section underscore the importance a variety of individual factors have in explaining the evolution in age at menarche. There is a strong, positive association with the age of the woman, rural residence, and residence in a poorer state. This is likely to reflect the important improvements in nutritional and health status that have occurred over time and with economic development. Further, the community-level variables as a group have a significant impact on age at menarche. Personal health services have a negligible impact, while public services and housing are the dominant determinants.

Health and the Determinants of Wages

This section estimates the impact on wages of investment in women's health and nutrition by using variation in the age at menarche as the proxy for health and nutrition. While earlier sections developed arguments to support the existence of an important link among childhood and adolescent nutritional status, labor market productivity, and age at menarche, the functional form of the associations is unclear. While there may be a linear relationship between wages and menarche, it is also possible that this linkage involves returns to scale or is mediated by complementarity between health and education. Given this potentially nonlinear relationship, this section explores a variety of specifications for the menarche variable in the wage equation. In particular, the regressions are run using a linear, a logarithmic, a quadratic, and an interaction-with-education specification for the menarche variable.

The results of the instrumental variable estimates of the impact of age at menarche on the wage equations for the full sample are given in Table 2.5. As a point of reference, the OLS result is given for the double logarithmic specification, and the menarche variable is insignificant.[12] This, contrasted

[12] Other OLS regressions were also run, although they are not reported. The relationship between menarche and wages is negative and significant in a simple OLS regression with no other control variables. Adding education and experience to the human capital equation reduces the impact of menarche and renders the coefficient statistically insignificant.

with the significant results for the instrumental variable regressions, suggests the presence of downward attenuation bias due to errors of measurement.

The first column of the instrumental variable regressions refers to the double logarithmic, the second to the linear, and the third to the quadratic specification of the functional relationship between menarche and wages. The fourth column presents the results, including a linear term for menarche and an interaction term with years of education. The final column includes both a quadratic specification of the menarche variable and the education interaction term.

The effect of age at menarche, in both the double (column I) and the semilogarithmic specifications (column II), is negative and significant. The quadratic function (column III) also shows an inverse relationship between menarche and wages. Further, the three functions give very similar results in terms of the marginal impact of menarche on wages. The coefficients in the semilogarithmic equation indicate that a decrease of one year in the age at menarche is associated with an increase of 26 percent in hourly wages. The results of the double logarithmic specification suggest that a fall of 1 percent in menarcheal age results in an increase of 3.5 percent in wages. This figure is very similar to the coefficient for the linear specification in that a decline of one year in menarcheal age, equivalent to a change of 7.6 percent from the mean of 13.2 years, results in an increase of 27 percent in wages. Further, the quadratic specification suggests that a change of one year in menarcheal age is associated with a 26 percent difference in wages from the level at the mean age of menarche.

The last two specifications of the wage function include the interaction of age at menarche with years of schooling. The instrument set remains the same, and the linear term on years of education is treated as exogenous, while the menarche variable and the interaction term are treated as endogenous. The interaction term is added under the hypothesis that improvements in nutrition and health may operate through the individual's capacity to obtain educational capital (Moock and Leslie, 1986). The results of introducing this term are much weaker (columns IV and V, Table 2.5). The menarche variable retains its negative sign and significance in both the semilogarithmic (column IV) and in the quadratic specifications (column V). The magnitude of the coefficient increases with the addition of the interaction term in the semilogarithmic case, yet the coefficients for the

Table 2.5 Earnings Equations for Women with Health Status Variable

Independent Variables	OLS		Instrumental Variable Regressions (I)		(II)		(III)		(IV)		(V)	
Menarche												
Ln Menarche	0.089	*0.56*	–3.524	*2.44*								
Menarche					–0.261	*2.35*	–7.258	*2.26*	–0.569	*2.31*	–6.688	*0.24*
Menarche²							0.266	*2.18*				
(Menarche) x (Years of educ.)									0.047	*1.44*	0.019	*0.49*
Other Human Capital Variables												
Education (years)	0.135	*29.51*	0.130	*24.50*	0.131	*25.06*	0.117	*13.04*	–0.484	*1.13*	–0.127	*0.26*
Experience (years x 10)	0.340	*6.60*	0.375	*6.41*	0.375	*6.41*	0.363	*5.18*	0.336	*5.06*	0.349	*4.78*
Experience² (x 1,000)	–0.414	*3.44*	–0.438	*3.34*	–0.435	*3.33*	–0.503	*3.25*	–0.324	*1.97*	–0.452	*–2.46*
Control for ethnicity and residence												
Non-Spanish-speaking (%)	–0.527	*2.24*	–0.503	*2.11*	–0.498	*2.09*	–0.655	*2.45*	–0.546	*2.23*	–0.659	*2.51*
- Non-Spanish share missing[a]	–0.033	*0.24*	–0.164	*1.09*	–0.165	*1.09*	–0.007	*0.04*	–0.060	*0.35*	0.02	*0.10*
Altitude (1,000 meters)	0.012	*0.72*	–0.013	*0.79*	–0.013	*0.79*	–0.012	*0.68*	–0.003	*0.14*	–0.008	*0.42*
- Altitude missing values	0.077	*0.46*	–0.098	*0.53*	–0.091	*0.49*	–0.107	*0.47*	–0.151	*0.79*	–0.13	*0.57*
Rural	–0.025	*0.56*	0.062	*1.06*	0.061	*1.03*	0.007	*0.10*	0.064	*1.05*	0.014	*0.19*
Oversampled states[b]	–0.132	*2.56*	–0.038	*0.57*	–0.041	*0.62*	–0.046	*0.54*	–0.046	*0.69*	–0.047	*0.59*
Distance to market (km)	0.026	*0.17*	–0.020	*0.12*	–0.030	*0.17*	0.259	*1.13*	–0.105	*0.57*	0.2	*0.77*

Table 2.5 (continued)

Independent Variables	OLS		Instrumental Variable Regressions (I)		(II)		(III)		(IV)		(V)	
- Market distance missing[a]	−0.163	*0.75*	−0.145	*0.65*	−0.140	*0.62*	−0.300	*1.21*	−0.267	*1.07*	−0.335	*1.33*
Constant	0.021	*0.05*	9.170	*2.51*	3.546	*2.53*	49.147	*2.35*	7.599	*2.36*	46.251	*2.18*
N = 3,155												
R^2	0.260											
F-statistic	99.610		83.120		84.220		51.290		73.410		53.81	
Hausman test:			6.714		6.177		5.648		7.242		9.503	
(Prob > Chi^2, Degrees of Freedom)			(0.01, 1)		(0.01, 1)		(0.01, 2)		(0.03, 2)		(0.001, 3)	
Overidentification Test			33.370		33.920		34.510		32.650		23.58	
(Prob > Chi^2, Degrees of Freedom = 13)			0.000		0.000		0.000		0.000		−0.04	

F Test for Joint Significance of instruments using sample of positive wage earnings: 3.77 (Prob > F = 0.00, d.f. = 13)

Notes: OLS regression applied to the sample of women ages 18 to 54 years who reported age at menarche between 10 and 17 years old. Dependent variable is log of hourly wages of salaried and unsalaried workers using weekly wages as the base. Absolute t-values reported in italics. Standard Errors are calculated using robust (Huber, White, Sandwich) estimator of variance provided with STATA. There is no correction for sample selection because the inverse mills ratio is insignificant using available identifying variables. The instrumental variables are listed in Table 2.4

a/ Applies only to rural areas because distance is assumed to be 0 for urban areas.

b/ Oversampled states: Chiapas, Estado de México, Guanajuato, Guerrero, Hidalgo, Michoacán, Oaxaca, Puebla, and Veracruz.

quadratic are similar to column III without the interaction term. The coefficients on the interaction term are insignificant. Further, the coefficient on the exogenous years-of-education variable becomes insignificant. These results may reflect the need to treat both education and menarche as endogenous, as well as to use additional variables to measure health inputs. Neither of these techniques is feasible with the existing data set.

The results for the other variables in the regressions are consistent with human capital theory. The experience terms display increasing returns to scale in all of the regressions. The returns to education vary between 12 percent and 13 percent. The insignificance of the rural and the oversampled-state dummies is surprising given the results of the first stage of the regression. Still, this may be due to differential rural-to-urban migration or to the special characteristics of the sample of the NFPS. When the regressions are repeated on the urban and rural samples separately, the impact of menarche is stronger for the urban areas (not shown). Given that the urban areas and the more urbanized states are underrepresented in the sample, it is possible that these types of differences are understated. It is also possible that the other independent variables explain a large part of the rural–urban differences in wages.

The Hausman tests are generally significant and reject the hypothesis of exogeneity. This is a particularly strong finding given that Hausman tests may be indecisive if the OLS estimates are imprecise, or if the instrumental variables do not explain a significant portion of the variance in the endogenous variable of interest (Staiger and Stock, 1997; Murrugarra and Valdivia, 1999; Thomas and Strauss, 1997). Still, the overidentification tests reject the equality of coefficients, suggesting some misspecification of the instrumental variables.

Although the set of instruments is jointly significant even for the restricted sample of positive wage earners, it is important to highlight the limitations of these variables. First, the explanatory power of the overall regression is very low. While this is not unusual in estimates of health production functions (e.g., Schultz and Tansel, 1997), it is worrisome. Studies on the validity of instrumental variable estimation with weak instruments suggest that results may be biased (Bound et al., 1995). Furthermore the available instruments for this study are all indicators of community-level factors and refer to current conditions. In the absence of information on migratory histories, it is impossible to analyze how closely the result is correlated with the condi-

tions in the place where, and at the time when, the woman was growing up. Given these considerations, the robustness of the findings are tested by repeating the analysis using a variety of instrument sets, including state-level fixed effects, information on the mother for a subset of the women under study, and restriction of the sample to younger cohorts. The results of these robustness tests are summarized below.

The sign of the menarche variable is quite robust to varying the instrument set, although the magnitude of the impact of menarche on wages increases when the number of instruments is reduced. These results are presented for the double logarithmic specification in Table 2.6. Further, comparing the differences in the strength and magnitude of the coefficient underscores the unimportance of the personal health service variables (line 3) and the contrasting importance of the public service and housing variables (line 2) as determinants of age at menarche. This coincides with the findings from the estimates of the health production function presented in the previous section. Limiting the instruments to the four public-service and housing variables (line 2), the effect of a 1 percent increase in age at menarche increases to 6.4 percent. The effect is 6.5 percent if only earthen flooring and drainage facilities are used as instruments (line 5). Including personal health services and educational capital, the effect is 4.7 percent (line 10).

The results of the quadratic specification are also robust to varying the instrument set, although they are much more sensitive than in the case of the linear or logarithmic specification. Specifically, the coefficients lose individual significance with the exclusion of a large number of the instruments. They are robust in terms of significance and magnitude to the exclusion of any part of the instrument set other than the educational access and capital variables, as well as to dividing the sample between rural and urban areas.

To test the robustness of the instrumental variable wage equation, a full set of state dummies was added to the equations.[13] This latter specification provides a test for the validity of the instrumental variable instruments. Although the full set of dummies absorbs a substantial degree of the geographic variation that is not attributable to the accessibility of health services and to local levels of education, the coefficient on the menarche variable is stable in sign, magnitude, and significance.

[13] See Table 9 in Knaul (1999).

Table 2.6 Impact of a One-Percent Increase in Menarche on Hourly Wages by Instrument Set

Instrument Set	Coefficient of menarche	t-statistic
1. Complete set	–3.52	–2.44
2. Public Services and Housing Quality	–6.36	–3.25
3. Personal Health Services	–1.68	–0.36
4. Education Level and Educational Services	–5.63	–2.53
5. Public Services and Housing Quality (including only flooring and drainage)	–6.53	–3.17
6. Public Services and Housing Quality (including only flooring)	–8.42	–3.33
7. Public Services and Housing Quality (including only drainage)	–4.65	–2.18
8. Public Services and Housing Quality and Education Level and Educational Services	–3.75	–2.50
9. Public Services and Housing Quality and Personal Health Services	–5.61	–3.16
10. Personal Health and Education Level and Educational Services	–4.67	–2.35

Notes: The model is identical to column II of the instrumental variable estimates presented in Table 2.5, with the exception of the variation in the instrument set. The dependent variable is the natural logarithm of hourly wages. Menarche is also presented as the natural logarithm. Unless noted otherwise, all instruments in the given category are used.

To further test the strength of the model, both the analysis of the determinants of age at menarche and the wage equations were repeated for the sample of women who live with their mothers. For this small sample, it is possible to identify education of the mother and, for the further-reduced sample of those whose mother is aged between 15 and 54, to identify her age at menarche. While these are very select groups, the analysis provides additional insight into the importance of both family-level genetic and socioeconomic determinants of age at menarche. The sign and significance of the menarche variable is robust to this respecification.[14]

[14] These results are available from the author.

The analysis was also repeated for the restricted sample of younger cohorts. This provides a strategy for testing the sensitivity of the results to issues related to differential migration. In particular these regressions provide insight into the importance of using instruments based on current conditions at the community level, which are likely to differ from the situations experienced by the women under study when they were children and adolescents. Furthermore since the probability of migration increases over time, it is also more likely that younger cohorts reside in the place where they experienced puberty. The results of the wage regressions are very stable both to restricting the sample to women aged 44 and younger, as well as to a further reduction to include only women between the ages of 18 and 30. The signs, magnitudes, and levels of significance of the impact of menarche on wages are similar to the results for the complete sample. This is true for all five specifications of the instrumental variable wage regression and for both age groups. Considering the quadratic specification for example, the coefficients are –9.9 (t-statistic = 2.4) and .36 (t-statistic = 2.3), and –8.9 (t-statistic = 2.5) and .33 (t-statistic = 2.5) for women aged 18 to 44 and aged 18 to 30, respectively.

The findings of this section support the hypothesized relationship between investments in health and nutrition (as measured through age at menarche) and labor market productivity. The finding of higher wages among women who are younger at menarche is robust to the inclusion of a number of control variables as well as to changes in functional form and in the instrument set.

Conclusions

Age at menarche can be used to estimate labor market returns to childhood investments in health and nutrition. Measurement error combined with simultaneity, however, demonstrate the need to use instrumental variable techniques when estimating such wage functions.

The retrospective recall data available for this study show that average age at menarche has been decreasing in Mexico over the past 40 to 50 years. The decline has been somewhat slower than in developed countries. Factors associated with this decline include urbanization, increased levels of education, and improved living conditions. In particular, variables that measure access to public services and the quality of housing appear to have

an important impact. The proportion of the community with earthen flooring in their homes and the proportion that lack toilet or drainage facilities are particularly strong correlates of menarche. Access to personal health services appears to have little marginal impact on the age at menarche.

The findings reported in this chapter suggest that nutrition and cumulative health status, as measured by age at menarche, have a significant effect on the labor market productivity of Mexican women. Younger ages of menarche are associated with higher wages. The overall effect is masked in an OLS wage equation due to the errors of recall, rounding by year, and misreporting of the variable. The instrumental errors-in-variables model suggests that a decline of one year in menarche is associated with an increase of 26 percent in wages. This figure is consistent using a double logarithmic, semilogarithmic, or quadratic specification of the menarche variable.

Future research should use other data sets to include additional human capital inputs. It will also be interesting to broaden the conceptualization of female health by considering additional measures of health and nutrition. These should be compared and combined with age at menarche. Further, it will be important to include more information on the place of origin and migration patterns of the women in order to better identify the impact that health, education, and other public services, as well as poverty and living conditions, during infancy and childhood have on adult health outcomes.

The results of this chapter support the importance of investing in health and early nutrition, particularly through better sanitation and housing conditions, in order to improve individual and family well-being and to reduce poverty. Health has an important impact independent of education as an investment in human capital. Furthermore, the findings suggest that for the purposes of economic analysis, age at menarche should be considered a complement to adult height as a measure of secular changes in the health and nutritional condition of women.

CHAPTER 3

Elderly Health and Salaries in the Mexican Labor Market

Susan W. Parker[1]

This chapter investigates the determinants of health among the elderly in Mexico, and how health indicators are related to earnings in the labor market. In analyzing determinants, several different measures of health status are considered in order to gauge the impact on income. The results show that health indicators have a strong negative effect on wages for elderly male workers. The lowest estimates demonstrate that poor health reduces hourly earnings by 58 percent. These are sizable effects, particularly in a developing country that does not have a universal social security system, suggesting that many elderly individuals may continue working whether or not their health status makes it advisable to do so. Poor health also may prevent others from working and thereby contribute to high poverty rates among the elderly.

Introduction

One of the most important public policy issues in developed and developing countries is the aging of the population. Complex issues are involved that affect public policy, pension systems, and the demographics of the labor force. From the public policy perspective, government needs to understand how and why health costs will change. Pension systems must be assessed

[1] Dr. Susan W. Parker is Advisor to the National Coordinator of the Program for Education, Health, and Nutrition (PROGRESA), Secretary of Social Development, Mexico.

to see if the current structure will remain financially viable and whether pension levels will be sufficient to meet beneficiary needs. Ironically, viability may be unintentionally undermined if pension systems promote early retirement and create a shortage in the labor force generating revenue to support the system.

All these larger public policy questions require an understanding of how individuals behave as they confront the aging process. From the individual's perspective, aging may raise questions about health and its effect on daily activities, how to care for oneself in the event of illness, and how to pay for its costs. Decisions about when or whether to retire may hinge on the need to support one's family, one's physical ability to continue working, and how the pension system does or does not reward previous years of participation.

In developing countries and for the poor, these questions become even more pressing. Many developing countries have no social security systems, or limited coverage that applies only to workers in the formal sector and provides pension levels insufficient to finance retirement. The lack of security plus high rates of poverty imply that the elderly participate in the labor force at much higher rates and for longer periods of time than in developed countries, even though the population generally has poorer health and a much lower life expectancy (World Bank, 1994a).

In spite of the importance of these issues, there is a very small literature on elderly health, labor force participation, and retirement in developing countries. This chapter uses a recent data set—the Mexican National Aging Survey of 1994, which contains detailed self-reported indicators of health as well as labor market information—to evaluate how these factors may be interrelated and assess the impact on the income of elderly workers. This study applies recently developed models from the new literature that has developed on the importance of health as a human capital investment and therefore as an important determinant of wages and economic growth (Fogel, 1994). Empirical estimation of these models has focused on the possible endogeneity of health to productivity and wages (Schultz, 1997; Thomas and Strauss, 1997). They emphasize that health indicators may be endogenous and/or subject to measurement error, which would reduce the estimated impact of health on wages. This justifies the use of an instrumental variables technique to measure the effect of health on wages.

This chapter also puts substantial emphasis on the determinants of elderly health. There have been few studies on adult health status in developing countries, and it is not clear that studies on developed countries necessarily apply, given the differences in context. Existing health studies in developing countries have tended to focus on the health and nutrition of children. Nevertheless it has been shown that the relationship between child and adult mortality is not particularly close in many developing countries (Phillips et. al, 1993), which justifies the study of adult health on its own. Because the population share of older people is growing in most developing countries at a much faster pace than in developed countries, Smith (1994) comments that "aging and health are the emerging policy issues in the Third World."

Mexico provides an interesting case study for aging. While still a relatively young country, it is beginning the process of rapid aging. Whereas the population growth rate for children is effectively 0 percent and is now about 2 percent and declining for the working age population, the elderly population is growing at a rate of 4 percent annually. These trends imply that by the year 2030, the elderly population will quadruple in size.

The following sections explore what this will mean for Mexico. Discussion opens with a descriptive summary of labor force participation and health status of the elderly in Mexico. This is followed by an overview of some relevant literature on aging, health, wages, and labor supply. Third, the theoretical model and the data used for the analysis in this chapter are characterized more fully. Findings regarding the determinants of elderly health status are then presented, followed by the instrumental variable estimation of the impacts of elderly health on wages. The closing section discusses the implications of the results and offers suggestions for focusing future research.

Aging and Health in Mexico

This section examines recent trends in aging and health in Mexico, then briefly discusses the current state of the country's health systems. Table 3.1 shows the dramatic increases in life expectancy and declines in infant mortality that have occurred since 1950. Education levels and other indicators of development, such as the percentage of households with running water, have shown similar rises. The Mexican economy grew steadily between 1940

and 1980, with the gross national product more than tripling. The table suggests health and economic growth are highly correlated, although health conditions continued to improve in the 1980s despite low economic growth.

Table 3.2 shows life expectancy in Mexico since 1930. The climb in life expectancy during the past half century has been matched by steep declines in mortality rates. A man or a woman born in 1930 could expect to live 35.5 or 37 years, respectively (Gómez de Leon and Parker, 1998). By 1995 the figures had doubled. Most of the improvement can be attributed to the decline in mortality rates—from 26 deaths per 1,000 inhabitants in 1930 to 4.4 per 1,000 in 1995.[2]

Yet despite the substantial progress, these indicators still lag behind what might be expected given Mexico's level of per capita income. From that perspective Mexico fairs slightly worse in life expectancy than other Latin American countries, and Latin America itself fairs worse on average than other regions (Inter-American Development Bank, 1996).

While still a relatively young country, Mexico's elderly population is expected to grow at an increasing rate. Individuals 65 and older represented 4.2 percent of the population in 1990, a proportion expected to almost double to 7.3 percent by the year 2020 (INEGI, 1993).

Participation by elderly men in the labor market is relatively high in Mexico (43.5 percent of those over 65 worked in 1994 while only 15 percent of male seniors in the United States did). Although the rate for women is quite low, this is not surprising since female labor force participation in Mexico is far below the levels in more-developed countries.

As in many other countries, labor force participation by the elderly in Mexico has been decreasing over time. The rate for men aged 60 and over fell from 72.1 percent in 1970 to 53.3 percent in 1990. The rate for women fell from 12.6 percent to 6.7 percent[3] (INEGI, 1993).

---

[2] This is relevant to the analysis in this chapter since the individuals in the data sample were at least 60 years old at the time of the survey (that is, they were born in 1934 or earlier). This means that most of them have lived to twice the age of their life expectancy at birth, which indicates a strong sample selection effect. See Strauss et al. (1993) for an analysis of how selection by death—that is, the likelihood that the least healthy will die earlier—may affect the estimated determinants of health in a population.

[3] The fall in elderly female labor force participation is particularly notable since the percentage of workingwomen increased tremendously, rising from 17 percent to 30 percent between 1970 and 1993 (Gregory, 1986; and INEGI, 1993).

Table 3.1 Measures of Health and Well-Being in Mexico, 1950–1990

Indicator	1950	1960	1970	1980	1990
Infant mortality rate per 1,000 registered births	96.2	74.2	68.5	38.9	23.9
Life expectancy (years)	49.6	59.0	62.0	66.2	69.6
For men	48.1	57.6	60.0	63.2	66.4
For women	51.0	60.3	63.9	69.4	73.1
Literacy rate (%)[a]	56.8	66.5	76.3	83.0	87.6
For men	60.4	70.5	79.5	86.2	90.3
For women	53.4	62.7	73.1	79.9	85.0
Households with water (%)			61.0	70.7	79.4
Households with sewerage (%)			41.5	51.0	63.6
Households with electricity (%)			58.9	74.8	87.5
Per capita GNP (1980 dollars)	1,408	1,547	2,180	2,096	2,708

Sources: Secretaría de Salud, *Compendio histórico de estadísticas vitales, 1893–1993*; Nacional Financiera, *La economía Mexicana en Cifras 1990*; and INEGI, *Estadísticas Históricas 1993*.

a/ Figures for 1950–1970 include individuals 10 years and older; 1980–1990 figures include individuals 15 years and older.

As Table 3.1 demonstrates, overall health in Mexico has been improving sharply. Yet these improvements are not distributed equally between poorer and richer groups. Incidence of acute diseases is highest among the poorest subgroups of the population (Lozano et al., 1993), which tend to be rural; reside in shoddy, overcrowded dwellings; and have heads of households with little schooling. The main causes of death among the rural poor are infections and malnutrition, while injuries and chronic and degenerative diseases are the most common causes of death among the wealthier urban population (National Academy of Medicine, 1993).

The health care system in Mexico is oriented toward the public sector. Although the underlying philosophy is that the government should protect all individuals and households, coverage is provided through a patchwork of programs of uneven quality. Some public health care institutions are geared for people working outside the formal economy and the uninsured. These include the Ministry of Health (SSA), the National Institutes of Health, the

Table 3.2 Cross-Sectional Life Expectancy for Selected Years, 1930–2050

	Total	
Year	Male	Female
1930	35.5	37.0
1943	41.5	43.8
1956	53.4	56.6
1995	71.3	75.9
2000	73.1	77.6
2020	78.4	82.3
2050	82.0	85.4

Source: Gómez de Leon and Parker (1998).

Solidarity Program of the Social Security System (IMSS), the National System for Integral Family Development (DIF), and the Health Services of the Federal District Department (DDF). The public sector also runs several social security systems for those inside the formal economy. These include the IMSS, the Institute of Social Security and Services for State Workers (ISSSTE), the Armed Forces Social Security (ISSFAM), the Mexican Oil Workers social security system run by the national petroleum industry PEMEX, as well as other health services for state and federal government employees.

Despite the prominent role played by government, private services are available. The private sector includes a variety of medical practitioners and institutions providing a range of traditional and alternative medical services through mobile units, hospitals and clinics, and private practices. Expenses are covered by out-of-pocket payments from patients and by private health insurance plans. In 1995 almost half the population was covered by a public social security institution, 40 percent by institutions for the uninsured, 5 percent used private services, and 11 percent had no access to the health system's facilities (Secretaría de Salud, 1994).[4]

[4] *Editors note:* Yet these figures can be misleading since Mexicans covered by public insurance frequently use private services, partially explaining why some 40 percent of health spending is in the private sector.

Studies of Elderly Health and Productivity

Before discussing the specifics of this study of the impact of health on wages among the elderly in Mexico, it is useful to review other relevant studies that provided contextual information for assessing the challenge and shaping the analytical model and its application to the available data sets.

The Elderly in the Labor Market

Labor market participation among the elderly varies enormously by country and culture. Clark and Anker (1993) analyze the rates in 151 nations, concluding that participation by individuals 55 and older is much higher in developing countries, including Latin America, than in more-developed countries. The differences are particularly large for men, as might be expected given that some developing countries lack social security systems, and even those that have systems provide less-comprehensive coverage with generally lower pension levels, making it more likely that individuals will need to work later in life.

Few studies have analyzed the wage profiles of the elderly; in fact most studies of wages exclude the elderly from their design. An exception comes from Johnson and Neumark (1996) who examine the relationship between aging and wages for older men in the United States. Their research tests the human capital theory developed by Becker that postulates depreciation in human capital with advancing age, thereby resulting in declines in productivity and wages. The study showed wage declines seemed to begin for workers in their 60s, but the authors stress that this may be related to interactions with social security. That is, workers shift from full-time to part-time work when they start to receive benefits, and this results in lower reported wages. The authors emphasize that the sample of workers ineligible for social security demonstrates even weaker evidence that wages decline at older ages.

Posner (1995) emphasizes that productivity profiles for the elderly differ, depending on their occupations. Profiles vary across occupations by the age of peak earnings and whether or not that peak is sustained. For instance, occupations such as painting are characterized by early but sustained peaks, whereas corporate managers have late peaks that are not sustained. Posner claims that most studies of the issue of age and productivity find no age-related declines. A partial explanation is attributed to the fact that jobs usu-

ally tap only a portion of an individual's physical and mental capabilities. Therefore, Posner says, "it may be many years before the ... [worker's] ability to do his job declines to a point at which he either cannot do it at all or cannot do it without a costly (to him) increment of effort. Until that point is reached, he may be able to compensate for diminution in occupationally relevant capabilities with small increases in effort." He adds that the elderly are less likely to change jobs and may be more careful on the job since they are aware that leaving could be very costly in terms of accrued benefits (such as pensions) and that alternative employment is harder to find at older ages.

For elderly individuals, the decision to work is generally considered to be the same as a decision not to retire. Yet retirement is notoriously difficult to define and is likely to be a more ambiguous concept in a country such as Mexico where a very low percentage of the population receives a social security pension. In the elderly sample for this case study, for instance, only 12 percent reported receiving a pension, and a large fraction remain outside social security.[5] Additionally, a significant percentage of those receiving pensions (18 percent as compared with 30 percent of those without pensions) reported working in the previous week, indicating that retirement is not an all or nothing condition.

There is a large literature estimating the impacts of health on work and retirement decisions of the elderly in the United States and other developed countries, although fewer studies analyze the impact of health on wages. Most of these studies find that health status is a significant predictor of retirement. Many of the earlier studies assumed that health was exogenous to retirement decisions, and simply included a measure of individual health on the right-hand side of the model. More recent studies (Bound, 1992; and Stern, 1989) have considered health to be potentially endogenous to labor supply and have proposed corrective models. Studies have also discussed potential problems with self-assessed health indicators because individuals may be more likely to report health reasons as their motivation for retiring than other more-stigmatized reasons. Even worse, many self-assessed indicators of health are measured in terms of the ability to work, which clearly make them endogenous to a labor supply model.

[5] The recent reform in pensions at the Mexican Social Security Institute (IMSS) should eventually increase the percentage of individuals with pensions.

The theoretical impact of health on work and retirement decisions is, in general, ambiguous. Increases in health status may be expected to increase potential wage offers, but the income and substitution effects of this increase will work in opposite directions. Income effects will tend to reduce the amount of labor supply, while substitution effects will tend to increase it. Nevertheless, good health may have its own effect, independent of wages, which would be expected to increase the labor supply of individuals.

The focus in the analysis that follows will be more on the relationship between health and wages than on health and labor supply. The case study hypothesizes that sample selection may be an important factor in the accuracy of that analysis because the elderly who work may not be a representative sample of all elderly Mexicans. Consequently wage equation estimations would be biased unless a correction is included. Therefore the labor force participation decisions of the elderly are analyzed in order to correct for potential selection bias.

Measures of Health in Older Individuals

The success of the study depends critically on the extent to which the variables used to measure health status actually reflect the health of the individual. There is a fairly extensive epidemiological literature measuring health among the elderly population in developed countries, particularly in the United States. Much of it emphasizes Activities of Daily Living (ADL) as an indicator of health status. For example Dunlop et al. (1997) analyze measures of disability and physical functioning among the elderly in order to define a hierarchy of disabilities that set in with old age. The authors argue that a person's ability to perform basic tasks of daily living is an indicator of morbidity and a significant predictor of health service use. The study also concludes that while women live longer than men do, they spend more time disabled. Clark (1997) measures chronic disability in his study of whites and blacks in the United States, defining it as the inability to perform one of six activities of daily living for at least three months without assistance. While these indicators appear to be widely accepted in the United States and other developing countries as measures of elderly health, there is little evidence about their validity in developing countries.[6]

[6] An exception is Strauss et al. (1993).

Another set of indicators is derived by asking respondents to evaluate their own health. These indicators have in some cases been shown to be more accurate predictors of mortality than clinical examinations (Schultz and Tansel, 1997). In the literature on labor supply and retirement, substantial disagreement exists about whether self-reported health measures produce more-accurate estimates of the impact of health on labor supply than do more-objective measures of health (Bound, 1992). The main concern is that self-reported indicators of health may be biased if individuals who do not work are more likely to report health problems. This may occur if an individual feels that ill health is the only socially acceptable reason for being retired, or if the person believes some financial penalty may come from not declaring a current impairment when prior disability was used as the grounds to qualify for early retirement benefits.[7]

Schultz and Tansel (1997) propose an alternative measure of adult health within the context of two developing countries in Africa. Using the number of days disabled as an indicator of morbidity to estimate the impact of health on wages and labor supply, they report an important significant negative effect of health both on wages and labor supply.

All of these health indicators will be used here to sharpen the analysis by comparing how age results vary or not, depending on the choice of indicator. If all the health indicators show consistent results, it will suggest that some common element of the individual's health status is being measured.

Theoretical and Empirical Framework

The model that is used here incorporates health production and productivity in an integrated human capital framework following Schultz (1996) and Schultz and Tansel (1997). Cumulative health status is produced over the individual's lifetime and begins with parental and personal investments in nutrition, disease-preventing interventions and practices, and in health-conserving behaviors. These health inputs (*HI*) combine with heterogeneous endowments of the individual (*G*) that are unaffected by family or

[7] This may be less of a problem in the Mexican case since all of the health questions in the survey are asked under a separate section entitled "Health" and none are related explicitly to work behavior of the elderly.

individual behavior, to determine the individual's cumulative health status (h^*):

$$h^* = h^*\left(HI, G, e\right). \qquad [3.1]$$

Since health status is self-reported, it may differ from actual health status by a measurement error (ε), so that

$$H = h^* + \varepsilon \qquad [3.2]$$

where ε is assumed to be a random variable uncorrelated with other determinants of health.

Presumably, the individual maximizes a single-period utility function over a lifetime that includes health, the nonhealth-related consumption bundle, and annual time allocated to nonwage activities, subject to the budget, time, and health production constraints.

The individual's hourly wage (W_i) is a function of cumulative health status (h^*); other reproducible forms of human capital such as education, experience, and migration (C); the vector of exogenous variables (X) that are included additively; and other unobserved forms of human capital transfers and genetic endowments (y):

$$W_i = W_i\left(h^*, X, C, y\right). \qquad [3.3]$$

The econometric strategy addresses the possible endogeneity of health status to wages. The wage function is identified by the exclusion of community health variables (prices are not available), and the associated labor force participation equation by the exclusion of family wealth (proxied by characteristics of the home) and life cycle measures (number of living sons and daughters, and marital status).

The health production function in equation [3.1] cannot be estimated directly because many potentially relevant health inputs that have accumulated over the course of a lifetime are unavailable, as well as the prices of these inputs. Rather, reduced-form health equations are estimated with the health status measures as the first stage of our wage estimations, as follows:

$$H_i = g + h_j O_{ji} + r_k P_{ki} + t_i \qquad [3.4]$$

where O represents the vector of individual and family education, wealth, and resource opportunities and P represents the vector of k community health infrastructure variables for individual i.

The empirical specification of the wage equation is given as follows:

$$W_i = a + b_j H_{ij} + c_k X_{ki} + d_h C_{hi} + f_i \qquad [3.5]$$

where H represents health status indicators; X represents the vector of exogenous endowments such as age and sex, which are not modified by the individual or his/her family; and C represents the vector of reproducible forms of human capital, including years of schooling and migration, that can be increased by the investment of time and resources. As wages are only observed when the elderly individual participates in the labor market, the probability of participating is estimated with a probit model, which is then used to correct the wage equation [3.5].

There are at least two reasons why an instrumental variables approach to health status measures and wages is necessary. First, health for the elderly represents a lifetime of accumulated decisions and investments that are jointly determined with their productivity. It is likely that previous earnings and labor supply have affected the actual health status of the elderly. Second, the problem of inaccurate and incorrect answers, which is present in all surveys, may be even worse among the elderly, despite efforts at the beginning of the interview to establish the individual's capacity to answer questions.

Two community variables are used to identify the impact of health on wages in these estimations. The first variable is the number of hospital beds per capita in the municipality where the elderly respondent resides. This variable is expected to be positively related to health status. The second variable used as an instrument is the percentage of households in the person's community of residence that have an earthen floor. This variable is associated with poverty and living conditions that are expected to have a negative effect on health status.

Data and Definitions

The 1994 National Mexican Aging Survey provides the data for this analysis. This nationally representative data set carried out interviews of households

that included at least one member 60 years old or older. The questionnaire generated information about health, economic, and sociodemographic characteristics as well as support networks. The health information is particularly useful for this analysis since it permits construction of a number of different health indicators. The survey collected information on sick days, hospital days, and accident days and also asked questions based on the activities of daily living (ADL), self-reported health status measures (How would you rate your health? How does it compare to other individuals your age?), and disability measures. The total sample size was 5,159 individuals. Analysis was limited to those between 60 and 79 years old, or a total of 4,358 persons. Missing-data problems excluded 100 of these, leaving 4,258 individuals for the regression analysis sample.

A principal problem for the analysis of wages is that the income variable includes all sources, making it difficult to isolate wage income.[8] Fortunately the survey design asks respondents about their primary source of income for cost-of-living expenses, and then gives them the opportunity to list as many as four additional sources of supplementary income. This makes it possible to identify which workers have only labor income.[9]

For the analysis, three samples of individuals were considered. The first is limited to all individuals who reported working in the past week and defines their total income as their wage income, excluding all those who reported no labor market earnings. The second sample includes only those workers who reported labor market earnings as their primary source of income. The third includes only workers who reported labor market earnings as their sole income source.

All three samples suffer from potential bias, but the second sample is the most reasonable for analysis. Consequently the results from the second sample, comprised of workers who report wages as their principle source of income, are featured in the sections that follow. These results may, however, skew the results downward since the subsample is healthier than the sample of all workers, making it more likely that one will find a lower impact of

[8] The income question is phrased as follows: "¿Contando todas las formas de ingreso que tiene, me puede indicar por favor, en cuanto calcula sus ingresos mensuales?" (Including all sources of income, how much would you calculate your monthly income to be?).

[9] For more details, including the wording of particular questions and the characteristics of excluded observations, see Parker (1999).

health on wages. The results therefore should be interpreted as conservative estimates of the true effect.[10]

An additional problem is that the National Aging Survey reports income as a categorical variable (0 pesos, 0–500 pesos, 500–1000 pesos, etc.). To gauge hourly wages for all workers, then, the midpoints of the income categories are used to construct total earnings, which are then divided by hours worked in the previous week.

The survey has sufficient information to consider a number of different health indicators. After weighing the options, this research finally settled on three categories of indicators for health in the elderly population:

- *Disabled days.* The total number of disabled days in the sample is equal to the number of sick days, hospital days, and accident days during the previous 180 days. Thought was given to excluding hospital days since they may partly be determined by whether or not hospitals exist in the area where the individual lives, therefore making this measurement endogenous to the health indicators. It was decided to include this measurement, however, because of the very high correlation that exists among the three kinds of disabled days. Since most of the sample reports no disabled days, a dummy variable is used in the empirical analysis to represent whether or not disabled days were incurred.[11]
- *Self-reported ordinal indicators of health.* There are two such measures in the survey. The first measures how a respondent's health compares—on a scale of 1 to 5—to the health of other individuals the same age. The second, also measured on a scale from 1 to 5, indicates whether respondents consider their health to be very bad, bad, all right, good, or very good. Since correlation between these two variables is very high, only results from the first measure are included here. For ease of exposition in the descriptive analysis, a dummy variable set equal to 1 was also used for respondents who

[10] Estimates for the third sample can be found in Parker (1999, Appendix B). The conclusions are substantially the same, although the specific point estimates differ.
[11] It is important to note that most disabled-days indicators in other data sets are defined over a much shorter reference period (for instance two weeks or a month).

considered their health to be better or much better than individuals the same age, and 0 if otherwise. The five distinctive categories were retained in the regression analysis, however.

- *Functional limitations.* This variable ranges from 0 to 4, defined as the sum of the number of following activities that can only be performed with difficulty or not at all: walking up stairs; walking at least 300 meters; carrying a heavy object for 100 meters; or doing light domestic tasks such as washing dishes, sweeping, cooking, etc.[12]

The survey also includes information on migration. Respondents were asked how long they had lived in their present residence. More than 43 percent replied, "Always." This variable can be used to divide the sample into "movers" and "stayers." Migration is an important variable in this analysis for at least two reasons. First, it can be considered a type of human capital investment in and of itself. Second, it can be expected to affect some of the critical variables used in the analysis. For example, the health service supply variables for the place of current residence would be expected to be less relevant to the portion of the population that had migrated from somewhere else.[13]

In addition to the information available from the community segment of the National Aging Survey, this research uses two sources of municipality-level information. First is the *Indicadores Socioeconómicos e Indice de Marginación Municipal* (Socioeconomic Indicators and Index of Marginalization at the Municipal Level) generated by the National Population Council, or Consejo Nacional de Población (1993). The Consejo, also known as CONAPO, used results from the national population census of 1990 to develop an indicator of socioeconomic exclusion for each of Mexico's municipalities. This database records the percentage of inhabitants in each

[12] This classification follows the one in Davis et al. (1997), which in turn was based on the Nagi disability scale. In this classification, tasks are sorted in three groups of daily living activities (such as getting out of bed, getting dressed, etc.). See, also, Strauss et al. (1993) for similar health measures used in Jamaica.

[13] Unfortunately no information is available on where the elderly lived before their current residence.

municipality who are adult illiterates, are adults without a complete primary education, lack electricity in their homes, live in dwellings with earthen floors, lack toilet and drainage facilities, are without running water, live in overcrowded homes, reside in *localidades* with less than 5,000 people, and earn less than twice the minimum salary per month. The second source of information is a database jointly developed by researchers at the Colegio de México, CONAPO, and Johns Hopkins University. Based on records of the Secretariat of Health and the Mexican Social Security Institute, it includes data on doctors, clinics, and hospitals of the Mexican health system at the municipal level. The analysis in this chapter merges both of these data sets with municipal level results from the Aging Survey.[14]

Characteristics of Mexico's Elderly

Table 3.3 shows the labor force participation rates of the elderly. The first pair of columns for males and females measures overall labor force participation, whereas the second and third pairs represent subsamples of workers. The table clearly shows that men participate much more in the labor force than do women. A significant share of elderly Mexican men continue to work well into their 80s. More than 22 percent report that they are still working, much higher than comparable figures for the United States and other more-developed countries.

Only about one-third of elderly Mexicans report having been sick or disabled within the last 180 days (see Figure 3.1). Of these, the sample is fairly uniformly distributed between 1 and 180 days, although there is some bunching at 1–10 disabled days and at 180. The histogram suggests it may be inappropriate to assume that disabled days is a continuous variable. Thus in the estimations below, a dummy variable indicator will be used to measure disabled days. The other health variables show better-behaved distributions. All of the health status variables show that women tend to be less healthy than men.

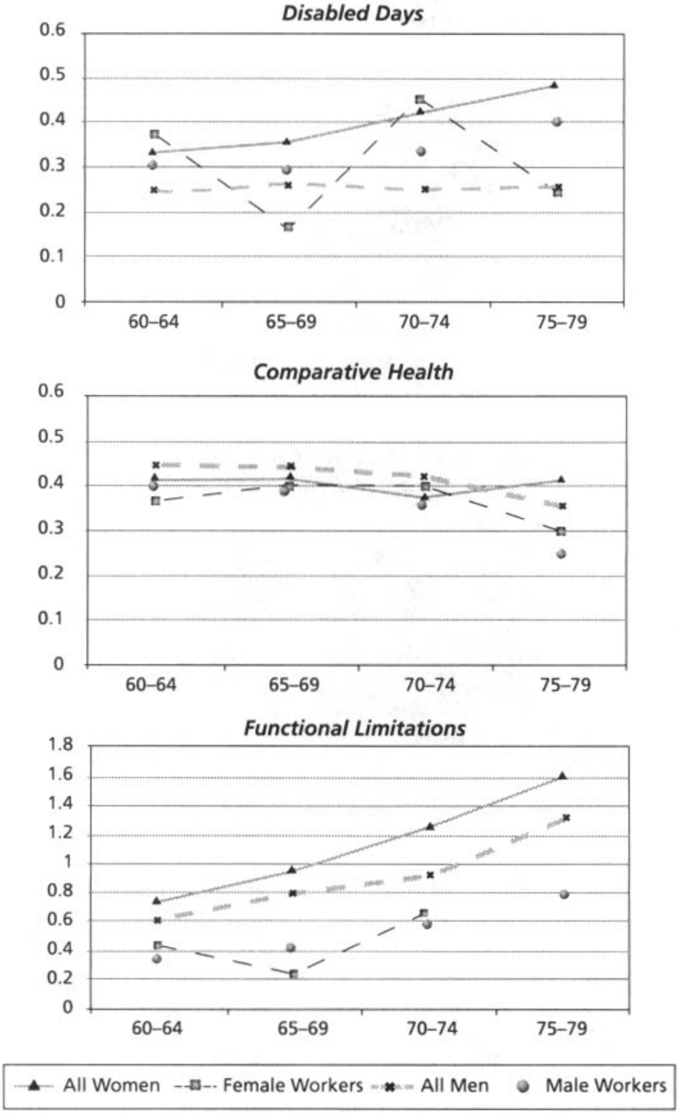
[14] Because of coding problems, it was only possible to identify about half of the rural municipalities in the Aging Survey. For the other half, community-level indicators were developed for each state using an average from all its municipalities in the sampling framework of the Aging Survey. For 685 of the 4,358 individuals, an average of several (ranging from 2 to 10) municipality characteristics was used for their community characteristics.

Figure 3.1 Health Status Measures of the Elderly in Mexico

Source: 1994 Mexican National Aging Survey.

Note: Possible functional limitations involve self-reported difficulty/inability in (1) walking upstairs, (2) walking 300 meters, (3) carrying a heavy object, or (4) performing light domestic tasks.

Figure 3.2 Health Indicators for Individuals Aged 60–79



Sources: 1994 Mexican National Aging Survey, and author's calculations. See Parker (1999) for more details in Tables 4 and 5.

Notes: The following definitions apply. *Disabled Days* refers to whether a disabled day was incurred in the last 180 days, including sick days, days spent in hospital, and days injured due to an accident. *Comparative Health* is equal to 1 if respondent reported his/her health was better or much better than the health of others one's own age (otherwise = 0). *Functional Limitations* record the number of functional limitations, from 0 to 4, including self-reported difficulty/inability in (a) walking upstairs, (b) walking 300 meters, (c) carrying a heavy object, or (d) performing light domestic tasks.

Table 3.3 Percent of Elderly Workers by Age, Gender, and Alternative Measures of Work Status

Age group	All elderly workers		Elderly workers whose primary source of income is labor earnings[a]		Elderly workers whose only source of income is labor earnings	
	Male	Female	Male	Female	Male	Female
60–64	63.6	14.0	54.2	10.7	41.0	7.1
65–69	57.6	13.0	45.2	10.0	28.7	6.9
70–74	44.5	11.3	34.4	8.8	23.4	5.4
75–79	40.2	3.4	27.9	1.5	19.0	0.7
80+	22.5	4.2	17.4	3.3	12.3	2.6
N	1,990	2,268	1,990	2,268	1,990	2,268

Source: 1994 Mexican National Aging Survey.

a/ Sample used for main estimation models.

Looking at health status among the elderly by age demonstrates that women, again, uniformly seem worse off than men in all cohorts (see Figure 3.2). As expected, all health status indicators worsen as the population ages. However, the figure shows that elderly *workers* display better health status than the overall elderly population since the rate of decline is lower. Health status appears to be worse among female than among male workers, but the very small sample size of working women makes any conclusion speculative.

Since a main contribution of this study is the evaluation of several different health indicators, it is of interest to know how correlated they are among themselves and how correlated they are with the other potential human capital indicators of education and migration. Higher correlation among the health status indicators would be reassuring since similar and consistent results could indicate they are measuring from different vantage points an "objective" and common health characteristic.

All three health indicators indeed are correlated significantly and have the expected signs (see Table 3.4). Education levels are also very highly correlated with the health indicators. Perhaps surprisingly, migration does not

Table 3.4 Correlation between Health and Other Human Capital Indicators

Indicator	Disabled days (1)	Better health (2)	Limits (3)	Migrated (4)	Educ. (5)
With disabled days (1)	1.000				
Better or much better health than others one's own age (2)	−0.253 (0.000)	1.000			
Number of functional limitations (3)	0.193 (0.000)	−0.167 (0.000)	1.000		
Migrated (4)	0.042 (0.117)	0.013 (0.659)	−0.004 (0.887)	1.000	
Education level (5)	−0.0802 (0.028)	0.203 (0.000)	−0.040 (0.167)	0.020 (0.481)	1.000
N	1,098	1,098	1,098	1,098	1,098

Note: Figures are correlation coefficients, with significance level in parentheses. See Figure 3.2 for sources, and definitions of disabled days, comparative health, and functional limitations.

appear to be especially correlated with health.[15] Despite the significant correlation among health measures, only one health indicator—the one based on whether individuals report they are in better or much better health than their peers—is significantly correlated with log wages (with a correlation coefficient of 0.15 and significant at the 0.1 percent level). Log wages are also significantly correlated with whether or not an individual has migrated to his or her current residence and with educational level.[16] Finally, the means and standard deviations of the independent variables used in the analysis can be found in Table 3.5.

[15] Migration is coded as to whether or not an individual has always lived in the current residence. Approximately 44.3 percent of the sample reported having always done so. Another 37 percent reported having lived in their current house for 10 years or more. Unfortunately there is no information on previous place of residence.

[16] The specific correlation measures and significance levels can be found in Table 7 of Parker (1999).

Table 3.5 Means of Variables for Individuals Aged 60–79 (Standard Deviations)

Variable	All elderly individuals		Elderly workers[a]	
	Men	Women	Men	Women
Individual characteristics				
Age 60–64	.367	.409	.455	.483
	(.482)	(.491)	(.498)	(.448)
Age 65–69	.275	.259	.284	.289
	(.446)	(.438)	(.451)	(.454)
Age 70–74	.224	.215	.175	.209
	(.417)	(.411)	(.381)	(.407)
No education	.299	.412	.295	.421
	(.458)	(.492)	(.456)	(.495)
Primary education	.564	.482	.295	.450
	(.496)	(.499)	(.456)	(.499)
Migrated (= 1)	.549	.563	.542	.516
	(.497)	(.496)	(.498)	(.501)
With savings (= 1)	.054	.042	.048	.061
	(.226)	(.197)	(.215)	(.241)
Number of sons living	2.84	2.66	2.76	1.99
	(2.10)	(2.09)	(2.11)	(1.91)
Number of daughters living	2.76	2.67	2.67	2.25
	(2.07)	(2.10)	(2.06)	(2.07)
Widowed	.144	.380	.144	.459
	(.351)	(.485)	(.351)	(.499)
Owns car	.122	.037	.164	.071
	(.328)	(.189)	(.371)	(.287)
Log wage (1994 pesos)[b]	—	—	2.77	2.65
			(1.01)	(1.02)
Household characteristics				
Urban–rural residence	.661	.703	.629	.706
(urban = 1)	(.473)	(.456)	(.483)	(.456)
Household dwelling has access	.731	.749	.731	.725
to running water	(.443)	(.433)	(.443)	(.447)
Community characteristics				
% of municipal households	20.9	19.5	22.4	19.9
with dirt flooring	(17.9)	(16.64)	(16.9)	(18.6)
Per capita hospital beds in municipality	.0014	.0014	.0013	.0013
	(.0011)	(.0011)	(.0014)	(.0013)
N	1,990	2,268	879	209

Source: 1994 Mexican National Aging Survey.

Note: Specific information for the 74–79 age group is not presented because of the small number of observations.

a/ Sample includes workers whose primary source of income is labor earnings.

b/ The 1994 exchange rate was approximately 3.5 pesos per dollar.

The Determinants of Health among Elderly Mexicans

This section evaluates the determinants of health, as measured by the different health indicators described previously. Although primarily intended for use in the second-stage regressions, these estimations are interesting per se. They shed light on which factors affect elderly health status and the extent to which their impacts differ by sex. They also shed some light on the effects of health policy variables, such as the supply of health services.

For self-reported health status and number of functional limitations, ordered-probit-regression and ordinary-least-squares estimates were performed. Ordinary least squares may not be appropriate for estimating ordinal health indicators since it assumes that the difference between ranks is identical (e.g., it assumes the difference between "bad" and "very bad" is identical to the difference between "bad" and "all right"). Ordered probit models are more appropriate for estimating the relationship between an ordinal (and ordered) dependent variable and other independent variables. Nevertheless, ordered probit estimation in this first stage complicates substantially our subsequent instrumental variable estimates. Ordinary-least-squares estimation is more computationally convenient.[17] For these two ordinal health indicators, the threshold point parameters from the ordered probit estimation were used to evaluate whether it was reasonable to use the linear specification based on the ranking of 1 to 5. Since they both appeared to be fairly linear, the 1-to-5 ranking was retained for computational considerations.

The main determinant variables included in the health status equations are age, education, and urban–rural residence of the individual, along with wealth measures, including whether the household dwelling has indoor running water and whether the individual reports having savings. Higher economic status is expected to have a positive impact on the health status of the elderly.

Disaggregation of the health status determinants by migration suggests that the effects of variables such as education and municipality measures differ for those who have moved to an area and those who were born

[17] The problem can be expressed as follows: $y^* = b + e$, where y^* is unobserved, and $y = 0$ if $y^* \leq 0$; $y = 1$ if $0 < y^* < m_1$; $y = 2$ if $m_1 < y^* < m_2$;...; $y = J$ if $m_{J-1} < y^*$. The threshold parameters (m_J) are estimated in the model (for fuller exposition see Greene, 1997).

and still live there. That is, many in the sample of elderly have moved from where they were born and grew up and from the environments that helped shape many of their human capital investments. Thus the characteristics of the current area of residence may be expected to have had less impact on their health than on the health of native residents.[18] To test this hypothesis, interactions were included of all variables whose impact could possibly be affected by migration.

Each table reports the joint significance tests for the identifying variables. This is critical to demonstrating whether the use of instrumental variable estimates is justified in the next stage of analysis. If the set of identifying variables is insignificant, the use of instrumental variables to adjust for potential endogeneity of health is not justified.

In examining the results, there appear to be some differences depending on which health status indicator is used, although some important similarities are retained. For the sample of men, all of the determinant health status regressions (Tables 3.6 to 3.8) show that, as expected, health status clearly declines with age. Education generally has an important impact, with higher education leading to better health status. The household wealth indicators for men (whether the dwelling has running water and whether the individual has savings) also show positive effects on health.

For women, the results differ substantially between health measures. It is interesting to note that while health status worsens with age according to the disabled-days indicator and the functional limitations measure, it does not worsen according to the self-reported indicator. Nevertheless for the rest of the independent variables, there are few consistent results. For the self-reported health variables, education is positively related to health status, as is living in an urban area. Wealth measured by whether the household has running water and whether the women have savings also positively affects self-reported health status. Yet the regression results for the determinants of the probability of disabled days and the functional limitations indicator show few significant variables apart from age.

[18] Of course even when individuals are still living where they grew up, the local conditions will have changed over time. Unfortunately, one cannot say much about these changes since information about the development of Mexico's social infrastructure is sparse. There has, however, been a historical tendency for health services of IMSS to be overly concentrated in urban areas, particularly in Mexico City (González and Parker, 1998).

Table 3.6 Determinants of Self-Reported Health Status for People Aged 60–79

Variable	OLS Regression Men	OLS Regression Women	Ordered Probit Regression Men	Ordered Probit Regression Women
Constant	2.96	3.13		
	(23.4)	(22.2)		
Age 60–64 (Coef. x 100)	17.7	0.100	22.3	–0.300
	(2.95)	(0.02)	(2.89)	(–0.05)
Age 65–69 (Coef. x 10)	1.24	–0.210	1.51	–0.290
	(1.98)	(–0.34)	(1.87)	(–0.36)
Age 70–74 (Coef. x 10)	1.19	–0.960	1.53	–1.23
	(1.85)	(–1.50)	(1.85)	(–1.50)
No education	–0.197	–0.164	–0.254	–0.213
	(–2.07)	(–1.66)	(–2.06)	(–1.68)
Primary education (Coef. x 10)	–0.220	–0.110	–0.250	–0.190
	(–0.27)	(–0.12)	(–0.23)	(–0.16)
Urban–rural residence (urban = 1)	0.277	0.237	0.363	0.301
	(4.02)	(3.51)	(4.08)	(3.47)
Per capita hospital beds[a]	44.4	3.45	59.2	7.07
	(2.14)	(0.16)	(2.21)	(0.26)
Dirt floors[a] (Coef. x 100)	0.100	0.048	0.100	0.058
	(0.58)	(0.24)	(0.55)	(0.23)
With savings (= 1)	0.187	0.154	0.245	0.203
	(2.29)	(1.73)	(2.25)	(1.77)
Indoor running water	0.106	0.088	0.137	0.114
	(2.27)	(1.98)	(2.28)	(2.01)
Migration (migrator = 1)	0.540	0.480	0.720	0.622
	(3.47)	(2.84)	(3.58)	(2.87)
Interaction of migration	–0.600	–0.600	–0.700	–0.800
with dirt floors[a] (Coef. x 100)	(–2.40)	(–2.52)	(–2.43)	(–2.48)
Interaction of migration with	–37.8	–45.2	–48.7	–58.3
per capita hospital beds[a]	(–1.35)	(–1.55)	(–1.34)	(–1.56)
Interaction of migration	–0.297	–0.201	–0.389	–0.253
with urban residence	(–3.00)	(–2.13)	(–3.05)	(–2.09)
Interaction of migration	–0.184	–0.142	–0.255	–0.197
with no education	(–1.47)	(–1.12)	(–1.58)	(–1.21)
Interaction of migration	–0.256	–0.161	–0.348	–0.219
with primary education	(–2.25)	(–1.33)	(–2.35)	(–1.40)

Table 3.6 (continued)

Variable	OLS Regression Men	OLS Regression Women	Ordered Probit Regression Men	Ordered Probit Regression Women
Threshold-1 Coef. (Std. Err.)			−1.60 (0.17)	−1.81 (0.19)
Threshold-2 Coef. (Std. Err.)			−0.631 (0.16)	−0.857 (0.18)
Threshold-3 Coef. (Std. Err.)			0.673 (0.16)	0.442 (0.18)
Threshold-4 Coef. (Std. Err.)			2.03 (0.16)	1.81 (0.18)
Sig. test of identifying variables	F(4, 1973) = 3.28	F(4, 2251) = 3.34	Chi² (4) = 13.83	Chi² (4) = 12.82
Sig. test of migration and migration interaction terms	F(6, 1973) = 2.83	F(6, 2251) = 1.59	Chi² (6) = 17.60	Chi² (2) = 9.51
N; R^2	1,990; 0.061	2,268; 0.046	1,990	2,268
Model statistic	F(16, 1973) = 8.0	F(16, 2251)= 6.8	Chi² (16) = 128.22	Chi² (16) = 107.52

Source: 1994 Mexican National Aging Survey.

Notes: The omitted education category is "more than primary schooling"; the omitted age category is ages 75–79. For the OLS regression, t-statistics are reported in parentheses; for the ordered probit, Z-statistics are reported in parentheses.

a/ Identifying variables for health in the wage equation.

The total effects of migration depend both on the migration dummy as well as the interaction of migration with other community variables. In the results here, the effects of migration on health vary, depending on the health status model. In the case of men, only in the model in which personal health is compared to that of other individuals does migration have a significant (positive) effect; whereas for women, the effect of migration is only significant in the model of disabled days.[19] In the remaining models, migration has no significant effect on health status.

Finally, the F tests of our identifying variables are generally significant, except for the disabled-days model in which the set of identifying variables is insignificant for women. Related to this, the health service indicator (hospi-

[19] The total marginal impact of migration on male health in the comparative health status model is 0.091, whereas for women in the disabled-days model it is 0.039. The total effects of migration on health are calculated by summing the marginal effects of migration and the other migration interaction terms, which are evaluated at the means of all the variables interacted with migration.

Table 3.7 Probability of Illness/Injury in Past 180 Days of People Aged 60–79

Independent Variables	Probit Regression	
	Men	Women
Age 60–64	–0.098	–0.131
	(–3.03)	(–3.95)
Age 65–69	–0.089	–0.114
	(–2.66)	(–3.29)
Age 70–74	–0.066	–0.050
	(–1.92)	(–1.39)
No education	0.095	0.030
	(1.65)	(0.52)
Primary education	0.087	0.004
	(1.70)	(0.07)
Urban-rural residence (urban = 1)	0.023	–0.084
	(0.60)	(–2.10)
Dirt floors[a]	0.002	0.0002
	(2.22)	(0.18)
Per capita hospital beds[a]	–20.1	–7.34
	(–1.64)	(–0.59)
Migration (migrator = 1)	–0.019	–0.152
	(–0.22)	(–1.52)
Interaction of migration with dirt floors[a] (Coef. x 10)	–0.0002	0.0009
	(–0.15)	(0.62)
Interaction of migration with per capita hospital beds[a]	23.0	27.07
	(1.42)	(1.57)
Interaction of migration with urban residence (Coef. x 10)	0.090	0.590
	(0.16)	(1.08)
Interaction of migration with no education	0.049	0.146
	(0.66)	(1.89)
Interaction of migration with primary education	0.003	0.105
	(0.05)	(1.41)
With savings (= 1)	–0.169	–0.003
	(–3.55)	(–0.06)
Indoor running water (Coef. x 10)	0.210	–0.080
	(0.83)	(–0.35)
Sig. test of identifying variables	Chi^2 (4) = 12.2	Chi^2 (4) = 3.67
Sig. test of migration and migration interaction terms	Chi^2 (6) = 4.83	Chi^2 (6) = 13.17
N	1,990	2,268
Model statistic	Chi^2 (16) = 58.92	Chi^2 (16) = 63.72

Source: 1994 Mexican National Aging Survey.

Notes: Omitted education category is "more than primary schooling"; omitted age category is ages 75–79. Z-statistics are in parentheses.

a/ Identifying variables for health in the wage equation.

tal beds per capita) seems to be a much more important and significant determinant of the level of health for men than for women. These results appear consistently in all of the health status equations. One can only speculate about the reasons for these gender differences. They may reflect differential access to or usage of health services. That is, elderly men may be more likely to use available services (e.g., they may be more likely to have social security health insurance than women). Or perhaps the quality of services offered differs between male and female patients. Nevertheless, the fact that the identifiers are significant (with the exception noted above) implies that we may cautiously proceed to the estimation of the full instrumental variable results.

The Impact of Health on Wages

One of the main concerns in estimating the impact of health on wages is the possible endogeneity of health. In this section, wage estimations that consider health status to be endogenous are compared with wage estimations in which health status is exogenous.

Sample selection bias is also addressed. Given the low labor force participation rates of the elderly, it is reasonable to hypothesize that the sample of elderly workers is not necessarily representative of those who do not work. For instance, the elderly who work may be those who are most able to do so, and therefore the most productive. In such a case, the main impact of health may be to permit people to enter the labor market and find employment, rather than affect their wages directly. On the other hand, if elderly labor market participation is largely determined by economic need (due to lack of other sources of income), one may find that the sample of elderly workers is less productive than is the population of those who do not work. Presumably, being needy is associated with low education and other factors that may reduce productivity.

There is an additional restriction in the selection model due to the decision to use only workers whose primary source of income is through labor market earnings (rather than all workers). Therefore a selection correction is made for both being in the labor force and having earned income as the primary source of income.[20]

[20] The level of health would be expected to have strong positive effects on the labor supply of elderly workers (and in probit models of working, in which health is assumed to be exog-

Table 3.8 OLS and Ordered Probit of Number of Functional Limitations for Individuals Aged 60–79

Variable	OLS Regression[a]		Ordered Probit Regression[b]	
	Men	Women	Men	Women
Constant	1.17	1.34		
	(6.61)	(6.36)		
Age 60–64	–0.733	–0.839	–0.649	–0.699
	(–8.69)	(–9.56)	(–7.87)	(–8.94)
Age 65–69	–0.516	–0.623	–0.440	–0.501
	(–5.87)	(–6.71)	(–5.17)	(–6.08)
Age 70–74	–0.394	–0.350	–0.289	–0.246
	(–4.33)	(–3.66)	(–3.33)	(–2.95)
No education	2.71	0.189	2.17	–0.304
(Coef. x 10)	(2.01)	(0.13)	(1.62)	(–0.23)
Primary education	0.910	–0.043	0.490	–0.830
(Coef. x 10)	(0.76)	(–0.31)	(0.40)	(–0.66)
Urban residence	2.02	1.07	2.14	0.820
(Coef. x 10)	(2.08)	(1.06)	(2.23)	(0.89)
Dirt floors[a]	0.500	0.800	0.600	0.670
(Coef. x 100)	(2.14)	(2.89)	(2.44)	(2.54)
Per capita hospital beds[c]	–48.0	–20.9	–52.8	–23.2
	(–1.65)	(–0.67)	(–1.77)	(–0.80)
Migration	0.273	–0.182	0.239	–0.212
	(1.25)	(–0.72)	(1.08)	(–1.04)
Interaction of migration	–0.500	–0.0001	–0.500	0.0003
with dirt floors[a] (Coef. x 100)	(–1.58)	(–0.04)	(–1.66)	(0.10)
Interaction of migration with	–3.79	42.0	–19.5	38.7
per capita hospital beds[c]	(–0.09)	(0.984)	(–0.47)	(0.96)
Interaction of migration with	–23.6	–1.90	–20.6	–0.200
urban residence (Coef. x 100)	(–1.69)	(–0.14)	(–1.48)	(–0.17)
Interaction of migration	–0.242	0.245	–0.193	0.257
with no education	(–1.37)	(1.30)	(–1.09)	(1.49)
Interaction of migration with	–0.700	17.3	1.90	21.8
primary education (Coef. x 100)	(–0.05)	(0.95)	(0.12)	(1.30)
With savings	–0.199	–0.029	–0.015	–0.015
(Coef. x 10)	(–1.69)	(–0.22)	(–1.27)	(–0.13)
Indoor running water	–1.60	0.210	–1.250	0.305
(Coef. x 10)	(–2.44)	(0.32)	(–1.93)	(0.51)

Table 3.8 (continued)

	OLS Regression[a]		Ordered Probit Regression[b]	
	Men	Women	Men	Women
	Threshold Tests			
Threshold 1 (Std. Err.)			–0.086 (0.18)	–.284 (0.19)
Threshold 2 (Std. Err.)			0.567 (0.18)	0.290 (0.19)
Threshold 3 (Std. Err.)			0.935 (0.18)	0.711 (0.19)
Threshold 4 (Std. Err.)			1.27 (0.18)	1.17 (0.19)
	Significance Tests			
Sig. test of identifying variables	F (4, 1973) = 2.90	F (4, 2251) = 4.80	Chi^2 (4) = 16.3	Chi^2 (4) = 17.0
Sig. test of migration and migration interaction terms	F (6,1973) = 1.65	F (6,2251) = 0.53	Chi^2 (6) = 9.90	Chi^2 (6) = 3.33
N, R^2	1,990; 0.0570	2,268; 0.0625	1,990	2,268
Model statistic	F (16,1973) = 7.54	F (16,2251) = 9.47	Chi^2 (16) = 105.5	Chi^2 (16) = 132.8

Source: 1994 Mexican National Aging Survey.

Notes: Maximum number of functional limitations is 4, defined as having difficulty with 1) walking up stairs, 2) walking 300 meters, 3) carrying a heavy object 100 meters, and 4) performing light domestic tasks. Omitted education category is "more than primary schooling"; omitted age category is ages 75–79.

a/ T-statistics are reported in parentheses.

b/ Z-statistics are reported in parentheses.

c/ Identifying variables of health for wage equation.

To test for the possible impact of sample selection bias, Heckman (1979) sample selection models are estimated, using the number of still-living sons and daughters and whether the individual has been widowed. Given the Mexican custom of family support (and the general lack of governmental welfare programs such as unemployment insurance), it is hypothesized that the number of living children will be an indicator of likely income transfers

enous, the effects are large and significant). Nevertheless it is also likely to be an endogenous variable to labor supply, and it is beyond the scope of this chapter to estimate a model of labor supply and wages with endogenous health measures. For this reason, health has not been included as an independent variable in the probit participation equation.

and therefore be negatively related to the probability of elderly parental participation in the labor market. Being widowed may imply fewer dependents to support with labor market income, or it may have the opposite effect and imply an increased need to work given the absence of spousal support.

The results of the probit model of labor force participation are reported in Table 3.9. As expected, the table shows that older individuals are less likely to be working. The education variables show no impact on the participation of women, whereas for men, those with lower levels of education are less likely to be working than are higher-educated individuals. Men in rural areas are more likely to be working, whereas there is no impact of residence on female labor force participation.

Turning to the identifying variables, being a widower reduces the probability of working for men, but widowhood increases the probability for women. This difference may occur because for a man to lose his spouse might imply a reduction in financial dependents, whereas for women, who are not traditionally the main source of family income in Mexico, to lose a spouse means needing to find new means of support.

It is interesting to note that the number of children, both males and females, has a negative effect only on the probability of women's labor force participation, whereas there is no significant effect for males. Additionally, the negative effect of having sons is much higher on women's labor force participation than is the effect of having daughters. If sons tend to give more monetary support to their mothers than do daughters, this would be consistent with the lower rates of labor force participation among women in general. That is, women in Mexico may be less able to transfer resources to their parents since they are less likely to be working and earning their own disposable incomes. Finally, greater wealth, as proxied by the existence of indoor running water, is negatively related to the probability of men working. The effect for women, however, is insignificant.

Table 3.10 contains results from regressions for men, using all three health status measures and treating health as either exogenous or instrumented. The coefficients on health in the OLS equations that treat health exogenously are significant, with the exception of the functional limitation measure. By contrast, the instrumental variation (IV) estimation models show that the impacts of health on wages are much larger and much more significant for all three of the health measures. The table also reports the Hausman tests of exogeneity in health, showing exogeneity to be strongly refuted in all cases.

Table 3.9 Probit Estimates of Probability of Working for Individuals Aged 60–79

Variable	Men[a]	Women[a]
Constant	0.023	–1.93
	(0.171)	(–7.76)
Age 60–64	0.655	1.08
	(6.83)	(5.07)
Age 65–69	0.447	1.02
	(4.51)	(4.73)
Age 70–74	0.170	0.883
	(1.65)	(4.03)
No education	–0.028	–0.054
	(–2.88)	(–0.42)
Primary education	–0.317	–0.090
	(–3.62)	(–0.73)
Urban–rural residence (urban = 1)	–0.157	–0.027
	(–2.30)	(–0.30)
Migrated (= 1)	–0.013	–0.069
	(–0.22)	(–0.89)
Number of sons living[b]	–0.004	–0.085
	(–0.269)	(–3.90)
Number of daughters living[b]	–0.016	–0.034
	(–1.07)	(–1.71)
Widowed[b]	–0.197	0.274
	(–2.33)	(3.49)
Indoor running water (Coef. x 10)	–1.83	–0.069
	(–2.73)	(–0.77)
Chi^2 (12)	117.0	76.2
N	1,990	2,268

Source: 1994 Mexican National Aging Survey.

Notes: Sample is limited to workers whose principal earnings are from labor income. Omitted education category is "more than primary schooling"; omitted age category is ages 75–79.

a/ Z-statistics are in parentheses.

b/ Identifying variables for wage equation.

Table 3.10 Selection-Corrected Wage Equations with Exogenous and Instrumental Variable Estimates of Health Measures (Men Aged 60–79)

	Disabled days		Health compared to others		Number of functional limitations	
Variable	Exogenous	IV estimates	Exogenous	IV estimates	Exogenous	IV estimates
Constant	1.89	2.77	1.61	–.0.611	1.89	2.96
	(3.35)	(3.15)	(2.87)	(–0.65)	(3.32)	(4.26)
Age 60–64	0.463	0.284	0.419	–0.37	0.452	–0.145
	(1.54)	(0.66)	(1.43)	(-0.11)	(1.50)	(–0.39)
Age 65–69	0.370	0.275	0.333	–0.038	0.361	–0.098
	(1.57)	(0.82)	(1.44)	(–0.14)	(1.52)	(–0.33)
Age 70–74	0.136	0.035	0.117	-0.093	0.134	–0.084
	(0.793)	(0.14)	(0.69)	(–0.45)	(0.78)	(–0.39)
No education	–0.987	–0.702	–0.945	–0.460	–0.996	–0.913
	(–6.43)	(–2.92)	(–5.60)	(–2.18)	(–6.48)	(–6.19)
Primary education	–0.818	–0.493	–0.797	–0.489	–0.828	–0.733
	(–5.13)	(–1.93)	(–5.11)	(–3.85)	(–5.20)	(–4.66)
Urban–rural residence	0.142	0.073	0.127	–0.030	0.148	0.262
	(1.21)	(0.44)	(1.11)	(–0.23)	(1.26)	(2.15)
Migrated	0.154	0.301	0.152	0.188	0.147	0.132
	(1.85)	(2.26)	(1.87)	(2.30)	(1.77)	(1.63)

Table 3.10 (continued)

	Disabled days		Health compared to others		Number of functional limitations	
Variable	**Exogenous**	**IV estimates**	**Exogenous**	**IV estimates**	**Exogenous**	**IV estimates**
Disabled days	**−0.126**	**−3.29**				
	(−1.69)	**(−2.44)**				
Health compared to others one's own age			**0.099**	**0.998**		
			(2.21)	**(3.21)**		
Number of functional limitations					**−0.023**	**−0.867**
					(−0.58)	**(−2.65)**
Lambda	1.31	0.998	1.23	0.353	1.32	1.44
	(2.31)	(1.21)	(2.21)	(0.54)	(2.30)	(2.60)
N	879	879	879	879	879	879
		(0.0000)		(0.0001)		(0.0000)
Hausman test of exogeneity[a]		5.82		3.93		5.06
Model statistic	F(9,869) = 14.0	F(9,869) = 14.3	F(9,869) = 14.4	F(9,869) = 14.0	F(9,869) = 13.7	F(9,869) = 13.9

Source: 1994 Mexican National Aging Survey.

Notes: The sample is limited to workers whose principal earnings are from labor income. Omitted education category is "more than primary schooling"; omitted age category is ages 75–79. T-statistics are in parentheses.

a/ Hausman test is absolute value of t-statistic on coefficient of residual health status measure when actual value and residuals are included in the wage equation.

The impact of education on wages appears to be significantly reduced when IV estimations are done for disabled days and comparative health status, relative to when health is assumed to be exogenous. In both the comparative health status model and the disabled-days models, the effects of education on wages are reduced as much as 50 percent in some cases, although there is no significant change in the functional limitations regression. Overall, these results may suggest that estimations of the effect of education on wages that do not take account of health may overestimate the returns to education.

The impacts of health on wages for elderly men implied by the estimations are quite large. The coefficient on disabled days suggests that poor health, as measured by workdays lost during the previous 180 days, reduces hourly wages by 3.29 log points or 96 percent.[21] The functional limitations health measure indicates that having an additional functional limitation reduces salaries by 0.867 log points or 58 percent. Finally, while perhaps more difficult to quantify, improving one's health relative to others (for instance from a similar to a better health level) is associated with receiving a wage that is 0.998 log points or 172 percent higher, controlling for observable characteristics.

All of these estimated effects are large and should be treated as upper-bound estimates of the impact of health. A more conservative estimate of the impact of health can be derived from the lower-bound 95 percent confidence interval estimates. These would imply that poor health in the case of disabled days is associated with a wage reduction of 40 percent, and in the case of functional limitations, with a reduction of 26.9 percent. Good health for the comparative health measure, meanwhile, is associated with an increase in wages of 58.2 percent. Clearly, even extremely conservative estimates demonstrate that health has large effects on wages.

The other variables have the expected effects. Education is positively related to wages, as is urban residence. Migration has an important signifi-

[21] The estimated impact of disabled days on wages is improbably large. No other study using this indicator as a dummy variable was available for comparison. Schultz and Tansel (1997) found that one disabled day is associated with as much as 33 percent and 26 percent reductions in hourly wages in Côte d'Ivoire and Ghana, respectively, although the magnitudes decreased as disabled days increased. In the sample for this chapter, workers who incurred disabled days reported an average number of approximately 22.

cant and positive impact on wage levels both for men and women. The impact is perhaps surprisingly strong since many of these individuals may have migrated decades earlier. One interpretation is that the migration variable may be a proxy for greater investments in human capital over the individual's entire lifetime, which are not adequately captured with age or education.[22]

Finally, the sample selection correction coefficients show ambiguous results, with generally positive significant effects in the exogenous health wage equations and generally insignificant effects in the endogenous health equations. It is important to note, however, that the sample selection coefficients for men are extremely sensitive to the inclusion and exclusion of some variables, such as that of running water, so the results on sample selection bias should be evaluated cautiously. For comparison, the results with no sample selection correction can be found in Parker (1999, Appendix).

In contrast to the findings for men, the findings for women are disappointing (see Table 3.11). There is virtually no measured impact of health on women's wages. This may be due to several factors. First, we have a very small sample subset since female elderly labor force participation is less than 10 percent. Moreover the participation rate has never been large throughout their lifecycle.[23] It is to be hoped that future data on the well-being of the elderly will include larger samples so that the important topic of health, aging, and women's earnings can be studied. Second, health identifiers in this study were weaker at explaining health status for females than for males. Further study is needed to find why health service indicators and overall development seem to have less impact on the health of elderly females than on elderly males.

A final aspect deserving further comment is that wages do not appear to be declining by age in our sample. This could be due to bias in our income variable if, for instance, older workers are more likely to receive transfers from other sources that are contaminating our income measures. Nevertheless, the results from the sample limited to workers whose sole source of income is from labor also show similar relationships between age and salaries (Parker, 1999, Appendix). Additionally, this analysis was compared

[22] I am grateful to T. Paul Schultz for suggesting this interpretation.

[23] For instance, the labor force participation rate in 1950 of women in Mexico was approximately 12 percent (Gregory, 1986).

Table 3.11 Selection-Corrected Wage Equations with Exogenous and Instrumental Variable Health Measures (Women Aged 60–79)

Variable	Disabled days		Health compared to others		Number of functional limitations	
	Exogenous	IV estimates	Exogenous	IV estimates	Exogenous	IV estimates
Constant	3.61	3.60	3.11	2.37	3.59	3.61
	(4.10)	(4.21)	(3.40)	(1.24)	(4.11)	(4.16)
Age 60–64	–0.506	–0.553	–0.514	–0.518	–0.556	–0.498
	(–0.97)	(–1.09)	(–0.98)	(–1.03)	(–1.06)	(–0.97)
Age 65–69	–0.548	–0.532	–0.564	–0.589	–0.563	–0.541
	(–1.05)	(–1.14)	(–1.08)	(–1.15)	(–1.08)	(–1.10)
Age 70–74	–0.374	–0.436	–0.372	–0.359	–0.467	–0.355
	(–0.74)	(–0.88)	(–0.74)	(–0.74)	(–0.92)	(–0.69)
No education	–1.34	–1.36	–1.33	–1.31	–1.36	–1.34
	(–6.50)	(–5.34)	(–6.53)	(–5.77)	(–6.65)	(–5.95)
Primary education	–1.12	–1.16	–1.13	–1.13	–1.16	–1.11
	(–5.50)	(–4.00)	(–5.60)	(–4.81)	(–5.74)	(–4.64)
Urban–rural residence	0.316	0.335	0.287	0.242	0.324	0.316
	(2.20)	(2.02)	(2.00)	(1.36)	(2.27)	(2.32)
Migrator	0.263	0.234	0.286	0.325	0.237	0.267
	(1.99)	(1.25)	(2.18)	(2.04)	(1.81)	(2.00)
Disabled days	**–0.027**	**0.245**				
	(–0.20)	**(–0.23)**				

Table 3.11 (continued)

	Disabled days		Health compared to others		Number of functional limitations	
Variable	**Exogenous**	**IV estimates**	**Exogenous**	**IV estimates**	**Exogenous**	**IV estimates**
Health compared to others one's own age				**0.131**	**0.326**	
			(1.83)	**(0.69)**		
Number of functional limitations					**0.141**	**−0.041**
					(1.76)	**(−0.16)**
Lambda	0.147	−0.221	0.184	0.242	0.162	0.173
	(0.45)	(−1.08)	(0.57)	(0.67)	(0.504)	(0.50)
N	209	209	209	209	209	209
Hausman test of exogeneity[a]		4.51		4.21		1.71
		(0.00)		(0.00)		(0.0984)
Model statistic	F(9,199) = 7.22	F(9,199) = 6.67	F(9,199) = 7.69	F(9,199) = 6.65	F(9,199) = 7.65	F(9,199) = 6.44

Source: 1994 Mexican National Aging Survey.

Notes: Sample is limited to workers whose principal income is labor market earnings. Omitted education category is "more than primary schooling"; omitted age category is ages 75–79. T-statistics are in parentheses.

a/ Hausman test is absolute value of t-statistic on coefficient of residual of health status measure when actual value and residuals are included in the wage equation.

to the trends in wages of the elderly population in the Mexican National Employment Survey of 1995. This data set also showed no decline in wages between the ages of 60 and 80, the age group used in the analysis for this chapter.

Conclusions

This is one of the first studies to explore the relationship between health and wages among the elderly population of a developing country. It finds that a number of health measures have an effect on wages for male elderly workers. Although there is some variability in the results depending on the health measure used, this finding is consistent and strong. The point estimates demonstrate that poor health lowers the hourly earnings of elderly males by no less than 58 percent; and even with more-conservative 95 percent confidence intervals, the lowest estimated effect of poor health is 27 percent. These are important factors, particularly within the context of a developing country that lacks a widespread social security system and in which many elderly individuals are therefore required to work whether or not they are healthy. Health problems may also, of course, prevent poor people from working and contribute to high poverty rates among the elderly. Future research should more explicitly incorporate the impact of poor health on the work behavior of the elderly.

The most important econometric implications of this study are that health cannot be treated as an exogenous factor in influencing wages and that the measured impact of health on wages increases tremendously when proper corrections are introduced. The Hausman tests uniformly reject the hypothesis of exogeneity of health to wages for the elderly, confirming the appropriateness of using an instrumental variable estimation approach. It is also important to mention that the education coefficient tends to decrease in the instrumental variable specifications in two of the three health models used. This implies that the returns to education may be overestimated when health is not included in wage analyses. Those returns may include impacts that should actually be attributed to health.

In setting priorities for future research, the relationship between available health services and the health status of the elderly warrants closer attention, as do the overall determinants of health status among the elderly. Mexico is undergoing a number of important health reforms within its health

sector. Given the extent to which the elderly population in Mexico will grow in coming decades, further research on health and the elderly is crucial if effective policies are to be designed to assess their health needs and dedicate sufficient resources to meet them.

It is also likely that poverty and elderly health status are closely linked and that these relationships affect labor force decisions and salary levels. Mexico's social security retirement system does not yet have universal coverage of the elderly, a pattern that can be expected to continue due to the country's large informal sector. This makes it all the more urgent to better understand how poverty and health interact to determine labor force participation by the elderly and the salaries they receive.

This page intentionally left blank

CHAPTER 4

Adult Morbidity, Height, and Earnings in Colombia

Rocio Ribero and Jairo Nuñez[1]

In this chapter, both public and private investments in health in Colombia are analyzed in relation to future earnings of individuals. The magnitude of labor market returns from good health status, as identified through instrumental variables, shows that each additional day of disability decreases earnings between 13 percent and 33 percent, and that each additional centimeter of stature increases earnings by about 7 or 8 percent. The returns to height are larger than those found in some other countries and suggest that investments in nutrition may be as important as investments in education for future increases in productivity and growth. Estimations of health production functions show that it would be desirable to increase social security coverage in rural areas in order to lower the incidence/duration of illness there. However in urban areas, where coverage is more developed, the degree of benefit is less certain because improvements in the social security system may only increase the tendency to report illness. Policies to improve housing appear to translate into better health status, productivity, and earnings for individuals. There is also some evidence that expanded access to social security in rural areas will have a significant and positive effect on both health status and earnings.

[1] Dr. Rocio Ribero is a Professor at the Universidad de los Andes in Bogotá and a Researcher at that university's Centro de Estudios de Desarrollo Económico. She is also a Postdoctoral Fellow at the Economic Growth Center at Yale University. Jairo Nuñez is a Researcher in the Colombian Departamento Nacional de Planeación.

Introduction

This chapter examines how public and private health investments affect the earnings of individuals in Colombia. It is the first study to analyze the links between primary health indicators and individual labor productivity in the country, and to look at how additional public expenditures might improve the health status of Colombians.

It is well established that investing in education is associated with higher productivity and earnings. Such investment is considered a form of "human capital" (Becker, 1964). But human capital has many forms besides schooling. Investments in health also affect individual productivity. Such investments include personal decisions about resource allocation and consumption that determine the nutritional intake and health of children and adults within the household. They also include public investments in health services and environmental improvements.

Research in recent decades has expanded the framework that Mincer (1974) devised to analyze the impact of schooling on wages, by including additional human capital determinants such as migration, fertility decisions, and health status. For example Schultz (1997) analyzes how state and family investments influence the formation of reproducible human capital and how these in turn affect labor earnings and growth. He asks two primary questions: What determines the household demand for human capital, and what are the wage returns of human capital stock in the labor market?

Schultz (1996) finds that health status, as measured by adult height, is an important determinant of adult productivity. Height is a measure of adult health status that reflects cumulative health investments beginning with prenatal care and continuing through the conditions of early childhood and the nutritional intake provided by parents as a child matures (Strauss and Thomas, 1995; Martorell and Habicht, 1986). Adult height has been shown to be inversely correlated with chronic health problems among the middle-aged and elderly. Moreover, studies show that height is inversely related to mortality and consequently directly related to the length of productive life. Fogel (1994) also finds that height and the body mass index (BMI) are related to male mortality at older ages, and to chronic diseases for those between 20 and 50 years of age. The increasing average height of populations over time has been well documented and may be attributed to changes in reproducible human capital investments or in disease environments (Fogel, 1994).

In line with Schultz (1996), Thomas and Strauss (1997) used Brazilian urban household data containing height and BMI, and found that height has a large and significant effect on wages for males and females. This chapter confirms that height is positively related to individual earnings in Colombia. It then explores this relationship in order to identify policies that can improve health outcomes and promote economic growth efficiently and equitably.

The analysis that follows proceeds in four stages. First, the data sources and the main characteristics of the surveys that were used are described, including the indicators that were chosen to measure health status. Second, the sample characteristics are discussed. Third, results of econometric analyses and policy simulations are presented and explained. Finally, conclusions and suggestions for future action are offered.

The Data

Two major surveys conducted by the Departamento Administrativo Nacional de Estadística (DANE) were used for this study. Other sources were consulted to obtain regional data on environmental factors. That data was merged with the individual household survey data so that each individual was linked to the characteristics of his or her community.

The first survey is the 1993 Encuesta de Caracterización Socioeconómica (CASEN), which was national in scope and includes specific modules on health, education, and child mortality. Interviews were conducted in a total of 27,271 households—22,257 urban and 5,014 rural. The sample includes 33,395 individuals between 18 and 70 years old with positive wages or earnings, of which 64 percent are male and 74 percent live in urban areas.

From this survey two indicators of health were used: disability and days disabled. Disability was measured as a dichotomous variable, equal to 1 if the respondent reports that he or she was unable to work in the month before the survey because of an illness. Specifically, respondents were asked, "During the last month did you not go to work or not do your ordinary activities because of the illness or health problem mentioned above [in the preceding question]?"[2] The answer recorded was either "yes" or "no."

[2] The question asked in Spanish was, "¿Durante el último mes dejó usted de asistir al trabajo o realizar sus actividades ordinarias debido a la enfermedad o problema de salud señalado antes?"

Days disabled equaled the actual number of workdays lost because of the specific illness (as reported by the respondent). Specifically, individuals were asked, "For how many days during the last month did you not go to work or did you stay in bed?"[3]

The survey also includes a rich set of variables describing each individual's socioeconomic characteristics, such as age, education, hourly earnings, whether or not the person is a salaried worker, sources of nonlabor income, whether the individual lives in a house or an apartment, and the characteristics of the dwelling.

In order to analyze health outcomes, it was also necessary to have information on the individual's environmental conditions. Such variables were constructed from CASEN by averaging the observations in rural and urban areas in each *departamento*, or department.[4] Using this approach, the following community characteristics were linked to each individual:

- Availability of credit (either from the public or private sector) in the *departamento*, by rural and urban subareas;
- Education level in the *departamento* by rural and urban subareas, including illiteracy rates and coverage of primary and secondary schooling;
- Percentage of persons with social security in the *departamento* by rural and urban subareas; and
- Infrastructure conditions (water, electricity) in the *departamento* by rural and urban subareas.

The Instituto Geográfico Agustín Codazzi (1996) provided the following additional information regarding environmental factors that affect health outcomes:

- Temperature, altitude, and rainfall ranges in each municipality;
- Distance from each town to the department capital, where major hospitals are located;

[3] In Spanish, "¿Cuántos días estuvo incapacitado o en cama durante el último mes?"
[4] Colombia is divided into 26 departments.

- Availability of schools in the municipality; and
- Availability of hospitals and hospital beds in the municipality.

The Ministry of Health provided information regarding coverage of vaccination programs by municipality, the number of hospitals in each municipality and their "quality,"[5] and other relevant topics. Another external source provided an index of the kilometers of paved roads per capita and square kilometer in each *departamento*.[6]

The second survey utilized in this chapter is the Encuesta Nacional de Hogares—Etapa 74 (ENH-91). Collected in December 1991, this household survey covers Colombia's 11 major cities: Bogotá, Cali, Medellín, Barranquilla, Bucaramanga, Manizales, Pasto, Cúcuta, Pereira, Ibagué, and Montería. Surrounding metropolitan areas are also included. These urban centers represent almost 40 percent of the national population and about 70 percent of the urban population, with the smallest of the 11 cities having had at least two hundred thousand residents at the time of the survey.

The urban part of ENH-91 is the only survey in Colombia that includes the height of individuals. Fortunately this makes it possible to analyze health status and earnings in urban areas, using adult height as the measure of health. Unfortunately the survey does not include any information about recent illnesses or lost workdays, making direct comparisons to the CASEN impossible. The sample population from ENH-91 was between ages 18 and 70 years old, but the age range is restricted in some estimations and figures (the wage equations and health equations are estimated for those between 18 and 60 years old). This restriction is made because individuals after the age of 60 may grow shorter and individuals younger than 18 may not yet be full-grown (in both cases understating favorable childhood nutritional status). From all persons who earn positive wages or earnings, those with unreasonable heights (less than 135 cm) were excluded, leaving a working sample of 23,910 adults.

As in the CASEN, ENH-91 includes many socioeconomic variables that describe the individual, such as age, education, hourly earnings, whether the person is a salaried worker, nonlabor income, and whether the respondent

---

[5] Institutions were graded 1–6 to indicate the service level provided (from attending minor wounds to performing major medical interventions).

[6] Económica Consultores (1996).

or the respondent's family owns the dwelling in which he/she lives (e.g., owner-occupied housing).

In order to explain adult height, two environmental health factors were derived from ENH-91. First was the percentage of households in the community with access to basic services: water, sewerage, and electricity. In constructing the dummy variable, housing with all three services was considered to have access to basic services while housing without one or more was considered to lack access. The second factor was the percentage of households in the community with favorable population density according to poverty standards (i.e., the percentage of households in the community without overcrowding as defined by DANE).

The characterization of the respondent's resident community was made first by city and then, within each city, by socioeconomic strata. In Colombia, the populations of major cities are classified into six socioeconomic strata depending on the economic capacity of the households in each area of the city (thereby yielding 66 values for community variables for the 11 cities). These classifications are used to charge differential rates for public utilities such as water, electricity, and telephone service; and individuals are well aware of which classification applies to their housing.[7]

It is assumed that the individual's place of residence is exogenous, although people may have migrated to a specific area or community as a consequence of variables that are treated here as exogenous. This introduces potential bias into study estimates (Rosenzweig and Wolpin, 1988).[8]

Descriptive Statistics of the Colombian Surveys

The first sample (CASEN) includes persons who lived in rural or urban areas in 1993, were earning a wage or had positive labor earnings, and were between 18 and 70 years old. This sample is 36 percent female and 64 percent male. Only 9 percent have more than 13 years of schooling, and 8 percent have 0 years of schooling. Some 46 percent of the sample have partial or complete primary schooling, and 74 percent live in urban areas. This sample

[7] Sometimes the interviewer is ordered to ask the respondent for an electricity, phone, or water bill to check that the information of strata is accurate.

[8] This study does not try to explain migration decisions because the surveys provide insufficient information on migration histories and the height of rural populations.

is used to estimate the models with disability and the number of days disabled.[9]

The second sample (ENH-91) covers only urban areas. Only wage earners or individuals with positive labor earnings between 18 and 60 years old were included in this study. About 60 percent of this sample is male, and 40 percent female. Only 4 percent have 0 years of schooling, and 13 percent have more than 13 years of education. This survey includes measurements of adult height, which are used as a measure of health status for this analysis.

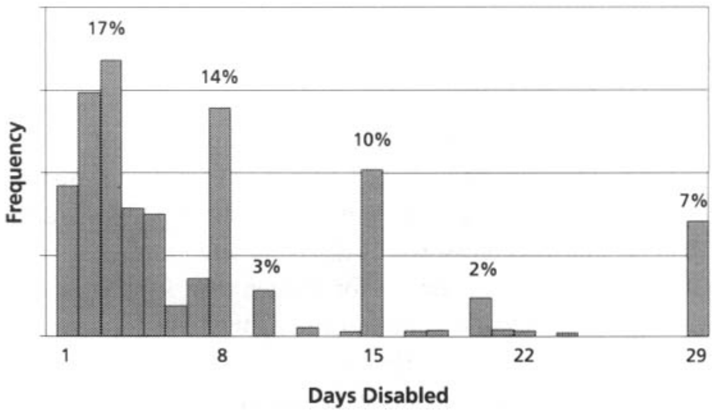
In general, illness as measured by disability and number of days disabled in the CASEN is more frequent for women than for men, and it increases with age. Illness is more common among the less educated than the more educated (within rural and urban populations), and occurs more frequently for rural than urban residents at all levels of education. The patterns for disability and number of days disabled observed in Tables 4.1 and 4.2 are similar. In each case, illness declines with education with the single exception of an upturn in number of days disabled for urban men and women with 13 or more years of schooling. This contrasts with findings for Ghana and Côte d'Ivoire by Schultz and Tansel (1997), where the propensity of adults to report illness was positively related to education. The average number of days disabled may increase in Colombia for urban residents with more than 13 years of schooling because they may have higher expectations about their health, be more able to perceive illness, or be more willing to seek professional advice (Johansson, 1991). In addition, more-educated individuals may have more resources to "indulge" their illnesses and be able to take more days off work when they are ill.

The distribution of the number of days disabled for the population who reported some disability is shown in Figure 4.1. The bulk of this sample (78 percent) has fewer than 10 days of disability, 10 percent have 15 days disabled, and 7 percent have been disabled for the entire past month (they may be chronically disabled).

The ENH-91 survey shows that the youngest cohort (18–24) is about 2.9 cm taller than the oldest cohort (60–70) for both men and women (see

[9] *Editors' note:* For clarity to the reader, the figures in this paragraph relate to the entire sample, while the information provided later in Table 4.1 is only for the population that has experienced some disability.

Figure 4.1 Distribution of the Number of Days Disabled

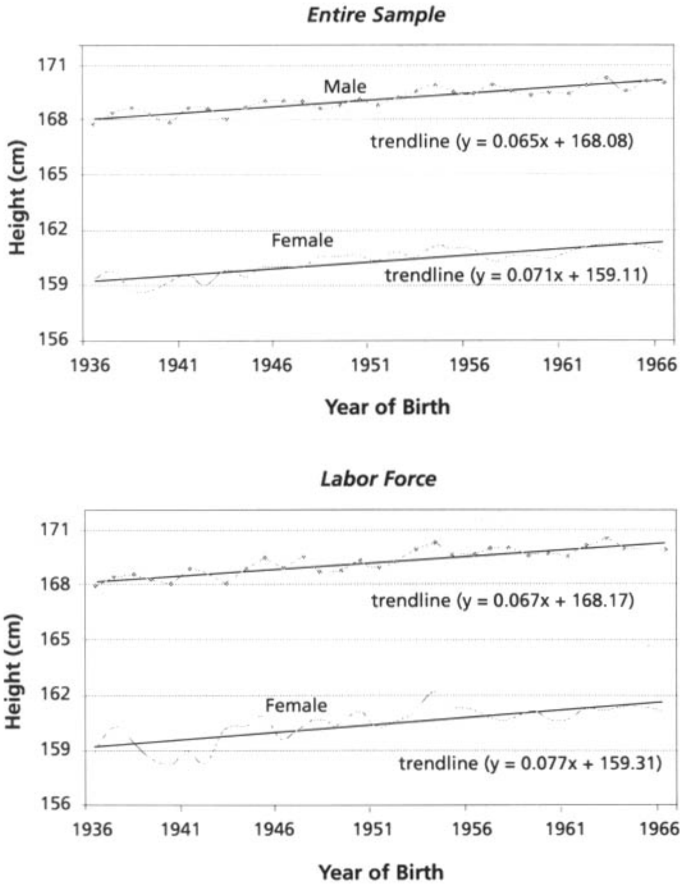


Source: CASEN.

Table 4.3). These patterns of height, summarized in Table 4.3, refer to the whole sample between 18 and 70 years old, not only to the labor market participants. Most of the height differential is between age groups 45–59 and 60–70, suggesting that gains from nutrition to height may be subject to sharply diminishing returns. However since part of the change observed between these two age ranges may be due to the fact that old people shrink for biological reasons, the environmental causes may be less significant than they seem at first glance. The best-educated subsample has an 8.4 cm per person advantage over those with no schooling, although this varies by age. There is about a 1 cm difference within each of the education subgroups and across ages (between youngest and oldest age groups), except for the group with no schooling, for whom there was a 2 cm difference. Within each cohort, the more highly educated are significantly taller. This gap between the least- and most-educated groups (0 versus 13+ years of schooling) is about 5 cm for the youngest group (18–24 year olds) and is about 9 cm for the oldest cohort.

Figure 4.2 shows the trends in height for the entire population aged 25 to 55 in 1991, in relation to their dates of birth. Figure 4.3 shows the results for labor force participants from the same pool. There is a secular increase in height similar in shape and size to the one observed in Brazil by Strauss and Thomas (1998). The slope in the trendline is steeper for females than males, as can be deduced from Table 4.3. Additionally, comparing the two

Figure 4.2 Mean Height by Year of Birth (Ages 25–55)



Source: ENH-91.

parts of Figure 4.2, the slopes for labor force participants are higher than the slopes for the entire population. This indicates that a larger share of individuals with better health entered the urban labor market. In Colombia, the height gains per decade are approximately 0.65 cm for urban men and 0.71 cm for urban women.[10]

[10] The age ranges in these Figures are restricted to avoid bias at the ends of the age range. Younger people may still be growing and therefore have a lower height than their peak adult height, and older people may shrink. The growth reported here is free of biological growth/shrinkage.

Table 4.1 Number and Share of Individuals with Disability

	Area		Sex		
Education	**Rural**	**Urban**	**Male**	**Female**	**All**
0 years	52,184	24,606	53,164	23,626	76,790
	9.6	6.5	8.0	9.1	8.3
1–6 years	139,580	222,164	237,664	124,080	361,744
	6.8	6.8	6.5	7.6	6.8
7–12 years	18,315	237,096	150,262	105,149	255,411
	4.8	6.3	6.1	6.2	6.2
13+ years	1,298	42,179	13,723	29,754	43,477
	3.6	4.3	2.4	6.6	4.3
All	211,377	526,045	454,813	282,609	737,422
	7.0	6.3	6.2	7.0	6.5

	Education			Sex		
Age	**1–6 yrs.**	**7–12 yrs.**	**13+ yrs.**	**Male**	**Female**	**All**
18–24	49,186	64,010	3,139	70,496	48,305	118,801
	5.4	5.8	3.2	5.0	6.2	5.5
25–34	85,494	87,591	18,020	118,245	80,457	198,702
	6.2	5.3	4.1	5.2	6.1	5.5
35–44	81,775	68,222	15,559	98,094	80,226	178,320
	6.4	7.4	4.8	5.9	7.6	6.5
45–59	98,301	28,814	6,544	104,983	55,392	160,375
	7.3	7.2	4.9	7.0	7.9	7.3
60–70	46,988	6,774	215	62,995	18,229	81,224
	11.5	8.8	1.3	12.1	9.8	11.5
All	361,744	255,411	43,477	454,813	282,609	737,422
	6.8	6.2	4.3	6.2	7.0	6.5

Source: CASEN.

Notes: The sample includes all persons between ages 18 and 70 in the labor force. Share of total population is indicated below the number of people.

To our knowledge, this is the first time secular height gains have been quantified in Latin America, except for Brazil. Such information is also available for the United States, some European countries, some African countries, and Vietnam. Strauss and Thomas (1998) show that in the United States

Table 4.2 Mean Number of Days Disabled by Sex, Education, and Area

	Area		Sex		
Education	**Rural**	**Urban**	**Male**	**Female**	**All**
0 years	8.54	10.55	9.38	8.74	9.18
	(8.24)	(9.52)	(8.93)	(8.21)	(8.72)
1–6 years	8.77	7.37	8.45	6.87	7.91
	(8.34)	(7.49)	(8.32)	(6.77)	(7.86
7–12 years	6.62	6.22	7.07	5.06	6.24
	(5.15)	(6.7)	(7.07)	(5.66)	(6.60)
13+ years	3.67	9.66	9.15	9.63	9.48
	(0.95)	(10.29)	(9.67)	(10.41)	(10.19)
All	8.49	7.18	8.12	6.65	7.56
	(8.09)	(7.62)	(8.09)	(7.17)	(7.78)

	Education				Sex		
Age	**0 years**	**1–6 yrs.**	**7–12 yrs.**	**13+ yrs.**	**Male**	**Female**	**All**
18–24	6.153	5.862	5.51	11.244	6.87	4.28	5.82
	(4.66)	(5.9)	(5.02)	(10.25)	(6.25)	(4.23)	(5.66)
25–34	7.821	6.943	5.866	5.521	6.38	6.36	6.37
	(6.5)	(6.54)	(6.14)	(6.24)	(6.44)	(6.26)	(6.37
35-44	8.026	6.936	6.213	13.551	7.53	7.05	7.31
	(8.81)	(6.77)	(8.03)	(12.99)	(8.35)	(8.37)	(8.36)
45–59	8.922	9.549	7.672	9.737	10.17	7.12	9.12
	(8.83)	(9.35)	(6.66)	(6.41)	(9.47)	(6.8)	(8.76)
60–70	10.627	10.057	12.304	13.046	10.30	10.95	10.44
	(9.14)	(8.97)	(5.26)	(13.08)	(8.57)	(9.61)	(8.82)
All	9.18	7.907	6.244	9.479	8.12	6.65	7.56
	(8.72)	(7.86)	(6.6)	(10.19)	(8.09)	(7.17)	(7.78)

Source: CASEN.

Notes: Standard deviations are in parentheses; the sample includes all persons between 18 and 70 in the labor force with at least one day of disability during the month prior to being surveyed.

the mean male stature increased 1.25 cm per decade between 1910 and 1950. The relative figure in Vietnam was 1.05 cm per decade, and in Brazil 0.77 cm per decade. Fogel (1994), using historical European data on stature and weight, reports that mean male height in Sweden increased 0.81 cm per

Table 4.3 Mean Height in Centimeters by Sex, Age, and Education

	Sex		Education				
Age	**Male**	**Female**	**0 years**	**1-6 yrs.**	**7-12 yrs.**	**13+ yrs.**	**All**
18–24	169.32	160.40	160.37	162.11	164.72	165.76	164.23
	(11.43)	(9.98)	(14.81)	(11.06)	(10.97)	(14.03)	(11.49)
25–34	169.50	160.38	158.27	162.69	164.75	166.84	164.55
	(10.10)	(9.93)	(15.98)	(10.82)	(10.97)	(9.97)	(10.96)
35–44	169.13	160.21	160.28	162.50	164.73	167.61	164.30
	(9.57)	(9.75)	(8.27)	(10.53)	(11.20)	(8.15)	(10.57)
45–59	168.05	158.63	158.28	161.62	164.47	168.08	162.94
	(8.51)	(10.86)	(13.19)	(10.53)	(10.86)	(8.56)	(10.91)
60–70	166.41	157.52	157.59	160.92	162.85	166.10	161.32
	(10.67)	(10.62)	(10.74)	(10.41)	(12.44)	(19.11)	(11.47)
All	168.89	159.81	158.58	162.04	164.63	166.94	163.89
	(10.14)	(10.17)	(12.46)	(10.68)	(11.05)	(10.74)	(11.08)

	Sex		
Education	**Male**	**Female**	**All**
0 years	164.66	155.25	158.58
	(9.52)	(12.79)	(12.46)
1–6 years	167.01	158.30	162.04
	(9.10)	(10.28)	(10.68)
7–12 years	169.44	160.72	164.63
	(10.63)	(9.80)	(11.05)
13+ years	171.64	162.01	166.94
	(10.25)	(9.05)	(10.74)
All	168.89	159.81	163.89
	(10.14)	(10.17)	(11.08)

Source: ENH-91.

Notes: The sample includes all persons between 18 and 70 in and out of the labor force; standard deviations are in parentheses.

decade between the third quarter of the nineteenth century and the third quarter of the twentieth century. During the same period, the increase in male stature was 0.64 cm per decade in France, 0.57 cm in Norway, and 1.07 cm in Denmark. Schultz (1996) reports that the gain of height per decade has been almost 1.33 cm for men and 1 cm for women in Côte d'Ivoire, and

Table 4.4 Log Hourly Earnings by Education and Health Indicators

	Years of schooling									
	Men					Women				
Indicator	0	1–6	7–12	13+	All	0	1–6	7–12	13+	All
Disability										
No	5.64	5.95	6.37	7.20	6.16	5.41	5.68	6.22	6.99	6.05
	(0.88)	(0.85)	(0.82)	(0.78)	(0.92)	(1.14)	(1.02)	(0.86)	(0.73)	(1.03)
Yes	5.55	5.99	6.42	7.10	6.11	5.37	5.65	6.24	7.13	5.99
	(1.12)	(0.90)	(0.74)	(0.81)	(0.96)	(1.08)	(1.01)	(0.78)	(0.64)	(1.03)
All	5.63	5.96	6.37	7.20	6.16	5.41	5.68	6.22	7.00	6.05
	(0.90)	(0.85)	(0.81)	(0.78)	(0.92)	(1.14)	(1.02)	(0.86)	(0.72)	(1.03)
Days disabled										
0	5.64	5.95	6.37	7.20	6.16	5.41	5.68	6.22	6.99	6.05
	(0.88)	(0.85)	(0.82)	(0.78)	(0.92)	(1.14)	(1.02)	(0.86)	(0.73)	(1.03)
1–7	5.41	5.97	6.41	6.96	6.11	5.36	5.57	6.24	7.12	5.99
	(1.21)	(0.86)	(0.71)	(0.80)	(0.94)	(1.27)	(1.04)	(0.79)	(0.65)	(1.06)
8–14	5.77	6.04	6.49	7.42	6.16	5.49	5.80	6.26	6.97	6.01
	(0.90)	(0.96)	(0.72)	(0.46)	(0.93)	(0.77)	(0.95)	(0.68)	(0.60)	(0.89)
15–29	5.72	5.95	6.38	7.78	6.09	5.30	5.62	6.11	7.44	5.83
	(1.05)	(0.91)	(0.90)	(1.01)	(1.00)	(0.96)	(1.01)	(0.74)	(0.77)	(1.04)
30	5.58	6.04	6.56	7.02	6.10	5.28	6.18	6.35	7.18	6.33
	(1.15)	(1.04)	(0.80)	(0.55)	(1.06)	(1.13)	(0.73)	(1.00)	(0.56)	(0.92)
All	5.63	5.96	6.37	7.20	6.16	5.41	5.68	6.22	7.00	6.05
	(0.90)	(0.85)	(0.81)	(0.78)	(0.92)	(1.14)	(1.02)	(0.86)	(0.72)	(1.03)
Height (centimeters)										
135–154	6.90	7.14	7.43	7.94	7.24	6.59	6.73	7.15	7.93	7.02
	(0.61)	(0.71)	(0.78)	(0.47)	(0.74)	(0.72)	(0.72)	(0.72)	(0.60)	(0.81)
155–159	6.78	7.14	7.41	8.20	7.29	6.58	6.83	7.27	8.02	7.19
	(1.00)	(0.65)	(0.66)	(0.73)	(0.72)	(0.63)	(0.73)	(0.72)	(0.66)	(0.82)
160–164	6.94	7.16	7.38	8.19	7.33	6.65	6.88	7.31	8.01	7.28
	(0.65)	(0.63)	(0.70)	(0.67)	(0.72)	(0.59)	(0.65)	(0.65)	(0.69)	(0.77)
165–169	6.96	7.16	7.47	8.26	7.43	6.90	6.86	7.34	8.03	7.36
	(0.60)	(0.66)	(0.62)	(0.75)	(0.74)	(0.69)	(0.73)	(0.64)	(0.68)	(0.79)
>169	6.98	7.21	7.51	8.34	7.58	6.56	6.84	7.34	8.07	7.39
	(0.62)	(0.59)	(0.63)	(0.78)	(0.77)	(0.94)	(0.70)	(0.63)	(0.65)	(0.79)
All	6.95	7.18	7.48	8.31	7.50	6.65	6.83	7.29	8.02	7.26
	(0.64)	(0.62)	(0.64)	(0.76)	(0.76)	(0.68)	(0.71)	(0.67)	(0.67)	(0.80)

Sources: CASEN (excluding domestic servants) for disability and number of days disabled, ENH-91 for height.

Note: Standard deviations are in parentheses.

0.66 cm for men and 0.33 cm for women in Ghana. Although the figures for these countries are not strictly comparable because they cover different time periods and represent different historical and economic stages of national development, the measurements provide useful perspective. They suggest that the order of magnitude of height changes in Colombia is similar to what has occurred elsewhere in the world.

The main question addressed by this study is whether or not health and productivity are related at the level of an individual. Table 4.4 shows the mean values of the logarithm of hourly earnings for males and females by education ranges and by different values for health indicators. Although some cells of the table contain too few observations for the means to be statistically representative of the subsamples, taller individuals (males and females) generally earn higher wages, and individuals with more disability earn lower wages.

Estimates of the Productivity of Health Investments

To assess the returns to health investments, the health indicators are included in standard earnings equations, with corrections to address potential bias in those indicators. The analysis proceeds through several stages. First, Mincerian log earnings equations are estimated under the assumption that health is exogenous to wages. This is the simplest specification and serves as a basis for evaluating the impact of correcting for different kinds of bias. Next, the selection bias introduced by considering only individuals with positive labor earnings is adjusted with a standard Heckman (1979) correction for labor force participation. Subsequently, health equations are estimated to understand the determinants of health. These estimates show the connection between local health policy instruments and adult health outcomes, and are used to construct instrumental variable estimators of health—corrected for endogeneity. This procedure generates estimators of health that are free of "noise" and better indicate the relationship between the part of health status that is influenced by individual choice and policy and labor earnings. Finally, the earnings functions are reestimated with these corrected health status measures.[11]

[11] The sample means and standard deviations of the variables in both surveys can be found in Ribero and Nuñez (1999, Tables A1 and A2).

The Determinants of Earnings—A Simple Model

The following earnings function is estimated:

$$\log(w_i) = a + \Sigma b_j X_{ji} + \Sigma c_k C_{ki} + \Sigma d_h H_{hi} + f_i \qquad [4.1]$$

where w_i is the productivity measure (hourly earnings), X_{ji} contains only exogenous endowments that are not modified by the individual or family, C_{ki} are reproducible forms of human capital, and H_{hi} are the health status indicators. In this subsection, the health status indicators are assumed to be exogenous to the hourly earnings function and not correlated with the errors (f) in equation [4.1]. The parameters a, b, c, and d are estimated; the error-term (f) is assumed to have 0 mean and be independently distributed. The subscript i refers to individuals; and subscripts j, k, and h refer to the specific variables in the sets denoted by X, C, and H, respectively. The sample includes wage earners and also nonwage workers with positive earnings.

Among the exogenous endowments (X), age and age squared are included. The variable in C_{ki} is the number of years of schooling. Although a dummy variable for migration (equal to 1 if the person lives in a different place from the residential site five years prior to the survey, and 0 otherwise) was initially included in set C, this did not affect substantially the coefficients of health or education in any manner. Therefore these results are not reported.

As health status indicators (H_{hi}), three variables are considered in separate regressions:

- *Disabled*—a dummy variable that is 1 when the person did not go to work at least one day in the previous month because of illness;
- *Days disabled*—the number of days that the person was disabled in the previous month;[12] and the
- *Height* of the individual—measured in centimeters.

The number of days disabled uses the threshold of inability to work to provide an objective measure of morbidity, just like the binary variable, disability. However the number of days disabled adds information about the

[12] The persons who were not disabled in the previous month had a value of 0 in this variable.

severity of the illness based on its duration. As the empirical results show, both variables explain more or less the same facts. On the other hand, adult height—used as an indicator of child nutritional status, exposure to diseases, and variation in other environmental factors (Schultz, 1997)—demonstrates significantly stronger effects on earnings.

To check the robustness of the findings to differences in the sample, the equations were estimated with and without domestic servants. The coefficients of all the variables except for the intercept were effectively the same when domestic servants were included or excluded from the sample. Similarly the model was estimated separately for wage earners and the self-employed without revealing any substantive differences. In the results presented below, introducing two dummy variables—one for domestic service and one for wage earners—controls the level effects of these different categories of worker.

Although in estimating equation [4.1] the human capital variables may be correlated with the error, education is treated as an exogenous variable. The earnings functions were also estimated separately for men and women, taking into account that some of the health status and control variables may differ by sex, height in particular. The earnings function was also estimated separately for rural and urban areas, though they are linked by the choice of migration.

The hourly earnings regressions are shown in Table 4.5. A surprisingly weak correlation is observed between wages and disability and the number of days disabled. The variables are not significant and do not even have the expected signs. With regard to the other included variables, the results are plausible. The weak measured effect of health in these simple estimations was expected because the health variables are themselves simultaneously determined and measured with error. Hence it will be necessary to correct for downward biases introduced by measurement error and endogeneity.

The regressions with height show that, even in this simple model, height has a significant impact on earnings and has the correct sign. Height benefits men's earnings more than women's (comparing the coefficient of the linear term for males and females). Quadratic terms in height and in the number of days disabled are included to check for nonlinearities, and only height and height squared are significant.

Estimations of the model for the whole sample with a gender dummy confirm that women systematically earn less than men, a result previously

found in other studies (Ribero and Meza, 1997). Similarly earnings in rural areas are lower, a result also previously documented by Leibovich et al. (1997). The age variables are significant and have the expected signs.

Being a salaried worker appears to have different effects on wages in the two data sources. In the CASEN survey of 1993 that includes both urban and rural sectors, a salaried worker generally has higher wages than one who is nonsalaried. The variable is positive and significant for rural males and females and urban females, but not significant for urban males. By contrast in the 1991 ENH-91 survey that covers only urban areas, salaried workers earn less if they are male and more if they are female. These varying results may be due to changes in labor market conditions between 1991 and 1993, to differences in the sampling process, or to real earnings differences for salaried and nonsalaried employment. In the case of domestic servants, earnings are systematically lower than for the rest of the sample.

The model was also estimated excluding the health variables from the right-hand side of equation [4.1]. Notice that the returns to education are not altered by the inclusion of health variables in the regressions. The high degree of correlation between education and health was apparent in additional regressions (not reported) that used height but not education. Without education, the earnings coefficients for height were 0.21 and 0.15 for men and women, respectively. When education was included, these coefficients drop to 0.072 and 0.048, respectively.

Hourly Earnings with Corrections for Selection Bias

When an earnings equation is estimated to calculate the returns to human capital in the population based only on a sample of individuals who are market participants, the estimated returns may be biased (Heckman, 1979). The selection bias may be particularly serious in estimations of female earnings because significantly fewer women choose to enter the labor force. In such cases, the individual will probably take her potential earnings into account when she decides whether or not to enter the labor force. By finding variables that simultaneously affect the decision to seek employment and are unrelated to the observed wage, it is possible to obtain corrected estimates of the returns to human capital through joint estimation of the probability of receiving positive earnings and the earnings equation itself.

Table 4.5 Determinants of Hourly Earnings

	Men										
	Urban							Rural			
Variables	1	2	3	4	5	6	7	8	9	10	11
Age	0.069	0.069	0.069	0.069	0.057	0.056	0.057	0.036	0.036	0.036	0.036
	25	*25*	*25*	*25*	*16*	*16*	*16*	*6.9*	*7*	*7*	*6.9*
$Age^2/1000$	–0.674	–0.675	–0.675	–0.675	–0.507	–0.502	–0.504	–0.349	–0.352	–0.352	–0.350
	20	*20*	*20*	*20*	*11*	*11*	*11*	*5.6*	*5.7*	*5.7*	*5.6*
Schooling (years)	0.087	0.087	0.087	0.087	0.098	0.095	0.095	0.078	0.078	0.078	0.078
	61	*61*	*61*	*61*	*76*	*71*	*71*	*19*	*19*	*19*	*19*
Salaried worker	0.003	0.003	0.003	0.003	–0.070	–0.068	–0.067	0.210	0.210	0.210	0.210
	0.23	*0.24*	*0.24*	*0.23*	*6*	*6*	*6*	*8*	*8*	*8*	*8*
Domestic servant					–0.326	–0.311	–0.312				
					3	*2.9*	*2.9*				
Days disabled (/100)		0.200	0.865						0.400	–0.221	
		0.89	*1.38*						*0.99*	*0.21*	
Days disabled² (/1000)			–0.292							0.278	
			1.14							*0.63*	
Disabled				0.028							0.021
				1.07							*0.45*
Height/100						0.72	–9.02				
						8.7	*3.7*				
$Height^2/10^4$							2.88				
							4				
Intercept	4.160	4.159	4.158	4.159	5.397	4.201	12.430	4.496	4.492	4.493	4.495
Joint significance tests:											
Age and Age^2	607	607	607	607	698	715	712	49	49	49	49
Days disabled and $Days^2$			1.88								0.05
Height and $Height^2$							46				
Adjusted R^2	0.20	0.20	0.20	0.20	0.33	0.33	0.33	0.08	0.08	0.08	0.08
Observations	18,666	18,666	18,666	18,666	13,721	13,721	13,721	4,966	4,966	4,966	4,966

Table 4.5 (continued)

	Women										
	Urban							Rural			
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Age	**0.074**	**0.074**	**0.074**	**0.074**	**0.047**	**0.047**	**0.047**	**0.047**	**0.047**	**0.047**	**0.047**
	16	*16*	*16*	*16*	*10*	*10*	*10*	*3*	*3*	*3*	*3*
Age²/1000	**–0.705**	**–0.705**	**–0.704**	**–0.705**	**–0.429**	**–0.424**	**–0.420**	**–0.421**	**–0.421**	**–0.420**	**–0.41**
	13	*13*	*13*	*13*	*6.8*	*7*	*7*	*2.5*	*3*	*2.5*	*2.5*
Schooling (years)	**0.106**	**0.106**	**0.106**	**0.106**	**0.096**	**0.095**	**0.095**	**0.102**	**0.102**	**0.102**	**0.102**
	47	*47*	*47*	*47*	*55*	*53*	*53*	*11*	*11*	*11*	*11*
Salaried worker	**0.211**	**0.211**	**0.211**	**0.211**	**0.141**	**0.139**	**0.139**	**0.276**	**0.276**	**0.274**	**0.274**
	11	*11*	*11*	*11*	*8*	*8*	*8*	*4*	*4*	*4*	*4*
Domestic servant					**–0.322**	**–0.317**	**–0.303**				
					14	*13*	*13*				
Days disabled (/100)		0.297	–0.601						0.004	–1.532	
		0.92	*0.73*						*0.0*	*0.63*	
Days disabled² (/1000)			0.411							0.738	
			1.18							*0.69*	
Disabled				0.010							–0.055
				0.30							*0.5*
Height/100						**0.48**	**10.08**				
						4.7	*3.1*				
Height²/10⁴							**–2.99**				
							2.9				
Intercept	3.488	3.487	3.490	3.487	5.294	4.542	–3.143	3.890	3.889	3.896	3.899
Joint significance tests:											
Age and Age²	**607**	**607**	**607**	**328**	**262**	**265**	**267**	**17**	**17**	**17**	**17**
Days disabled and Days²			1.88							0.50	
Height and Height²							**15**				
Adjusted R^2	0.25	0.25	0.25	0.25	0.35	0.35	0.35	0.12	0.12	0.12	0.13
Number of observations	10,464	10,464	10,464	10,464	9,332	9,332	9,332	1,299	1,299	1,299	1,299

Sources: ENH -91 for columns 5, 6, and 7. CASEN (excluding domestic servants) for all others.

Notes: The dependent variable is the log hourly labor income measured in pesos of the year of the survey. Boldface indicates statistically significant at the 10% level. T-statistics are italicized. A 1991 peso is equivalent to 1.53 pesos of 1993.

The model, incorporating the selection correction, then has two parts. The first is a probit estimation for labor force participation (selection mechanism):

$$z_i^* = g'p_i + u_i$$

$$z_i = 1 \text{ if } z_i^* > 0 \text{ and } z_i = 0 \text{ if } z_i^* < 0$$

$$\text{Prob}(z_i = 1) = F(g'p_i) \text{ and } \text{Prob}(z_i = 0) = 1 - F(g'p_i), \qquad [4.2]$$

where $z_i = 1$ when the individual (i) participates in the labor market, and $z_i = 0$ when i does not participate in the labor market. The symbol F represents the standard normal cumulative distribution function. The error term (u) is assumed to be distributed with the mean 0, and the variance 1. The subscript i refers to individuals, and g' are the parameters estimated in the probit model. The variables p_i determine the decision to participate and are exogenous to the market wage offer. In theory, the individual will enter the market if the wage offer he or she receives is higher than his or her reservation wage. It is theoretically appealing to consider variables such as nonlabor income as determinants of the probability of working because those variables determine the reservation wages of individuals and may affect their entrance into the labor market.

The second part of the model is the hourly earnings equation:

$$\log(w_i) = a + \Sigma b_j X_{ji} + \Sigma c_k C_{ki} + \Sigma d_h H_{hi} + f_i, \text{ observed if } z_i = 1$$

$$(u_i, f_i) \sim \text{Bivariate Normal } (0,0,1,s_f^2,r). \qquad [4.3]$$

The hourly earnings equation [4.3] is equal to equation [4.1], but it is observed only when the individual is a participant in the market. The symbol s_f represents the standard deviation of the error term (f), and r is the correlation coefficient between the error terms (u) and (f). The variables z_i and p_i are observed for a random sample of individuals, but log (w_i) is observed only when $z_i = 1$. The model with selection correction included is then:

$$E[\log(w_i) | z_i = 1] = a + \Sigma b_j X_{ji} + \Sigma c_k C_{ki} + \Sigma d_h H_{hi} + rs_f \lambda(g'p_i) \quad [4.4]$$

where $\lambda(g'p_i)=j(g'p_i)/F(g'p_i)$, and j is the standard normal probability density function.

Besides age and education, the additional variables used to explain labor force participation (p_i) were nonlabor income, the dummy for living in a house or an apartment, a dummy variable for having adequate house flooring, and a dummy for owning the house in which one lives.[13] These variables are all proxies for the individual's assets and are expected to reduce his or her likelihood of participating in the labor force.

The estimates from this model for females are shown in Table 4.6. The variables explaining labor force participation are significant and have the expected signs. The nonlabor income proxy and the other proxies for wealth diminish the probability of participation for the rural and urban samples, except for the case of adequate flooring in the rural sample.

The returns to schooling in the earnings equations for women are similar to those found in Table 4.5 for men in urban areas, but they are smaller (less than one-third) for the rural areas. The significance of the coefficient for λ indicates that the returns to schooling estimated without the Heckman correction are not biased for the urban sample, but they are biased for the rural sample. According to the sign of λ, rural women who work, holding the observables in equations [4.2] and [4.3] constant, are those who would be paid less, suggesting that they are "pushed" rather than "pulled" into the labor force.

For women, the productivity effects of disability remain insignificant after the correction for the labor force participation decision. Table 4. 6 shows the coefficients of height and height squared change when selection bias is corrected, but the derivatives of earnings with respect to height evaluated at the sample mean remain equal, with and without the correction. The parameter λ is significant in the specifications of columns (5), (7), and (8), but remains insignificant for the others. Although the correction for selection bias was found relevant in some cases, it is ignored elsewhere in this study because it is computationally difficult to implement instrumental variables together with selection bias correction. This does not affect the main conclusions of the study because the selection correction appears to have no significant impact on estimates for the variables that are of primary interest.

[13] For more details on the definitions of these variables, see Ribero and Nuñez (1999).

Table 4.6 Determinants of Women's Earnings Corrected for Choosing to Work

	Urban					Rural		
Variables and Statistics	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Age	**0.127**	**0.083**	**0.083**	**0.286**	**0.099**	**0.066**	–0.016	–0.016
	35.63	*12.80*	*12.82*	*41.37*	*24.97*	*8.20*	*1.04*	*1.05*
$Age^2/1000$	**–1.558**	**–0.826**	**–0.826**	**–3.551**	**–1.040**	**–0.693**	0.244	0.248
	35.26	*10.17*	*10.18*	*38.60*	*20.10*	*7.22*	*1.35*	*1.37*
Education (years)	**0.051**	**0.106**	**0.106**	**–0.014**	**0.090**	**0.062**	**0.032**	**0.032**
	22.64	*37.90*	*37.93*	*4.24*	*66.64*	*9.43*	*2.78*	*2.79*
Salaried worker		**0.224**	**0.224**		**–0.032**		**0.155**	**0.155**
	11.72	*11.70*		*2.79*		*2.25*	*2.25*	
Domestic servant		**–0.700**	**–0.700**		–0.029		**–0.580**	**–0.579**
	25.49	*25.46*		*0.26*		*6.33*	*6.31*	
Nonlabor income/10^6	**–0.946**			**–1.160**		0.161		
	6.69			*11.37*		*0.28*		
Occupy complete house or apartment	**–0.135**					**–0.366**		
	3.63					*4.11*		
Adequate floors	–0.009					**0.103**		
	0.49					*2.02*		
Owner-occupied home	0.021							
	1.24							
Days disabled/100		–0.480					–1.543	
		0.61					*0.71*	
Days $disabled^2/1000$		0.302					0.821	
		0.92					*0.87*	
Disabled			0.004					–0.032
			0.14					*0.33*
Height/100					**–0.986**			
					5.89			
$Height^2/10^4$					**0.515**			
					7.98			
Intercept	–2.653	3.225	3.222	–3.970	4.764	–1.865	7.099	7.102
Lambda		0.113	0.114		**0.404**		**–1.291**	**–1.292**
		(0.06)	(0.06)		(0.02)		(0.08)	(0.08)
Rho		0.131	0.132		0.628		–0.851	–0.851
Sigma		0.865	0.865		0.643		1.518	1.519
Log. likelihood		–32,807	–32,807		–19,395		–5,200	–5,200
Observations	27,292	11,956	11,956	16,974	9,824	5,390	1,472	1,472

Sources: ENH–91 for columns 4 and 5. CASEN for all others.

Notes: Italics indicate t-statistics for the wage regressions using log of hourly earnings and Z-statistics for probit regressions of whether or not the woman is working (participating in the labor force). Figure in parentheses for lambda is standard error. Boldface figures indicate coefficients that are statistically significant at the 10% level.

The model was also estimated for males, but these results are not reported. The negative effect of the wealth proxies on participation that was found for females holds for males. However for males the parameter λ was insignificant, implying no sample selection bias. In the absence of selection bias, the uncorrected estimates are more efficient as well as consistent (Heckman, 1979).

The Determinants of Health Status

In this section the determinants of the observed health outcomes are explored. Using information on individuals' education and wealth, local prices (O), and the community health infrastructure prices and policies (P), the model tries to account for variation in the individual indicators of health status (H). The estimated equation is:

$$H_i = g + \Sigma\, c_l X_{li} + \Sigma\, h_j O_{ji} + \Sigma\, r_k P_k + t_i \qquad [4.5]$$

where g, c, h, and r are estimable parameters; t is the error term; l, j, and k index the sets of exogenous endowments to the individual (X), private opportunities (O), and public policies (P), respectively; and i indexes the individual. Equation [4.5] was estimated with probit models when the health variable is the dichotomous variable for disability. The model was estimated with ordinary least squares when the health variables were the number of days disabled in the last month, and height.[14]

In the 1993 CASEN survey, the sample includes urban and rural individuals, and the health variables are disability and the number of days disabled in the last month. The age of the individual is specified as an exogenous endowment. Nonlabor income and a dummy to indicate the type of housing were included as indicators of individual private opportunities (O) since these forms of wealth might shift health outcomes positively by providing more resources for health maintenance and improvements. Variables to describe the community-specific environment (P) were derived from various sources and matched to sample clusters. The variables of P are defined for 52 regions (approximately twice the number of *departamentos* since most

[14] Because the number of days disabled is truncated at 0 and 30, a tobit analysis was also used, but it did not significantly alter the conclusions.

of the regions have rural and urban areas). At the municipality level, characteristics that were expected to be related to the health outcomes were climate, availability of health centers, enrollment in social security,[15] transportation infrastructure, transportation time to reach hospitals, transportation time to reach schools, and availability of water and electricity. Few of these, however, result in significant correlations and some have a counterintuitive sign. At the departmental level, the only significant variable was the number of yearly transfers from the central government to the *departamento* for health.[16]

Regional differences in Colombia are very important. Levels of earnings, formality of labor markets, education levels, and health status are generally worse in coastal regions than in the interior. The Pacific Coast, in particular, is known as the poorest region in the country. Although the altitude variable may capture some of these regional variances, there are persistent cultural, racial, and institutional differences that are not captured by climate. To capture some differences due to these elements, two additional dummy variables were introduced: one for living in a *departamento* on the Pacific coast and one for living in a *departamento* on the Atlantic coast.[17]

In the ENH-91 survey, the data covers only the urban population, and the only health indicator is adult height. The model includes the following characteristics of the individual: age, nonlabor income, and a dummy to indicate owner-occupied housing. The variables that capture environmental health risks (P) are two indicators constructed at the city level. The first measures availability of basic services in the households of the community where the respondent lives. An aggregate indicator of the availability of water, sanitation, and electricity was used because the variables entered separately proved to be insignificant. The second community variable mea-

[15] In 1993 approximately 25 percent of urban residents and 8 percent of rural residents were affiliated with or beneficiaries of the Social Security Institute for health services; 10 percent of the Colombian population used private health care; and 5 percent were covered by other services.

[16] Other variables that were available and tested but not included in this final model were: the number of primary and secondary schools per capita, the number of hospital beds per capita, average times to reach schools, and average daily hours in schools.

[17] The reference category was residency in the interior.

sures the percentage of houses in the community that are not overcrowded. City and strata define the communities or sample clusters for these urban areas.

Estimation results show that age is an important factor in explaining the three health indicators, with older individuals tending to have worse health (see Table 4.7). The coefficients of age and age squared are individually significant for the number of days disabled and jointly significant for the probability of having a disability. Estimation of this model using only a linear term for age (not reported) showed that one more year of age increases the probability of having disability by 0.6 percent and 0.8 percent for urban males and females, respectively, and by 1.1 percent and 1.3 percent for rural males and females, respectively.

The negative effects of age on health are larger in rural than urban areas and for females than males. Nonlabor income is not significant, but the wealth proxy of living in a house or apartment is negatively related to the number of days disabled and to the incidence of disability. This wealth measure is significant in the urban samples. Home ownership is also positively related to height. These results coincide with the intuition that wealthier individuals tend to have better health, controlling for the individual and community characteristics listed in Table 4.7.

The number of hospitals or clinics per capita is not significant, except in the regressions for the number of days disabled using the urban male sample. In this case, however, the variable has a sign that contradicts our expectations, implying that more hospitals or clinics per capita led to more days disabled in this subsample.

In rural areas, the model shows the expected negative sign for the variable "percentage of people directly enrolled in or beneficiaries of social security," and it is significant when the health outcome is the number of days disabled. The econometric model reveals that a 10 percent increase of the population enrolled would reduce the mean disability for a rural male by more than 0.2 days. At the same time, the incidence of disability for rural females would be reduced by two percentage points. In urban areas, however, the "percentage of people directly enrolled in or beneficiaries of social security" has a positive coefficient on disability, contradicting the expectation that access to social security should improve health. This could be due to the fact that significantly more urban individuals are affiliated with social security and therefore have greater contact with medical diagnostic services,

Table 4.7 Determinants of Health, by Gender, Indicator, and Area

	Men									
	Disabled				Days Disabled				Height	
	Urban		Rural		Urban		Rural		Urban	
Variables	1		2		3		4		5	
Age	–0.004	*0.62*	0.005	*0.42*	–0.013	*1.46*	–0.029	*1.59*	**0.088**	*2.31*
Age²/1000	0.136	*1.58*	0.075	*0.55*	**0.311**	*2.84*	**0.608**	*2.80*	**–1.989**	*4.04*
Occupy complete house or apartment	**–0.215**	*3.58*	–0.182	*1.14*	–0.131	*1.62*	–0.270	*1.04*		
Owner-occupied home									**0.29**	*2.41*
Nonlabor income/10⁶	0.012	*0.09*	0.447	*0.59*	0.167	*0.97*	–1.256	*0.93*	–7.21	*0.40*
Hospitals/clinics per capita (* 1000)	0.592	*1.53*	0.270	*0.86*	**1.425**	*2.78*	–0.030	*0.06*		
Share of community covered by Social Security	**0.497**	*1.96*	–0.731	*1.24*	0.482	*1.57*	**–2.115**	*2.30*		
Share of households with electricity	0.151	*0.51*	0.130	*0.90*	–0.034	*0.11*	0.239	*1.04*		
Altitude (meters/100,000)	–0.57	*0.25*	–6.67	*1.65*	–2.201	*0.79*	**–11.98**	*1.91*		
Distance to *departamento* capital (km/1000)	0.206	*1.14*	0.441	*1.17*	0.004	*0.02*	**1.251**	*2.16*		
Share of houses with electricity, water and sewerage									**7.002**	*4.40*
Share of houses with adequate space									**10.84**	*14.14*
Nonlabor income / 10⁶ * Share of houses with adequate space									9.60	*0.49*
Govt. transfer to *departamento* (pesos per capita)	–2.474	*0.82*	16.368	*0.70*	–6.678	*1.88*	14.257	*0.40*		
Atlantic Coast	0.052	*1.19*	–0.084	*1.02*	–0.019	*0.36*	**–0.257**	*2.07*		
Pacific Coast	0.047	*1.02*	**0.211**	*2.79*	0.019	*0.33*	**0.314**	*2.59*		
Intercept	–1.780		–1.816		0.527		0.827		159.37	
Prob > F or Prob > Chi²	0.01		0.10		0.08		0.01		0.00	
Log likelihood	–3695		–1224							
Adjusted R^2	0.007		0.021		0.004		0.014		0.033	
Observations	18,666		4,966		18,666		4,966		13,721	

Table 4.7 (continued)

	Women									
	Disabled				Days Disabled				Height	
	Urban		Rural		Urban		Rural		Urban	
Variables	1		2		3		4		5	
Age	0.013	*1.32*	−0.004	*0.16*	0.013	*0.97*	−0.070	*1.74*	**0.119**	*2.47*
Age²/1000	−0.058	*0.49*	0.204	*0.74*	−0.009	*0.05*	**1.205**	*2.56*	**−2.514**	*3.90*
Occupy complete house or apartment	**−0.280**	*3.54*	−0.166	*0.58*	**−0.355**	*2.89*	−0.060	*0.13*		
Owner-occupied home									**0.32**	*2.11*
Nonlabor income (10⁶)	0.308	*1.19*	1.049	*0.69*	0.809	*1.90*	−0.893	*0.30*	−66.1	*1.85*
Hospitals/clinics per capita (* 1000)	0.045	*0.08*	0.014	*0.02*	−1.167	*1.57*	0.704	*0.63*		
Share of community covered by Social Security	**0.79**	*2.56*	**−2.05**	*1.98*	−0.109	*0.26*	−3.101	*1.87*		
Share of households with electricity	0.011	*0.03*	−0.086	*0.36*	0.243	*0.44*	−0.306	*0.71*		
Altitude (meters/100,000)	−2.08	*0.79*	−10.56	*1.54*	−3.95	*1.10*	**−27.04**	*2.42*		
Distance to *departamento* capital (km/1000)	0.272	*1.32*	−0.052	*0.07*	0.262	*0.87*	−1.098	*0.94*		
Share of houses with electricity, water and sewerage									0.495	*0.32*
Share of houses with adequate space									**11.70**	*12.51*
Nonlabor income / 10⁶ * Share of houses with adequate space									**76.50**	*2.04*
Govt. transfer to *departamento* (pesos per capita)	−4.803	*1.25*	−28.591	*0.61*	−2.524	*0.55*	23.118	*0.32*		
Atlantic Coast	0.079	*1.45*	−0.001	*0.01*	−0.083	*1.13*	−0.432	*1.65*		
Pacific Coast	**0.147**	*2.89*	**0.574**	*4.34*	**0.249**	*3.42*	**0.872**	*4.03*		
Intercept	−1.780		−1.816		0.527		0.827		159.37	
Prob > F or Prob > Chi²	0.00		0.00		0.00		0.00		0.00	
Log likelihood	−2613		−356							
Adjusted R²	0.011		0.056		0.005		0.038		0.032	
Observations	10,464		1,299		10,464		1,299		9,332	

Sources: ENH-91for height regressions; CASEN for all others.

Notes: Italics indicate Z-statistics for probit regressions that use Disabled as the dependent variable. Italics indicate t-statistics for all other regressions (which are OLS). Monetary values for the height regression are in 1993 pesos; all others are in 1991 pesos. A 1991 peso is equivalent to 1.53 pesos of 1993.

therefore leading to more frequent reports of disabilities even when the actual incidence is not greater.

Health outcomes were expected to be affected positively by access to electricity because it allows households to store food more safely (e.g., with refrigeration) and increases exposure to public health campaigns through radio and television. Furthermore access to electricity, unlike potable water, is significantly less widespread in rural areas. Nevertheless electrification showed no significant impact on health status in these estimations. Climate variables such as altitude were significant in the rural samples for the number of days disabled, yielding a negative sign, which implies that people residing at higher altitude enjoyed better health. The negative sign of the coefficient for altitude was expected since the regions closer to sea level are more humid and have a higher prevalence of communicable diseases. This coincides with a general perception that the lowlands in Colombia are less healthy (Rosenzweig and Schultz, 1982). Estimations with altitude and altitude squared (these regressions are not reported) suggested that nonlinearities in the effect of altitude on the study's health indicators are not strong.

The distance in kilometers between the municipality and the *departamento* capital is an approximation of travel time, or the cost of using public health services. This variable was significant only for the explanation of the rural male subsample's number of days of disability, with greater distance being associated with more days of disability. It is intuitive that distance is significant for a rural sample because urban areas usually have at least one health center, while rural individuals often have to travel very far to find one. However, it is hard to find an economic explanation for the gender difference in this result.

In addition to the other factors taken into account in Table 4.7, living in the Pacific Coast region contributes positively to disability and to the number of days disabled for all females (rural and urban) and for rural males. People living in that region have on average between 0.25 and 0.87 more days disabled. The dummy variable for residency in the Atlantic Coast region was significant in reducing the number of days disabled for the rural male sample, but it did not contribute significantly to individual health in the other samples. When the models were estimated without the regional dummies, the other coefficients were very similar.

The estimations for height by gender are shown in column 5 of Table 4.7. Disaggregating by sex, a regression is first reported on age as a linear and

then a quadratic function to quantify the trend in height and nutrition. The trend indicates that each cohort has, on average 0.06 cm less height than the cohort born one year later, and this is more or less the same for males and females. A woman with one more year of schooling is expected to be 0.3 cm taller, while a man with more schooling is 0.4 cm taller, holding only age constant. When other individual characteristics—such as nonlabor income, owner-occupied housing, and community characteristics—are taken into account, the partial association between schooling and height decreases.

Owner-occupied housing is a significant determinant of height and has the expected positive sign. The wealthier an individual is, the better his/her health status is. The percentage of houses in the community with basic services is correlated with individual height and has a positive sign. Residing in a community with a high level of basic services is associated with greater individual height, although only the male coefficient is significant. The supply of adequate housing, measured by the percentage of houses in the community with adequate number of rooms per person, is also associated with better health outcomes for individuals (males and females) as measured by height.

When wealth is interacted with public policies to analyze the personal distribution of health benefits, the product of nonlabor income and the percentage of community houses with adequate space per person is significant and positive. This suggests that nonlabor income and the adequacy of community housing are complementary. In other regressions, run with the same explanatory variables but excluding the interaction term, nonlabor income was significant and positive, and the effect was greater for females than for males (not reported).

Tests of joint significance of the identifying variables imply that they are jointly significant. The hypothesis that the coefficients of the identifying variables in each model are jointly equal to 0 can be rejected at the 10 percent level.

Earnings Equations with Corrected Health Indicators

In this section, the earnings function [4.1] is reestimated because the human capital health stocks (H) may be correlated with the earnings error or be measured with error, imparting bias to single equation estimates. These problems are addressed by estimating equation [4.1] using instrumental variable (IV) methods. The estimated equation is:

$$\log(w_i) = a + \Sigma b_j X_{ji} + \Sigma c_k C_{ki} + \Sigma d_h H^*_{hi} + f_i \qquad [4.6]$$

where w_i is the hourly earnings, X_{ji} contains only exogenous endowments, C_{ki} are forms of human capital, and H^*_{hi} are the endogenous health status indicators. The health status variables are assumed to be endogenous because they result from a process that is systematically related to individual resource opportunities (O), and community prices and policies (P). The fitted health indicators (H^*_{hi}) are computed using the estimated parameters from the equations of health determination in the previous section.

$$H^*_i = \hat{g} + \Sigma \hat{c}_l \hat{X}_{li} + \Sigma \hat{h}_j \hat{O}_{ji} + \Sigma \hat{r}_k P_k \qquad [4.7]$$

The identifying instruments used to predict H^* are included in the environmental variables (P) and (O).[18]

Table 4.8 reports the estimation results for the model using instrumental variables for health. The pattern of health effects changes significantly when using the IV method to correct for bias. The health variables are now more statistically significant and appear to affect wages in the expected directions.

For the dummy variable "disability," the impact of health is negative and significant for all the samples, although they are more significant for males than females. This is the expected sign for this variable, and it was not observed in the estimations without IV methods that are shown in Table 4.5. Having a disability in a given month decreased the earnings of urban men by 25 percent and urban women by 13 percent. In rural areas, earnings appear to be reduced by 34 percent and 17 percent, respectively.[19] In each case, the effect is statistically significant for men but not for women.

The effect on earnings of the number of days disabled is negative and significant for men and women in the rural samples when the health indicator enters linearly. The coefficients indicate that one more day of disability

[18] The squared endogenous variables are computed by running first-stage regressions on the quadratic term and using that auxiliary regression to predict the squared endogenous variable. These auxiliary regressions can be found in Ribero and Nuñez (1999, Appendix, Table A-4).

[19] Note that the percentages reported here differ from the coefficients reported in Table 4.8 due to the log-linear nature of the specification. The percentage change is equal to ($e^b - 1$), where b is the reported coefficient.

experienced in rural areas reduces male earnings by about 28 percent and female earnings by about 13 percent. When entered quadratically, the marginal impact of an additional day ill in rural areas for men and women who are healthiest (i.e., no days ill reported) is less than 5 percent, but is associated with more than a 50 percent reduction in earnings for individuals who reported seven days disabled. The results in urban areas are less reliable. For the estimates in which days disabled are entered linearly, the coefficients are insignificant; whereas the quadratic specification yielded implausibly large effects, on the order of 90 percent reductions in earnings. In sum, these results indicate that health does have an impact on earnings in rural areas, and that days disabled is an effective health measurement in rural areas when properly instrumented. For urban areas, the results are less clear—either because health has a smaller impact on wages or because the days-disabled measure is less accurate as a proxy.

The instrumented models in columns (4) and (5) of Table 4.8 show that height is significant in the determination of wages, in both the linear and quadratic specifications, for both males and females. The size of the coefficients in the linear specification of column (4) is much larger than the corresponding OLS estimates from Table 4.5 (the male coefficient is 11 times larger, and the female coefficient is 14.5 times larger). This indicates that when the endogenous character of height is taken into account, the "true" effect is substantially greater.

According to these estimates, a taller man receives hourly earnings that are 8 percent higher per centimeter, and a woman receives hourly earnings 7 percent higher per centimeter. The size of the returns in Colombia are in line with those found in Ghana (Schultz, 1996), where a 1 cm increase is associated with a wage gain of 5.7 percent for males and 7.5 percent for females, using instrumental variables and holding constant for BMI and migration. In Côte d'Ivoire, the effect of male height is not significantly associated with a wage gain based on instrumental variable estimates, holding constant for BMI and migration. These estimated returns to height in Colombia reveal that investments in childhood nutrition may be of great importance for future labor productivity and growth.

Unlike other studies, returns to education appear to be unaffected by the correction for bias in the health measures, and are approximately the same as those shown in Table 4.5. The returns to education change only slightly from 9.8 percent without height to 9.1 percent with height for ur-

Table 4.8 Determinants of Hourly Earnings with Corrected Health Indicators, by Gender and Area

	Men							
	Urban					Rural		
Variables and Statistics	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Age	**0.070** *21.3*	**0.070** *21.3*	**0.067** *22*	**0.049** *13*	**0.056** *13*	**0.026** *4.49*	**0.026** *4.50*	**0.038** *6.9*
$Age^2/1000$	**–0.702** *14.8*	**–0.714** *15.1*	**–0.634** *16*	**–0.357** *7.1*	**–0.497** *8*	**–0.143** *1.85*	**–0.150** *1.93*	**–0.313** *4.7*
Schooling (years)	**0.087** *62.1*	**0.087** *61.6*	**0.087** *62*	**0.091** *62*	**0.091** *62*	**0.075** *18.1*	**0.074** *17.97*	**0.076** *19*
Salaried worker	0.003 *0.2*	0.002 *0.11*	**0.003** *0.21*	**–0.069** *5.51*	**–0.072** *5.74*	**0.191** *7.32*	**0.190** *7.19*	**0.201** *8*
Domestic servant				**–0.381** *3.97*	**–0.372** *3.92*			
Fitted days disabled (/100)	8.756 *0.98*	**–96.027** *3.98*				**–32.930** *5.04*	–3.555 *0.09*	
Fitted days $disabled^2$ (/1000)		**51.104** *4.45*					–12.968 *0.79*	
Fitted disabled			***–0.281*** *3.17*					**–0.410** *3.35*
Fitted height (cm/100)				**7.973** *14.2*	**–477.4** *3.63*			
Fitted $height^2$ (cm^2/10,000)					**142.92** *3.69*			
Intercept	4.1	4.2	3.7	–8.0	403.4	4.8	4.7	3.8
Adjusted R^2	0.20	0.20	0.20	0.34	0.34	0.09	0.09	0.08
Joint significance tests:								
Age and age^2	**450**	**416**	**526**	**736**	**404**	**59**	**48**	**44**
Height and $height^2$		**13**			**119**		**15**	
Number of observations	18,666	18,666	18,666	13,721	13,721	4,966	4,966	4,966

Table 4.8 (continued)

Variables and Statistics	Women: Urban 1	2	3	4	5	Rural 6	7	8
Age	**0.075**	**0.068**	**0.076**	**0.038**	**0.037**	**0.036**	**0.034**	**0.045**
	15.5	*13.2*	*16*	*7.2*	*6.7*	*2.29*	*2.09*	*3.1*
Age^2 /1000	**−0.704**	**−0.601**	**−0.713**	**−0.260**	**−0.240**	−0.238	−0.201	**−0.366**
	11.4	*9.0*	*11*	*3.5*	*3.0*	*1.22*	*1.00*	*2.1*
Schooling (years)	**0.106**	**0.105**	**0.106**	**0.090**	**0.090**	**0.100**	**0.101**	**0.101**
	47.9	*47.2*	*48*	*47*	*46*	*11.8*	*11.73*	*11.9*
Salaried worker	**0.211**	**0.209**	**0.212**	**0.141**	**0.140**	**0.262**	**0.259**	**0.268**
	9.6	*9.53*	*10*	*7*	*7*	*3.77*	*3.67*	*3.88*
Domestic servant				**−0.370**	**−0.370**			
				14	*14*			
Fitted days disabled (/100)	−6.780	**−90.633**			**−13.475**	0.545		
	1.14	*4.36*			*2.04*			
Fitted days disabled² (/1000)		**42.780**				−7.965		
		4.22					*0.74*	
Fitted disabled			**−0.144**					−0.188
			1.73					*1.75*
Fitted height (cm/100)				**6.888**	56.10			
				9.28	*0.65*			
Fitted height² (cm²/10,000)					−15.36			
					0.57			
Intercept	3.5	3.7	3.2	−5.6	−45.0	4.1	4.1	3.6
Adjusted R^2	0.25	0.25	0.25	0.35	0.35	0.12	0.13	0.13
Joint significance tests:								
Age and age²	**232**	**226**	**237**	**288**	**262**	**16**	**16**	**15**
Height and height²		**11**			**48**			**2**
Number of observations	10,464	10,464	10,464	9,332	9,332	1,299	1,299	1,299

Sources: ENH-91 for all regressions with height; CASEN for all others.

Notes: Instrumental variables for health indicators are based on models in Table 4.7. Italics indicate t-statistics for robust standard errors. Boldface indicates statistically significant at the 10% level. See Ribero and Nuñez (1999) for more details.

ban men and from 9.6 percent without height to 9.0 percent with height for urban women. They are approximately equal when the instrumental variables for disability or the number of days disabled are included in the earnings equation.

Strauss and Thomas (1995) used survey data from urban Brazil and found that relative to the returns to education without controlling for health, the estimated returns to education with health controls were 45 percent smaller for literate men and 30 percent smaller for men with secondary education or more. Schultz (1996) finds that the estimated wage returns to schooling are reduced between 10 percent and 20 percent with the addition of three other human capital inputs in the regression: migration, BMI, and height. Parker's study in Chapter 3 also finds reductions in the returns to education for elderly Mexicans of as much as 50 percent when health variables are included; however, the returns to education are almost invariant to the introduction of health in the earnings equations in this Colombian data.

Policy Simulations for Health Status and Earnings

The last step of the research combines the estimates from the earnings function and the health outcome equations to simulate how changes in policy variables are likely to affect lifetime earnings. In order to apply this procedure, the simplifying assumption is introduced that the effects of health on wages are uniform over the life cycle. The effects of policy changes on the probability of having a disability, on height, and on productivity are presented. These calculations rely on the parameters estimated in Tables 4.7 and 4.8.

Most of the policy variables were not significant for the explanation of disability (see Table 4.7), which makes it difficult to draw many inferences from the model. Therefore, the only variable considered for simulations with disability was the coverage of affiliation with social security. Table 4.9 shows the impact of affiliation levels on the probability of being disabled when the share of individuals with social security coverage in the *departamento* and area (urban or rural) is increased by 10 percent, 20 percent, and 30 percent. The second part of Table 4.9 shows the associated changes expected in log earnings from this policy.

The simulation demonstrates how greater affiliation with social security in rural areas would decrease the probability of being disabled and would

Table 4.9 Policy Simulations Affecting Disability, by Area and Gender

	Incidence of Disability			
	Urban Male	Urban Female	Rural Male	Rural Female
Model Prediction of Incidence	0.050	0.069	0.068	0.077
Change in coverage of social security:				
Increase by 10%	0.051	0.071	0.067	0.075
Increase by 20%	0.052	0.073	0.067	0.073
Increase by 30%	0.054	0.076	0.066	0.071
	Mean Hourly Labor Earnings			
	Urban Male	Urban Female	Rural Male	Rural Female
Model Prediction (pesos per hour)	535.500	448.855	293.536	262.146
Change in coverage of social security:				
Increase by 10%	533.415	447.332	294.241	263.038
Increase by 20%	531.339	445.813	294.919	263.934
Increase by 30%	529.271	444.300	295.627	264.860

Sources: Calculations from Tables 4.7 and 4.8 and CASEN.

increase productivity, while the opposite would hold in urban areas. According to the model described by equation [4.5] and controlling for the other variables included in columns (1) to (4) of Table 4.7, an increase of 20 percent in social security coverage in rural areas could reduce the probability of being disabled by 5.3 percent for rural women and by 2.1 percent for rural males. This particular change would be associated with a productivity increase of 0.7 percent for rural women and 0.5 percent for rural men, after controlling for the other variables included in columns (3) and (8) of Table 4.8. This conclusion must be viewed cautiously since the link between social security and better health may not be causal. In urban areas, social security may be simply associated with a higher tendency to report illness, so that the indicator may only be reflecting the formality of labor markets in these subregions.

Table 4.10 Policy Simulations Affecting Height and Earnings, by Gender

		Height						ln Hourly Earnings			
		Female			Male			Female		Male	
		Mean	Change		Mean	Change		Mean	Change	Mean	Change
Simulations			(cm)	(%)		(cm)	(%)		(%)		(%)
Actual (mean values)		160.90			169.40			7.2784		7.4982	
		(1.21)			(1.22)			(0.476)		(0.440)	
Change in share of households with:											
All basic services	Adequate space										
Increase of 33%	Increase of 33%	164.42	3.53	2.2	172.63	3.22	1.9	7.5216	27.5	7.7553	29.3
		(1.70)			(1.45)			(0.491)		(0.451)	
Increase of 50%	Increase of 50%	166.24	5.34	3.3	174.28	4.88	2.9	7.6469	44.7	7.8876	47.6
		(2.02)			(1.59)			(0.500)		(0.458)	
Increase of 33%	—	160.90	0.00	0.0	169.45	0.05	0.0	7.2786	0.00	7.5025	0.4
		(1.21)			(1.17)			(0.476)		(0.438)	
—	Increase of 33%	164.42	3.52	2.2	172.57	3.17	1.9	7.5214	27.5	7.7511	28.8
		(1.70)			(1.49)			(0.491)		(0.454)	
Increase of 50%	—	160.90	0.01	0.0	169.48	0.08	0.1	7.2787	0.0	7.5045	0.6
		(1.20)			(1.16)			(0.476)		(0.437)	
—	Increase of 50%	166.23	5.34	3.3	174.21	4.80	2.8	7.6465	44.5	7.8813	46.7
		(2.03)			(1.65)			(0.501)		(0.462)	

Sources: Calculated from Tables 4.7 and 4.8.

Note: Standard deviations in parentheses.

The effect on height and the consequent changes in earnings from performing diverse simulations are shown in Table 4.10. The first row shows the estimated model without variations. Subsequent lines show diverse simulations with "percentage of houses in the community with basic services" and "percentage of houses in the community with adequate number of rooms per person." Simulations confirm that both policies would be positively associated with height and earnings, as expected from Tables 4.7 and 4.8. Almost all the simulations produced a greater impact on female height than on male height, but the related increases in earnings to each policy are higher for males than for females. This is consistent with the fact that the productivity effects of height for males are higher than for females. The effect on earnings of increasing the provision of adequate housing would be greater than that of increasing the provision of public services. Holding constant all the other variables included in the models and assuming that it was possible to increase by one-third the "percentage of houses with adequate number of rooms per person," height would increase by 2.2 percent for females and by 1.9 percent for males. The implied health status improvements could be associated with substantial increases in hourly earnings of 27 percent and 29 percent for women and men, respectively.

Conclusions

The main finding of this study is that health does have a significant and sizeable impact on individual earnings. The study identified the magnitude of the returns to having good health status through the direct effect of health variables on earnings of individuals. One more day of disability is associated with a decrease in earnings by rural men of 28 percent and by rural women of about 13 percent. Having any disability in a given month was associated with lower earnings for urban men and women of 25 percent and 13 percent, respectively. Of all the health measures, adult height was the most robust and consistently significant. In Colombia, after controlling for other factors, men earn 8 percent more per centimeter, while women earn about 7 percent more. These effects appear to be stronger than those found in studies of two African countries (Schultz, 1996).

A general result, whichever health status measure is chosen, is that wealthier individuals (those who have higher nonlabor incomes, own their dwellings, or live in a house or apartment) tend to have better health, even

controlling for age, community characteristics, and geographic location. Wealthier individuals also appear to be healthier when they reside in wealthier communities (as measured by the community average for rooms per person), confirming the expectation that community resources complement individual resources in improving health status.

Analyzing health status determinants also demonstrated ways to improve health through public policy. Social security coverage in rural areas was one of the most important factors in explaining disability. Greater coverage was associated with fewer reported illnesses. By contrast, greater urban social security coverage is associated with worse health status. This relationship may, however, be spurious since social security coverage could be correlated with some negative and unobserved factor or with an unexpected sample bias. Rather than individual health improving in tandem with social security coverage, at some threshold increased coverage may simply lead to respondents reporting illnesses more often, accounting for the findings in urban areas. A more troubling conclusion would be that urban social security services are so poor that they threaten health status, leaving beneficiaries worse off than those who don't qualify for or use the system.

In urban areas, individual wealth and favorable environmental conditions such as the provision of public services and adequate community housing were the most important determinants of health. In analyses using adult height as the health measure, an increase in the supply of adequate housing was strongly associated with better health and earnings. Housing policies generally would benefit the health status of men and women comparably, but would generate slightly larger earning gains for men probably due to the structural differences of reward and opportunity for men and women in the labor market.

Finally, it should be noted that the quality of available information about public health interventions limits the study. Despite a major effort to collect social and environmental data at the *departamento* and municipality levels and merge it with results from individual household survey data for analysis, most of these general indicators could not account for the variation in individual health indicators. Although several patterns are suggestive, variables that were expected to be highly correlated with health outcomes—such as the coverage of vaccination programs for different diseases, the regional supply of hospitals, the number of hospital beds per region, and the number of primary and secondary schools—were not

significant. This, however, may simply reveal the lack of reliable information and/or the need for better indicators of the standards and costs of health services. While one could conclude that health services have negligible impact on the indicators used in this study, it is also possible that what is being measured is simply the poor quality of services presently available.

Future research that utilizes adult height as a measure of health looks particularly promising. This kind of research should be extended by collecting data that measures acute and chronic illnesses and weight in relation to height, and improving the database on local policy and environmental conditions. This would permit a firmer case to be made for targeting investments in particular health programs and policies that could raise labor productivity by improving the health status of Colombians.

This page intentionally left blank

CHAPTER 5

The Returns to Health for Peruvian Urban Adults by Gender, Age, and across the Wage Distribution

Edmundo Murrugarra and Martín Valdivia[1]

This chapter analyzes the determinants of health status for urban adults in Peru and estimates the impact of health status on earnings. Using the reported numbers of days ill as an indicator of an individual's health status, it was possible to show that household infrastructure and the associated costs of reaching health centers have a significant effect on health. Schooling effects on health are also found important, especially among older people. When the former variables are used as instruments to correct for endogeneity and measurement error in the health variable, health status has a strong and positive impact on wages, especially for males. Using quintile regression techniques, this paper finds that the largest effects of illness on productivity are estimated among those in the bottom of the wage distribution. The lowest-paid workers have their hourly wages reduced by 3.8 percent due to one day of sickness. Other important effects are found by type of employment, where the effects on the self-employed are larger, especially among the eldest (–4.3 percent). Private-sector workers have a smaller but significant reduction due to illness (–1.8 percent). These results suggest that the impact on wages is stronger for jobs in which productivity and health are more closely connected or better observed, as in the private sector and among the self-employed.

[1] Dr. Edmundo Murrugarra is with the World Bank, and Dr. Martín Valdivia is with the Centro de Investigación y Docencia Económicas (CIDE) in Mexico and the Grupo de Análisis para el Desarrollo (GRADE) in Lima, Peru.

Introduction

The empirical connection between investments in education and wages has been widely analyzed in both developed and developing countries. The labor market returns to investments in health have been studied much less, even though the connection has long been implicit in development literature through concepts such as efficiency wages.[2] Health effects have not been explored enough because databases seldom include both health variables and measures of wages or productivity, and also because of the empirical difficulties in identifying them.

Two problems arise in estimating the impact of health on earnings. First, measurement error appears since adequate measures of health status are generally scarce in developing countries, especially for adults. The commonest available indicators involve self-reported morbidity (days ill or days disabled), but they face serious limitations due to self-reporting biases (Schultz and Tansel, 1997). Second, endogeneity of health represents potential biases in the estimates. It is not only that healthier individuals tend to be more productive, whether at work or school, making them more likely to receive higher wages in the labor market. Richer individuals will also tend to be healthier, either because they have more resources to spend on health or because they have more knowledge about how consumption choices and behaviors affect personal health. In that context it is difficult to infer the impact of health policy interventions on individuals' earnings.

This chapter estimates the effects of health—as measured by the days ill that adult respondents report for the four weeks prior to being surveyed—on the average hourly income of wage earners and self-employed individuals in urban Peru. The database comes from the 1994 Peruvian Living Standards Measurement Survey (PLSMS). In principle, it is assumed that an equation representing the determination of wages includes an unobservable health status variable, which must be approximated by the observed morbidity indicator. Using these tools, this study analyzes the nature of the interaction between the two classical measures of human capital, education and health. Secondly, the study examines the effects of health on wages using different subsamples based on gender, age, income rank, and type of

[2] For an overview, see Strauss and Thomas (1995).

employer (public or private; large or small) and employment (blue collar or white collar; wage earner or self-employed).

To properly estimate these effects, a predicted value for health status from a health equation is used to control for the endogeneity and measurement error in the available health indicator. The health equation is obtained by solving a household model that has health in the utility function, and a particular production function for health. Individual health outcomes are thus a function of individual, family, and regional characteristics. To identify health in the wage equation, sanitary infrastructure and the availability of health services are used as variables that directly affect health but not wages.

The analysis will proceed in six stages. First, we describe the model used to derive the wage and health equations. Second, the nature of the data in the 1994 PLSMS is discussed, focusing on the relationship between health and wages. Third, we describe the econometric model used to estimate both equations and discuss the strategy for correcting bias. Fourth, the relationship between health and earnings is examined more deeply by looking at different subpopulations. The final section summarizes the findings and weighs the implications for policy and research.

The Conceptual Model

The first step is to describe how health affects individual and household decisions. This analysis is based on a household model with constrained maximization of a joint utility function, following the framework initiated by Becker (1965). It is assumed that a household with n members is run by the household head who maximizes a utility function (U), which depends on the consumption, health, and leisure of all members,[3]

$$U = U\left(C^i, h^i, l^i\right) \qquad i = 1,2,\ldots,n \qquad [5.1]$$
$$C^i = \left(C^i_1,\ldots,C^i_j,\ldots,C^i_j\right) \qquad i = 1,2,\ldots,n$$

[3] This is equivalent to assuming that household members have identical preferences, that a dictator rules the household, or generally, that it is a unitary household model. Assuming that bargaining determines intrahousehold allocations complicates the results without providing additional insights for the goals of this paper.

where, i.e., C^i is a J dimensional vector, with elements corresponding to a commodity group, h^i denotes the health status, and l^i denotes the leisure of member i. It is assumed that the utility function is continuous, strictly increasing, strictly quasi-concave, and twice-continuously differentiable in all its arguments. Also, it satisfies the Inada conditions, i.e., the marginal utility $U_x \to \infty$ as $x \to 0$, for $x = c^i, h^i, l^i$, for all examples of i.

The health status of each household member is determined by a general production function:

$$h_i = h_i\left(C^i, Y^i, l^i, Z^i, X^{-i}, Z^{-i}, F, u^i, u^{-i}\right) \qquad i = 1, 2, \ldots, n \qquad [5.2]$$

where Y^i denotes the consumption of health-related inputs by the individual (i), Z^i denotes the member's observed characteristics, F denotes the access to sanitary/medical infrastructure, and u denotes the vector of unobserved characteristics. Also, X^{-i} denotes the consumption, health, and leisure of the other members of the family; and Z^{-i} and u^{-i} denote their vectors of observed and unobserved individual characteristics, respectively. The specific variables that appear in the health production function change if the i^{th} member is an adult, a child, or an infant. For instance, in a child's health production function, milk consumption and parental education are important components in C^i and Z^{-i}, respectively, although they would probably be unimportant in an adult's health production function. Since adults tend to take care of themselves, only their own education would matter. In the case of adults, the set of unobservable characteristics (u) includes the health and nutritional status in earlier years, especially childhood.

The household also faces a full-income constraint, which is derived from the time and income constraints:

$$\sum_{j=1}^{J} \sum_i P_j C^i_j = \sum_{k=J+1}^{k} \sum_i P_k Y^i_k + \sum_i w l^i = \sum_i w T^i + V = S \qquad [5.3]$$

where P_j represents the price of the j^{th} consumption good, P_k represents the price of the k^{th} health-related input, V is nonlabor income, w is the wage rate, T^i is the total time available for the i^{th} adult member, and S is the full income. Nonlabor income (V) includes net profits of any home enterprise, as well as other rents.

The reduced-form health demand function for adults would be as follows:[4]

$$h^{i^*} = h\left(P_c, P_y, S, F, Z^i, u^i\right) \qquad [5.4]$$

Although the conditional demand functions have the usual properties, this is not true for the reduced-form demand equations in [5.4]. The key point is that consumption affects health too, and substitution effects may attenuate some of the direct effects. For instance, as pointed out by Pitt and Rosenzweig (1986), a decrease in the price of health services (P_y) or an improvement in sanitary or housing infrastructure (F) could generate substitution in consumption patterns that can reinforce or attenuate the positive health effects of such changes. Consequently nothing conclusive can be said about the effect of prices or even of sanitary infrastructure upon health status, prior to empirical analysis.

Finally, following Mincer (1962) or Mincer and Polacheck (1974), the hourly wage equation is defined as a function of the individuals' working experience (A), education (E), health status (h), and regional variables that characterize local labor markets (L), as in:

$$w_{ij} = w\left(A_{ij}, E_{ij}, h_{ij}, L_j; \varepsilon_{ij},\right) \qquad [5.5]$$

The specific functional forms for equations [5.4] and [5.5] to be used in the empirical analysis are discussed after the description of the data.

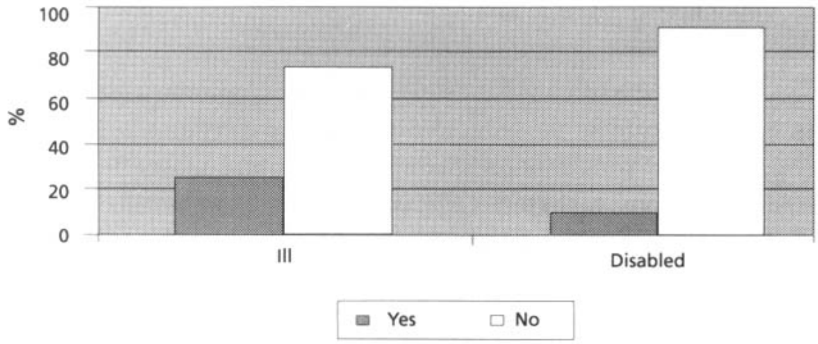
The Peruvian Data

The 1994 PLSMS contains information on adult morbidity and net income for wage earners and the self-employed, as well as characteristics of 19,284 individuals in 3,623 urban and rural households.[5] The analysis in this chapter

[4] The health production functions are assumed to be twice-continuously differentiable, strictly increasing, and strictly concave in all arguments. Then the constraint set formed by [5.2] and the full-income constraint [5.3] is convex, and the optimization of [5.1] yields a unique solution. Assuming that the health production functions satisfy the Inada condition guarantees the solution to be interior.

[5] See Grosh and Glewwe (1995) for more details on LSMSs.

Figure 5.1 Self-Reported Ocurrences of Illness and Disability for Peruvian Adults (%)



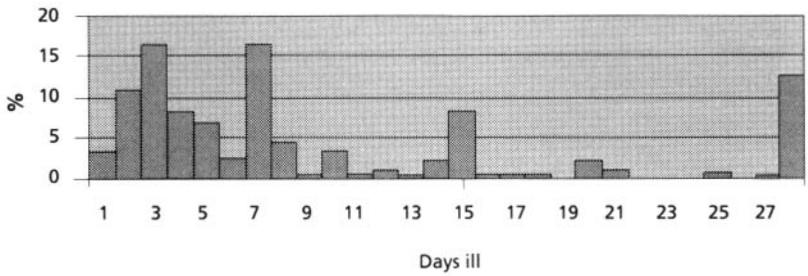
is restricted to urban areas because rural labor markets have a very different pattern of wage determination that would probably require a more complicated model.[6] Further restricting the analysis to adults between the ages of 16 and 60, the final sample tqJe is 6,610 individuals, of whom 3,102 are males.

The health measure used in this chapter is the number of days sick or injured reported by respondents for the four weeks prior to being surveyed. Typically LSMSs include a sequence of questions on health and sickness. They begin by asking, "Have you had any illness or injury during the past four weeks?" and then ask about the number of days sick, the number of days disabled due to sickness, and other related information. Among these variables, the number of days sick was chosen because it was presumed to contain more information about the latent health status of individuals. That is, it was presumed that an individual with fewer days sick in the last four weeks is generally *healthier* than one who reports more.

The typical problem with self-reported health status indicators is that they are contaminated by other individual characteristics such as education and other measures related to the ability to detect illnesses. Given two similar individuals, the one with more unobserved skills to detect an illness would

[6] Valdivia and Robles (1997) look at Peruvian labor markets, providing information on rural and urban differences.

Figure 5.2 Distribution of Days Ill for Individuals Reporting Illness (%)

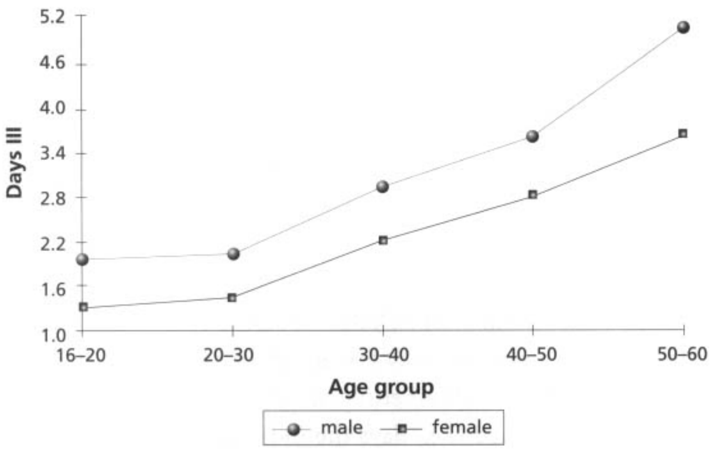


tend to report a larger number of sick days than the one who is less skilled. This would generate a positive bias for health variable coefficients. Thus Schultz and Tansel (1997), in their study of Ghana and Côte d'Ivoire, used the number of days *disabled* (i.e., days the individual was unable to work due to illness), assuming that this indicator is less biased since it is arguably more objective. However it also loses considerable morbidity information since those sick but not disabled are considered to be just as healthy as are those who experienced no illness at all. In this sense, there is a tradeoff between using a stricter measure that loses information, or using a measure that is more subjective but also more informative.

In the PLSMS, 27 percent of the individuals between 16 and 60 years old reported having been ill or in an accident during the four weeks before the survey (see Figure 5.1). About a third of those individuals also report having been disabled.

The days-ill variable is not only censored at 0 but also at 28 days since the question refers to the number of days that the individual was ill during the prior 4 weeks. As mentioned before, 73 percent of surveyed individuals did not report any sickness. Figure 5.2 shows the distribution of individuals by the number of days they reported being ill, including only those who had at least one reported day. Note that 14 percent of this subsample reported being sick for all 28 days. Overall, the average number of days ill is about 2.5, but the number is positively correlated with age, as can be seen in Figure 5.3. For instance the average number of days ill is less than 2 for individuals under 30, but rises above 4 for individuals over 50. This trend is not only the result of a higher probability of getting sick as one ages, but also of older people tending to get sick for longer periods.

Figure 5.3 Average Reported Days Ill, by Gender and Age Group (Includes respondents with no days ill)



The main question here is the connection between health and productivity, with the latter measured by earnings in the labor market. Thus one must understand how features of this particular labor market are likely to affect wage determination. One cannot ignore the fact that a considerable fraction of the labor force is self-employed, and the percentage has been growing over time. In the 1994 PLSMS, about 33 percent of individuals aged 16 to 60 were employed in the wage sector, while almost 27 percent were self-employed.[7] These proportions represent 56 percent and 44 percent of all working individuals, respectively. The rest were unemployed, working as unpaid family workers, or not participating. The self-employment rate also differs by gender, accounting for 49 percent of working women, but only 41 percent of working men.

Earnings are measured by the average hourly wage or net income during the 12 months prior to the interview. This indicator varies significantly between sectors, with the self-employed earning more, on average, than

[7] Descriptive statistics for the full sample can be found in Appendix B of Murrugarra and Valdivia (1999).

wage earners. When differences are examined by gender, self-employed men seem to be better paid than their wage sector counterparts. Among women, however, wage earners fare slightly better, but the difference is insignificant. This may reflect a gender gap in the accumulation of human capital since self-employed women have four fewer years of education on average. Differences can also be observed in morbidity indicators. The self-employed report the most days ill during the previous four weeks, followed by those who are not participating in the labor force. Wage earners report the fewest days ill.

Given the observable differences among individuals in different employment sectors, other unobserved variables in the wage equation could also affect the sector selection process, potentially generating some self-selection bias. Therefore one must control for such selectivity in order to estimate wage equations for each sector. The next section discusses the econometric techniques used to address some of the potential biases in this study.

The Econometric Methodology for Resolving Bias

This section discusses the entire system of regressions that are necessary to address the selectivity, endogeneity, and measurement error biases that were discussed above. This system is composed of three equations: the health status equation, the labor force participation equation, and the wage equation. First, the model specification for each of these equations is presented. Then, the simultaneity issues that complicate the estimation are discussed. Finally, a complete strategy to estimate the desired unbiased parameters is described.

The Health Status Equation

First, we specify a latent health variable (h^*) that is determined by a linear specification of equation [5.4] from the earlier discussion of the conceptual model. In particular,

$$h_i^* = H_i'\beta = u_i, \text{ where } H_i = \left(P_c, P_y, S, F, Z^i\right).$$

As we observe only the proxy, number of days ill (h), the relationship between h^* and h is given by the following conditions:

$$h_i = \begin{cases} 0 \text{ if } h_i^* = H_i'\beta + u_i \le 0, \\ h_i^* \text{ if } c > h_i^* = H_i'\beta + u_i > 0, \\ 28 \text{ if } h_i^* = H_i'\beta + u_i \ge c. \end{cases} \quad [5.6]$$

The double censoring of the proxy *(h)* requires that h^* be estimated through a two-limit tobit model. Following Lee (1981), predicted values $\hat{h}_i = H_i'\hat{\beta}$ are obtained, and this constructed variable is inserted in the wage equation. To identify health (*H*) in the wage equation, food prices, housing infrastructure, and indirect health cost are used as instrumental variables. These variables are assumed to be highly correlated with individuals' health status and uncorrelated with individual wages. Both assumptions will be tested with appropriate specification tests.

According to equation [5.4], self-reported individual health status depends on prices of consumption and health-related goods, total household income, access to sanitary infrastructure, and observed and unobserved individual characteristics. Here, self-reported morbidity events for the four weeks prior to the survey are used as a proxy for (negative) health status.

For measures of full income (*S*), the household's productive and nonproductive assets are used. The values of the house and other durable goods are included as nonproductive assets. A dummy for the operation of a household business is included as a measure of productive assets. Reported profits from household businesses are not used since there is considerable reporting error in the data, a phenomenon previously observed in other developing countries (Thomas and Strauss, 1997).

Individual observed characteristics include the age of the individual and his/her education, among others. Measures of sanitary infrastructure (*F*) include the existence of indoor connections for potable water and sewerage, and whether the dwelling has proper floors, walls, and roofing. Dirt floors tend to provide less-healthy environments. For prices of health services, proxies are constructed using the costs of service access, e.g., the distance to the health center and the waiting time before treatment. Since these indicators are reported only by those who were sick and received medical care, the district averages are used. Given intradistrict correlation of errors, the corrected covariance matrix suggested by Moulton (1986, 1990) is applied.

The Sector Choice Model

As discussed previously, the estimated wage equations in each sector must be corrected for potential selectivity biases. The choice of sector is estimated as a multinomial logit (MNL) model, from which correction terms are constructed following Lee (1982, 1983). Specifically, we identify three employment status choices (0 = nonparticipation, 1 = wage earners, and 2 = self-employment),[8] and estimate an MNL in which nonparticipation is the base category and in which the underlying latent variable is a linear function of a set of observable characteristics of the individual (B_i). That is,

$$b^*_{Si} = B_i^{'} \gamma_S + \sigma_S \nu_{Si}, \qquad [5.7]$$

where ν_{Si} is an unobserved error term and σ_S is the variance of this term. Then, an individual (i) chooses to be in sector (k) if:

$$b^*_{ki} = \max \left\{ b^*_{Si} \right\}, \text{ for } S = 0, 1, 2. \qquad [5.8]$$

This choice model has a utility interpretation in which b^*_{ki} is interpreted as the indirect utility for an individual (i) of choosing sector S. A utility-maximizing individual will behave as in equation [5.8]. This equation includes controls for individual, household, and community characteristics. The variables used to identify labor force participation are community labor market characteristics, household assets, and other nonlabor income, as defined for the health status equation.

The Wage Equation

Given the employment status choice, each individual has an associated wage in each sector, which is described by the following equation:

$$\ln\left(w_{Si}\right) = \delta_{S0} + \delta_{S1} E_{ij} + \delta_{S2} A_{ij} + \delta_{S3} \hat{h}_{ij} + \delta_{S4} L_j + \varepsilon_{Si}, \text{ for } S = 1, 2 \qquad [5.9]$$

[8] The self-employment sector includes both independent workers and owners of firms.

since no equation is estimated for the nonparticipants. The earnings equations for wage earners and self-employed individuals are estimated, taking into account the necessary selection correction that is analogous to the Heckman two-step estimation procedure (see Lee, 1983).

The selection correction term variable is identified in the wage equation through the inclusion of household assets and nonlabor income in B, which are assumed to be highly correlated with individual participation in the wage labor market but uncorrelated with individual wages. Both assumptions will be tested with appropriate Wald tests.

The controls include potential work experience and its square, defined as the number of years after leaving school. The terms for experience and education in [5.9] could be included in different ways. Age and health, for instance, could include a linear as well as a quadratic term. Schooling is included in a spline specification in order to capture differential returns to education at each education level. Three terms are included. First is the simple number of years of schooling, which is interpreted as the returns to primary education. The second term, the secondary spline, is the product of a dummy that equals 1 if the individual reached some level of secondary education, multiplied by the number of years. It captures the *additional* returns to secondary schooling, defined as years 7 to 11 of education. The third term, the higher spline, is defined analogously for postsecondary schooling and captures the returns to higher education beyond results obtained at the other two levels. Consequently, the returns to an additional year at the university, for instance, would be obtained by adding the value of all three coefficients.

Firm-size controls were also included in the earnings equations for wage earners, following the research by Anderson Schaffner (1998). A firm is classified as a micro firm if it has between 6 and 10 workers. Firms having between 11 and 20 workers are classified as small; those with more than 20 but no more than 200 are medium-sized; and those with more than 200 are large. The omitted category is the firm that has one to five workers. Even though wage effects of employer size are not totally explained by unionization, regulatory compliance, or higher worker quality, the wage premium associated with working for a larger employer is still found to be important in developing countries, and in Peru in particular (Anderson Schaffner, 1998). In order to capture the potential employer-size wage premium, indicators were included for the four firm categories described above: micro, small,

medium, and large. A control was also included for the degree of market arrangements in each province that may affect the hourly payment in either wage or self-employment.

The community variables included in the model are the local unemployment rate and the percentage of hired labor in the locality. For each province, the share of labor days worked in exchange for a wage is the indicator of local labor market development. Presumably when local labor markets are more developed, the number of labor days worked for a wage rises relative to the number of labor days worked by the self-employed and by unpaid family members. The selection correction term is reported as "IMR" in the tables. Regional dummies are also included to control for geographic differences among the regions compared to Lima (the omitted category).

Some Empirical Issues in the Estimation of the Effects of Health Status on Productivity

In the 1994 PLSMS, a relatively small percentage of individuals reported being ill. This presents a problem when estimating the relationship between health and productivity because the health indicator contains less information. That is, the health indicator cannot distinguish health status differences among those who did not report any sickness. This problem is circumvented by using the predicted estimate from a tobit model as an index for the latent health variable. By using the estimated index, it is assumed that a *healthier* person has a lower index (more negative), providing a basis for evaluating whether this implies a higher productivity for the individual. This estimated index is used in the second-stage wage equation.[9]

Some further caveats regarding the econometric strategy should be mentioned. First, unobserved heterogeneity in the health status regression could cause several problems. If those unobserved characteristics are correlated with the regressors, the estimates might be biased. On the other hand, if these unobserved components are not correlated with observed characteristics, heteroscedasticity patterns might appear. Even though this problem does not affect consistency in linear models (OLS), it is a

[9] A referee pointed also to the possibility of using a censored estimated index. The large censoring fraction made this inappropriate. Moreover, to the extent that the symmetry assumption in the tobit model holds, the estimated index recovers the censored latent variable.

potential source of inconsistency in nonlinear models (see Maddala, 1981, p. 179). The conditional expectation, however, is still consistent, and this is the instrument used in the second-stage estimation. Second, the tobit equation actually identifies the parameters of the model based on the uncensored section of the data. The conditional expectation used in the second stage is based on those estimates to predict the latent health status variable for the entire working sample. The assumption is that the estimated parameters from the health equation [5.6] are common for all individuals. Third, to obtain the conditional expectation of the health status through the tobit estimates, a normality assumption is imposed. While this is a direct approach to the problem, the conditional expectation loses any nonnormal behavior that could be present in the latent health variable. If these data peculiarities are important, a more flexible approach should be examined. It must be mentioned, however, that these caveats could not be circumvented when the censored portion is relatively large, as in the case of this study.[10]

Health and Productivity

The Health Status Equations: A Two-Limit Tobit Model

The health status equation is estimated as a function of individual characteristics (age, schooling), household characteristics (such as assets and housing infrastructure), cost of health services (distance and waiting time for medical care), and regional and date-of-interview controls. In addition to age and education, the individual characteristics include rural background through a dummy that indicates who migrated from the countryside. Some interaction terms are also included: first, the interaction between schooling and age and the implicit health prices; and second, the interaction between rural background and age.

Table 5.1 shows the results of the estimation of the tobit model for males and females.[11] Since days ill is negatively associated with health status,

[10] For more details on alternative estimation strategies and qualifications of these results, see Murrugarra and Valdivia (1999).

[11] Regressions with alternative dependent variables and specifications can be found in the Appendix of Murrugarra and Valdivia (1999).

Table 5.1 Two-Limit Tobit Health Regressions of Days Ill for Adults by Sex and Region

Variables	Males	Females
Individual Characteristics	*83.59***	*69.6***
Age	0.675**	0.243**
	5.67	*2.96*
Years of schooling	0.720	–0.735
	1.18	*–1.51*
Schooling x Age (x 10^{-2})	–3.996**	–0.066
	–3.51	*–0.09*
Rural migrant	11.539**	–1.982
	3.55	*–0.59*
Rural migrant x Age	–0.236**	0.076
	–3.04	*0.88*
Household Assets	*5.98*	*0.77*
Nonlabor income	0.132	0.056
	1.24	*0.64*
Nonproductive assets	–0.093	0.024
	–0.51	*0.14*
Home business	2.471**	0.572
	2.13	*0.59*
Housing Infrastructure	*6.12***	*7.19***
Adequate roofing	–2.707**	–2.864**
	–2.29	*–2.63*
Adequate floors	–0.533	0.394
	–0.45	*0.33*
Health Infrastructure	*7.84**	*4.45*
Distance time to health service	–13.375**	–9.651*
	–1.80	*–1.64*
Waiting time for medical attention	–0.015	–2.649
	–0.01	*–1.25*
Travel time to health service x Schooling	1.314**	0.681
	2.40	*1.33*
Waiting for medical attention x Schooling	0.123	0.246
	0.50	*1.20*
Food Prices	*5.59**	*14.27***
Potato price	7.265**	5.087
	2.36	*1.35*
Milk price	3.688	13.330
	0.83	*3.58*
Log likelihood	–3,399.62	–4,863.72
Global Chi^2	281.00	387.27
Number of observations	3,083	3,486

Notes: A constant and control variables for regions and interview months are included in all regressions but not reported; t-statistics are in italics.

* Statistically significant at 10% level of confidence.

** Statistically significant at 5% level of confidence.

a positive coefficient means that the corresponding variable affects the individual's health status negatively. Similarly, the positive coefficient for age indicates that the health of individuals deteriorates with age. The negative coefficient for the housing infrastructure variables, on the other hand, indicates that individuals living in houses with better quality tend to be healthier. It is important to notice the negative coefficient found with regard to distance to the closest health facility. This indicates that the more accessible the health facility is, the worse the health indicator is. Presumably this is due to the fact that people with access to health facilities are more likely to report illness rather than being a sign that health facilities aggravate illness. A similar, counterintuitive finding can be seen for education, an effect that is discussed below.

Household permanent income—proxied by the per capita value of family assets (see Table 5.1)—is included in the regression analysis, but none of the variables appear to be significant. Although any theoretical framework would seem to require some measures of household income, previous studies have not succeeded in obtaining a strong empirical relationship (e.g., Schultz and Tansel, 1997). Two alternative measures—per capita household expenditures (exogenous) and per capita full income (instrumented)—were also tried, but neither was significant.

The analysis presents consistent and robust results for men but not particularly for women. Nevertheless the proposed instruments for endogenizing health in the wage equation are strongly significant for both subsamples. Recall from section 2 that key instrumental variables are the indicators for food prices, housing infrastructure, and implicit prices for health inputs. Access to a house with adequate floors and roofing does improve the health status of adults.

Differences of Human Capital Effects over the Life Cycle

The interaction between the different forms of human capital is of particular interest for public policy. Human capital can include age, completed years of schooling, and migration experience, over the life cycle (across cohorts). The first finding is that even after controlling for other factors, the health of Peruvian adults clearly deteriorates with age, an effect that is stronger for male adults. When included, the quadratic term for age was not found significant, a result consistent with the strong linear correlation observed in Figure 5.3.

Somewhat surprisingly Table 5.1 seems to indicate that education only significantly affects health for males. Nevertheless, when calculating net schooling effects (Table 5.2), it can be seen that this variable is also significant for women at particular ages.[12] But there are, indeed, important gender differences in the role of human capital endowments for the determination of individual health status. First, more-educated men have more reported days ill when they are younger than 33, but the effect turns increasingly negative for older men. This result differs from those of Strauss et al. (1993), in which anthropometric measures are used, and would indicate differences in the magnitude of the self-reporting bias over the life cycle. That is, young noneducated males may be less inclined to report illnesses than their more educated counterparts. At older ages, though, the seriousness of the illnesses may eliminate subjectivity in reporting. On the other hand, the schooling effect for females is constantly around –0.25 over the life cycle. This implies that the schooling effect for individuals around the age of 30 is larger (more negative) for women than for men, but much smaller afterwards. These results were corroborated with regressions by age groups (Murrugarra and Valdivia, 1999, Appendix).

The other indicator of human capital included for the regressions for urban areas was the dummy variable indicating whether the individual had migrated from a rural area. The positive coefficient in Table 5.1 indicates that, other things being equal, a rural background implies poorer health status for adult males. This could be a consequence of poorer environmental or economic conditions in childhood. It is also possible, though, that such an indicator could be capturing some ethnic differences since it is limited to migration events from rural areas where there is a strong concentration of indigenous peoples. But if that were the case, the effect would be persistently negative (or positive) throughout the life cycle. That is not the case, though, since the coefficient of the interaction term is clearly negative. Table 5.2 shows that the net effect is indeed negative for males over 45 years of age. These findings

[12] Net effects reported in Table 5.2 refer to the variation of the schooling and migration effects due to the interaction terms with age and implicit health prices included in the tobit model of Table 5.1. Table 5.2 also includes the estimated net effects associated to a regression without the interaction terms for education and implicit health prices. It is interesting to note that the individual coefficients for education change significantly, but do not do so for the estimated net education effects.

Table 5.2 Net effect on Health by Gender and Age Group

Age	Male (1)	Male (2)	Female (3)	Female (4)
Net schooling effect on health				
16 years	0.70**	0.69**	–0.25	–0.24
	(0.01)	(0.01)	(0.29)	(0.30)
30 years	0.13	0.13	–0.25*	–0.25*
	(0.37)	(0.38)	(0.09)	(0.10)
45 years	–0.47**	–0.47**	–0.26**	–0.26**
	(0.00)	(0.00)	(0.02)	(0.03)
60 years	–1.08**	–1.07**	–0.27	–0.27
	(0.00)	(0.00)	(0.12)	(0.12)
Net migrant effect on health				
16 years	7.59**	7.76**	–0.63	–0.77
	(0.00)	(0.00)	(0.77)	(0.72)
30 years	4.44**	4.46**	0.40	0.30
	(0.00)	(0.00)	(0.74)	(0.80)
45 years	1.06	0.92	1.50	1.44
	(0.40)	(0.47)	(0.18)	(0.20)
60 years	–2.31	–2.62	2.60	2.58
	(0.25)	(0.19)	(0.22)	(0.23)

Notes: Results in columns (1) and (3) refer to regressions reported in Table 5.1. Results in columns (2) and (4) are based on a regression that includes interaction terms between schooling and housing infrastructure. Numbers in parentheses are the corresponding p-values. For values of (*) and (**) see Table 5.1.

support the interpretation that the migration indicator captures previous sanitary and economic conditions that put migrants at an initial disadvantage compared to urban natives. The other implication is that evidence suggests that the gap in health conditions *closes* significantly as time passes.

Work Status Choice: Nonparticipation, Wage Earning, and Self-Employment

The potential selection bias in the estimation of the earning equations is corrected here using the MNL-OLS two-stage correction method proposed by Lee (1983). In Table 5.3, MNL sectoral choice estimates are shown for

both men and women. Education is controlled in the sectoral choice model by including three dummies for each level of formal education, and interactions of these dummies with the number of years of schooling. This flexible functional form is intended to capture any nonlinearity in the effects of education on choice of employment sector. Results indicate that education increases the participation of individuals in a nonlinear way. Moreover, as education increases, the likelihood of participation increases faster for joining the wage sector than the self-employment sector.

Wealth and income control variables included measures of assets, nonlabor income, and the existence of a household business. The accumulation of nonproductive assets has an important negative effect on joining any sector, and is particularly strong for wage employment, probably reflecting some complementarity between household nonproductive assets (e.g., a stove) and self-employment activities (e.g., cooking). The home business variable also can be interpreted as the existence of another income source, which clearly increases the propensity to join the self-employment sector as a firm owner. However this finding differs by gender. There is little effect among men, probably because they are strongly attached to the wage sector. The effect is negative among women, which is consistent with traditional norms that women are more commonly attached to home-related activities. Possible discrimination or labor market segmentation cannot be excluded as contributing factors.

The effect of the local unemployment rate is positive for both genders but weaker for males. Individuals living in provinces with larger unemployment rates tend to participate more, especially among women. In other words, provinces with weak economic growth have higher labor force participation by women, possibly driven into the labor market to find additional household income. By contrast, the proportion of wage employees in each province does not have a significant effect on the sectoral choice. There is only an imprecise negative effect on the probability of joining the wage sector among women.

The estimates from this sectoral selection model are used to construct selection correction terms for use in the wage equation as explained below.

Wage Equations with Corrections for Endogeneity and Selection

Using the information obtained from the health status and sectoral choice equations, it is now possible to analyze the determinants of wages, including

Table 5.3 Participation Equation for Urban Adults: A Multinomial Logit Model

	Males		Females	
Variables	Wage Earner	Self-Employed	Wage Earner	Self-Employed
Individual Characteristics	*493.22***	*581.56***	*323.59***	*279.81***
Age	0.594**	0.702**	0.228**	0.393**
	(18.29)	(18.85)	(8.48)	(14.26)
Age² (x 10⁻²)	–0.734**	–0.824**	–0.341**	–0.469**
	–(16.69)	–(16.65)	–(8.84)	–(12.88)
Dummy primary school	2.130**	1.649**	–0.403	–0.145
	(3.06)	(2.29)	–(0.92)	–(0.45)
Years of school (Primary)	–0.130	–0.015	–0.041	0.043
	–(1.16)	–(0.13)	–(0.53)	(0.78)
Dummy secondary school	0.976	1.038	–1.580**	0.033
	(1.26)	(1.21)	–(2.29)	(0.05)
Years of school (Secondary)	0.061	0.048	0.142**	–0.003
	(0.94)	(0.66)	(2.28)	–(0.05)
Dummy higher education	–1.695*	–0.395	–5.427**	–1.219*
	–(1.63)	–(0.33)	–(6.29)	–(1.04)
Years of school (High)	0.204**	0.093	0.441**	0.077*
	(3.04)	(1.20)	(7.77)	(0.97)
Household Assets	*28.89***	*344.27***	*12.41***	*611.09***
Nonlabor income (x 10⁻²)	–0.467	–2.770**	0.688	0.246
	–(0.42)	–(2.12)	(0.77)	(0.25)
Nonproductive assets	–0.155**	–0.144**	–0.046**	–0.077**
	–(5.25)	–(4.45)	–(2.27)	–(3.80)
Home business	–0.007	21.330**	–0.237**	20.785**
	–(0.06)	(18.35)	–(2.44)	(24.67)
Labor Market Variables	*6.24***	*1.89*	*7.41***	*4.96**
Unemployment (% by prov.)	1.556*	1.426	2.073**	1.764**
	(1.74)	(1.37)	(2.65)	(2.21)
Hired rate (by province)[a]	0.641	–1.019	–1.384*	–0.964
	(0.69)	–(0.85)	–(1.85)	–(1.02)
Log likelihood	–2290.5	–2290.5	–2770.3	–2770.3
Global Chi²	546.4**	205851**	349.6**	338244**
Chi²	33.18**	798.07**	17.78**	1657.94**
R^2	0.305	0.305	0.208	0.208
Number of observations	3,102	3,102	3,508	3,508

Note: See notes to Table 5.1, including values of (*) and (**).

a/ Refers to the proportion of the employed individuals that work for a salary and not independently.

the effects of health status. After making these corrections, there appears to be a robust positive effect of health status on productivity and wages, and also some evidence of interaction between health and education.

In Tables 5.4a and 5.4b, different specifications for the earnings equations are presented separately for wage earners and the self-employed, respectively. The dependent variable in all cases is the logarithm of net earnings. The tables show the estimated earnings equations without health (columns 1 and 4), which are discussed first in order to establish the consistency of the specification with previous estimations. For instance, an important finding in terms of the estimated returns to schooling is that they are higher for women and increase with the level of educational attainment, especially among wage earners. The use of a spline specification for the returns to schooling allows this pattern to be discerned. The first coefficient in column 1 of Table 5.4a means that each year of primary schooling is associated with an increase of 5.2 percent in male earnings. The second coefficient is not significantly different from 0, which means that the return for each year of secondary school is about the same. Finally, the third coefficient means that each year of postsecondary school would increase wages for men by about 13 percent (5.2 + 1.2 + 6.6).

Among women, returns to primary and secondary schooling go as high as 8.7 percent, which is 3.5 percentage points above the returns for men. Nevertheless this gender difference almost vanishes when individuals reach higher education. The results in Table 5.4b for the self-employed show a very similar pattern, although less precisely. The return to primary schooling is 5.0 percent and 7.7 percent among these men and women, respectively, reproducing the gender difference observed among wage earners. Returns to higher education increase for self-employed men, but no evidence of such a pattern is apparent for women.

In terms of firm size, the results indicate a significant positive effect, consistent with Anderson Schaffner (1998). Among male wage earners, the premium is about 16 percent for those employed by micro firms (6–10 workers), increases to 23 percent for small firms (11–20 workers), 36 percent for medium firms (21–200 workers), and remains almost constant afterwards. Among women, the estimated firm size premium shows a similar pattern, although no significant effect is found for micro firms.

Tables 5.4a and 5.4b also show the wage regressions including health, treated as both exogenous and instrumented (IV). When health is included

Table 5.4a Equation for Log Hourly Earnings for Urban Wage Earners by Gender: Instrumenting Health

	Male			Female		
Variable	**No Health**	**Exog. Health**	**IV Health**	**No Health**	**Exog. Health**	**IV Health**
Individual Human Capital Variables	*94.45***	*70.99***	*72.43***	*28.39***	*22.24***	*23.32***
Years of schooling (x 10^{-2})	5.161**	5.186**	4.757**	8.709**	8.753**	8.678**
	(2.31)	(2.33)	(2.13)	(2.94)	(2.94)	(2.90)
Years of schooling minus 6 (x 10^{-2})	1.206	1.177	1.753	−2.295	−2.657	−2.566
	(0.42)	(0.41)	(0.60)	−(0.53)	−(0.61)	−(0.58)
Years of schooling minus 12 (x 10^{-2})	6.609**	6.607**	6.384**	7.020**	7.298**	7.539**
	(3.48)	(3.48)	(3.37)	(2.35)	(2.44)	(2.51)
# of days sick, instrumented (x 10^{-2})		0.081	−1.216**		−0.552	−2.410**
		(0.20)	−(2.23)		−(1.28)	−(2.32)
Other Individual Characteristics	*37.24***	*36.69***	*33.72***	*22.91***	*23.03***	*26.64***
Potential experience	0.029**	0.029**	0.029**	0.030**	0.030**	0.036**
	(4.86)	(4.87)	(4.97)	(3.79)	(3.80)	(4.39)
Potential-experience squared (x 10^{-3})	−0.313**	−0.314**	−0.268**	−0.279	−0.277	−0.277
	−(2.44)	−(2.45)	−(2.08)	−(1.36)	−(1.35)	−(1.35)
Firm Size and Local Labor Market Characteristics	*12.20***	*12.19***	*11.91***	*7.81*	*7.81***	*7.88***
Micro-scale firm	0.156**	0.156**	0.155**	5.4E-03	7.6E-03	8.1E-03
	(2.76)	(2.77)	(2.74)	(0.07)	(0.10)	(0.11)
Small-scale firm	0.232**	0.231**	0.233**	0.213**	0.216**	0.212**
	(3.69)	(3.67)	(3.71)	(2.62)	(2.66)	(2.61)

Table 5.4a (continued)

Variable	Male			Female		
	No Health	Exog. Health	IV Health	No Health	Exog. Health	IV Health
Medium-scale firm	0.355**	0.355**	0.347**	0.327**	0.328**	0.328**
	(7.20)	(7.21)	(7.04)	(4.62)	(4.63)	(4.70)
Large-scale firm	0.399**	0.400**	0.401**	0.371**	0.375**	0.367**
	(5.36)	(5.36)	(5.39)	(3.79)	(3.83)	(3.75)
Hired rate (by cluster)	–0.019	–0.018	–0.052	0.294*	0.289*	0.301**
	–(0.16)	–(0.15)	–(0.43)	(1.76)	(1.74)	(1.81)
Selection term	–0.042	–0.042	–0.020	–0.072	–0.075	–0.025
	–(0.81)	–(0.80)	–(0.38)	–(0.56)	–(0.59)	–(0.19)
Number of observations	1,543	1,543	1,543	844	844	844
F-test	39.45**	37.17**	37.52**	33.00	32.22**	31.90**
R^2	0.281	0.281	0.284	0.351	0.352	0.355
Exogeneity test (Hausman)			11.79			3.83
			(0.00)			(0.05)
Overidentification test			19.37**			18.16**
			(0.02)			(0.03)

Notes: A constant and regional control variables are included in all regressions but not reported; numbers in italics refer to the Wald statistics for the joint-significance tests; t-statistics are in parentheses and italics.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

Table 5.4b Equation for the Log Hourly Earnings of the Urban Self-Employed by Gender: Instrumenting Health

	Male			Female		
Variable	**No Health**	**Exog. Health**	**IV Health**	**No Health**	**Exog. Health**	**IV Health**
Individual Human Capital Variables	*54.04***	*40.85***	*42.77***	*18.66***	*14.01***	*15.47 ***
Years of schooling (x 10^{-2})	5.006	4.901*	3.527	7.665**	7.576**	7.841**
	(1.52)	(1.50)	(1.06)	(3.35)	(3.30)	(3.39)
Years of schooling minus 6 (x 10^{-2})	4.819	4.901	5.829	–0.464	–0.309	–0.758
	(1.14)	(1.17)	(1.39)	–(0.13)	–(0.09)	–(0.21)
Years of schooling minus 12 (x 10^{-2})	5.553**	5.520**	5.647*	1.420	1.336	1.555
	(1.83)	(1.82)	(1.87)	(0.36)	(0.34)	(0.40)
# of days sick, instrumented (x 10^{-2})		–0.761	–3.147**		–0.366	–2.295*
		–(1.23)	–(3.24)		–(0.71)	–(1.70)
Other Individual Characteristics	*1.14*	*1.29*	*3.85***	*3.82***	*3.97***	*5.70***
Potential experience (x 10^{-3})	3.698	2.780	–12.657	7.010	6.699	9.806
	(0.19)	(0.14)	–(0.61)	(0.28)	(0.27)	(0.39)
Potential-experience squared (x 10^{-3})	0.067	0.090	0.440	0.063	0.072	0.123
	(0.22)	(0.30)	(1.30)	(0.16)	(0.18)	(0.30)
Local Labor Market Characteristics	*0.57*	*0.72*	*0.27*	*3.05***	*2.92***	*3.24***
Hired rate (by cluster)	0.157	0.176	0.108	0.425*	0.414*	0.439**
	(0.75)	(0.85)	(0.52)	(1.75)	(1.71)	(1.80)
Selection term	–0.418	–0.436	–0.675**	–0.231	–0.238	–0.252
	–(1.49)	–(1.55)	–(2.20)	–(0.78)	–(0.81)	–(0.85)

Table 5.4b (continued)

Variable	Male			Female		
	No Health	Exog. Health	IV Health	No Health	Exog. Health	IV Health
Number of observations	1,144	1,144	1,144	843	843	843
F-test	19.54**	18.43**	18.40**	9.55**	8.82**	9.58**
R^2	0.162	0.164	0.170	0.123	0.124	0.126
Exogeneity test (Hausman)			10.08**			2.39
			(0.00)			(0.12)
Overidentification test			18.34**			12.02
			(0.03)			(0.21)

Note: See notes to Table 5.4a.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

exogenously there is no significant effect, nor does it change any of the estimates for the control variables for any gender or sector.

When health is instrumented, using the predicted value from the health equations, the number of days ill do have a significant effect on wages in both sectors, although this effect is small. Among men, an increase of one day in the morbidity indicator is associated with a 1.2 percent decrease in earnings for wage earners and a 3.1 percent decrease in hourly net income for the self-employed. One possible explanation for this result is related to sectoral differences in the observability of individual productivity. In the wage sector, the employer cannot observe individual productivity (effort) by looking at the final output when the production function is stochastic and the external shock not observed. Even in a deterministic environment, employers face an observability problem if tasks are performed in large teams.[13] The results among women show no significant difference across sectors. The effect of the morbidity indicator is –2.4 percent among wage earners and –2.3 percent among the self-employed.

The inclusion of the corrected health measure also has different implications for the estimated returns to education. In particular, estimated returns to primary education are slightly smaller after properly controlling for health. Returns to primary education for males drop from 5.2 percent to 4.8 percent in the wage sector, and the drop is even stronger among the self-employed (from 5.0 percent to 3.5 percent). Nevertheless the effect of including health for those with higher education is negligible in both sectors. Among the self-employed, for instance, returns to schooling drop only from 15.4 percent to 15.0 percent.[14]

The returns to schooling among women, though, are not particularly affected by the inclusion of health status. Perhaps the interaction between the two dimensions of human capital, education and health, is more important for men because of the kinds of work they do. It could also be argued that this result is related to some sectoral sorting of individuals based on unobserved characteristics. However, the estimated selection correction co-

[13] For a more detailed discussion of the possible explanations for the observed differences between sectors, see Murrugarra and Valdivia (1999).

[14] As indicated above, given the spline specification, the returns to education are constructed adding the estimates for each education level.

efficients indicate an insignificant selection problem between these sectors, with and without health in the wage equation (see the coefficients for the selection term in Tables 5.4a and 5.4b).

The results of estimating the wage equations with the instrumented health variable are consistent with the stories of endogeneity and measurement error discussed earlier. This is confirmed further by various diagnostic tests. Results of a Durbin-Hausman-Wu specification test and an overidentification test are shown at the bottom of Tables 5.4a and 5.4b. The Hausman specification tests suggest that the endogeneity of health status is important in the sense that the coefficient for the health indicator from the OLS estimates is significantly different from that of the IV estimates. This can be clearly observed for men, but is only true for women who are wage earners.

On the other hand, a regression-based overidentification test suggested by Hausman (1983), rather than the Basmann version, was used to test the validity of the used instruments. The results indicate that the overidentification restriction is rejected at the 5 percent level for men in both sectors and for female wage earners. These values, however, are close to the critical ones for both genders. As others have mentioned (Thomas and Strauss, 1997) overidentification and exogeneity test results must be taken carefully when the identifying instruments are weak. In this case, the corresponding pseudo-R^2 from the tobit estimates for the health equations are about 0.039 for males and 0.029 for females. The low correlation between instruments and health and the censoring of the health variable make interpretation of these tests difficult. Moreover according to Staiger and Stock (1997), weak instruments would identically affect any version of the overidentification tests, whether regression-based or Basmann, making any choice between them of little or no relevance.

Evidence by Age Groups

There are numerous reasons to expect that the impact of health on earnings would differ across different population groups, especially by age. Table 5.5 presents the wage equation estimates for wage earners and the self-employed, but this time disaggregated by age. For each sector, the sample was divided into two age groups: those who are 16 to 25 years old and those who are 26 and older.

Table 5.5 Equation for the Log Hourly Earnings of Urban Workers: Comparing Differences by Age Groups

Variable	Wage Earners		Self-Employed	
	< 26 years	> 26 years	< 26 years	> 26 years
Males				
Years of schooling (x 10^{-2})	6.161	4.029*	7.201	4.964
	(0.76)	(1.72)	(0.82)	(1.49)
Years of schooling minus 6 (x 10^{-2})	−3.635	2.722	7.524	3.081
	–(0.38)	(0.87)	(0.62)	(0.70)
Years of schooling minus 12 (x 10^{-2})	7.255*	5.854**	0.140	5.321*
	(1.78)	(2.53)	(0.02)	(1.64)
# of days sick, instrumented (x 10^{-2})	1.203	−2.008**	0.124	−4.274**
	(1.12)	–(2.84)	(0.06)	–(3.95)
Potential experience (x 10^{-3})	−0.451	25.521**	−13.668	4.804
	–(0.01)	(2.78)	–(0.20)	(0.32)
Potential-experience squared (x 10^{-3})	0.600	−0.175	3.471	0.253
	(0.26)	–(0.97)	(0.80)	(0.88)
Number of observations	443	1,100	191	953
F-test	4.04**	27.16**	6.24**	14.42**
R^2	0.115	0.281	0.171	0.161
Females				
Years of schooling (x 10^{-2})	22.475**	5.617*	−7.501	8.036**
	(3.70)	(1.76)	–(0.20)	(3.43)
Years of schooling minus 6 (x 10^{-2})	−19.524**	3.523	14.596	−2.068
	–(2.29)	(0.81)	(0.34)	–(0.54)
Years of schooling minus 12 (x 10^{-2})	13.298**	3.400	0.969	0.784
	(2.35)	(1.01)	(0.08)	(0.19)
# of days sick, instrumented (x 10^{-2})	−2.510	−2.778**	−3.484	−2.061
	–(1.22)	–(2.32)	–(0.81)	–(1.44)
Potential experience (x 10^{-3})	−43.382	36.388**	32.729	15.667
	–(0.89)	(2.59)	(0.30)	(0.99)
Potential-experience squared (x 10^{-3})	6.224*	−0.287	−2.491	−0.015
	(1.81)	–(0.97)	–(0.31)	–(0.06)
Number of observations	297	547	115	728
F-test	8.90**	19.84**	1.97**	7.93**
R^2	0.240	0.347	0.187	0.115

See notes to Table 5.4a.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

The results show that, although the size of the health effect may not be robust across age groups, the previously reported finding of larger effects among the self-employed persists. A precise negative effect of about –2 percent was found among those workers 26 or older, while no significant effect was found among younger workers. A very similar pattern occurs among the self-employed, but the –4.3 percent effect on the wages of older individuals is even stronger. Among women, the pattern is roughly the same, although the results are only statistically significant for older wage earners. There are several possible explanations for these results. First, younger workers are less likely to be sick, hence health effects on wages are harder to capture; and this may be a reason for the large standard errors in the estimates. Second and more importantly, the reporting bias identified to be stronger among the young increases the imprecision of estimates for this group. Probably due to a combination of these factors, the stronger effects are found among the older group. Nevertheless, the radical changes in the estimated coefficients for the younger group would indicate that the model is incapable of capturing the variability in wages and productivity for them.

Comparative Evidence between Public and Private Workers

In addition to age, there are other differences among workers that might affect the magnitude of the relationship between health and productivity. If the relation between health and productivity is related to the nature of the tasks performed, then the type of job or the contractual nature of the job would also be expected to have an impact. The division between public-sector and private-sector workers is a feature of the labor market that probably has an impact on the health-and-productivity relationship.

In the previous subsamples, it was hypothesized that a stronger health effect was found among self-employed workers since health (therefore productivity) and net income in this sector are more closely associated. In the wage sector, though, the connection between health and wages depends on the employer's ability to observe productivity. This subsection explores a quite different hypothesis. Health premiums in the public sector might be lower, or even negligible, compared with the private sector because of a different remuneration policy. Public-sector employers may not be encouraged to monitor their employees' effort because their objective is not to

maximize output, but includes such factors as keeping the employment level constant.[15]

The differences between the public and private sector were estimated, as well as the distinction between blue-collar and white-collar workers, in order to examine the alternative (but not exclusive) argument that health has a stronger impact among those performing more physically demanding tasks. The estimated equation is:[16]

$$\ln(w_i) = X_{ij}\beta_S + \delta_{S3}\hat{h}_{ij} + \delta_{S5}\hat{h}_{ij}JOB_{ij} + \delta_{S6}JOB_{ij} + \varsigma_{S_i} \qquad [5.10]$$

where *JOB* is an indicator that takes the value of 1 if the worker is in the public sector (or performing a blue-/white-collar job). The results of this exercise for wage earners are shown in Table 5.6. Column 1 shows that despite the small and significant effect of health on the overall population (–0.015) no significant differences exist between the health effects for blue- and white-collar workers (see variable "# of days sick for blue collar"). A similar finding among women is found (column 3), where the negative effects of illness (–2.49) are not affected by the imprecise effect for blue-collared workers. This evidence suggests that the impact of health on productivity and wages is not entirely due to differences in the physical demands represented by different jobs, since blue-collar returns to health would be higher (the impact of health would be more negative).

On the other hand, the effects of health among private workers are larger than in the public sector. Even though these differences are small, they are very precise. The estimated effect of worse health among private male workers is –1.8 percent lower productivity, slightly larger but significantly different than the net effect among male public employees (–1.6 percent). The same pattern is found among women. It must be emphasized that the "public worker" effect is not due to firm size since that effect has already been controlled for.

[15] Recent labor market deregulation may have strengthened the connection between productivity and wages in Peru, but the added flexibility has been slower and more difficult to implement in the public sector.

[16] We are aware of the problems these classifications introduce due to potential selectivity biases. A fully specified participation equation with separate wage regressions for each sector could compensate for this, but it would require additional identifying instruments that were unavailable.

Table 5.6 Equation for the Log Hourly Earnings of Urban Workers: Comparing Effects by Employment and Employer Types

Variable	Males: White collar vs. blue collar	Males: Public vs. private	Females: White collar vs. blue collar	Females: Public vs. private
Individual Human Capital Variables	*42.80***	*75.07***	*31.34***	*38.71***
Years of schooling (x 10^{-2})	0.048**	4.513**	8.815**	8.140**
	(2.14)	(2.02)	(3.25)	(2.87)
Years of schooling minus 6 (x 10^{-2})	0.010	2.154	–4.909	–2.688
	(0.33)	(0.74)	–(1.22)	–(0.67)
Years of schooling minus 12 (x 10^{-2})	0.057**	6.695**	8.670**	7.815**
	(2.97)	(3.55)	(2.91)	(2.67)
# of days sick, instrumented (x 10^{-2})	–0.015**	–1.812**	–2.499**	–2.813**
	–(2.52)	–(3.25)	–(2.45)	–(2.57)
Other Individual Characteristics	*33.04***	*41.84***	*26.20***	*22.56***
Potential experience	0.028**	0.032**	0.037**	0.033**
	(5.11)	(5.74)	(4.63)	(4.13)
Potential-experience squared (x 10^{-3})	–0.269**	–0.304**	–0.337*	–0.237
	–(2.15)	–(2.46)	–(1.73)	–(1.22)
Occupation Variables	*6.22***	*8.87***	*8.68***	*4.63***
# of days sick for blue collar (x 10^{-2})	0.606		1.421	
	(1.08)		(1.38)	
Blue-collar dummy	–0.051		–0.093	
	–(0.47)		–(0.58)	
# of days sick for public worker		0.021**		0.016**
		(3.56)		(2.04)
Public-worker dummy		0.220**		0.336**
		(2.11)		(2.92)
F (21, 1521)	37.34	36.82	30.55	32.26
R^2	0.290	0.291	0.371	0.360
Net effect on health (x 10^{-2})	–0.911	0.305	–1.078	–1.212
	–(1.48)	(0.46)	–(0.81)	–(1.13)
Number of observations	1,543	1,543	844	844

Note: See notes to Table 5.4a.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

Health Effects on Productivity across the Wage Distribution

Having analyzed different health effects among specific employment or employer types, this subsection turns to differences in health effects over the wage distribution. It is not unreasonable to think that low-paid jobs imply a stronger connection between health and productivity (therefore wages) due, for instance, to more physically demanding tasks. To avoid splitting the sample by wage groups, Buchinsky (1994) is followed to estimate a quantile regression. The quantile regression model is:

$$y_i = x_i^{'}\beta_\theta + u_{\theta i}, \qquad [5.11]$$

where θ denotes the log wage quantile being estimated, and $x_i'\beta_\theta$ denotes the θ^{th} quantile. The parameters for the health status variables are interpreted as the effects on the θ^{th} quantile of *ln*(wage), not on the mean *ln*(wage) as in the OLS case. Conditional on $x_i'\beta_\theta$ this is equivalent to examining the θ^{th} quantile of the residual $u_{\theta i}$. Here, the effects on the 10th, 25th, 50th, 75th, and 90th percentiles are presented. This method is useful for several reasons. First, by providing information about the effects at each part of the wage distribution, it makes it possible to describe changes in the wage distribution when a bad health shock occurs, using the Interquartile Range (IQR) as an estimate of the wage variance (inequality). Second, the quantile estimator, when computed at the 50th percentile, provides the least absolute deviation (LAD) estimator, which is more robust than is the OLS estimator in the presence of outliers.

The results are shown in Tables 5.7a and 5.7b for wage earners and the self-employed, respectively. In each table, the top panel shows the results for men and the bottom for women. Before discussing the health effects, it should be noted that returns to primary schooling are larger among the lower bottom of the wage distribution. As before, returns to secondary schooling are not any different from primary schooling. On the other hand, the premiums to higher education are larger for those around the third quartile. Note also that the returns to potential experience are lower for the top quartile. Corresponding estimates for the self-employed are less precise.

Table 5.7a shows that the effects of bad health shocks on wages are larger for men who are at the bottom of the wage distribution scale (–2.6 percent at the 10th percentile). This result indicates that men who earn the

Table 5.7a Equation for the Log Hourly Earnings of Urban Wage Earners: Comparing Effects across the Wage Distribution

Variable	Quantile 10	25	50	75	90
Males					
Years of schooling (x 10^{-2})	7.200**	5.592*	8.066**	3.708	0.712
	(2.20)	(1.58)	(3.23)	(1.00)	(0.12)
Years of schooling minus 6 (x 10^{-2})	–0.723	0.915	–3.488	1.529	5.975
	–(0.18)	(0.22)	–(0.92)	(0.39)	(0.96)
Years of schooling minus 12 (x 10^{-2})	5.040**	6.468**	7.042**	8.813**	6.664**
	(1.83)	(3.59)	(2.94)	(4.11)	(2.31)
# of days sick, instrumented (x 10^{-2})	–2.619**	–1.471**	–1.469**	–0.866	–0.436
	–(3.79)	–(1.91)	–(2.31)	–(1.32)	–(0.43)
Potential experience	0.047**	0.035**	0.032**	0.029**	0.031**
	(5.25)	(5.48)	(7.22)	(4.79)	(3.31)
Potential-experience squared (x 10^{-3})	–0.577	–0.339**	–0.326**	–0.268**	–0.366*
	–(2.92)	–(2.14)	–(3.13)	–(2.00)	–(1.69)
Number of observations	1,543	1,543	1,543	1,543	1,543
R^2	0.167	0.180	0.174	0.174	0.167
Females					
Years of schooling (x 10^{-2})	13.754**	11.812**	4.833	3.659	9.432**
	(2.74)	(3.03)	(1.22)	(0.97)	(2.80)
Years of schooling minus 6 (x 10^{-2})	–9.809	–4.298	1.058	1.071	–6.089
	–(1.43)	–(0.79)	(0.17)	(0.17)	–(1.19)
Years of schooling minus 12 (x 10^{-2})	12.522**	6.939**	8.147**	7.487**	7.665*
	(2.15)	(1.97)	(2.44)	(1.80)	(1.72)
# of days sick, instrumented (x 10^{-2})	–0.860	–2.134	–2.877**	–3.868**	–3.610**
	–(0.33)	–(1.44)	–(2.46)	–(2.56)	–(2.20)
Potential experience	0.043**	0.039**	0.032**	0.035**	0.022*
	(2.93)	(4.85)	(4.06)	(2.78)	(1.53)
Potential-experience squared (x 10^{-3})	–0.499	–0.252	–0.211	–0.199	0.067
	–(1.40)	–(1.10)	–(1.06)	–(0.72)	(0.23)
Number of observations	844	844	844	844	844
R^2	0.222	0.250	0.255	0.205	0.164

Note: See notes to Table 5.4a.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

Table 5.7b Equation for Log Hourly Earnings of the Urban Self-Employed: Comparing Effects across the Wage Distribution

	Quantile				
Variable	10	25	50	75	90
Males					
Years of schooling (x 10^{-2})	1.376	2.559	4.726	7.289**	9.531
	(0.23)	(0.72)	(1.28)	(2.04)	(1.29)
Years of schooling minus 6 (x 10^{-2})	8.936	6.995*	3.813	0.730	0.105
	(0.96)	(1.76)	(0.84)	(0.18)	(0.01)
Years of schooling minus 12 (x 10^{-2})	–0.682	3.893	4.871	9.116**	7.479
	–(0.11)	(1.25)	(1.37)	(4.04)	(1.19)
# of days sick, instrumented (x 10^{-2})	–3.752**	–2.147**	–0.930	–2.225*	–3.160**
	–(2.00)	–(2.25)	–(1.08)	–(1.82)	–(2.75)
Potential experience	0.027*	0.029**	0.032**	0.025*	0.022
	(1.80)	(2.15)	(3.59)	(1.76)	(0.80)
Potential-experience squared (x 10^{-3})	–0.321	–0.256	–0.303	0.026	0.201
	–(0.86)	–(0.87)	–(1.53)	(0.08)	(0.41)
Number of observations	1,144	1,144	1,144	1,144	1,144
R^2	0.099	0.089	0.104	0.107	0.104
Females					
Years of schooling (x 10^{-2})	3.718	6.604*	7.101**	5.939**	8.962**
	(0.70)	(1.80)	(2.81)	(3.01)	(2.83)
Years of schooling minus 6 (x 10^{-2})	3.331	–0.114	–1.203	2.026	0.530
	(0.43)	–(0.02)	–(0.35)	(0.64)	(0.09)
Years of schooling minus 12 (x 10^{-2})	1.808	–3.233	2.636	5.488	1.066
	(0.24)	–(0.48)	(0.52)	(1.13)	(0.18)
# of days sick, instrumented (x 10^{-2})	–2.047	–2.275	–3.239**	–3.117*	–2.016
	–(0.98)	–(0.89)	–(2.13)	–(1.76)	–(0.78)
Potential experience	0.038*	0.044**	0.029**	0.029**	0.018
	(1.73)	(2.66)	(2.34)	(2.78)	(1.18)
Potential-experience squared (x 10^{-3})	–0.288	–0.460**	–0.254	–0.105	0.136
	–(0.99)	–(1.90)	–(1.38)	–(0.47)	(0.55)
Number of observations	843	843	843	843	843
R^2	0.096	0.064	0.067	0.079	0.098

Note: See notes to table 5.4a.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

least are more vulnerable to negative shocks in health. That is, the greatest reduction in wages from an exogenous change in health status that led to one more day of illness for all men would occur among the poor. On a more positive note, this result would imply that a health policy that improves health conditions equally for all men in the labor force would reduce wage inequality, as measured by the log variance of wages, since the lower-income workers would benefit more than proportionally. Clearly, the program's impact on wage inequality would be even larger if targeted exclusively among all the poor. But even if some "leakage" to upper-income groups occurs, the net impact will be inequality reducing. Among the self-employed, the results show the same decreasing pattern for health effects up to the median (from –3.8 percent at the lowest decile to –0.9 percent at the median). These effects increase again for the upper tail, although they remain below those found for the lowest decile.

Surprisingly the results for women show the reverse pattern. The largest health effects are found among the top of the wage distribution scale. This finding is consistent across sectors but accentuated among wage earners. While the result of the quantile regressions for men are consistent with the hypothesis that the health effect is due to the extent of physical tasks associated with jobs, the results for women are not. Notice that, as in all previous estimations, health effects tend to be larger among the self-employed than among wage earners, although the difference is not as precise for women.

Conclusions

This chapter has demonstrated that health does have a significant impact on productivity as measured by earnings, particularly in the case of urban men. Furthermore it showed that these effects vary across different sectors, across wage distribution, across public and private employment, by age, and by sex.

Estimates of the factors that influence health status were very robust for men, but key instrumental variables also were found to be significant for explaining women's reported health conditions. For men, the effects of schooling on morbidity differ by age. Older males tend to get sick more often or for longer periods. Access to adequate housing infrastructure was a significant factor in explaining reported illness, an effect that is very consis-

tent across area (urban/rural) and gender divisions. On the other hand, access to local health centers appeared to have no positive impact on health; if anything access was associated with worse health status for men.

Table 5.8 summarizes the main results. First, health effects are found to be stronger for men who are self-employed than for those employed as wage earners. Second, health effects are weak and negligible among younger individuals. For both age and employment types, imprecision in the estimates for women make firm conclusions impossible. Third, the larger effects found among the self-employed and the negligible effects found among public workers suggest that health has a greater impact on wages for jobs in which productivity and wages are closely connected, either by observability or by the design of the remuneration policy. Finally, when examining the effects across the distribution of wages, the largest effects are found at the bottom of the scale for both wage-earning and self-employed men. Surprisingly, the reverse pattern is found among women.

The key caveat in the analysis presented here is the overall inability to explain the relationship among health, productivity, and wages among women. This suggests it would be important for future research to target how women choose to enter the labor force and how they are selected into particular sectors and occupations. In addition, better estimates might be derived from less-subjective health indicators, such as anthropometric measures (see chapter 2).

Other interesting findings emerged regarding the relationship between health and education. Returns to primary schooling were slightly reduced when instrumented health measures were added into the wage equation, suggesting that when health is omitted, part of the educational effect on wages is due to health status and its associated productivity. The potential effects of better health are significant, but are especially large among those less endowed with human capital and hence earning lower wages. Thus public health policies should contribute to reduced inequality and poverty—effects that generally may be overlooked in traditional program evaluations. Moreover, if policymakers consider health to be a relevant human capital measure, health packages could be included as part of a program package designed to reduce poverty and improve living conditions. Focusing policy on education as the sole human capital variable and ignoring health conditions may lead to less-effective policies for development and social progress for the less-favored people of Peru.

Table 5.8 Summary of Health Effects on Hourly Earnings

	Male		Female	
	Wage Earners	**Self-Employed**	**Wage Earners**	**Self-Employed**
Full sample	−1.22%**	−3.15%**	−2.41%**	−2.29%*
	−(2.23)	−(3.24)	−(2.32)	−(1.70)
Young	1.20%	0.12%	−2.51%	−3.48%**
	(1.12)	(0.06)	−(1.22)	−(0.81)
Old	−2.01%**	−4.27%**	−2.78%**	−2.06%
	−(2.84)	−(3.95)	−(2.32)	−(1.44)
Quantile 10	−2.62%**	−3.75%**	−0.86%	−2.05%
	−(3.79)	−(2.00)	−(0.33)	−(0.98)
Quantile 25	−1.47%**	−2.15%**	−2.13%	−2.28%
	−(1.91)	−(2.25)	−(1.44)	−(0.89)
Quantile 50 (median)	−1.47%**	−0.93%	−2.88%**	−3.24%**
	−(2.31)	−(1.08)	−(2.46)	−(2.13)
Quantile 75	−0.87%	−2.23%*	−3.87%**	−3.12%*
	−(1.32)	−(1.82)	−(2.56)	−(1.76)
Quantile 90	−0.44%	−3.16%**	−3.61%**	−2.02%
	−(0.43)	−(2.75)	−(2.20)	−(0.78)
Public	0.30%		−1.21%	
	(0.46)		−(1.13)	
Private	−1.81%**		−2.81%**	
	−(3.25)		−(2.57)	

Note: Numbers in parentheses are t-statistics.

*Statistically significant at 10% level of confidence.

**Statistically significant at 5% level of confidence.

This page intentionally left blank

CHAPTER 6

Health and Productivity in Peru: Estimates by Gender and Region

Rafael Cortez[1]

This chapter seeks to measure the effect of health on the hourly wages of adult men and women in the rural and urban areas of Peru. A health indicator based on the number of reported days of illness is utilized to verify the association of health with the wages of individuals, and to measure the returns of good health to productivity. The principal finding of the research is that the health indicator has a significant positive effect on the level of productivity as measured by wages. The study reports the results of health indicator impacts on the wages of four groups—by gender and across urban and rural populations. It also describes the most important differences in the determinants of wages by population group. It demonstrates that the rates of return to education may be overestimated when the instrumented health variable is not included, principally in the cases of urban women and rural men. The significant estimates of returns to health indicate that one less day of reported illness in a month increases the wage rate of urban and rural women by 3.4 percent and 6.2 percent, respectively. For men the increase is higher—4.7 percent and 14.2 percent in urban and rural areas, respectively. Consequently, public and private investment in health should be recognized as mechanisms for increasing household income, principally in rural areas where rates of return to health are high.

[1] Dr. Rafael Cortez is a Professor and researcher with the Research Center of the University of the Pacific and a Postdoctoral Fellow of the Economic Growth Center of the Department of Economics, Yale University.

Introduction

The study of economic growth and the distribution of wealth increasingly emphasizes various kinds of human capital as factors in the determination of economic growth and wage rates. However the returns on investment in health have only recently begun to be studied in developing countries. Recent studies (Schultz, 1997; Schultz and Tansel, 1997; Thomas and Strauss, 1997) confirm the idea that health is a form of human capital that influences the wage levels of individuals and their capacity to generate sustained income over time. This, in turn, should have immediate positive consequences for the level of spending and the standard of living of household members.

By including health indicators in equations that seek to explain wages, various studies have measured the returns to health in the labor market and, simultaneously, facilitated an evaluation of the effects of public investment policies on the condition of health and incomes. For example, Thomas and Strauss (1997) utilize such a framework to analyze a Brazilian household survey that contains information on the height of adults and conclude that height has a positive effect on individual productivity. Interestingly, they also say that estimated returns on education, which include the health variable, were 45 percent lower for men with no education and 30 percent lower for men with secondary or tertiary education. Schultz and Tansel (1997) utilized estimates of instrumental variables for days of disability to estimate wage equations in Ghana and the Ivory Coast; their principal finding is that health conditions are a determinant of wage levels and the length of productive life.

This chapter measures the association between health and wages in Peru, and then estimates the impact of health on individual productivity. The consequences of omitting the health variable in the estimates of other variables included in the wage equation are evaluated, and the public policy implications of spending priorities and the availability of public health services are explored.

To conduct this analysis, we use an indicator based on days of illness reported by adults in the 15 days prior to an interview for the Peruvian Household Survey of 1995. The analysis proceeds in several stages. First, an overview is given of survey data, which contains demographic, social, economic, and health information from almost 20,000 households. It also gives a basic

Table 6.1 Conditions of Health, Medical Care, and Poverty in Peru for Ages 18 to 70

Poverty levels	Population (%)	Rate of illness reported (%)	Average number of days of illness reported	Rates of health care (%)
Nonpoor	64.0	25.9	2.58	42.9
Poor	22.8	27.5	2.74	32.8
Extremely poor	13.2	31.6	2.88	26.2
All	—	27.0	2.65	38.0

Source: 1995 ENAHO, author's own calculations.

description of the distribution of income and health in Peru. Second, the choice of indicators for measuring health status is discussed. Third, issues that make statistical analysis difficult are examined, including the simultaneity between health and wages, omitted variables, and measurement error. Fourth, an analytical model based on Becker (1964) is presented to analyze the data, followed by an outline of the approach to estimating the relationship among public policy variables, health, and income. Fifth, the empirical results are shown. Finally, conclusions summarize the main findings and comment on the implications for policy.

The ENAHO Survey of 1995 and Peruvian Population Characteristics

This analysis utilizes the National Household Survey (ENAHO) of 1995, which contains socioeconomic and demographic information on 19,975 households and 98,984 persons. The data was collected between October and December of that year by the National Institute of Statistics and Informatics (INEI). Of interest here are the 51,545 individuals who are between 18 and 70 years of age and can be considered potential members of the labor force. This group will be the focus of the analysis that follows.

Using information from ENAHO, Table 6.1 gives an initial approximation of poverty and health conditions in Peru. Approximately 36 percent

Figure 6.1 Distribution of Days Ill

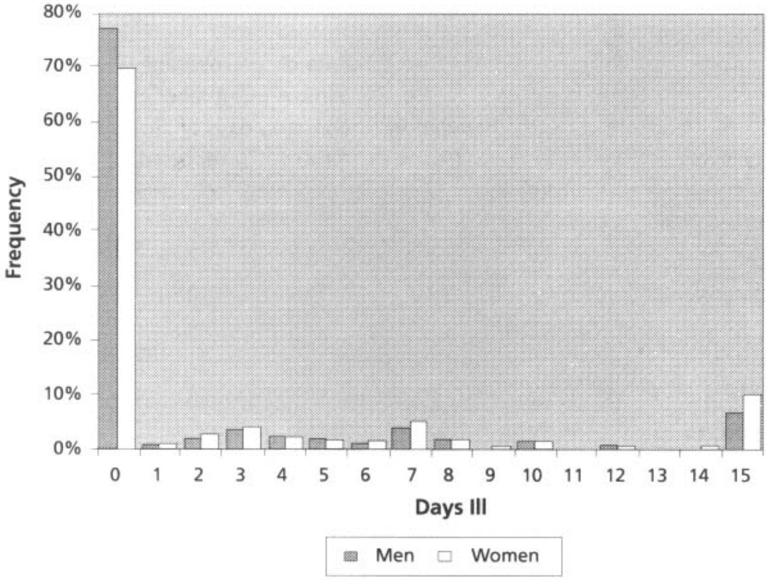


Figure 6.2 Reported Rate of Illness by Income Quintile

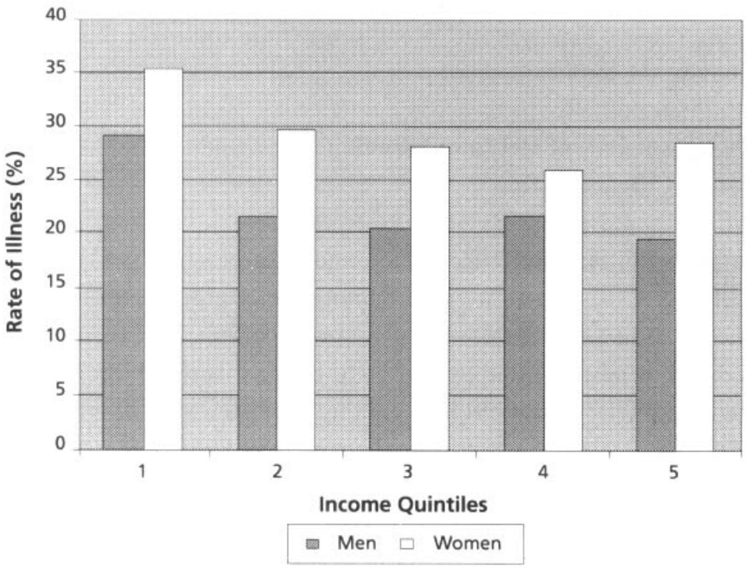


Table 6.2 Health Service Providers by Patients' Level of Poverty (% of Ill Individuals Receiving Medical Care)

Provider of Health Service	Total	Poverty Levels		
		Nonpoor	Poor	Extremely poor
Public providers	60.0	57.9	63.2	66.0
Health Ministry post and center	26.3	22.5	31.1	38.7
Social Security hospital	16.6	18.0	15.3	11.2
Health Ministry hospital	12.6	12.4	12.9	13.3
Others	4.5	5.0	3.8	2.8
Private providers	40.0	42.1	36.8	34.0
Pharmacy	17.8	17.6	19.7	15.1
Private consultation	9.9	11.5	7.7	4.8
Private clinic	4.1	4.9	2.7	2.6
Folk healer	0.9	0.8	1.1	1.2
Others	7.3	6.3	5.6	10.3
All	100.0	100.0	100.0	100.0

Source: 1995 ENAHO, author's own calculations.

of this population live in poverty and 13 percent in extreme poverty.[2] The health status of this population is measured by the frequency of illness and by the average length of illness during the 15 days prior to the survey, with the latter illustrated in Figure 6.1. In both cases, there is a negative correlation between poverty and health (see Table 6.1). The rate of reported illness for the population in extreme poverty (32 percent) exceeds the rate for the nonpoor population (25.9 percent). The average rate of illness for the working-age population is 27 percent.

The association between greater poverty and poor health constitutes the subject matter of this study. Even in terms of a first approximation to the problem, Table 6.1 indicates that poverty is not only associated with poor health but also associated with a lower rate of seeking medical care when ill.

[2] Considering the entire population results in slightly lower rates: 32.6 percent for poverty and 12.6 percent for extreme poverty.

This rate falls notably from 43 percent for the nonpoor to 26 percent for the extremely poor. These differential utilization rates for health services could explain part of the negative association between poverty and health.

Table 6.2 illustrates the importance of the public sector in the delivery of health services. On average, public establishments handle 60 percent of medical consultations received by the sick. This percentage is higher for the poor (63 percent), and even higher for those in extreme poverty (66 percent). Thus, on the few occasions when they seek health care, these households rely preponderantly on public health establishments.

Another way of seeing the relationship between health and earnings is highlighted in Figure 6.2. Lower rates of illness are observed in the population at the top of the wage distribution. For men, the rate of illness falls from 29 percent in the lowest quintile to less than 20 percent in the highest quintile. Among women, the differences across the distribution are not as strong, fluctuating between 35 percent and 28 percent.

Measuring Health

In comparison with other forms of human capital, health conditions are particularly difficult to measure. The state of an individual's health (H^*) can be considered a hidden variable, a variable that is not observed and is only approximated by health indicators (H) such as days of illness and days when an individual is too ill to work. These indicators, obtained from household surveys, are reported values and, as such, are contaminated by measurement errors. Such measurement errors bias the estimated impact of health towards 0.

Measurement errors are especially significant when the available information about health status is self-reported by individual respondents. Perception of personal health (or of illness) could be inconsistent across a population set in ways that are systematically correlated with other characteristics. For example, individuals with more education or greater access to health services may be more likely to detect and report symptoms of illness. In an equation designed to explain the determinants of health, these effects could confuse the direct impact of education or medical care.[3] The subjectivity of self-reported data can also generate heteroskedasticity in the health

[3] Wolfe and Behrman (1984) find with data from Nicaragua that more-educated women report fewer occurrences of parasitic diseases.

equation because the variability of the measurement error could also depend on some of its explanatory variables.

In addition, the severity of recent illnesses will necessarily be underestimated since the particular episode may not have concluded at the time of the survey interview. Even if the illness duration (D) were a perfect reflection of its severity, a survey would only observe the following relationship:

$$D^* = \mathrm{Min}\{D; L\} \qquad [6.1]$$

where L is the period of time from the start of the illness to the time of the survey. A dichotomous variable recording episodes of recent illness avoids this potential distortion; however all sensitivity to the severity or type of illness is lost. In any event, the longer the interval defined in the survey (e.g., the past month vs. the past week), the more reliable is its approximation to the true condition of health. In a sense, the length "increases" the number of observations, which reduces the sensitivity of the indicators to temporary or random factors.

The usefulness of turning to the past is also explained by the importance of previous health conditions. According to the intra-personal variations proposed by Behrman (1990), the human organism can, in the short term, maintain productivity despite adverse health conditions. With data from the south of India, Deolalikar (1988) observes that current nutrition indicators lose significance when other indicators are included that reflect nutritional status over a prolonged period.

In practice, the selection of the best indicator of state of health is normally "resolved" by the absence of better alternatives in most of the available surveys of households (or individuals). Even so, most studies find that the most frequently used indicators are usually highly correlated. Haddad et al. (1994) compare databases from the Philippines, Brazil, Ghana, and Mexico and find that the occurrence of illness, as well as the days of duration, are "useful" approximations of indicators based on weight and height. Other studies in this volume arrive at similar conclusions (see chapter 3).

The 1995 ENAHO offers two possible indicators for health: a dichotomous variable on the recent occurrence of illness and the number of days that the illness or complaint affected the individual. Presumably the reported illnesses were more likely to have occurred in individuals who were more

susceptible to poor health. In other words, if H^* is the real and unobserved measure of health status, then:

$$G = 1 \text{ if } H^* < H^C; \text{ and } G = 0 \text{ if } H^* \geq H^C \qquad [6.2]$$

where G is the proposed dichotomous variable and H^C is a certain critical level of strength of the individual's constitution. Below H^C, the individual falls ill.

Use of the number of days ill as an indicator of health status requires an additional assumption: that the illness lasts longer when the individual's health is weaker (when recovery capacity is lower). In this case, the length of the illness (D) depends negatively on H^*:

$$D = D(H^*), \quad \text{if } H^* < H^C, \quad \text{where } D' < 0,$$

$$\text{and} \quad D = 0, \quad \text{if } H^* \geq H^C. \qquad [6.3]$$

It is important to take into account that the description of health based on days of illness would be censored for values greater than H^C. Despite differences between the health conditions of individuals who have not fallen ill, the indicator D attributes the same value (0) to all of them (see Figure 6.1). Even so, days of illness is used in the following analysis as the indicator of health status because it captures more of the interpersonal health differences than does the dichotomous variable.

An additional distortion associated with days of illness is that individuals tend to round off their responses, thereby erasing differences between individuals whose illnesses are similar in length. The tendency to round off is stronger for longer illnesses. Accurate responses are more probable for short illnesses.

The relationship previously cited in equation [6.1] between health status and days of illness may not be linear. For this reason, several transformations of D must be tested to generate a health indicator (H). Since the expected relationship between the health indicator (H) and the underlying health status (H^*) is an inverse one, the censorship of the health indicator at 0 would imply an upper limit for health status defined as:

$$H^S = D^{-1}(0). \qquad [6.4]$$

The Simultaneity between Health and Productivity

The estimate of the impact of health on productivity is complex because the relationship between wages and health is not one-way. While health is a form of human capital that promotes productivity, wages (W) also affect health in at least three ways. First, and most obviously, higher income favors positive health conditions when these are considered as normal consumer goods. That is, the capacity to earn more permits greater consumption of health "inputs" like food or medicines.

Second, higher productivity can generate incentives to alter specific behaviors. For example, productivity can lead an individual to overwork until it undermines the person's health. Alternatively, the family might decide to dedicate a greater portion of its available resources to strengthening the health of the most productive member of the household.

Third, individuals may have specific characteristics that affect both their health and productivity. These might be unalterable traits (e.g., physical constitution) that are exogenous and random. If these characteristics are not observed in the estimation and controlled for, then they will enter into the respective error terms of the estimated equations—e_W for the wage equation and e_{H^*} for the health equation—and create no particular problem. However when the equations are viewed as simultaneous and estimated together, these two errors are correlated and the resulting estimates are biased. There is a problem of simultaneity whenever $\mathrm{Cov}(e_W, e_{H^*})$ is different than 0.

The errors may be systematically correlated if the same "endowment" or characteristic that facilitates greater productivity also permits the individual to conserve a better state of health (a particular ability or physical capacity, or some psychological trait). Such a systematic correlation can also arise from individual choices. For example, any characteristic that increases productivity and earnings also creates incentives to make additional investments in strengthening health. Those characteristics that are unobserved will influence the error term in the health equation (e.g., a special concern for promoting health) and introduce further bias.

A Household Model Incorporating Health

Extending Becker (1964), the decisions of a household can be usefully analyzed as the result of maximizing the utility that is derived from consumer

goods (C^i), consumer goods that improve health (I), state of health (H^i), and level of leisure (L^i). For simplicity, household decisions are analyzed as if the household were a single undifferentiated unit, and the household is constrained by time restrictions and total income. The model can be presented as a household of n persons, with a head who seeks to maximize its utility function:

$$U = U\left(C^i, I^i, H^i, L^i\right), \text{ and } i = 1, 2, \ldots, \text{n.} \qquad [6.5]$$

The utility function is considered to have the desired conditions; that is, it is continuous, strictly growing, quasi-concave, and differentiable in the second order in all its explanatory variables.

The first restriction is the health production function:

$$H^i = H^i\left(C^i, I^i, L^i, X^{-i}, Z^i, Z^{-i}, F, u^i, u^{-i}\right) \text{ and } i = 1, 2, \ldots, n \qquad [6.6]$$

where C^i, I^i, L^i represent the level of consumption of goods, health inputs, and leisure of an individual (i). X^{-i} denotes the level of consumption, health, and leisure of other members of the household; and Z^{-i} and u^{-i} are the vectors of characteristics of these individuals that are observed and unobserved, respectively. F denotes the availability of welfare or health programs and community infrastructure.

The second restriction is total income (S), which is exhausted by all the household's consumption of goods, services, and leisure activities:

$$\sum_{j=1}^{J}\sum_{i} p_j c_j^i + \sum_{k=J+1}^{k}\sum_{i} p_k I_k^i + \sum_{i} wl^i = \sum_{i} wT^i + V = S \qquad [6.7]$$

where V represents nonlabor income; p_j and p_k represent the prices of consumer goods and health inputs; T^i is the total quantity of time available; and w is the market wage.

The reduced form for the health demand function is:

$$H^i = h\left(P_c, P_y, S, F, Z^i, u^i\right), \qquad [6.8]$$

where P_c and P_y represent the prices of consumer goods and health inputs, respectively.

Because equation [6.8] is a demand function for health inputs, not a supply function, the vector X_H must include the effects of income and the prices of relevant health inputs. These variables influence the quantity of health inputs (e.g., nutrients, medical services) that the family unit consumes.

Due to (a) the errors in the measurement of health and (b) their simultaneity with respect to wages or the endogenous character of the health variable, it is necessary to use this reduced-form relationship to correct the estimation with instrumental variables. Health is estimated by the equation:

$$H = b_0 + b_1 X_H + b_2 X_W + e_H, \qquad [6.9]$$

where the error term e_H includes the measurement error $k(e_H = e_{H^*} + k)$. Furthermore, the explanatory variables are divided into those (X_H) that affect only health status and not wages and those (X_W) that affect both health status and wages. Unless sufficiently good variables are found for X_H, the instrumental variable technique will be ineffective.

Then, extending the framework of Mincer (1974) to include health as a form of human capital in the wage equation leads to:

$$\ln(W) = \alpha_0 + \alpha_1 X_W + \alpha_H H^* + \varepsilon_w, \qquad [6.10]$$

in which X_W is a set of variables that affect wages, H^* denotes the corrected state of health of the individual, and e_W is a random error term. A semilogarithmic specification is used since it is the most common specification found in the empirical studies of returns to human capital.

The wage equation [6.10] presents selectivity bias that has to be corrected by the Heckman procedure or two-stage estimate (Heckman, 1979; Lee, 1983). The equation that predicts the decision to participate in the labor market (*LM*) includes the explanatory variables of market wage, health, and a set of variables (X_L) that identify the system. The wage is not directly included in the estimate because it is not observed when the individual does not participate in the labor market and it is endogenous when the individual is employed. Instead it is replaced by its predicted value based on the variables X_W. Similarly health is replaced by its expectation based on a set of variables (X_H.) Thus:

$$LM = LM(W, H^*, X_L), \quad \text{or} \quad LM = LM(X_W, X_H, X_L). \qquad [6.11]$$

The set of equations [6.9], [6.10], and [6.11] form the system of equations to be estimated. Health is predicted in equation [6.9], which is estimated with a tobit model due to censorship of the health indicator (*H*). The fitted health variable is then included in the wage equation [6.10], which is also corrected for selection by the Heckman term derived from equation [6.11].

The Strategy of Econometric Estimation

Following the usual method in previous empirical studies, the sample is divided by gender to estimate separate wage and health equations for men and women. To consider the differences between urban and rural areas, the wage functions for men and women are also estimated separately for each geographical area. The determination of the explanatory variables X_W, X_H, and X_L also follow the usual method in the literature. Table 6.3 reports the definitions and the sample moments of all the variables utilized here.

The health variable is critical. It is used both as the dependent variable in the first-stage regression and as a regressor in the wage equation. The variable used to indicate health status is the number of days ill in the 15 days preceding the survey interview. Since this variable is truncated at 0, it is impossible to distinguish health status among the respondents who reported no days ill. Rather than assuming a linear relationship between days ill and the explanatory factors, and extrapolating these relationships through the entire sample, a transformation with declining marginal impact was applied. The transformation is:

$$H = 1/(1+D) \qquad [6.12]$$

where *D* is equal to the reported number of days ill. For 0 days ill, this health status variable is equal to 1. As the number of reported days ill increases, it declines to a minimum in this survey of .0625 (i.e., 1/(1+15)). Since the variable is necessarily censored at 0, a censored tobit estimation is utilized.

The proposed health indicator has the advantage of showing a marginally decreasing impact for days of illness. This characteristic is convenient because of the tendency to round off when individuals cannot

remember the exact number of days of illness, which as previously mentioned, is more likely for prolonged illnesses. The indicator (H) reduces the importance of variations in the reported days of illness as this number increases.[4]

In addition to the health variable, the wage equation includes age, years of study, and the quadratic terms for both. These terms consider possible nonlinear effects. X_W also incorporates two additional variables: residence in the capital city and the local unemployment rate. The latter variable is intended to capture interdistrict differences in labor markets.

Initially the wage equation is estimated without the health variable. Then the health variable is included without and with corrections. In the empirical estimates the following qualifications need to be kept in mind.

Limitations in the survey that prevent introducing all the relevant variables mean that some bias may remain. Recall that families acquire health inputs for their members with the awareness, among other things, of their unobserved endowments. As a result, the consumption of these inputs is probably correlated with the error term (e_H). This will bias the estimates. By contrast, factors that appear in the demand for health and that are not determined by the household, at least in the short run, can be considered exogenous and will not yield biased estimates. Some examples of these latter factors are age, education, food prices, and the supply of public services.

Within the explanatory variables of health, access to medical care by the state is important for analyzing the impact of public health investments on individuals. Access to health services is not measured in terms of the services received by each individual because the person or family endogenously determines this. Using such services results from a decision based

---

[4] Other transformations were tested, and showed that the highest likelihood ratios were registered by analyses that used transformations of the days-ill variable that presented decreasing marginal impact. In particular, $H_a = -D^{1/2}$, $H_b = -\ln(1+D)$, and $H = 1/(1+D)$ all obtained the highest likelihood ratios. The last of these transformations produced the maximum values for the logarithm of the likelihood function in all the samples and was the preferred alternative. The results of the following sections were robust in relation to changes in the selection of the health indicator when the characteristic of the marginal reduction of the impact of D was maintained (for more details see Cortez, 1999). It should be noted that using such nonlinear transformations of the health variable in the final wage equation has its own serious limitations for statistically extracting unbiased estimates, unless it is jointly estimated as a nonlinear system. Nevertheless it was judged important to address directly the problems of representing health status with these modifications.

Table 6.3 Definitions and Sample Moments of the Variables (population aged 18–70)

Variable	Definition	Mean	Standard deviation
Dependent variables			
Days of illness reported	Number of days of illness during the 15 days prior to the interview.	2.21	4.49
Rate of illness reported	1 = reported complaint or illness in the 15 days prior to the interview, 0 = otherwise.	0.27	0.44
Ln(Wage)	Natural logarithm of individual hourly wage (in new soles) (calculated on weekly hours of work and monthly wages, weekly wages and semiannual benefits).	0.34	1.15
Independent variables			
Age	Age (not including fractions of year).	23.22	82.92
Years of education	Years of study (calculated on grades approved).	8.07	4.08
Automobile	Dichotomous: Owner of automobile = 1; otherwise = 0.	0.10	0.31
Other vehicle	Dichotomous: 1, if the household has a vehicle other than an automobile; 0, otherwise.	0.32	0.47
Nonlabor income	Incomes in new soles (labor and nonlabor) received by all household members in the preceding month, excluding labor income of the individual in observation, divided by family size.	6.23	57.37
Residence on coast	Dichotomous: Residence on coast =1; otherwise = 0.	0.44	0.50
Residence in Lima	Dichotomous: Residence in capital = 1; otherwise = 0.	0.14	0.35
Hours of water supply	Hours of supply of potable water from the public system during the last week.	76.22	98.78
Appropriate system of drains	Dichotomous: Indoor household access to public sewerage = 1; otherwise = 0.	0.56	0.50
Surfaced flooring	District rate of households with surfaced floors.	0.60	0.25
Health establishments per capita	District number of hospitals, posts, or centers of Health Ministry, IPSS, local government or other state agencies, per 10,000 inhabitants.	4.06	8.72

Table 6.3 (continued)

Variable	Definition	Mean	Standard deviation
Unemployment rate	District unemployment rate.	7.76	3.13
Poverty rate	Provincial rate unsatisfied basic needs calculated by FONCODES.	2.22	2.22
Head of household	Dichotomous: Individual is head of the household = 1; otherwise = 0.	0.35	0.48
Price of rice	Price in new soles of one kilo of rice in November 1995 in province.	1.27	0.15
Price of milk	Price in new soles of one large can of evaporated milk in November 1995 in province.	1.57	0.11
Price of tomatoes	Price in new soles of one kilo of tomatoes in November 1995 in province.	1.12	0.33
Urban area	Dichotomous: Residence in urban or semi-urban area = 1 (according to the INEI classification included in the survey); otherwise = 0.	0.78	0.42

Source: 1995 ENAHO, author's own calculations.

on personal income, the opportunity cost of time, personal health status, and perceptions of the quality and effectiveness of the service, among other things known by the individual that are unknown to us. By contrast, the number of public health establishments per person in each district indicates the availability of public services independent of the individual decisions to utilize those services or not. This community-average measure is the variable introduced into health demand equation [6.9]. The quadratic term of this variable is also included to observe possible nonlinear effects.

Using the community average of public health establishments in equation [6.9] solves the problem of endogeneity with regard to families and individuals, but it presents another problem of endogeneity at the community level. If the state distributed services randomly, as in a social experiment, then there would be no bias. However it is unlikely for the state to distribute services without reference to some criterion—systematically benefiting the rich or the poor or a particular economic sector or political

Table 6.4 Censored Tobit Estimates of Health Determinants by Gender and Region

	Men		Women	
Independent Variables	**Urban**	**Rural**	**Urban**	**Rural**
Constant	3.020*	1.408***	2.367***	1.572***
	(10.54)	(3.74)	(10.84)	(4.98)
Individual characteristics	***180.3****	***70.7****	***280.5****	***128.5****
Age (10^{-2})	–0.578	0.105	–1.163***	–1.537**
	(–1.11)	(0.13)	(–2.93)	(–2.25)
Age^2 (10^{-4})	–0.871	–1.584*	–0.192	0.093
	(–1.36)	(–1.68)	(–0.39)	(0.12)
Variables of Human Capital	***49.7****	***22.5****	***56.1****	***13.2****
Years of education (10^{-2})	3.827*	5.899***	2.296***	–0.480
	(2.62)	(3.26)	(2.63)	(–0.36)
Years of education squared (10^{-1})	–0.071	–0.275**	–0.024	0.178
	(–0.76)	(–2.05)	(–0.39)	(1.64)
Household assets				
Nonlabor income	0.018	–0.017	0.041***	–0.097*
	(1.35)	(–0.53)	(3.55)	(–1.85)
Housing infrastructure	***16.4****	***6.1***	***26.4****	***8.6****
Hours of water supply (10^{-4})	1.538	3.762*	–0.155	2.907*
	(1.25)	(1.86)	(–0.16)	(1.75)
Adequate drainage system	0.006	0.049	0.013	–0.113
	(0.22)	(0.57)	(0.61)	(–1.63)
Nondirt flooring	0.276*	0.093	0.283***	0.124*
	(3.66)	(1.22)	(4.96)	(1.93)
Regional variables	***13.9****	—	***21.2****	—
Residence at coast	–0.130*	0.131**	–0.101***	0.228***
	(–2.98)	(2.31)	(–3.05)	(4.86)
Residence in Lima	0.182*	—	0.176***	—
	(3.64)		[4.60]	
Community variables	***9.7****	***6.2***	***17.4****	***6.7***
Poverty indicator	–0.978**	–7.095**	–0.711*	–1.214
	(–1.99)	(–2.20)	(–1.76)	(–0.45)
Unemployment rate	–1.095**	–1.036*	–1.361***	–1.236**
	(–2.22)	(–1.74)	(–3.65)	(–2.57)

Table 6.4 (continued)

Independent Variables	Men		Women	
	Urban	Rural	Urban	Rural
Health infrastructure	***2.3***	—	***6.7****	—
Health centers per capita	–0.321	0.139**	–0.797**	0.089*
	(–0.66)	(2.15)	(–2.18)	(1.70)
Health centers per capita²	0.224	–0.014**	0.699**	–0.012***
	(0.48)	(–2.50)	(1.99)	(–2.73)
Food prices	***101.2****	***17.0****	***97.4****	***14.4****
Rice	–0.977*	–0.410***	–0.718***	–0.322***
	(–9.32)	(–2.92)	(–9.08)	(–2.77)
Tomatoes	–0.285*	–0.092	–0.198***	–0.088
	(–6.64)	(–1.37)	(–6.03)	(–1.53)
Milk	0.276**	0.607***	0.286***	0.475***
	(2.08)	(3.26)	(2.77)	(2.92)
HI (Health centers)	–0.304	0.135**	–0.740**	0.085*
	(–0.68)	(2.14)	(–2.20)	(1.67)
Log Likelihood Ratio	–12,209	–4,174	–16,107	–4,721
Chi²	483.3***	—	755.4***	—
Prob ($H^* < 1$)	22.7%	29.1%	32.7%	38.0%
Number of observations	18,787	5,633	20,435	5,671

Notes: The dependent variable in all cases is $H = 1/(1 + \text{number of days ill})$; t-statistics are in parentheses; and tests of joint significance are in bold face and italics.

* Statistically significant to 10% confidence level.

** Statistically significant to 5% confidence level.

*** Statistically significant to 1% confidence level.

supporters.[5] In the case of Peru, public health posts appear to have been built in localities with the worst health endowment, leading to a positive correlation between the error term (e_H) and the number of public centers.

[5] Sen (1995) describes the economics of the distribution of social program benefits. Beyond the interest in benefiting the most needy—as suggested by Rosenzweig and Wolpin (1986) and Pitt et al. (1995)—there are problems of political viability and the strength of powerful interest groups. The well-known theory of pressure groups models this phenomenon, while the theory of altruism formalizes the desire to compensate the most needy.

This would generate a downward bias in the estimated impact of public health services on health. It was not possible to correct for this bias in the analysis because effective instrumental variables could not be found to make the required adjustments. Therefore, the results have to be viewed cautiously due to the possibility of downward bias.

Results: Health Affects Productivity

The sample used in the regressions includes only individuals aged between 18 and 70; that is, adults who are potential participants in the labor market. Corrections were made for participation in the labor force using a probit equation (see Appendix II in Cortez, 1999). Wald tests indicate that the set of instruments (X_L) is significant in the four samples: men and women from urban and rural areas.

The results of the health equation (see Table 6.4) generally follow our expectations. In each of the samples, age is associated with a decline in health status. Both the linear and quadratic coefficients are negative in all cases except for rural men. But in this latter case, the negative coefficient on the quadratic term leads to a negative relationship with age over the relevant range (18 to 70 years). The estimated impact of age is always negative with respect to health.

The variables that indicate wealth are only significant in some cases. Nonlabor family income per capita is significant only for women, and rural women obtain an unexpectedly negative coefficient. This result is difficult to explain. Hours of access to the public system of potable water, adequate drainage, and surfaced floors in the home are jointly significant and have positive effects on the health of the urban population and of rural women. These variables may be measuring the direct impact on health status of better hygiene and environment and/or acting as proxies for wealth. The interaction of these variables with nonlabor income might account for the negative coefficient on nonlabor income found for rural women.

Education has a positive impact that could be associated with better use of knowledge and available inputs, leading to better health care. In general the coefficients of education in Table 6.4 would be biased downwards if there were a systematic correlation between education and reported health status. For example, in our sample it appears that more-educated individuals tend to report illnesses more often than do those who are less educated.

In relation to the impacts of health input costs, increased food prices (with the exception of milk) significantly reduce health status in all the samples. The positive coefficient of the milk price may be explained by its correlation with the prices of other foods not included in the estimate.[6]

Community conditions are relevant determinants of individual health status. People living in provinces with higher poverty rates or local unemployment rates tend to have poorer health. After controlling for these factors, living in Lima and at the rural coast is associated with better health status, while living at the urban coast is associated with poorer health status. These tendencies could be the result of greater access to better-quality health care (in the case of Lima), to climatic factors, or to other health-enhancing resources in generally wealthier areas.[7]

In rural areas, the availability of public health posts has a positive and significant impact on health status. The negative coefficients on the quadratic terms indicate declining marginal returns on public investment in health. That is, for each additional health post provided, the marginal improvement in expected health status will be smaller. In urban areas the results are counterintuitive. The impact of more health posts in urban areas is negatively associated with health status, and this effect is statistically significant for women. One possibility is that health posts may be associated with poorer sections of urban areas, while in wealthier sections people make greater use of private services or public hospitals.

Using the results of the health equation, it is now possible to see the effect of introducing a corrected health measure into an earnings function (see Tables 6.5 through 6.8). Three different cases are shown for the wage function: (i) excluding the health variable, (ii) including the reported values of health status, and (iii) including the corrected values of the health indicator. The latter are obtained from the regressions of Table 6.4 and are presumed to be free from the biases of simultaneity and measurement error.

As expected, health conditions have a significant and positive effect on productivity. An improvement in health status raises wages across all

[6] Only some of the consumer prices were available for all the departments (provinces) in the country. The data in this case are obtained from the 24 Departmental Statistical Compendiums for 1995 by the National Institute of Statistics and Informatics (INEI).

[7] Note that rural coastal Peru is relatively prosperous, with large, irrigated, and productive commercial plantations.

population groups, by gender and area, whether using the reported health variable or the corrected variable. However both the coefficient and level of significance of the health variable are much greater in the latter case.

The magnitude of the impact of health on productivity varies by population group. Improvements in health conditions have a greater impact on productivity in rural areas. This difference in favor of rural areas is greater among men than women. To see this more clearly, Table 6.9 shows the impact on hourly wages of an additional day of health per month.[8]

The results for the remaining variables in Tables 6.5 through 6.8 show that when health is not controlled, the coefficients on age and education are biased slightly upward. This suggests that when the health variable is omitted, part of health's impact on productivity is being attributed incorrectly to age and education.

When the instrumented health variable is utilized in the wage equation and compared with the specification for no health variable at all, the estimated returns to education for men are lower by 9.5 percent and 55.2 percent in urban and rural areas, respectively. For women, the proportion of overestimation is 15.7 percent and 12.5 percent in urban and rural areas, respectively. This upward bias is due to the expected positive correlation between both forms of human capital: education and health (Schultz, 1996). Education captures part of the impact of health when the health status variable is omitted from the equation.

The effect of an additional year of education depends on the coefficients of the linear and quadratic terms of the education variable, and on the number of years of study.[9] At the bottom of Tables 6.5 through 6.8, the marginal return to education is evaluated at the sample mean. The magnitude of the impact varies from one sample to another. Including the instrumented health variable, the correct returns to education in urban areas are 7.4 percent and 5.1 percent for men and women, respectively, and in rural areas 5.8 percent and 10.4 percent.

[8] Given that $H = 1/(1 + D)$, then $\partial W / \partial D = (\partial W / \partial H) \times (\partial H / \partial D) = -\alpha_H/(1 + D)^2$, where α_H is the coefficient of the indicator H in the wage equation and D is evaluated at the sample means.

[9] Given equation [6.1], then $\Delta \ln(W) = \Delta\% \; W = \alpha_{S1} + 2\alpha_{S2} S$ with one additional year of study, where α_{S1} and α_{S2} represent the coefficients of the linear and quadratic terms of the years of study. Specifically for urban men, the return ($\Delta\% \; W$) is equal to 0.074 = 0.094 + 2 (-0.00101) (9.13), where 9.13 is the sampling average of the years of study.

Table 6.5 Wage Determinants of Urban Men

Variables	(i) Without health	(ii) Exogenous [IV]	(iii) Health
Constant	−1.374***	−1.421***	−2.103***
	(−8.50)	(−8.77)	(−8.79)
Individual characteristics			
Age	0.052***	0.051***	0.051***
	(7.12)	(6.94)	(6.96)
Age^2 (10^{-2})	−0.053***	−0.051***	−0.048***
	(−5.98)	(−5.77)	(−5.39)
Human capital variables			
Years of education	0.108***	0.107***	0.094***
	(10.14)	(10.07)	(8.40)
Years of $education^2$ (10^{-2})	−0.143**	−0.140**	−0.101
	(−2.07)	(−2.02)	(−1.45)
Health indicator	—	0.090***	0.933***
		(4.19)	(4.13)
Variables of the local market			
Residence in Lima	0.265***	0.265***	0.276***
	(13.91)	(13.90)	(14.32)
Unemployment rate (to district level)	0.387*	0.398*	0.486**
	(1.65)	(1.70)	(2.06)
Selection term	−0.184***	−0.196***	−0.193***
	(−3.36)	(−3.57)	(−3.53)
Rate of return on education	8.1%	8.0%	7.4%
Return to health	—	0.5%	4.7%
Joint significance tests:			
Individual characteristics	197.1***	202.1***	181.6***
Education	1,175.3***	1,165.6***	621.6***
Human capitol variables	1,175.3***	1,192.8***	1,192.7***
Labor market variables	233.8***	236.3***	247.9***
Hausman test	—	—	184.1***
Log likelihood	−18,478	−18,469	−18,469
Chi^2	295.5***	261.1***	261.0***
R^2 adjusted	0.126	0.127	0.127
Number of observations	14,321	14,321	14,321

Notes: The estimates are OLS corrected by the Heckman two-stage estimate for participation in the labor force; the dependent variable is the natural logarithm of the hourly wage; t-statistics are in parentheses.

* Significance to 10% confidence level.

** Significance to 5% confidence level.

*** Significance to 1% confidence level.

Table 6.6 Wage Determinants of Rural Men

Variables	(i) Without health	(ii) Exogenous [IV]	(iii) Health
Constant	−1.496***	−1.551***	−3.944***
	(−6.03)	(−6.21)	(−9.17)
Individual characteristics			
Age	0.029**	0.028**	0.015
	(2.57)	(2.52)	(1.36)
Age^2 (10^{-2})	−0.038***	−0.037***	−0.006
	(−3.05)	(−2.97)	(−0.46)
Variables of human capital			
Years of education	0.024	0.022	−0.057***
	(1.25)	(1.17)	(−2.60)
Years of $education^2$ (10^{-2})	0.537***	0.545***	0.938***
	(3.79)	(3.85)	(6.16)
Health indicator	—	0.086*	3.464***
		(1.77)	(6.95)
Variables of the local market			
Residence in Lima	—	—	—
Unemployment rate (to district level)	1.180**	1.163**	0.787
	(2.35)	(2.32)	(1.57)
Selection term	0.147	0.143	0.129
	(1.61)	(1.56)	(1.42)
Rate of return on education	9.0%	8.9%	5.8%
Return to health	—	0.4%	14.2%
Joint significance tests:			
Individual characteristics	17.9***	16.2***	15.8***
Education	291.5***	288.6***	150.5***
Human capital variables	291.5***	294.8***	343.1***
Labor market variables	10.2***	9.8***	5.6*
Hausman test	—	—	122.1***
Log likelihood	−7,164	−7,162	−7,139
Chi^2	76.6***	66.1***	73.2***
R^2 adjusted	0.093	0.093	0.102
Number of observations	4,445	4,445	4,445

Notes: The estimates are OLS corrected by the Heckman two-stage estimate for participation in the labor force; the dependent variable is the natural logarithm of the hourly wage; t-statistics are in parentheses.

* Significance to 10% confidence level.

** Significance to 5% confidence level.

*** Significance to 1% confidence level.

Table 6.7 Wage Determinants of Urban Women

Variables	(i) Without health	(ii) Exogenous [IV]	(iii) Health
Constant	−1.752***	−1.784***	−2.577***
	(−10.44)	(−10.61)	(−9.81)
Individual characteristics			
Age	0.074***	0.073***	0.074***
	(10.26)	(10.18)	(10.26)
Age² (10⁻²)	−0.079***	−0.078***	−0.073***
	(−8.77)	(−8.64)	(−8.03)
Variables of human capital			
Years of education	0.105***	0.104***	0.091***
	(10.45)	(10.34)	(8.53)
Years of education² (10⁻²)	−0.279***	−0.275***	−0.243***
	(−3.96)	(−3.90)	(−3.43)
Health indicator	—	0.064***	1.060***
		(2.60)	(4.08)
Variables of the local market			
Residence in Lima	0.383***	0.381***	0.374***
	(15.45)	(15.37)	(15.05)
Unemployment rate (to district level)	−0.468	−0.435	−0.125
	(−1.43)	(−1.33)	(−0.37)
Selection term	0.002	−0.007	−0.011
	(0.04)	(−0.12)	(−0.19)
Rate of return on education	5.9%	5.9%	5.1%
Return to health	—	0.2%	3.4%
Joint significance tests:			
Individual characteristics	221.4***	225.7***	206.6***
Education	630.2***	637.3***	647.9***
Human capital variables	630.2***	620.4***	268.8***
Labor market variables	239.8***	237.7***	227.9***
Hausman test	—	—	229.3***
Log likelihood	−13,040	−13,036	−13,031
Chi²	174.3***	153.5***	154.8***
R^2 adjusted	0.112	0.113	0.114
Number of observations	9,598	9,598	9,598

Notes: The estimates are OLS corrected by the Heckman two-stage estimate for participation in the labor force; the dependent variable is the natural logarithm of the hourly wage; t-statistics are in parentheses.

* Significance to 10% confidence level.

** Significance to 5% confidence level.

*** Significance to 1% confidence level.

Table 6.8 Wage Determinants of Rural Women

Variables	(i) Without health	(ii) Exogenous [IV]	(iii) Health
Constant	–2.418***	–2.487***	–4.267***
	(–6.78)	(–6.95)	(–7.25)
Individual characteristics			
Age	0.052***	0.051***	0.054***
	(3.43)	(3.32)	(3.51)
Age^2 (10^{-2})	–0.058***	–0.055***	–0.043**
	(–3.27)	(–3.10)	(–2.40)
Variables of human capital			
Years of education	0.124***	0.124***	0.116***
	(4.78)	(4.78)	(4.44)
Years of $education^2$ (10^{-2})	–0.086	–0.093	–0.130
	(–0.42)	(–0.45)	(–0.63)
Health indicator	—	0.146*	2.247***
		(1.93)	(3.94)
Variables of the local market			
Residence in Lima	—	—	—
Unemployment rate (to district level)	3.694***	3.646***	3.384***
	(4.52)	(4.47)	(4.14)
Selection term	0.087	0.075	0.082
	(1.08)	(0.94)	(1.02)
Rate of return on education	11.7%	11.6%	10.4%
Return to health	—	0.4%	6.2%
Joint significance tests:			
Individual characteristics	12.0***	11.7***	23.4***
Education	171.5***	175.6***	188.4***
Human capital variables	171.5***	168.7***	114.4***
Labor market variables	23.5***	22.5***	19.8***
Hausman test	—	—	182.5***
Log likelihood	–3,254	–3,252	–3,246
Chi^2	41.5***	36.2***	38.1***
R^2 adjusted	0.113	0.114	0.120
Number of observations	1,908	1,908	1,908

Notes: The estimates are OLS corrected by the Heckman two-stage estimate for participation in the labor force; the dependent variable is the natural logarithm of the hourly wage; t-statistics are in parentheses.

* Significance to 10% confidence level.

** Significance to 5% confidence level.

*** Significance to 1% confidence level.

The impact of residing in Lima is always positive and robust to the inclusion of the reported or instrumented health conditions. Participation in the labor market in Lima implies a wage that is 28 percent higher for men and 37 percent higher for women. Local unemployment (measured at the district level) should control for differences in labor market conditions across districts. Unexpectedly, higher unemployment is associated with higher wages for all of the samples except urban women. These positive effects are also statistically significant. In a normal labor market model, slack labor markets (with high unemployment) would be associated with lower wages. The positive effect of unemployment could indicate significant migration to areas with stronger labor demand and higher wages, with queuing for those better-paying jobs. However verification of this hypothesis would require more detailed analysis of the labor market.

The correction term of the selectivity bias (λ) is significant for urban men and has a negative sign. This indicates that the unobserved characteristics that increase a worker's probability of labor participation negatively impact the level of wages received in the market.

Looking at the interaction between health and education further extends the analysis (Cortez, 1999, Appendix 2). After instrumenting, the interaction of health and schooling has a positive impact for men. Therefore for this group education and health appear to be complementary. The rate of return on education is higher when the individual has good health, and health is also more productive when the individual has more years of education. This result contradicts the common hypothesis that health status is of greater importance for physically intensive "blue collar" than for more intellectually intensive "white collar" occupations. Rather health appears to be worth more as a person's education, and presumably the intellectual content of his/her occupation, increases.

Finally, including an interaction term between instrumented health and age shows that productivity is more sensitive to changes in health status for older people. This lends some corroboration to the findings on health and productivity among the elderly in chapter 3. In all the samples analyzed here, a positive coefficient is obtained for the interaction of health and age, and the estimate is statistically significant for rural men and urban women. This suggests that policies designed to improve the health conditions of the older population would improve their wages more than similar policies aimed at younger population groups.

Table 6.9 Estimates of Returns to Health by Gender and Residence

Population group	Return to health (%)
Men, urban area	4.7
Women, urban area	3.4
Men, rural area	14.2
Women, rural area	6.2

Note: Returns to health are measured as the percentage impact on hourly wages of one less day of illness at the mean.

To demonstrate the uses of these findings for public policy, a series of simulations are shown in Table 6.10. The simulation focuses on the quality of housing conditions, which were shown to have a significant and positive effect on hourly wages through their impacts on health status. In each panel, the mean housing conditions are improved by 10 percent, 30 percent, and 50 percent, respectively. The results suggest that the wages of urban women and rural men are most sensitive to changes in housing conditions, although not significantly. For example, an increase of 50 percent in the hours of water supply, the quality of flooring, and the community drainage system would be associated with an increase in women's hourly wages of 3.5 percent and 1.8 percent in urban and rural areas, respectively. For men the improvement would be 2.1 percent and 2.3 percent, respectively.

Conclusions

This chapter has shown that health status can be usefully considered to be a component of human capital since it is systematically influenced by decision variables and it influences wage levels. It is shown that public policies, by improving individual health conditions, can raise household wages and thereby improve the standard of living. The techniques of instrumental variables reduce the bias due to measurement errors and potential endogeneity present in the reported health information in Peru's national household surveys.

The results of the 1995 national household survey show that the rate of illness and number of days of illness reported are negatively related to

Table 6.10 Impact on Wages of Improved Housing Infrastructure by Gender and Area (%)

Policies	Urban men	Rural men	Urban women	Rural women
Simulation 1				
Mean hours of water supply (per district)*1.1	0.0	0.1	0.0	0.1
Mean share with adequate drainage (per district)*1.1	0.1	0.0	0.0	-0.1
Mean quality of floors (per district)*1.1	0.4	0.3	0.7	0.3
Hourly wage increase due to 10% improvement in housing	*0.4*	*0.4*	*0.7*	*0.3*
Simulation 2				
Mean hours of water supply (per district)*1.3	0.1	0.4	0.0	0.3
Mean share with adequate drainage (per district)*1.3	0.0	0.1	0.1	-0.2
Mean quality of floors (per district)*1.3	1.2	0.9	2.0	1.0
Hourly wage increase due to 30% improvement in housing	*1.3*	*1.4*	*2.1*	*1.1*
Simulation 3				
Mean hours of water supply (per district)*1.5	0.1	0.7	0.0	0.4
Mean share with adequate drainage (per district)*1.5	0.0	0.1	0.2	-0.3
Mean quality of floors (per district)*1.5	2.0	1.5	3.3	1.7
Hourly wage increase due to 50% improvement in housing	*2.1*	*2.3*	*3.5*	*1.8*

individual wages and household income levels. Unexpectedly, individuals with more years of education report higher rates of illness and receive more health care services.

Health had a positive and significant impact on wages, with or without corrections for endogeneity and measurement error. Comparisons of regressions with and without corrections demonstrated empirically that, unless addressed, endogeneity and measurement error bias the health coefficient downward.

The results were also robust to choices of the dependent variable and alternative specifications. Comparing findings by gender and area, it was found that the productivity of urban men and rural women are most sensitive to health status. The health status that is associated with one less day of

reported illness has a larger impact on men's wages (4.7 percent in urban areas and 14.2 percent in rural areas) than for women (3.4 percent in urban areas and 6.2 percent in rural areas).

In all the samples, controlling for health reduces the estimated returns to age (i.e., experience) and education. As theory predicts, the estimates of these variables are overestimated when the health variable is omitted from the wage equation. The difference between the controlled and uncontrolled estimates shows that the upward bias for age and education are most problematic for rural men and urban women.

Health conditions appear to be sensitive to public decisions to build and run health posts in rural areas. The health of rural men is most favored by such an expansion of health services. In urban areas, however, the presence of public health services is associated with poorer health status, probably reflecting a greater concentration of public services in poorer urban areas. Other public services—such as the number of doctors/nurses for every 10,000 inhabitants, the number of beds per district, the level of Social Security Institute coverage by district—did not show any significant impact on health status. In these cases, it is not possible to distinguish whether the results were due to the low quality of the information, insufficient level of disaggregation, serious problems of measurement, or a simple lack of impact on health.

In addition to the supply of health infrastructure, housing and the community environment had significant effects on health and wages. In particular the results of simulations indicate that both the hourly wages and health status of urban men and rural women are particularly sensitive to investments in improved housing. Local labor market conditions also appear to have an impact on health since the level of poverty and unemployment had a negative impact on individual health status.

One of the problems that had to be dealt with in the research was insufficient information for other policy variables that could be incorporated into the health equation. Also, the survey only contained information on incidence of illness and the number of days of illness reported by individuals. Thus it was not possible to validate these health indicators against such measures as days of disability due to illness (i.e., an inability to work) or anthropometric measurements for adults. Validation of these health indicators is an important challenge for future research. It will also be important to include more information from other sources, principally indicators

associated with policy instruments for specific public health programs and food subsidies that are priorities in the state budget. An evaluation of how these policies affect individual health would make it possible to evaluate their indirect impact on wages. This could be very useful in measuring the benefits of such programs beyond the direct effects on health.

This page intentionally left blank

CHAPTER 7

Productivity and Health Status in Nicaragua

Jaime Espinosa Ferrando and
Carlos Hernández Alvarez,
with William D. Savedoff[1]

This chapter presents evidence on the determinants of health status in Nicaragua, the lowest-income country covered by this book. It discusses how these health status determinants affect individual productivity as measured by hourly income. Using self-reported illness and days of illness as measures of health status, the analysis shows that poor health does reduce productivity, by as much as 58 percent, although the estimates are not precise. Health status is significantly influenced by community and policy variables, with the most important being urban housing conditions, rural sanitation, and the supply of preventive care services by nurses. A community program, Casa Mujer, in which women organize to provide public services, also has a demonstrably positive impact on the health of rural men and both rural and urban women.

The study is based on data collected by the Living Standard Survey (ENV93) conducted in 1993. In addition to the known difficulties of measuring health status, data quality problems were encountered as a result of errors in the questionnaires, sampling, and data processing. The country's difficult political, economic, and social transition during the early 1990s, encompassing the time of the survey, may have contributed to the reduced explanatory power of the models and account for some of the odd findings of unexpected or reversed relationships.

[1] Dr. Espinosa and Dr. Hernández are Physicians and Consultants at the Fundación Internacional Para el Desafío Económico Global (FIDEG) in Managua, Nicaragua. Dr. William D. Savedoff, Inter-American Development Bank, assisted in the writing of this chapter and in refining the interpretation of results originally presented in a working paper by Espinosa and Hernández (1999).

Introduction

Nicaragua is the lowest-income country studied in this book. Its per capita income in 1997 was about US$491, compared to US$1,738 in Colombia, US$2,209 in Peru, and US$3,347 in Mexico. For various reasons, it is reasonable to expect that better health would have a larger and more significant impact on productivity and earnings in such a low-income context. This would be the case if health has declining marginal returns, or if occupations requiring physical strength or endurance are more affected by health status than the service occupations and intellectual activities that make up an increasing share of work in middle- and upper-income countries. In fact much of the literature that has analyzed the links between health and productivity has focused on lower-income countries or lower-income groups within those countries for these very reasons, even though the results have not been as dramatic as expected.[2] In terms of income levels, this analysis of Nicaragua may be more comparable to studies in Ghana and Côte d'Ivoire by Schultz (1996) and Schultz and Tansel (1997) than to studies of other countries in Latin America. In the two African countries, per capita income was US$370 and US$690, respectively.[3]

This chapter will first study the determinants of household health investments and then analyze how these investments contribute to the productivity, wages, and income of families in Nicaragua. In addition to understanding the relationship between health and productivity in Nicaraguan households, special attention is paid to regional differences and to identifying which health policy measures may have the greatest impact on health and earning capacities.

After a brief description of Nicaragua, the analysis proceeds in two stages. First, the determinants of the health of individuals are analyzed, differentiated by age and gender and socioeconomic characteristics such as urban–rural location, endowment of public services, housing conditions,

[2] See for example, Schultz and Tansel (1997) for Ghana and Côte d'Ivoire; Deolalikar (1988); Pitt et al. (1990); and Strauss (1986).

[3] The earlier per capita income figures for the Latin American countries are taken from the IDB (1998) and are not directly comparable to the figures presented for Ghana and Côte d'Ivoire, which are derived from the World Bank (1999).

region of residence, etc. Second, the returns to health status are analyzed using the income of wage-earning and self-employed workers as the measure of productivity. The remaining sections discuss policy implications and provide concluding remarks.

Nicaragua: A Low-Income Country in Transition

Nicaragua represents about one-third of the area of Central America. Topographically the country is divided into three regions. The Pacific region is the most densely populated and highly urbanized; the Central region is mountainous, less populated, and more sparsely settled; and the Atlantic region along the Caribbean coast is hot and humid, thinly populated, and characterized by an enclave economy whose activities are still predominantly rural and agricultural.

Several studies have demonstrated the high levels of poverty among the majority of the Nicaraguan population. According to a study by the World Bank (1995), 50 percent of the population was living below the poverty line in 1993, and 19 percent was living in extreme poverty. The most difficult situation is found in rural areas, where 76 percent of the population lives below the poverty line and 36 percent are in extreme poverty.[4]

Income distribution is profoundly unequal. According to a Country Economic Memorandum of the World Bank (1994b), the richest 20 percent of the population received 65 percent of national income, while the poorest 20 percent received only 3 percent—a ratio of 22 to 1. This ratio puts Nicaragua into a situation that is similar to the United States, but at significantly lower levels of average income.

The 1995 Nicaraguan census reported a population of approximately 4.1 million. The fertility rate between 1990 and 1995 was estimated to be 4.4 children per woman. A high fertility rate throughout the twentieth century has given the country a very young age structure: Fully 45 percent of the population is under 15 years old. Although population growth has slowed

[4] Based on the same 1993 survey but using a Basic Needs approach, the Ministerio de Acción Social (MAS), with assistance from UNDP, classified 44 percent of households as extremely poor and 31 percent as poor, leaving only 25 percent to be considered "not poor" (MAS et al., 1994b).

to 2.6 percent in recent years, the number of Nicaraguans is still expected to double during the next 30 years.[5]

Life expectancy at birth has been increasing in line with declining mortality. Life expectancy rose from 48.5 years in the 1960–65 period to 66 years in the 1990–95 period. In rural areas life expectancy was almost 10 years lower, but higher for women than for men. It is estimated that in 1990 the country had a general mortality rate of 10 per 1,000 inhabitants, with a high share accounted by infants and those over 50 years old.[6] Infant mortality accounted for 24 percent of all deaths in 1981. Diarrheic diseases are the primary cause of infant and preschool morbidity and mortality, and general mortality. The country also has high rates of infectious diseases, particularly malaria, dengue, and leishmaniasis. Among the immunogenic diseases susceptible to vaccines, measles and whooping cough reached epidemic levels in 1990.

Public policy in Nicaragua has been erratic. The mobilization of social programs under the Sandinista regime in the early 1980s reduced illiteracy and mortality, however this process was not sustained during the ensuing civil war and socioeconomic crisis. During the 1980s Gross Domestic Product (GDP) fell by an average of 1.8 percent annually; and by the end of this period, income per capita had fallen 36 percent. The accumulated crisis led to deterioration in general living and hygienic conditions. This spurred the reemergence of certain infectious and vaccine-preventable diseases as important causes of morbidity and mortality, and increased the demand for health services that were already stretched to the breaking point by thousands of refugees, repatriates, and demobilized soldiers. Following resolution of the internal conflict, the Nicaraguan economy underwent a series of structural changes intended to reduce inflation and reactivate economic growth, but without achieving substantial success.

Health policy in the 1980s was characterized by an effort to guarantee access, at public expense, to health services for the greatest number of people. This led to a serious fiscal deficit. Later, when economic policy focused on

[5] Figures in this and following paragraphs come from a variety of government data sources, including MAS et al. (1994a, 1994b), INEC (1992), Ministerio de Salud (1991, 1996), and Management Sciences et al. (1996).

[6] Deaths are substantially underreported in Nicaragua, so the data on mortality must be interpreted with considerable qualification.

reducing the deficit, public financing for health services fell. In the 1985–89 period, the public system represented approximately 75 percent of health spending. This ratio has nearly reversed, with the share of public outlays declining to 55 percent by 1994 and to a little more than 30 percent in recent years. Yet in 1993, the year of our data, the state still financed free medical care for 84 percent of all patients treated (Espinosa, 1996).

A Model for Analyzing Health and Productivity

The model used for this analysis follows closely the one presented in chapter 1. A two-stage least squares model was used to address endogeneity and measurement error in the health status variables. The first-stage regression explained health as a function of several individual, community, and policy variables—using the latter two sets of variables to identify the model. In the second-stage regression, wages were analyzed as a function of individual characteristics and a health status variable. The regressions were estimated without the health status variable, with the health status variable, and with the instrumented health status variable. The estimates were also done separately by gender, region, and type of employment to determine whether there were major differences among these categories.

In the first stage, several variables were selected and evaluated to represent health status. Different alternatives were studied, including days of illness, days disabled, and days in bed. Finally two models were selected: one utilizing the dichotomous variable of reported illness, and the other using reported days of illness. The model is necessarily incomplete since health status is the outcome of a complex process determined by a convergence of factors derived from personal characteristics, family, integration into the productive process and the market, the community in which the individual resides and works, and social exclusion.

The reduced form equation that was estimated for health status (H), using illness or days ill as dependent variables, is:

$$H = f\left(CI, CV, EC, OC\right) \quad [7.1]$$

in which CI represents individual characteristics such as schooling and labor experience in years, and CV represents family housing conditions (type of flooring, overcrowding, and sanitation) and the amount paid for electric

power as a proxy for household wealth. *EC* represents the level of social services available in the community, including the supply of health services (availability of personnel) and supply of primary education. *OC* represents the level of community organization, for which we use a specific community program called Casa Mujer.

In the second stage, an income function was estimated. The model of the income function was constructed taking into consideration variables representing individuals' capacities and opportunities in the labor market. The earnings function, as an extension of Mincer (1974), is:

$$W = f(E, EC, UT, H) \qquad [7.2]$$

in which *W* represents the log of hourly wages, explained by experience (*E*), schooling (*EC*), region (*UT*), and health status (*H*).

The Database and Sample

The database utilized here is the Living Standard Survey (ENV93). The survey was conducted in 1993 by the Instituto Nacional de Estadísticas y Censos (INEC), with international support. The survey collected data from 25,165 individuals in 4,554 households representative of rural and urban areas in all regions of the country.

The definition of the sampling framework was quite old, based on records prior to the programmed National Population Census of 1982 by INEC, and information from Vote Receiving Boards (JRV) covering the 1990 general elections. Given the lack of accurate data on the demographic distribution by region due to the internal and external migration of previous years, the sample size was standardized by regional strata. The defects in calculation of the sampling size make it impossible to make accurate projections for the nation as a whole.

The main questionnaire for individuals provided all the variables related to the personal characteristics of each respondent in the sample, as well as the respondent's labor and income data. Many variables had to be excluded because of inconsistencies, missing information, or an excessive reduction of the sample due to a very small number of valid records.

The household module of the 1993 survey provided all the data on housing conditions (flooring, roofing), water supply and sanitation, family

wealth, and others. The values of the variables included were assigned equally to all the individuals belonging to each household.

The information on public services was taken from a database, maintained by the Ministry of Social Action (MAS), on the supply of public and community services during the year covered by the living standards survey. The database is organized at a neighborhood level (i.e., the *comarca*, which represents the smallest political territorial unit, of which there are 7,500 nationwide), and has 162 variables at that disaggregated level. Given that both databases coincide only at the municipal level, the data was merged by municipality. That is, all the inhabitants of a municipality were assigned the same level of community facilities.

Finally, two mitigating factors must be mentioned that in combination had a significant impact on the study. First there is substantial evidence that the survey lacked adequate quality controls and that there were problems in the design of the questionnaire. Presumably these caused the many observations with missing data, and the inconsistency between many values in the same record. Furthermore, the definitions were frequently unclear, particularly in the case of defining income (e.g., Is the reported income figure net of costs for the self-employed? Is income produced by a rural family consisting of its agricultural production net of the family's own consumption in the rural areas?).

The second factor relates to the period of the survey, which was characterized by a dramatic transition as the armed conflict ended and there was a shift from a state-regulated to a market economy. The structural changes in the economy and in society affected thousands of Nicaraguans who either lost or gained jobs and associated benefits that supplemented their wages, or lost or gained ownership of housing and productive assets. These developments severely affected the labor market. The disequilibrium of markets undergoing dramatic institutional changes may account for some of the puzzling and unexpected relationships between key variables that will be described below.

The Sample

Of the 14,064 individuals of working age (18 to 65) who were surveyed in the sample, 9,439 were excluded because they did not work, did not report hourly wages, or did not report their type of employment. Of the 4,625 who remain, another 174 were excluded because of missing variables or responses

Table 7.1 Distribution of Study Sample by Gender, Area, and Economic Activity (%)

Gender	Area of Residence		Type of Employment		All
	Rural	Urban	Self-employed	Wage earner	
Women	23.6	76.4	37.6	62.4	27.0
Men	37.7	62.3	33.1	66.9	63.0
All	32.5	67.5	34.8	65.2	100.0
Number	1,446	3,005	1,548	2,903	4,451

Source: Nicaraguan Living Standard Survey, 1993.

that were outside a reasonable range of values.[7] The final sample, then, included 4,451 individuals, of whom 63 percent were men and 68 percent resided in urban areas (see Table 7.1). For other relevant characteristics of the sample, see table 7.2.

The sample was disaggregated by employment type into wage earners and the self-employed. Only the principle productive activity was taken into consideration because the data (particularly for income) related to secondary activity contained very few valid values. Wage earners are defined as respondents who stated they were workmen (*obreros*), employees, or domestic servants. The self-employed comprise workers who identified themselves as such or who were independent professionals. Individuals who stated they were nonremunerated relatives or employers were excluded.

Of the 4,451 individuals in the sample study, 65 percent were wage earners concentrated in the 25–34 age group, the largest of the age groups. The self-employed account for 35 percent of the sample, and are concentrated among individuals aged 35 to 44. This pattern is probably due to the drastic cut in state employment between 1987 and 1992 when there was weak demand for workers in the private sector, which led a large number of displaced public employees to seek income through their own economic activities.

[7] For more details regarding the sample and the selection criteria, see Espinosa and Hernández (1999).

Table 7.2 Living Standard Survey 1993 Sample Characteristics—Remunerated Individuals (18 to 60 Years Old)

Variables	Mean	Std. Dev.	Min	Max	Values
Individual Characteristics					
Area	0.68	0.47	0.00	1.00	0-Rural, 1-Urban
Gender	0.63	0.48	0.00	1.00	0-Women, 1-Men
Employment	0.65	0.48	0.00	1.00	0-Self-employed, 1-Wage earner
Age	34.85	11.32	18.00	65.00	Years
Schooling	5.76	4.69	0.00	18.00	Years
Experience	23.09	13.11	0.00	59.00	Years
Central region	0.33	0.47	0.00	1.00	0-No, 1-Yes
Pacific region	0.57	0.50	0.00	1.00	0-No, 1-Yes
Atlantic region	0.10	0.30	0.00	1.00	0-No 1-Yes
Health Status Indicators					
Illness in last 29 days	0.20	0.40	0.00	1.00	0-No, 1-Yes
Days of illness	1.61	3.97	0.00	28.00	Days
Income					
Income per hour	5.39	12.24	0.02	416.67	1993 cordobas
Logarithm income/hour	1.12	1.00	–4.17	6.03	Log (cordobas)
Housing Conditions					
Adequate floor	0.61	0.49	0.00	1.00	0-No, 1-Yes
Overcrowding – A	4.19	2.69	0.25	20.00	Persons per room
Overcrowding – B	0.35	0.48	0.00	1.00	0-<4.5 pers/room; 1->4.5 pers/room
Sanitation	0.87	0.34	0.00	1.00	0-No, 1-Yes
Electricity bill	45.48	76.08	0.00	800.00	Cordobas
Municipal Characteristics					
Nurses	1.94	1.76	0.24	9.84	1,000 inhabitants per nursing assistant
Doctors	3.58	3.09	0.28	26.18	1,000 inhabitants per doctor
Classrooms	0.06	0.03	0.02	0.18	1,000 children per primary classroom
Community Organization					
Casa de la Mujer	8.64	7.47	0.00	27.57	1,000 women per Casa de la Mujer

Note: Calculated for the sample of 4,451 people with earnings who were used in the study. Only 4,395 observations were used for the overcrowding variables due to missing data.

Table 7.3 Log Hourly Income by Gender, Residence Area, and Employment Type

Gender	Area of Residence		Employment Type		All
	Rural	Urban	Self-employed	Wage earner	
Women	0.76	1.27	1.26	1.08	1.15
Men	0.51	1.46	1.14	1.08	1.10
Average	0.58	1.38	1.19	1.08	1.12
Number	1,446	3,005	1,548	2,903	4,451

Source: Nicaraguan Living Standard Survey, 1993.

Hourly Income

To elicit information about income, respondents were asked how much they received for work done in the seven days prior to the survey, as well as payment or assistance in the form of in-kind benefits, meals at work, housing, uniforms, and bonuses. For the purposes of this study only cash income was considered—that is, payments in kind were excluded. However since bonuses are paid once a year and the income or wage was received in different installments (daily, weekly, bimonthly, or monthly), all forms of cash income were standardized by calculating an average monthly income and then calculating an implicit hourly wage.[8] As mentioned previously, the design of the questionnaire led to many ambiguities, and other researchers have also noted problems with the income figures reported in the 1993 survey.

Although the average income of urban workers is higher than that of rural workers, the gender differentials show that women in rural areas have higher earnings than men have (see Table 7.3). A study by the World Bank (1995) using the same 1993 survey reported a similar finding in the hourly and monthly income data, with higher income for women in several age groups, in rural areas, and in certain informal economic activities.

Women, both wage earners and the self-employed, have a slightly higher average income than do men overall. The lowest average income is recorded

[8] For more details on the definition of income, see Espinosa and Hernández (1999).

by the rural self-employed—who are primarily small-scale farmers with subsistence production. In general, wage earners have higher average incomes than the self-employed; only in urban areas do the latter exceed the average income of wage earners.

Health Status

The basic indicator of health status used in this analysis is based on self-reported illnesses in the 30 days preceding the survey. Admittedly this indicator fails to objectively capture the full range of somatic and psychic development and environmental relationships that are important to measuring an individual's health status. Nevertheless the self-reported measure is the indicator provided by the survey, and problems generated by the measurement error inherent in this variable can be addressed through the use of instrumental variable techniques.

Four questions in the 1993 survey were related to an individual's health status. The first asked whether the individual had been ill or in an accident in the 30 days prior to the survey. If the answer was affirmative, three additional questions were asked: How many days were you ill? Did you stay in bed because of the illness? How many days did you stay in bed because of this illness?

Of individuals who reported illness or an accident, 30 percent indicated that they had been forced to "stay in bed" (see Table 7.4). This response was more frequent among women (30.1 percent) and the self-employed (32.6 percent). The relatively small number of reported cases that involved the respondent staying in bed and the poor performance of this variable in our analysis caused it to be discarded as an indicator of health status for this study. The health status indicators that were used are (1) whether or not the individual reported an illness and (2) the number of days ill.

In reviewing the data, an appreciable number of individuals reported being ill for the full 30 days prior to the survey. Presumably many of these individuals were chronically and acutely ill. Retaining this group in the sample substantially impaired the performance of the models, and so all of these 134 individuals were subsequently excluded. The sample for the health indicator of days ill therefore is limited to respondents reporting between 1 and 29 days of illness. This raises a question of whether other studies would also benefit by distinguishing the qualitative difference between

Table 7.4 Distribution of Illness by Gender, Area, and Employment Type

Sample		All	Reported ill	Incidence (%)
by area				
Women	Rural	390	98	25.1
	Urban	1,262	280	22.2
	All	1,652	378	22.9
Men	Rural	1,056	242	22.9
	Urban	1,743	278	15.9
	All	2,799	520	18.6
Total	Rural	1,446	340	23.5
	Urban	3,005	558	18.6
	All	4,451	898	20.2
by Employment type				
Women	Self-employed	621	184	29.6
	Wage earners	1,031	194	18.8
	All	1,652	378	22.9
Men	Self-employed	927	208	22.4
	Wage earners	1,872	312	16.7
	All	2,799	520	18.6
Total	Self-employed	1,548	392	25.3
	Wage earners	2,903	506	17.4
	All	4,451	898	20.2

Source: Nicaraguan Living Standard Survey, 1993.

health status as indicated by reports of occasional illnesses from those due to chronic or acute conditions.

In the total sample, 898 individuals (20.2 percent) reported illness or an accident in the 30 days prior to the survey. The largest incidence was in the rural area with 23.5 percent, compared to 18.6 percent in urban areas. Women (22.9 percent) reported illness more frequently than did men (18.6 percent).

The gap in incidence of reported illness widens according to condition of employment: Among the self-employed, 25.3 percent suffered some kind of illness, higher than the figure for wage earners, which was 17.4 per-

cent. This situation is unexpected because, presumably, the self-employed are more susceptible to illness than wage earners who should be able to get medical care through their social security coverage. However, in Nicaragua social security covers no more than 15 percent of the economically active population; the majority of wage earners are not covered by the system either because they are linked to productive units in the informal sector or to seasonal activities (e.g., harvesting of agricultural export products).

The average number of days of illness reported by individuals is 8.0. The higher incidence of illness found among the self-employed is also reflected in a higher average number of days ill—an average of 8.4 (see Table 7.5). Women also report a higher average number of sick days than do men. The higher incidence rate and the higher average number of days ill for women occur among both wage earners and the self-employed.

Urban wage-earning men report the lowest average number of days ill, while self-employed rural women workers record the highest average number of days ill. These differences could have numerous potential causes related to such factors as access to medical services, norms regarding roles in the family, exposure to occupational hazards, or employer supervision and contractual forms.

Health Status and Income: A Preview

Despite the subjectivity of the perception of being ill, there is an empirical relationship between reported illness and average income in different groups. Without exception, income is lower on average in the groups that reported

Table 7.5 Average Days of Illness by Gender, Area, and Employment Type

	Self-employed			Wage earner			All		
Gender	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All
Women	9.9	8.4	8.9	7.5	7.7	7.6	8.8	8.0	8.2
Men	7.6	8.2	7.9	8.2	7.4	7.8	7.9	7.7	7.8
Average	8.4	8.3	8.4	8.0	7.5	7.7	8.2	7.9	8.0
Number	154	238	392	186	320	506	340	558	898

Source: Nicaraguan Living Standard Survey, 1993.

Table 7.6 Log Hourly Income by Health Status, Gender, Area, and Employment Type

Health status	Gender		Area of residence		Employment type		All
	Women	Men	Rural	Urban	Self-employed	Wage earner	
Not ill	1.16	1.13	0.59	1.38	1.23	1.09	1.14
Reported ill	1.11	0.99	0.53	1.35	1.08	1.01	1.04
All	1.15	1.10	0.58	1.38	1.19	1.08	1.12
Difference (%)	–4.9	–13.1	–5.8	–3.0	–13.9	–7.7	–9.5
Observations	1,652	2,799	1,446	3,005	1,548	2,903	4,451

Source: Nicaraguan Living Standard Survey, 1993.

less illness; however the size of the gap differs across groups. While income fell by 4.9 percent for women who reported being ill, the 13.1 percent drop for men who were ill was almost three times greater.

In terms of region of residence, equally significant differences are found (see Table 7.6). Individuals who are ill in the Pacific region report hourly income that is only 6.8 percent lower on average, compared to a difference of 11.3 percent in the Central region and as much as 18.1 percent in the Atlantic region (see Table 7.7).

Schooling

In general, income increases with years of schooling. However this pattern does not appear to hold in Nicaragua during the survey period for several reasons. For example, the predominance of export agriculture may account for the fact that individuals with a technical education (13+ years) have a higher income in rural than in urban areas. This occurs because rural technicians work on large-scale plantations oriented for foreign sales—the most dynamic economic sector in 1993. Also, among respondents with a primary education, women have higher incomes than men do. This pattern may be due to the fact that women tend to work in the independent informal sector in which they earn higher incomes than do unskilled workers in other sectors. However for those with at least a secondary education, it is more common for male earnings to exceed female earnings.

Table 7.7 Log Hourly Income by Health Status, Gender, and Region

Health status	Geographical region			All
	Pacific	Central	Atlantic	
Not ill	1.36	0.75	1.11	1.14
Reported ill	1.29	0.63	0.91	1.04
Average	1.35	0.73	1.08	1.12
Difference (%)	-6.8	-11.3	-18.1	-9.5
Observations	2,549	1,445	457	4,451

Source: Nicaraguan Living Standard Survey, 1993.

The relationship between illness and schooling is not as clear as the relationship of illness with income. Although days of illness decrease slightly as years of schooling increase, in the aggregate and when disaggregated by gender and area of residence, there are still numerous irregularities that make it impossible to generalize.[9]

Household Characteristics

Both public and family investments are responsible for the services available to families and the characteristics of their dwellings. Four variables representing household characteristics were extracted from the survey to represent living conditions that are likely to be related to health status. These are:

- *Overcrowding*, assuming a qualitative difference starting at 4.5 or more people per room.
- Housing conditions, expecting *dirt floors* to be associated with poorer health.
- Housing conditions, expecting the *lack of a sewerage system or toilet* to be associated with poorer health.

[9] See Espinosa and Hernández (1999) for more details.

- *Household wealth*, using *electricity costs* as a proxy for wealth since the amount is linked to household consumption and ownership of electrical appliances.

In simple cross tabulations, the number of days ill reported by individuals is higher in houses that have dirt floors and those that report more than 4.5 persons per room. Sanitation is particularly uneven when comparing countryside and city. About 39 percent of rural households lack proper sanitation, while only 2.7 percent of urban households report inadequate sanitation. Simple cross tabulations demonstrate that more days of illness are reported in households that lack sanitation. However sanitation services are also correlated with income, so a definitive statement on this and the other relationships awaits proper controls for income and for the interrelationship between income and health.

Access to potable water, which was expected to have a strong relationship to health status, was excluded from the reported estimations because, surprisingly, it was never significant in any of the specifications for explaining health status.

Access to Social Services

The 1993 survey did not contain information on the services available to the communities in which the respondents resided. Instead, information on social services for that year was obtained from the Ministry of Social Action (MAS). The MAS database, which is derived from institutional reports and local governments, measures the supply of health services by the availability of human resources (nurses and doctors) and primary schoolrooms (see Table 7.8).

In the Nicaraguan health system, nursing assistants play a very important role in community primary care services. They play key roles in disseminating information on health maintenance and disease prevention, promoting hygiene, and dispensing preventive care and immunizations. Priority is given to programs for maternal–infant care and the control of epidemics. Nurse assistants are distributed among the most peripheral units in urban and rural areas. The Central region has relatively few nurse assistants; the Atlantic region has a slightly higher demographic ratio of personnel to population; and the Pacific region has the highest concentration.

Doctors, who almost exclusively practice therapeutic medicine and organize their work in response to the demand for services, are largely concentrated in urban areas. The Pacific region contains the largest number of doctors. The Central and Atlantic regions are comparatively underserved. In the study analysis, the number of doctors per 1,000 inhabitants was excluded since a greater supply was associated with more days of illness. Although plausible explanations for this phenomenon exist—i.e., that doctors generate their own demand or individuals served by doctors are more likely to report being ill—it is impossible to determine the actual cause of this relationship. Apparently doctors do not have enough positive impact on health status to outweigh these other measurement problems.

As was the case with health services, the Pacific region is also relatively privileged with respect to the supply of education.

Table 7.8 Supply of Social Services by Region and Area

	Region I			
Area	**Pacific**	**Central**	**Atlantic**	**All**
	1,000 inhabitants per nursing assistant			
Rural	1.62	2.42	2.20	2.10
Urban	1.65	2.58	1.65	1.87
All	1.64	2.50	1.85	1.94
	1,000 inhabitants per doctor			
Rural	2.72	4.55	6.60	4.10
Urban	2.61	4.34	5.78	3.32
All	2.63	4.45	6.08	3.58
	1,000 children per classroom			
Rural	0.05	0.06	0.10	0.06
Urban	0.05	0.07	0.07	0.06
All	0.05	0.07	0.08	0.06
	1,000 women per Casa Mujer unit			
Rural	8.61	3.28	7.13	5.72
Urban	12.18	5.53	6.16	10.04
All	11.42	4.38	6.52	8.63

Source: MAS.

In response to the restriction on public health services since the late 1980s, a number of social organizations (trade unions, employer associations, neighborhood associations, and NGOs) with assistance from international organizations created alternative services for the most vulnerable sectors of the population. This was especially true of community services oriented toward women and children. These organizations operate in the most marginal municipalities and include the Women for Health Network and the national coordinating committee of organizations that work with children. The state has been forced to coordinate its policies and services with these movements.

The Casa Mujer phenomenon exemplifies this effort by civil society to fill some of the vacuum left by the contraction of public services. It is generally made up of private nonprofit service units, operated and funded by NGOs or civil associations. Their outreach tends to be oriented toward women, providing health promotion and preventive care to reduce risk during pregnancy and childbirth. Most units also include follow-up services for newborns and infants and extensive programs of education and community mobilization. Affiliates of Casa Mujer are distributed very unevenly across municipalities, and vary greatly in technological and decision-making capacity. The number of Casa Mujer service units was selected because it was the most suitable indicator available for representing the supply of such community-organized services.

In contrast to the distribution of public resources, social movements give priority to the most deprived areas of the country. As a result, community services are especially common in the Central and Atlantic regions and their rural areas.

The Determinants of Health Status in Nicaragua

Thus far, we have looked at health status determinants separately. Using a probit analysis, it is possible to estimate how much each of these factors contributes to the probability of falling ill. The results of this analysis are shown in Table 7.9. The dependent variable is dichotomous and discrete, equal to 1 if the individual reported being ill in the last 29 days and equal to 0 if the individual reported no illness. A positive coefficient indicates that the factor *increases* the probability of having been ill, while a negative coefficient indicates that the factor *reduces* the probability of having been ill.

In this Nicaraguan data, after controlling for other factors, age is still associated with a greater probability of reporting illness. The experience variable, which is simply a linear transformation of age and education, has a positive coefficient, although it is not statistically significant. The joint effect of experience and experience squared in the various subsamples is stronger, but still not significant at the 10 percent level. Nevertheless, the effect appears to be stronger when the sample is split into its various groupings.

Reported illness is also positively correlated with schooling, but again, the statistical significance is weak. As noted in other studies, the impact of education on reported illness can be ambiguous since it may be associated with different perspectives and norms regarding the definition of being "ill."

Unlike education and age, residence appears to have a strong impact on health. People in the Central and Pacific regions report illness more frequently than those in the Atlantic region. This is true for the Pacific region in all the subsamples, and is less consistent for the Central region across subsamples. This is somewhat counterintuitive because, as described previously, the Pacific and Central regions disproportionately benefit from the availability of health services and are economically more prosperous.

Housing conditions, indicated by whether or not the house has a dirt floor or adequate sanitation, show the expected signs as risk factors that would increase the probability of being ill. However, only in the case of dirt floors in the subsample of women is there a statistically significant effect.

With respect to the health policy variables, estimates that included the supply of doctors generally showed an inverse relationship—the more doctors, the more illness. Consequently the variable was excluded in the results presented here, as discussed earlier, since its interpretation is ambiguous. Nevertheless the other health service variable demonstrated a positive and significant effect. The number of nursing assistants is associated with better health status. The effect is strongest for women and in rural areas. This suggests that the work of nursing assistants may effectively improve the health of their charges since the main kinds of care they provide—preventive services and maternal and infant care—are more oriented toward women. It also suggests that nurses may be especially effective in rural areas, where they are often the only health professionals available; whereas in urban areas, their presence may be closely associated with, indeed an indication of, high concentrations of impoverished and underserved populations.

Table 7.9 Probit Estimates for Being Ill by Gender and Area

Health determinants	Total	Women	Men	Urban	Rural
Experience	0.005	–0.002	0.010	–0.001	0.008
	0.750	*–0.150*	*1.050*	*–0.070*	*0.950*
Experience^2	0.000	0.000	0.000	0.000	0.000
	0.930	*1.130*	*0.260*	*1.290*	*0.160*
Schooling	0.001	–0.005	0.007	0.017	–0.004
	0.200	*–0.470*	*0.760*	*1.310*	*–0.530*
Pacific region	0.224	0.281	0.193	0.139	0.268
	2.770	*1.960*	*1.940*	*1.080*	*2.520*
Central region	0.151	0.298	0.076	0.073	0.184
	1.810	*2.000*	*0.750*	*0.590*	*1.610*
Floor	–0.072	–0.144	–0.027	0.005	–0.120
	–1.360	*–1.690*	*–0.390*	*0.060*	*–1.800*
Sanitation	–0.058	–0.154	–0.039	–0.122	–0.049
	–0.800	*–1.060*	*–0.460*	*–1.450*	*–0.280*
Electricity bill	–0.001	–0.001	–0.001	0.000	–0.001
	–1.860	*–1.200*	*–1.510*	*–0.030*	*–2.100*
Inhab. per nurse (1,000s)	0.022	0.036	0.013	–0.030	0.046
	1.790	*1.770*	*0.850*	*–1.350*	*3.070*
Women per Casa Mujer (1,000s)	0.010	0.013	0.007	0.012	0.008
	2.970	*2.630*	*1.700*	*2.260*	*1.950*
Employee	–0.189	–0.258	–0.143	–0.156	–0.209
	–4.050	*–3.400*	*–2.380*	*–2.010*	*–3.530*
Gender	–0.167			–0.031	–0.213
	–3.680			–0.370	–3.900
Urban	–0.168	0.001	–0.265		
	–2.960	*0.010*	*–3.690*		
Constant	–0.904	–0.865	–1.081	–0.887	–1.034
	–6.120	*–3.460*	*–6.100*	*–3.550*	*–4.310*
Observations	4,451	1,652	2,799	1,446	3,005
Pseudo-R^2	0.03	0.04	0.03	0.03	0.04
Chi^2	133.31	65.50	67.61	41.62	97.06

Notes: The dependent variable in all cases is 1 if the individual reported being ill in the previous 29 days, and 0 otherwise; t-statistics are in italics.

The level of community organization, expressed by the number of care centers for women per municipality, reflects a clear link with improved health status. The preventive and educational work done by these community centers of organized women seems to reduce the probability of falling ill. The impact is statistically significant for all subsamples, but is strongest for women and in urban areas.

Although the model overall is statistically significant, as demonstrated by the Chi^2 and the pseudo-R^2, it is not very powerful in predicting the likelihood of reported illness. However, it appears to be strong enough to serve the purpose of demonstrating the impact of nurse assistants and the Casa Mujer community centers, which will serve as exogenous instruments for identifying the health variable when it is introduced into the second-stage earnings function.[10]

The Determinants of Wages: Including Health in the Equation

A standard Mincerian wage function was estimated with the data in three different specifications. First, the wage function is estimated without the health indicator, as is common in most studies of wages. Second, the health indicator is included to see whether it is significant and how it affects the performance of the model and the other included variables. Finally, a third estimation is done that includes the health indicator corrected for endogeneity and measurement error through the use of the instrumental variables as derived from the first-stage regressions.

The earnings equations are quite robust (see Tables 7.10 and 7.11). The first, fourth, and seventh columns of Table 7.10 and the first and fourth columns of Table 7.11 show the earnings equations without any health variable. The standard explanatory variables, experience and education, are strongly significant and have the expected signs. Being self-employed, a woman, or living in rural areas significantly reduces hourly earnings. On average, living in the Pacific region is associated with higher earnings than the Atlantic region, whereas living in the Central region is a disadvantage.

---

[10] More-detailed regressions for different subsamples, and also for the continuous variable "days ill" can be found in Espinosa and Hernández (1999). The results are generally less statistically significant, but the overall conclusions remain the same.

Table 7.10 Wage Regressions With and Without Health

Variable	All			Women			Men		
Constant	–0.01	–0.00	0.17	0.55	0.56	0.66	–0.17	–0.16	–0.09
	–0.13	*–0.01*	*1.23*	*4.33*	*4.33*	*4.01*	*–1.77*	*–1.66*	*–0.49*
Experience	0.03	0.03	0.03	0.01	0.01	0.00	0.05	0.05	0.05
	7.26	*7.26*	*6.84*	*0.84*	*0.84*	*0.61*	*8.59*	*8.60*	*8.40*
Experience^2	0.00	0.00	–0.00	0.00	–0.00	0.00	0.00	–0.00	–0.00
	–5.25	*5.22*	*–3.91*	*–0.05*	*–0.03*	*0.34*	*–6.74*	*–6.72*	*–6.30*
Years of school	0.08	0.08	0.08	0.07	0.07	0.07	0.09	0.09	0.09
	22.62	*22.61*	*21.08*	*12.54*	*12.52*	*10.97*	*18.29*	*18.32*	*17.44*
Pacific region	0.08	0.09	0.14	0.13	0.13	0.17	0.06	0.06	0.08
	1.90	*1.97*	*2.35*	*1.70*	*1.71*	*1.96*	*1.04*	*1.10*	*1.09*
Central region	–0.24	–0.24	–0.20	–0.15	–0.14	–0.10	–0.27	–0.27	–0.26
	–5.16	*–5.11*	*–3.53*	*–1.82*	*–1.80*	*–1.09*	*–4.82*	*–4.79*	*–4.12*
Employee	–0.21	–0.21	–0.26	–0.33	–0.33	–0.37	–0.14	–0.14	–0.16
	–7.46	*–7.53*	*–5.77*	*–7.37*	*–7.37*	*–6.01*	*–3.93*	*–3.99*	*–2.87*
Female	0.14	0.14	0.10						
	5.22	*5.14*	*2.70*						
Urban	0.45	0.45	0.41	0.23	0.23	0.22	0.56	0.56	0.53
	14.65	*14.57*	*9.50*	*4.40*	*4.39*	*4.19*	*14.54*	*14.38*	*6.58*
Ill		–0.04			–0.02			–0.07	
		–1.38			*–0.30*			*–1.60*	
Ill (Instrumented)			–0.87			–0.51			–0.51
			–1.61			*–1.02*			*–0.53*
Observations	4,451	4,451	4,451	1,652	1,652	1,652	2,799	2,799	2,799
R^2	0.27	0.27	0.16	0.19	0.19	0.14	0.33	0.33	0.30

Note: t-statistics are in italics.

Not only do men earn more on average than women do; they also have higher returns to experience and education. This may be the consequence of occupational segregation in Nicaragua, where women are more likely to be found in unskilled informal commerce (self-employment) than are men. It is also common for women to interrupt their working life, usually when they have children. Men, on the other hand, as well as having a higher probability of continuous years of service, have more opportunities to improve their proficiencies and become skilled workers or technicians. This level of technical and professional training is also found among self-employed men. An additional year of schooling is associated with an 8 percent increase in earnings. The returns for men are somewhat higher than for women, 9 percent and 7 percent respectively.

The regional variables reflect the wide territorial gaps existing in Nicaragua. To reside and work in the Pacific region means 8 percent more income than in the Atlantic region, while residence in the Central region means almost 24 percent less income than in the Atlantic region. This distribution coincides with regional patterns of poverty and underemployment.

When disaggregating the sample by rural and urban areas, all the same variables remain significant (see Table 7.11). Interestingly there is no apparent difference in returns to experience and schooling across rural and urban areas. The most important differences emerge for women and the self-employed. In urban areas, women earn less than men and the self-employed earn less than the employees. However, in rural areas these relationships are reversed; that is, women earn more than men and the self-employed earn more than employees.

The next step is to introduce the health indicator, in this case whether or not the individual reported being ill in the previous 29 days. The results of these estimations, without any corrections for endogeneity or measurement error, can be found in the second, fifth, and eighth columns of Table 7.10, and in the second and fifth columns of Table 7.11. In all of the cases, the coefficient is negative, indicating that reported illness is negatively associated with hourly earnings—even after controlling for the other factors. But the coefficient is never statistically significant, even though it comes close to being significant at the 10 percent level for the sample of men. The coefficient is statistically significant for subsamples of urban women and rural men (not shown, but reported in Espinosa and Hernández, 1999). The continuous variable of days ill had similar findings in terms of magnitudes and significance.

Table 7.11 Wage Equation with Health Variables by Area

Variable	Urban			Rural		
Constant	0.07	0.09	0.16	0.44	0.44	0.56
	0.47	*0.58*	*0.82*	*4.79*	*4.81*	*4.49*
Experience	0.04	0.03	0.03	0.03	0.03	0.03
	4.11	*4.10*	*4.00*	*5.82*	*5.82*	*5.59*
Experience^2	–0.00	–0.00	–0.00	–0.00	–0.00	–0.00
	–3.55	*–3.49*	*–2.81*	*–3.54*	*–3.54*	*–3.00*
Years of school	0.08	0.08	0.09	0.08	0.08	0.08
	10.50	*10.54*	*10.51*	*20.98*	*20.96*	*18.70*
Pacific region	0.03	0.03	0.05	0.11	0.11	0.17
	0.31	*0.36*	*0.53*	*2.13*	*2.16*	*2.57*
Central region	–0.40	–0.40	–0.39	–0.12	–0.12	–0.07
	–5.24	*–5.23*	*–4.88*	*–2.05*	*–2.02*	*–0.99*
Employee	0.08	0.08	0.06	–0.37	–0.37	–0.41
	1.69	*1.61*	*0.96*	*–10.96*	*–10.97*	*–8.54*
Female	–0.11	–0.11	–0.12	0.25	0.25	0.21
	–2.04	*–2.07*	*–2.15*	*8.40*	*8.34*	*5.10*
Ill		–0.08			–0.02	
		–1.45			*–0.54*	
Ill (Instrumented)			–0.47			–0.76
			–0.65			*–1.56*
Observations	1,446	1,446	1,446	3,005	3,005	3,005
R^2	0.15	0.15	0.13	0.19	0.19	0.09

Note: T-statistics are in italics.

The third step is to introduce the corrected health indicator into the model. Columns three, six, and nine in Table 7.10 and columns three and six in Table 7.11 report these results. For the most part, as seen in previous chapters, both the magnitude of and the statistical significance of the coefficient for the health indicator increase when the instrumental variable method is applied. This is not the case for the samples of men only or of the urban area only. In the rest, however, the precision and magnitude improve. These findings also appear when the days-ill indicator is used (see Table 7.12).

Table 7.12 Summary of the Impact of Health Status on Earnings

Variable	Wage function with OLS			Wage function with 2SLS		
	All	Women	Men	All	Women	Men
Dichotomous variable of health status						
Coefficient	–0.044	–0.015	–0.067	–0.873	–0.512	–0.508
t-statistic	*–1.38*	*–0.30*	*–1.60*	*–1.61*	*–1.02*	*0.53*
R^2	0.271	0.188	0.325	0.162	0.137	0.299
Continuous variable of health status (days ill)						
Coefficient	–0.004	–0.004	–0.005	–0.075	–0.086	–0.500
t-statistic	*–1.28*	*–0.96*	*–1.10*	*–1.61*	*–1.66*	*0.045*
R^2	0.271	0.189	0.338	0.194	0.040	0.306

Source: Table 7.10.

The most marked decline in income is measured for the entire sample. Due to the log-linear specification of the equation, the coefficient of –0.87 is equivalent to a 58 percent difference in hourly earnings. That is, having the health status associated with reporting illness in the survey is associated with a 58 percent reduction in hourly earnings for the entire sample. This is statistically significant at the 10 percent level, but the confidence interval is fairly large, so the precise value should be treated cautiously. The poor fit of the health regressions, and the rather imprecise estimates of the health impact, may indicate that health has an uncertain relationship with income, that the self-reported health indicators are weak measures of health status, or that the design problems in the survey generated too much "noise." Nevertheless, bearing in mind these qualifications, these findings do suggest that there may be very significant effects of health status on productivity as measured by hourly earnings.

The Potential Impact of Public Health Policies[11]

Based on the estimates of the determinants of health status in Nicaragua, a few factors were identified as having significant impact. These factors were

[11] This section summarizes results that are reported in much greater detail in Espinosa and Hernández (1999).

whether or not the family's dwelling had a floor or sanitation, the presence of nurse assistants, and the presence of a Casa Mujer unit.

Using the regression results, it is possible to do a thought experiment. How would health status improve if everyone had sanitation or the number of nurse assistants increased? Furthermore, what impact would the resulting changes in health status have upon hourly earnings?

Simulating the results of such policies on the sample shows that improving housing characteristics so that no one has dirt floors and everyone has sanitation could improve health status by the equivalent of a 5 percent reduction in days ill. The effect would apparently be greater for women (12 percent) than for men (3 percent). The subsequent effect on wages is rather small, less than a 1 percent increase in hourly earnings; but for women, the estimates would predict a 2.6 percent improvement in hourly earnings.

Following the same approach, an increase of 25 percent in the number of nurse assistants would be associated with a 2 percent reduction of days ill. This is roughly comparable for men and women at the mean. The associated wage improvement would be only 0.3 percent for the sample as a whole, and almost 0.6 percent for women. Similarly, a 25 percent increase in the number of Casa Mujer units would reduce days ill by about 3 percent, with a small impact (0.4 percent) on average hourly earnings.

Conclusions

The analysis of Nicaraguan data suggests that health status is significantly affected by housing conditions and the presence of particular health services (nurse assistants) and community services (Casa Mujer). Furthermore, better health status—as measured by self-reported illness—is associated with higher hourly income, on average. If the point estimates are valid, then an individual with the health status associated with reporting an illness in the last 30 days will have wages that are 58 percent lower, on average, than an individual with the health status associated with reporting no illness. We believe that the quality of the data is the main reason that the estimates are not more precise or significant. Future research could seek to evaluate these findings with other, and hopefully better, surveys.[12]

[12] Another Living Standard Survey was conducted in 1996. At the time this study was undertaken, the 1996 survey was not yet available.

The use of the 1993 Living Standards Survey also limited the number of variables that could be explored. It would have been useful to have a richer base of information on such things as nutrition, behavior (e.g., smoking), use of health services, and the quality of services available in the community. As a result, the policy analysis is necessarily restricted, and only limited inferences can be taken from the findings that are reported here.

Nevertheless, the findings do concur with the other studies in this book in providing evidence that health affects earnings. If the point estimates here hold up to further scrutiny, it may also suggest that there are declining marginal returns to health since the effects estimated here are much larger than in the other, higher-income countries. The results are also promising for future evaluations of public health policy. The presence of nurse assistants and the Casa Mujer program are strongly correlated with improved health status in ways that are consistent with the content of the services that they provide (i.e., having stronger effects for women and in rural areas). Clearly, better data and controls are needed to verify these results, but the findings are robust and suggest that some distinctions can be made between public services that have demonstrated effects and those that do not.

REFERENCES

Alves, D., et al. 1999. "Health, Development and Policies in a Warming Environment." São Paulo, Brazil, report for the Research Department of the Inter-American Development Bank, Washington, D.C.

Anderson Schaffner, J. (1998). "Premiums to Employment in Larger Establishments: Evidence from Peru." *Journal of Development Economics*. 55 (1): 81–113.

Avila-Curiel, A., T. Shamah-Levy, C. Galindo-Gómez, G. Rodrigo-Hernández, and L. M. Barragán-Heredia. 1998. "La desnutrición infantil en el medio rural." *Salud pública de México*. 40: 150–60.

Barro, R., and X. Sala-I-Martin. 1995. *Economic Growth*. Advanced Series in Economics. New York: McGraw Hill.

Bean, J. A., J. D. Leeper, and P. B. Wallace. 1979. "Variations in the Reporting of Menstrual Histories." *American Journal of Epidemiology*. 109 (2): 181–5.

Becker, G. 1964. "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education." Reprinted in G. Becker. *Human Capital*. New York: Columbia University Press, 1975.

______. 1965. "A Theory of the Allocation of Time." *The Economic Journal*. 75 (299): 493–517.

______. 1993. *Human Capital*. 3rd ed. Chicago: University of Chicago Press.

Behrman, J. R. 1990. "Macroeconomic Adjustment, Household Food Consumption, Nutrient Intakes, and Health Status." In J. R. Behrman. *Macroeconomic Reforms, Poverty and Nutrition: Analytical Methodologies*. Ithaca, New York: Cornell Food and Nutrition Policy Program.

Bergsten-Brucefors, A. 1976. "A Note on the Accuracy of Recalled Age at Menarche." *Annals of Human Biology*. 3 (1): 71–3.

Bielicki, T., A. Waliszho, and B. Hulanicka. 1986. "Social-Class Gradients in Menarcheal Age in Poland." *Annals of Human Biology*. 13 (1): 1–11.

Bliss, C., and N. Stern. 1978. "Productivitty, Wages and Nutrition." *Journal of Development Economics*. 5: 331–62.

Bobadilla, J. L., J. Frenk, and R. Lozano. 1993. "The Epidemiologic Transition and Health Priorities." In D. Jamison, A. R. Measham, W. H. Mosley, and J. L. Bobadilla, editors. *Disease Control Priorities in Developing Countries.* Oxford: Oxford University Press, for the World Bank.

Bojlén, K., and B. M. Weis. 1971. "Seasonal Variation in the Occurrence of Menarche in Copenhagen Girls." *Human Biology.* 43 (4): 493–501.

Bonilla-Chacin, M., and J. S. Hammer. 1999. "Life and Death Among the Poorest." Mimeographed document. Johns Hopkins University and the World Bank. Washington, D.C. (April)

Bound, J. 1992. "Self-reported versus Objective Measures of Health in Retirement Models." *Journal of Human Resources.* 26 (11): 106–38.

Bound, J., D. A. Jaeger, and R. M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables Is Weak." *Journal of the American Statistical Association.* 90 (430): 443–50.

Brundtland, G., and L. Walløe. 1973. "Menarcheal Age in Norway: Halt in the Trend towards Earlier Maturation." *Nature.* 241: 478–9.

Buchinsky, M. 1994. "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression." *Econometrica.* 62 (2): 405–58.

Cameron, N., and I. Nagdee. 1996. "Menarcheal Age in Two Generations of South African Indians." *Annals of Human Biology.* 23 (2): 113–9.

Clark, D. O. 1997. "Risk Factors for Disability and Functional Limitations among Blacks and Hispanics." AARP Andrus Foundation final report. Indianapolis: Regenstrief Institute for Health Care, Indiana University School of Medicine.

Clark, R. L., and R. Anker. 1993. "Cross-national Analysis of Labor Force Participation of Older Men and Women." 1993. *Economic Development and Cultural Change.* 13 (79): 489–512.

Consejo Nacional de Población (CONAPO). 1993. *Indicadores socioeconómicos e indice de marginación municipal, 1990.* México, D.F.: CONAPO.

Cortez, R. 1999. "Salud y productividad en el Perú. Un análisis empírico por género y región." Working Paper Series R-363. OCE, Inter-American Development Bank, Washington, D.C. (May).

Costa, D. 1996. "Health and Labor Force Participation of Older Men, 1900–1991." *Journal of Economic History.* 56 (1): 62–89.

Cravioto, P., J. Cravioto, and G. Bravo. 1987. "Age of Menarche in a Rural Population: Accuracy of Records Four Years Later." *Boletín médico del Hospital Infantil de México.* 44 (10): 589–93.

Cumming, D. C. 1990. "Menarche, Menses and Menopause: A Brief Review." *Cleveland Clinical Journal of Medicine.* 57 (2): 169–75.

Davis, M. A., D. J. Moritz, J. M. Newhaus, J. D. Barclay, and L. Gee. 1997. "Living Arrangements, Changes in Living Arrangements and Survival among Community Dwelling Older Adults." *American Journal of Public Health.* 87 (3): 371–7.

Davey Smith, G., M. J. Shipley, and G. Rose. 1990. "Magnitude and Causes of Socioeconomic Differentials in Mortality: Further Evidence from the Whitehall Study." *Journal of Epidemiology and Community Health.* 44: 265–70.

Delgado, H. L., and E. Hurtado. 1990. "Crecimiento físico y menarquia en adolescentes de Guatemala." *Archivos Latinoamericanos de nutrición.* 40 (4): 503–17.

Deolalikar, A. B. 1988. "Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India." *Review of Economics and Statistics.* 70 (3): 406–13.

Díaz de Mathman, C., V. Landa Rico, and R. Ramos Galván. 1968. "Crecimiento y desarrollo en adolescentes femeninos: Edad de la menarquia." *Boletín médico del Hospital Infantil de México.* 25: 787–93.

Dirección General de Estadística e Informática de la Secretaría de la Salud. 1998. "Aspectos relevantes sobre la estadística de deficiencias de la nutrición." *Salud pública de México.* 40 (2): 206–15.

Dreizen, S., C. N. Spirakis, and R. E. Stone. 1967. "A Comparison of Skeletal Growth and Maturation in Undernourished and Well-Nourished Girls before and after Menarche." *Journal of Pediatrics.* 70: 256.

Dunlop, D., S. Hughes, and L. Manheim. 1997. "Disability in Activities of Daily Living: Patterns of Change and a Hierarchy of Disability." *American Journal of Public Health.* 87 (3): 378–83.

Económica Consultores. 1996. "Estudio para la implementación del sistema de cargos a los usuarios de carreteras." Mimeographed document. Económica Consultores, Bogotá, Colombia.

Espinosa, J. 1996. "Pobreza y salud en Nicaragua." Mimeographed document. CRIES/NITLAPAN – UCA, Managua (March).

Espinosa, J., and C. Hernández. 1999. "Productividad de la inversión en salud de los hogares en Nicaragua." Working Paper Series R-362. OCE, Inter-American Development Bank, Washington, D.C. (May).

Eveleth, P. 1978. "Population Differences in Growth: Environmental and Genetic Factors." In F. Falkner and J. M. Tanner, editors. *Human Growth: A Comprehensive Treatise.* Vol. 2. New York: Plenum Press.

Falkner, F., and J. M. Tanner, editors. 1986. *Human Growth: A Comprehensive Treatise.* 2nd ed. Vol. 1–3. New York: Plenum Press.

Floud, R. 1994. "The Heights of Europeans since 1750: A New Source for European Economic History." In J. Komlos, editor. *Stature, Living Standards, and Economic Development: Essays in Anthropometric History.* Chicago: University of Chicago Press.

Fogel, R. W. 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.

Food and Agriculture Organisation (FAO). 1996. *The State of Food and Agriculture.* Rome: FAO, the United Nations.

Frenk, J. 1997. "Reformar sin deformar: La necesidad de una visión integral en la transformación del sistema de salud Mexicano." In J. Frenk, editor. *Observatorio de la salud: Necesidades, servicios y políticas.* México, D.F.: Fundación Mexicana para la Salud.

Frenk, J., J. L. Bobadilla, and J. Sepúlveda. 1989. "Health Transition in Middle-Income Countries: New Challenges for Health Care." *Health Policy and Planning.* 4 (1).

Frenk, J., J. L. Bobadilla, and C. Stern. 1994. "Elements for a Theory of the Health Transition." In L. Chen, A. Kleinman, and N. Ware, editors. *Health and Social Change in International Perspective.* Cambridge: Harvard University Press.

Frenk, J., I. Duran-Arenas, and A. Vázquez-Segovia. 1995. "Los médicos en México, 1970–1990." *Salud pública de México.* 37: 19–30.

Frenk, J., and R. Frisch. 1974. "A Method of Prediction of Age of Menarche from Height and Weight at Ages 9 through 13 Years." *Pediatrics.* 53 (3): 384–90.

Frenk, J., M. A. González-Block, and R. Lozano. 1994. *Economía y salud: Propuestas para el avance del sistema de salud en México. Informe final.* México, D.F.: Fundación Mexicana para la Salud.

Frenk, J., F. Knaul, L. A. Vázquez-Segovia, and G. Nigenda. 1999. "Trends in Medical Employment: Persistent Imbalances in Urban Mexico." *American Journal of Public Health.* 89 (7): 1054-8.

Frenk, J., R. Lozano, and J. L. Bobadilla. 1994. "La transición epidemiológica en América Latina." *Notas de Población.* 22 (60).

Frisch, R., and R. Revelle. 1970. "Height and Weight at Menarche and a Hypothesis of Critical Body Weights and Adolescent Events." *Science.* 169: 397–8.

Glewwe, P., and H. Jacoby. 1992. "Estimating the Determinants of Cognitive Achievement in Low-Income Countries: The Case of Ghana." Living Standards Measurement Study Working Paper No. 19. World Bank, Washington, D.C.

Gómez de Leon, J., and S. Parker. 1998. "Demographic Changes, Mortality Reductions, and the New Mexican Social Security Law." Paper presented at the 1998 Annual Meeting of the American Association for the Advancement of Science, Philadelphia, Pennsylvania: United States.

González, E., and S. W. Parker. 1998. "Provider Financial Risk under Alternative Payment Schemes: The Case of Mexican Social Health Insurance Policy." Mimeographed document. Instituto Mexicano de Seguor Social, México, D.F.

Graber, J., J. Brooks-Gunn, and J. Warren. 1995. "The Antecedents of Menarcheal Age: Heredity, Family Environment and Stressful Life Events." *Child Development.* 66: 346–59.

Greene, W. H. 1997. *Econometric Analysis.* Upper Saddle River, New Jersey: Prentice Hall.

Gregory, P. 1986. *The Myth of Market Failure: Employment and the Labor Market in Mexico.* Baltimore: Johns Hopkins University Press.

Griliches, Z. 1977. "Estimating the Return to Schooling: A Progress Report." *Econometrica*. 45 (1): 1–22.

Grosh, M., and P. Glewwe. 1995. "A Guide to Living Standard Measurement Study Surveys and Their Data Sets." Living Standards Measurement Survey Working Paper No. 120. World Bank, Washington, D.C. (September).

Haddad, L., E. Kennedy, and J. Sullivan. 1994. "Choice of Indicators for Food Security and Nutrition Monitoring." *Food Policy*. 19 (3): 329–43.

Hausman, J. 1978. "Specification Tests in Econometrics." *Econometrica*. 46 (6): 1251–72.

_____. 1983. "Specification and Estimation of Simultaneous Equation Models." In Z. Griliches and M. Intriligator, editors. *Handbook of Econometrics*. Amsterdam: North-Holland Publishing Company.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*. 47 (1): 153–61.

Hediger, M. L., and R. A. Stine. 1987. "Age at Menarche Based on Recall Information." *Annals of Human Biology*. 14 (2): 133–42.

Hulanicka, B., and A. Waliszko. 1991. "Deceleration of Age at Menarche in Poland." *Annals of Human Biology*. 18 (6): 507–13.

Instituto Geográfico Agustín Codazzi. 1996. *Diccionario geográfico de Colombia*. Bogotá, D.C., Colombia.

Instituto Nacional de Estadísticas y Censos (INEC). 1992. "Estimaciones de población al 30 de junio de 1992." Mimeographed document. INEC, Managua, Nicaragua (July). Table 11.

Instituto Nacional de Estadística, Geografía e Informática (INEGI). 1993. "La tercera edad en México." In *XI censo general de población y vivienda. 1990*. México, D.F.: INEGI.

_____. 1995. *Anuarios estadísticos estatales, 1995*. México, D.F.: INEGI.

Inter-American Development Bank (IDB). 1996. "Como organizar con éxito los servicios sociales." In *Progreso económico y social en América Latina. 1996*. Washington, D.C.: IDB.

_____. 1998. *Facing Up to Inequality in Latin America: Social and Economic Progress Report*. Washington, D.C.: IDB.

Jacobo, M., and J. M. Malacara. 1985. "Correlation of Menarche with Age and Various Somatometric Indexes." *Boletín médico del Hospital Infantil de México.* 42 (1): 37–41.

Jamison, D. 1993. "Disease Control Priorities in Developing Countries: An Overview." In D. Jamison, A. R. Measham, W. H. Mosley, and J. L. Bobadilla, editors. *Disease Control Priorities in Developing Countries.* Oxford: Oxford University Press, for the World Bank.

Johansson, S. R. 1991. "The Health Transition: The Cultural Inflation of Morbidity during the Decline of Mortality." *Health Transition Review.* 1: 39–68.

Johnson, R. W., and D. Neumark. 1996. "Wage Declines among Older Men." *Review of Economics and Statistics.* 78 (4): 740–8.

Knaul, F. 1999. "Linking Health, Nutrition and Wages: The Evolution of Age at Menarche and Labor Earnings among Adult Mexican Women." Working Paper Series R-355. OCE, Inter-American Development Bank, Washington, D.C. (April).

Knaul, F., S. Parker, and R. Ramírez. 1997. "El prepago por servicios médicos privados en México: Determinantes socio-económicos y cambios a través del tiempo." In J. Frenk, editor. *Observatorio de la salud: Necesidades, servicios y políticas.* México, D.F.: Fundación Mexicana para la Salud.

Komlos, J. 1989. "The Age at Menarche and Age at First Birth in an Undernourished Population." *Annals of Human Biology.* 16 (5): 559–62.

Landa Rico, V., C. Díaz de Mathman, and R. Ramos Galván. 1968. "Crecimiento y desarrollo en adolescentes femeninos: Menarquia y crecimiento." *Boletín médico del Hospital Infantil de México.* 25: 795–802.

Laska-Mierzejewska, T. 1970. "Effect of Ecological and Socio-Economic Factors on the Age at Menarche, Body Height and Weight of Rural Girls in Poland." *Human Biology.* 42 (2): 284–92.

Lee, Lung-Fei. 1981. "Simultaneous Equation Models with Discrete and Censored Dependent Variables." In C. Manski and D. McFadden, editors. *Structural Analysis of Discrete Data with Econometric Applications.* Boston: MIT Press.

______. 1982. "Some Approaches to the Correction of Selectivity Bias." *Review of Economic Studies.* 49: 355–72.

______. 1983. "Generalized Econometric Models with Selectivity." *Econometrica.* 51 (2): 507–12.

Leibenstein, H. 1957. *Economic Backwardness and Economic Growth: Studies in the Theory of Economic Development.* New York: Wiley and Sons.

Leibovich, J., L. A. Rodriguez, and O. Nupia. 1997. "El empleo en el sector rural Colombiano: ¿Qué ha pasado en los últimos años? ¿Qué se puede prever?" Documento CEDE # 97-08. Centro de Estudios de Desarrollo Económico, Universidad de los Andes, Bogotá, D.C.

Liestøl, K. 1982. "Social Conditions and Menarcheal Age: The Importance of Early Years of Life." *Annals of Human Biology.* 9 (6): 521–37.

Liestøl, K., and M. Rosenberg. 1995. "Height, Weight and Menarcheal Age of Schoolgirls in Oslo: An Update." *Annals of Human Biology.* 22 (3): 199–205.

Lindgren, G. 1976. "Height, Weight and Menarche in Swedish Urban School Children in Relation to Socio-economic and Regional Factors." *Annals of Human Biology.* 3 (6): 501–28.

López, A. 1993. "Causes of Death in Industrial and Developing Countries: Estimates for 1985–1990." In D. Jamison, A. R. Measham, W. H. Mosley, and J. L. Bobadilla, editors. *Disease Control Priorities in Developing Countries.* Oxford: Oxford University Press, for the World Bank.

Lozano, R., C. Infante, L. Schlaepfer, and J. Frenk. 1993. "Desigualdad, pobreza y salud en México." Mimeographed document. Consejo Consultivo del Programa Nacional de Solidaridad, México, D.F.

Maclure, M. 1991. "A Prospective Cohort Study of Nutrient Intake and Age at Menarche." *American Journal of Clinical Nutrition.* 54 (4): 649–56.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics.* Econometric Society Monographs, No. 3. Cambridge: Cambridge University Press.

Malcolm L. 1978. "Protein-Energy Malnutrition and Growth." In F. Falkner and J. M. Tanner, editors. *Human Growth: A Comprehensive Treatise.* Vol. 2. New York: Plenum Press.

Management Sciences for Health, Ministerio de Salud, and Inter-American Development Bank. 1996. "Estudio de la demanda de servicios de salud-Nicaragua." Mimeographed document. Managua, Nicaragua (August).

Manniche, E. 1983. "Age at Menarche: Nicolai Edvard Ravn's Data on 3,385 Women in Mid-19th Century Denmark." *Annals of Human Biology.* 10 (1): 79–82.

Manning, W. G., J.P. Newhouse, and J.G. Ware. 1982. "The Status of Health in Demand Estimation: Or Beyond Excellent, Good, Fair, Poor." In V. R. Fuchs, editor. *Economic Aspects of Health.* Chicago: University of Chicago Press.

Marshall, W. A. 1978. "Puberty." In F. Falkner and J. M. Tanner, editors. *Human Growth: A Comprehensive Treatise.* Vol. 2. New York: Plenum Press.

Martorell, R., and J. P. Habicht. 1986. "Growth in Early Childhood in Developing Countries." In F. Falkner and J. M. Tanner, editors. *Human Growth: A Comprehensive Treatise.* 2nd ed. Vol. 3. New York: Plenum Press.

Mincer, J. 1962. "On-the-job Training: Costs, Returns and Some Implications." *Journal of Political Economy.* 70 (5): 50–79.

_____. 1974. *Schooling, Experience and Earnings.* New York: Columbia University Press.

Mincer, J., and S. Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy.* 82 (2): part 2.

Ministerio de Acción Social (MAS)/United Nations Development Programme (UNDP)/United Nations Children's Fund (UNICEF). 1994a. "Mapas de pobreza." Project NIC/93/016. Report No.5. (June 30).

_____. 1994b. Estudio de la pobreza en Nicaragua. Project Nic/93/016. Preliminary report.

Ministerio de Salud. 1991. *Plan Maestro de Salud 1991–1996.* Managua, Nicaragua: Ministerio de Salud.

_____. 1996. *Estadísticas de producción de servicios 1990–1995.* Managua, Nicaragua: Ministerio de Salud.

Moisan, J., F. Meyer, and S. Gingras. 1990. "A Nested Case Control Study of the Correlates of Early Menarche." *American Journal of Epidemiology.* 132 (5): 953–61.

Montgomery, M. M., K. Burke, and E. Paredes. 1997. "Measuring Living Standards with DHS Data." Paper presented at the Annual Meetings of the Population Association of America, Washington D.C.

Moock, P., and J. Leslie. 1986. "Childhood Malnutrition and Schooling in the Terai Region of Nepal." *Journal of Development Economics*. 20: 33–52.

Moseley, W. H., J. L. Bobadilla, and D. Jamison. 1993. "The Health Transition: Implications for Health Policy in Developing Countries." In D. Jamison, A. R. Measham, W. H. Mosley, and J. L. Bobadilla, editors. *Disease Control Priorities in Developing Countries*. Oxford: Oxford University Press, for the World Bank.

Moulton, B. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*. 32: 385–97.

______. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics*. 40.

Murray, C. J., R. G. Feachem, M. A. Phillips, and C. Willis. 1992. "Adult Morbidity: Limited Data and Methodological Uncertainty." In R. G. Feachem, T. Kjellstrom, M. Over, and M. A. Phillips, editors. *The Health of Adults in the Developing World*. New York: Oxford University Press, for the World Bank.

Murrugarra, E., and M. Valdivia. 1999. "The Returns to Health for Peruvian Urban Adults: Differentials Across Genders, the Life-cycle and the Wage Distribution." Working Paper Series R-352. OCE, Inter-American Development Bank, Washington, D.C. (March).

National Academy of Medicine. 1993. "The North American Free Trade Agreement and Medical Services: Elements for Analysis and Negotiation." Mimeographed document. National Academy of Medicine, México, D.F.

Pan-American Health Organization (PAHO)/United Nations Development Programme (UNDP)/World Bank. n.a. *Health Systems Inequalities and Poverty in Latin America and the Caribbean*, Equilac/IHEP Project, Pan-American Health Organization. Washington, D.C. (forthcoming).

Parker, S. 1999. "Elderly Health and Salaries in the Labor Market." Working Paper Series R-353. OCE, Inter-American Development Bank, Washington, D.C. (January).

Phillimore, P., A. Beattie, and P. Townsend. 1994. "The Widening Gap: Inequality of Health in Northern England, 1981–1991." *British Medical Journal*. 308: 1125–8.

Phillips, M., R. Feachem, C. Murray, M. Over, and T. Kjellstrom. 1993. "Adult Health: A Legitimate Concern for Developing Countries." *American Journal of Public Health.* 83 (11): 1527–30.

Piras, C., and W. D. Savedoff. 1999. "Are We Developed Yet?" Mimeographed document. Inter-American Development Bank, Washington, D.C.

Pitt, M., and M. Rosenzweig. 1986. "Agricultural Prices, Food Consumption, and the Health and Productivity of Indonesian Farmers." In I. Singh, L. Squire, and J. Strauss, editors. *Agricultural Household Models: Extensions, Applications and Policy.* Baltimore: Johns Hopkins University Press, for the World Bank.

Pitt, M., M. R. Rosenzweig, and D. M. Gibbons. 1995. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." In Dominique van de Walle and Kimberly Nead, editors. *Public Spending and the Poor: Theory and Evidence.* Baltimore and London: World Bank.

Pitt, M., M. Rosenzweig, and M. D. N. Hassan. 1990. "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries." *American Economic Review.* 80 (5): 1139–56.

Posner, R. A. 1995. *Aging and Old Age.* Chicago: University of Chicago Press.

Pritchett, L., and L. H. Summers. 1996. "Wealthier Is Healthier." *Journal of Human Resources.* 31 (4): 841–68.

Ramos Galván, R., C. Mariscal, and A. Vinegra. 1969. *Desnutrición en el niño.* México, D.F.: Hospital Infantil de México.

Ribero, R., and C. Meza. 1997. "Earnings of Men and Women in Colombia: 1976–1995." Mimeographed document. Archivos de Macroeconomía, Departamento Nacional de Planeación, Bogotá, D.C.

Ribero, R., and J. Nuñez. 1999. "Productivity of Household Investment in Health: The Case of Colombia." Working Paper Series R-354. OCE, Inter-American Development Bank, Washington, D.C. (January).

Rosenzweig, M., 1988. "Labor Markets in Low-Income Countries." In H. Chenery and T. N. Srinivason, editors. *Handbook of Development Economics.* Vol.1. Amsterdam: North Holland Publishing Company.

Rosenzweig, M., and T. P. Schultz. 1982. "Determinants of Fertility and Child Mortality in Colombia: Interactions between Mothers' Education and Health and Family Planning Programs." Mimeographed document. Economic Growth Center, Yale University, New Haven, Connecticut.

_____. 1983. "Estimating a Household Production Function." *Journal of Political Economy*. 91 (5): 723–46.

Rosenzweig M. R., and K. Wolpin. 1986. "Evaluating the Effects of Optimally Distributed Programs." *American Economic Review*. 76 (3): 470–82.

_____. 1988. "Heterogeneity, Intrafamily Distribution and Child Health." *Journal of Human Resources*. 23 (4).

Schultz, T. P. 1982. "Lifetime Migration within Educational Strata." *Economic Development and Cultural Change*. 30 (3): 559–93.

_____. 1988. "Heterogeneous Preferences and Migration: Self-selection, Regional Prices and Programs, and the Behavior of Migrants in Colombia." In T. P. Schultz, editor, *Research in Population Economics*, Vol. 6. Greenwich, CT: JAI Press.

_____. 1993. "Investments in the Schooling and Health of Women and Men: Quantities and Return." *Journal of Human Resources*. 28 (4): 694–734.

_____, editor. 1995a. *Investment in Women's Human Capital*. Chicago: University of Chicago Press.

_____. 1995b. "Human Capital and Development." In G. H. Peters and D. D. Hedley, editors. *Agricultural Competitiveness*. Aldershot, England: Dartmouth Publishing Company.

_____. 1996. "Wage Rentals for Reproducible Human Capital: Evidence from Two West African Countries." Mimeographed document. Yale University, New Haven, Connecticut.

_____. 1997. "Assessing the Productive Benefits of Nutrition and Health: An Integrated Human Capital Approach." *Journal of Econometrics*. 77 (1): 141–58.

_____. 1999a. "Health and Schooling in Africa." *Journal of Economic Perspectives*. 13 (3): 67–88.

_____. 1999b. "Productive Benefits to Improving Health." Paper presented at American Economics Association Meetings, Boston, Massachusetts, 8 January 2000.

Schultz, T. P., and A. Tansel. 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 (2): 251–86.

Secretaría de Salud. 1994. "Visión de conjunto de la reforma del sistema nacional de salud." Mimeographed document. Subsecretaría de Planeación, México, D.F.

Sen, A. 1995. "The Political Economy of Targeting." In D. van de Walle and K. Nead, editors. *Public Spending and the Poor: Theory and Evidence.* Baltimore and London: World Bank.

______. 1998. "Mortality as an Indicator of Economic Success and Failure." *Economic Journal.* 108 (446): 1–25.

Smith, J. P. 1994. "Measuring Health and Economic Status of Older Adults in Developing Countries." *The Gerontologist.* 34 (4): 491–6.

Staiger, D., and J. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica.* 65 (3): 557–86.

Stark, O., C. S. Peckham, and C. Moynihan. 1989. "Weight and Age at Menarche." *Archives of Disease in Childhood.* 64 (3): 383–7.

Steckel, R. H. 1995. "Stature and the Standard of Living." *Journal of Economic Literature.* 33 (4): 1903–40.

Stern, S. 1989. "Measuring the Effect of Disability on Labor Force Participation." *Journal of Human Resources.* 24 (3): 361–95.

Stewart, A. L., and J. E. Ware, Jr., editors. 1992. *Measuring Functioning and Well-Being.* Durham, NC: North Carolina University Press.

Strauss, J. 1986. "Does Better Nutrition Raise Farm Productivity?" *Journal of Political Economy.* 94 (2): 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.

______. 1995. "Gender and Life-cycle Differentials in the Patterns and Determinants of Adult Health." In T. P. Schultz, editor. *Investment in Women's Human Capital.* Chicago: University of Chicago Press.

Strauss, J., and D. Thomas. 1995. "Human Resources: Empirical Modeling of Household and Family Decisions." In J. R. Behrman, T. N. Srinivasan, and H. B. Chenery, editors. *Handbook of Development Economics.* Vol. 3A. Amsterdam: North-Holland Publishing Company.

______. 1998. "Health, Nutrition and Economic Development." *Journal of Economic Literature*. 36 (2): 766–817.

Tanner, J. M. 1962. *Growth at Adolescence*. Oxford: Blackwell Scientific Publications.

______. 1994. "Introduction: Growth in Height as a Mirror of the Standard of Living." In J. Komlos, editor. *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*. Chicago: University of Chicago Press.

Thomas, D., and J. Strauss. 1997. "Health, and Wages: Evidence on Men and Women in Urban Brazil." *Journal of Econometrics*. 77 (1): 159–185.

Treloar, S. A., and N. G. Martin. 1990. "Age at Menarche as a Fitness Trait: Nonadditive Genetic Variance Detected in a Large Twin Sample." *American Journal of Human Genetics*. 47 (1): 137–48.

Trussel, J., and R. Steckel. 1978. "The Age of Slaves at Menarche and Their First Birth." *Journal of Interdisciplinary History*. 8 (3): 477–505.

Ulijaszek, S. J., E. Evans, and D. S. Miller. 1991. "Age at Menarche of European, Afro-Caribbean and Indo-Pakistani Schoolgirls Living in London." *Annals of Human Biology*. 18 (2): 167–75.

Valdivia, M., and M. Robles. 1997. "Decisiones laborales en las economías rurales del Perú." *Notas para el debate*. 14: 79–131, Grupo de Análisis para el Desarrollo (GRADE), Lima, September.

Valenzuela, C. Y., E. Nuñez, and C. Tapia. 1991. "Month at Menarche: A Reevaluation of the Seasonal Hypothesis." *Annals of Human Biology*. 18 (5): 383–93.

Victoria, C. F. Barros, S. Hutly, A. M. Texeira, and J. P. Vaughan. 1992. "Early Childhood Mortality in a Brazilian Cohort: The Roles of Birthweight and Socioeconomic Status." *Internists Journal of Epidemiology*. 21: 423.

Vijverberg, W. 1995. "Educational Investment and Returns for Women and Men in Côte D'Ivoire." In T. P. Schultz, editor. *Investment in Women's Human Capital*. Chicago: University of Chicago Press.

Weir, J., J. E. Dunn, and E. G. Jones. 1971. "Race and Age at Menarche." *Communications in Brief*. 111 (4): 594–6.

Wellens, R., R. M. Malina, and G. Beunen. 1990. "Age at Menarche in Flemish Girls: Current Status and Secular Change in the 20th Century." *Annals of Human Biology*. 17 (2): 145–52.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*. 48 (4): 817-38.

Wolfe, B., and J. Behrman. 1984. "Determinants of Women's Health Status and Health-Care Utilization in a Developing Country: A Latent Variable Approach." *Review of Economics and Statistics*. 56 (4): 703–20.

Wong, R., S. Parker, and S. de la Vega. 1997. "Household Health Expenditures in Mexico: The Influence of the Local Health Sector." Paper presented at the Annual Meeting of the Population Association of America, Washington, D.C.

World Bank. 1993. *Investment in Health, World Development Report*. Washington, D.C.: World Bank

______. 1994a. *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*. Oxford. Oxford University Press.

______. 1994b. "Country Economic Memorandum: Nicaragua." World Bank, Washington, D.C.

______. 1995. "Republic of Nicaragua: Poverty Assessment." Report No. 14038-NI. Vol. 1–2. Poverty and Human Resources Division, Policy and Research, World Bank, Washington, D.C.

______. 1999. *Global Economic Prospects and the Developing Countries, 1998/99*. Washington, D.C.: World Bank.

Wu, J. 1994. "How Severe was the Great Depression? Evidence from the Pittsburgh Region." In J. Komlos, editor. *Stature, Living Standards, and Economic Development:Essays in Anthropometric History*. Chicago: University of Chicago Press.

Wyshak, G. 1983. "Secular Changes in Age at Menarche in a Sample of U.S. Women." *Annals of Human Biology*. 10 (1): 75–7.

Wyshak, G., and R. Frisch. 1982. "Evidence for a Secular Trend in Age of Menarche." *New England Journal of Medicine*. 306 (17): 1033–5.

Yoneyama, K., H. Nagata, and Y. Sakamoto. 1988. "A Comparison of Height Growth Curves among Girls with Different Ages of Menarche." *Human Biology*. 60 (1): 33–41.

Zacharias, L., and W. Rand. 1986. "Adolescent Growth in Weight and its Relation to Menarche in Contemporary American Girls." *Annals of Human Biology*. 13 (4): 369–86.

Zurita B., G. Nigenda, and T. Ramírez. 1997. "Encuesta de satisfacción con los servicios de salud 1994." In J. Frenk, editor. *Observatorio de la salud: Necesidades, servicios y políticas*. México, D.F.: Fundación Mexicana para la Salud.

ECONOMICS/HEALTH

There is common sense in the notions that healthier people are more productive and that wealthier people can obtain things that make them healthier. Investigating the strength of these associations, ***Wealth from Health*** asks whether investments in health also affect productivity and how public policy can influence this relationship. These questions are probed through a series of Latin American case studies, using household survey data on individuals to analyze the relationships between efforts to improve health on the one hand, and the potential impact of health status on individual hourly earnings on the other. By analyzing these relationships together—health determinants and the impact of health on earnings—it becomes possible to assess the effectiveness of particular strategies for improving health status and to see the critical importance of health as a component of "human capital."

Inter-American Development Bank
1300 New York Ave., N.W.
Washington, D.C. 20577

www.iadb.org/pub

ISBN 1-886938-84-9

9 781886 938847 >