10-26-2007

Malaria, Labor Supply, and Schooling in Sub-Saharan Africa

Taiwo Abimbola
University of South Florida

Malaria, Labor Supply, and Schooling in Sub-Saharan Africa

by

Taiwo Abimbola

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Economics
College of Business Administration
University of South Florida

Co-Major Professor: Gabriel A. Picone, Ph.D.
Co-Major Professor: Kwabena Gyimah Brempong, Ph.D.
Philip K. Porter, Ph.D.
Christopher R. Thomas, Ph.D.

Date of Approval:
October 26, 2007

Keywords: Health, Human Capital Accumulation, School, Work

Dedication

This dissertation is dedicated to my father whose wisdom has been my motivation. I also dedicate this to my husband, Nicholas Michael Pallutto, my sister, Mojisola Abimbola, and my twin brother, Olatunji Abimbola.

Acknowledgements

I would like to thank my dissertation committee for their leadership and support. Especially, the co-chairs of the committee, Gabriel A. Picone and Kwabena Gyimah-Brempong who provided invaluable guidance throughout the entire process. I am very grateful for their time. Furthermore, I also want to thank the other committee members, Christopher R. Thomas and Philip K. Porter for their helpful comments and advice. I would have not been able to complete this dissertation without the mentorship of my entire committee.

Table of Contents

List of Tables

List of Figures

Malaria, Labor Supply, and Schooling in Sub-Saharan Africa

Taiwo Abimbola

Abstract

The purpose of this study is to evaluate the causal effects of malaria and poor health in general on economic outcome in Sub-Saharan Africa. This study uses panel data from the Living Standard Measurement Survey (LSMS) for Tanzania from 1991 to 2004. Three main hypotheses are tested. First, the study evaluates the effect of malaria and other chronic illnesses on labor supply using the number of hours worked per week as a measure of outcome. Second, it determines the impact of poor health on human capital accumulation by measuring the number of weekly school hours lost to illness. The third objective deals with the question of whether changes in preconditioning factors such as income levels and healthcare accessibility have improved the disease environment in Sub-Saharan Africa over time.

The study uses several identification strategies in the empirical estimation process. The first estimation strategy applies the standard Ordinary Least Squares (OLS) and Fixed Effects (FE) estimators to the schooling and labor supply models. In addition to OLS and FE, the preferred methods of estimating the causal effects of malaria on schooling and labor supply outcomes are Two Stage Least Squares (2SLS) and Limited Information Maximum Likelihood (LIML). Findings in this study suggest that malaria significantly increases school absenteeism. In particular, 2SLS and LIML estimates of

the number of school hours lost to malaria suggests that children sick with malaria are absent from school for approximately 24 hours a week. However, the results show the effect of malaria on work hours is inconclusive. Furthermore, difference in difference estimates of the disease environment show slight improvements in the disease environment resulting from changes in income levels. The study finds no statistically significant improvements in the disease environment due to increases in the number of health facilities over time.

Chapter 1

Introduction

1.1 Motivation

Recently, the debate over the influence of disease on social welfare in Africa has inspired a collection of empirical papers on the economic impact of illness on the continent. Most of these papers have focused on three major diseases: malaria, HIV/AIDS, and tuberculosis (TB) (Martine Audibert 1986; El Tahir Mohamed Nur 1993; Joseph K. Wang'Ombe and Germano M. Mwabu 1993; Martha Ainsworth, et al. 2005; Kathleen Beegle 2003; Anne Case, et al. 2002; Harsha Thirumurthy, et al. 2006). More than any region in Africa, the Sub-Saharan belt (the region of Africa below the Sahara desert) has suffered the biggest impact of malaria, HIV/AIDS, and TB. Table A1 compares the global impact of malaria, HIV/AIDS and TB to the Sub-Sahara burden of disease.

Although the region is home to only 11% of the world population, it is also home to 70% of all people infected with HIV (the Global Fund (GF)). Together, HIV and TB form a deadly force, as one-third of those infected with HIV will develop TB (GF – Social, Economic and Development Impact). On the other hand, malaria is prevalent in 105 countries, 45 of which are in the continent of Africa (GF – Facts and Figures on Malaria). Figure B1 shows a map of the estimated incidence of clinical malaria episodes by region. Sub-Saharan Africa also has the highest mortality rates from malaria. A disturbing 90% of malaria deaths occur in Sub-Saharan Africa every year (World Health Organization (WHO) - Roll Back Malaria Report (RBM) 2003).

The Sub-Sahara region of Africa is unique because of its disproportionate burden of disease and for its extremely high level of poverty; a combination that fuels further spread of disease. In a region where total health expenditure runs as low as $4 per capita, it is no surprise that poverty persists and high levels of mortality and morbidity continue to plague Sub-Saharan Africa (Table A2).

1.2 Malaria and Its Impact on Sub-Saharan Africa

Malaria is caused by four sub-species of a protozoon of the plasmodium genus namely, *falciparum*, *vivax*, *malariae*, and *ovale*. The f*alciparum* sub-species cause the greatest illness and death in Africa. Human malaria is only transmitted through mosquito bites by the females of the genus *Anopheles*. *Anopheles* mosquitoes prefer to feed on humans or animals and are more likely to transmit the malaria parasites from one person to another (Center for Disease Control (CDC)). A period of incubation takes place following the infective bite, which lasts up to 30 days. Shorter periods of incubation are observed with plasmodium falciparum compared to other sub-species.

Upon transmission, people often experience fever, chills, and flu-like illness. Severities of infection with malaria parasites range from mild symptoms to more severe disease and even death. The most severe manifestation of malaria in Sub-Saharan Africa is cerebral malaria, which is most common in children and persons without previous immunity. People without immunity to malaria or people from regions where transmission is absent (i.e. persons living in regions of high altitudes or those from countries where the disease is virtually extinct) who contract the disease are also more likely to die from malaria.

Malaria prevention is usually implemented by vector control or by preventive therapy. Vector control involves the use of indoor residual spraying of long-lasting insecticides such as Dichloro-Diphenyl-Trichloroethane (DDT) and long-lasting insecticide-treated nets (ITNs). DDT and ITNs are cost effective and efficient methods of malaria vector control. Successful malaria control efforts in the United States involved the use of DDT and anti-malarial treatment in the 1950s. By the 1970s, environmental concerns over the use of DDT led to its ban in the United States. Since then, DDT has been shown to be safe for humans and the environment and the WHO has once again recommended using DDT alongside ITNs and malaria drugs. According to the WHO, there are currently no alternatives to insecticidal vector control methods and the development of new methods will be an expensive long-term effort.

Trials of ITNs in the 1980s and 1990s showed a 20% reduction in deaths of young children (RBM). Unfortunately, malaria control by ITNs is very expensive for most households in Sub-Saharan Africa at risk of the disease. The cost of ITNs is not the only barrier to their effective use. Most people in the Sub-Sahara region are unfamiliar with ITNs and are uninformed regarding re-treatment, which entails replenishing the nets with insecticides in order to maintain their effectiveness.

Malaria is often very costly and deadly, yet it is a preventable disease. The World Malaria Report 2005 (WMR 2005) lists malaria as the leading reason for doctor visits in Sub-Saharan Africa. According to the World Bank (2003), malaria costs Africa more than

$12 billion annually. The lack of government resources for malaria prevention and treatment puts most of the financial burden of the disease on households in Sub-Saharan Africa. A study by M. Ettling, et al. (1994) of the financial burden of malaria on households estimated that over 25% of income in low-income households in Malawi is spent on malaria treatments.

In the past decade, reported cases of malaria have escalated significantly in Sub-Saharan Africa. Historical epidemiological data on reported malaria cases show an upward trend of the disease in many countries in Sub-Saharan Africa (Table A3). For instance, in Ghana the number of reported malaria cases shows an upward trend from 1993 to 2003 (Table A2). Also in Ghana, malaria accounts for 44% of outpatient visits and 22% of under-5 mortality (WMR 2005). The recent estimate on reported episodes of malaria for Tanzania is comparable to the number of cases reported over a decade ago (Table A3). Established reasons for the observed trends include malaria resistance to Chloroquine and Antifolate combination drugs and biological factors such as the human immunodeficiency virus (HIV), which may also affect response to treatment (Peter Bloland 2001).

The global pattern of malaria suggests that the disease is centered in regions with the highest levels of poverty. As noted by Jeffery Sachs and Pia Malaney (2002), "where malaria prospers the most, human societies have prospered the least" (pp. 681). How exactly does malaria limit economic prosperity in Sub-Saharan Africa? In the context of economic welfare, malaria can limit the number of healthy workdays for the adult population. For younger members of the society, malaria can impede the development of human capital thereby promoting continued poverty in economically deprived areas.

The prolonged devastation of malaria has prompted an increased international effort to fund treatment and prevention mechanisms in less developed countries plagued by the disease. For example, from 1998 to 2005 the United States Agency for International Development (USAID) increased funding for malaria from $22 million to $89 million. Even though the financial resources for fighting malaria are increasing, approximately 3,000 children in Africa continue to die from malaria each day (United Nations Children's Fund).

Most research on health and the economics of development have focused on the impact of chronic illness on economic prosperity with HIV/AIDS as the main subject of analysis. However, unlike HIV/AIDS, the quantitative effect of malaria on economic welfare of African countries has been largely overlooked. Only a few studies relate malaria to economic outcomes in Sub-Saharan Africa (Audibert 1986; Nur 1993; Wang'Ombe and Mwabu 1993; Leighton and Foster 1993). These studies focus on the effects of malaria on labor supply to agricultural production and they tend to find no relationship between malaria and economic outcomes. None of these studies empirically measure the quantitative economic impact of malaria on the accumulation of human capital.

1.3 Goals

The primary objective of this study is to evaluate the causal effects of malaria and poor health in general on labor supply and the accumulation of human capital in Sub-Saharan Africa. This study also evaluates the extent to which improvements in the socio-economic environment have altered the disease environment in the region. To arrive at the first goal, the implicit cost of poor health in Sub Sahara Africa is measured using time lost to labor supply and the loss in human capital formation as measures of outcome. In the analysis of improvements in the disease environment, the study explores how changes in income levels and increased accessibility to healthcare have influenced the incidence of malaria in the region.

Measuring the time lost in labor supply is generally favored over the income or output approach of evaluating the effect of health on labor supply (Janet Currie and Brigitte Madrian 1999). Empirical evidence regarding the effects of health on wages and hours of work in the United States suggests that health has a greater effect on hours worked than on wages. These effects are even stronger for developing countries where historical literature documents that improvements in health are linked to bigger changes in standards of living over time (John Strauss and Duncan Thomas 1998). Considering this evidence, this study will not evaluate loss in income or output due to illness; instead, it focuses on measuring the number of work hours lost due to illness.

The causal effects of malaria on human capital accumulation will be evaluated by examining child education. In this analysis, the acquisition of human capital is treated as a household choice. The household decision to invest in human capital depends on its future expectations about the returns to education. Therefore, malaria reduces the return to education, possibly reducing the incentive to educate a child in Sub-Saharan Africa.

Assessing the impact of illness on education is particularly important given the high prevalence of disease in young children in Africa. Although adults also suffer severely from malaria, infections in children are far more severe. Malaria can easily lead to seizures and coma in young children (GF). Repeated episodes of infection in children have also been linked to a reduction in social interaction and educational opportunities, thereby contributing to poor development. In addition, children who have had malaria infections affecting the brain (cerebral malaria) suffer from an increased likelihood of learning impairments and brain damage, including epilepsy (Sean Murphy and Joel Breman 2001).

To evaluate the extent to which the disease environment has changed over time, this study employs the approach suggested by P.L. Rosenfield (1984). The continuing increase in malaria rates demonstrates the importance of evaluating the changes in the disease environment in Sub-Saharan Africa. As mentioned earlier, Table A3 shows an increase in the number of reported malaria cases for several Sub-Sahara countries including Tanzania. This new trend calls for an assessment of how previous efforts to end the disease have benefited the population. To assess these benefits, the changes in income levels and changes in socio economic factors are related to the reported cases of malaria over time.

The data used in this study is derived from the Living Standards Measurement Study (LSMS) for the Kagera region of Tanzania. The panel design of the LSMS data controls for some of the limitations that hampered previous studies in the area. The data used in previous studies by Audibert (1986) and Leighton and Foster (1993) lacked sufficient information on the episodes of malaria and are based on unrealistic recollection periods[1]. Fortunately, LSMS Tanzania-Kagera contains five waves of health data capturing peak and off-peak prevalence of malaria.

[1] The recall period for Audibert (1986) is 3 to 9 months while the recall period of Leighton and Foster (1993) is as high as one year.

Previous studies on malaria and labor market outcomes (Audibert 1986; Nur 1993; Wang'Ombe and Mwabu 1993) limit their analysis to farm employment. Audibert (1986) and Wang'Ombe and Mwabu (1993) rely on an output measure of productivity. On the other hand, Nur (1993) examined the actual time lost in labor supply due to malaria. These studies all found no relationship between malaria and labor market outcomes. The output approach used in the existing literature attempts to capture reduced productivity by comparing output of healthy persons to diseased persons. This approach fails when an omitted variable bias arises. Specifically, such bias arises when a study fails to capture the likelihood that sick workers may work longer hours with lower productivity.

The lack of a significant relationship between malaria and labor supply in the existing literature is also driven by the fact that labor substitution is far more likely in farm employment (especially for self-employed farmers) than in other sectors of employment. It is reasonable to expect the incidence of malaria to vary by type of employment. Farmers in general tend to work outdoors, thus increasing the likelihood of contact with mosquitoes. Therefore, the likelihood of labor substitution in farm employment necessitates analyzing the causal link between malaria and non-farm employment where labor substitution is less likely. Unlike earlier studies, this paper will evaluate the causal effects of malaria on farm and non-farm labor supply in Sub-Saharan Africa.

In the analysis of health and labor supply, labor supply is measured as the number of hours worked per worker. The 'hours worked' approach employed here controls for the above mentioned limitations in the output and income methods. The remaining sections of the study are arranged in the following order. Chapter two details the theoretical framework supporting the analysis. Chapter three reviews the existing literature on the effects of malaria on labor supply and education. Chapter four explains the research design. Chapter five presents the main findings of the study and chapter six presents the conclusion.

Chapter 2

Theoretical Framework

The first section of this chapter reviews the theoretical background of the relationship between human capital, labor supply and health. The second section of this chapter introduces a two-period household utility maximization model that links health directly to human capital and labor supply. Mathematical solutions to the two-period household utility maximization model are presented in Appendix C.

2.1 Theoretical Framework of Health, Human Capital and Labor Supply

There have been many empirical studies in the past three decades on health and economic welfare. At first, most of the research focused on documenting the health status of households in developed economies as a part of a larger inventory of human capital components influencing labor supply and earnings. More recently, this field of research has branched into evaluating the implications of health on the economic welfare of households in developing countries, particularly in Africa. Given this recent interest in the economics of health in developing countries, a growing concern has been whether the traditional theoretical framework used in developed country literature applies to developing economies.

A well-developed theory that explains the role of health in labor productivity for developing countries, is presented in Strauss and Thomas (1998). According to Strauss and Thomas, there are several reasons why the relationship between health and labor market outcomes in developing economies should be of special interest. First is the existing tradition of theoretical models of nutrition-based efficiency wages in the development literature that established the link between nutrition of workers and productivity. A link that is highly relevant in the developing countries context where the marginal productivity of health is likely to be higher relative to developed economies. This coupled with the structure of employment in lower income economies, which often relies more on strength and endurance (i.e. good health) suggests that labor market consequences of poor health are more severe for least developed countries than for developed economies.

To describe the theoretical link between health and labor supply, Strauss and Thomas begin with the traditional single-person household utility maximization model and household production function as presented in Gary Becker (1965) and the seminal work of Michael Grossman (1972). The household production function presented in Strauss and Thomas assumes that health is increasing in all inputs except for labor supply, which decreases the stock of health. Health impacts labor supply by influencing the decision to work, choice of work and hours of work. If health affects productivity, both the market wage and the marginal rate of substitution (MRS) between consumption goods and leisure also depend on health outcomes.

A structural labor supply function conditional on health outcomes and earnings is also provided in Strauss and Thomas. Health affects labor supply by influencing offered wages with resulting income and substitution effects, and by affecting the MRS between consumption goods and leisure; a direct result of the assumption that health directly appears in the utility function. In addition to the single-person household utility model, the paper mentions applications of the described individual labor supply behaviors to household choices. However, an extensive application of the model at the household level was not presented in the paper.

The existing theoretical background on health and schooling only explains the causal link running from education to health. A well developed literature on the theory of human capital accumulation and health can be found in Becker (1964) and Grossman (1972). Other papers including Strauss and Thomas have noted that the individual or household utility function can be conditioned on schooling without providing the theoretical link between health and schooling.

The purpose of this study is not to develop new theoretical insights, but to provide a conceptual framework to guide the empirical analysis. This chapter borrows heavily from Becker (1964), Grossman (1972), Rosenfield (1984), Paul Schultz and Aysit Tansel (1993), and Strauss and Thomas (1998) to develop the theoretical framework for analyzing the causal effects of health on labor supply and the accumulation of human capital.

This chapter introduces a two-period household model for evaluating the causal link between health and labor supply and health and human capital. The two-period model captures the household decision to invest in human capital and supply labor given its health production function. In period one, the household can choose to invest in human capital through educating its children and/or by sending them into the labor market. However, the entire household is expected to supply labor in period two after all gains from investing in human capital have been realized. The labor supply model is formulated in period two when the entire household supplies labor.

2.2 A Model of Health, Education, and Labor Supply

The study begins by proposing a two-period household utility function, which assumes that the household derives indirect utility from investments in health. In this framework, utility depends on the amount of non-health goods, health goods, and labor supply per period. In period one, health is categorized by two components – overall health of the household and health of the child within the household. That is, it is assumed the household not only derives utility from its own personal health but also from the health of its children. Thus, our two-period utility function can be expressed as[2]:

$$U = U^1(C_H^1, H_{H-K}^1, H_K^1, L^1) + \beta U^2(C_H^2, H_H^2, L^2) \qquad (3.2.1)$$

Where $L^1 = L(L_{H-K}^1, L_K^1)$ and

$$L^2 = L(L_H^2)$$

L_{H-K}^1 and L_K^1 are household less child labor supply in period one and child labor supply in period one, respectively. L_H^2 is overall household labor supply in period two when all members of the household are assumed to participate in the labor market. As seen in this chapter, if the household makes the decision to invest in human capital by educating its children in period one it is rewarded in next period in form of accumulated human capital. In period two, when all returns from the investment in human capital are realized, the entire household supplies the amount of labor denoted by L_H^2.

[2] Appendix C provides a definition of variables in the theoretical framework.

The household's utility in period one depends on the household consumption of non-health goods, C_H^1. H_{H-K}^1 is the household health less the health of the child in period one, H_K^1 is the health of the child in period one, and L^1 is labor supply in period one. Notice that utility in period two only depends on the consumption of health and non-health goods and labor supply for the entire household denoted by C_H^2, H_H^2, and L_H^2, since all gains from investing in health must be realized by the second period. The utility function differs from traditional two-period utility formulations only by the added health components. It is expected that C_H^1, H_{H-K}^1, H_K^1, C_H^2, and H_H^2 will have a positive impact on utility, while L^1 and L^2 will negatively impact utility.

Unlike other consumption goods C, health as described in equation (3.2.1) cannot be purchased in the market and instead has to be produced. In addition, the individual also derives indirect utility from good health. A person who is in good health spends less time devoted to illness, which translates into more time to pursue activities that enhance his/her utility. With this in mind, the process through which the household produces health for itself, H_{H-K}^1, for a child, H_K^1, and as a whole, H_H^2, is represented by[3]:

$$H_{H-K}^1 = H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1; A, D, \mu_{H-K}^1, e_{H-K}^1) \tag{3.2.2}$$

$$H_K^1 = H_K^1(HI_K^1, L_K^1; A, D, \mu_K^1, e_K^1) \tag{3.2.3}$$

$$H_H^2 = H_H^2(HI_H^2, L_H^2; H_{H-K}^1, H_K^1, A, D, \mu_H^2, e_H^2) \tag{3.2.4}$$

Here, health given by (3.2.2), (3.2.3), and, (3.2.4), is produced using a vector of health inputs, HI, and labor supply, L, which are controlled by the individual and other

[3] The household consumption of non-health goods can also be defined in terms of non-health consumption inputs, labor supply, and prices: $C_H = C_H(CI_H, L_H, P; A, D, \mu_{H,} e_H)$.

uncontrollable factors such as previous health history. Health is increasing in inputs except labor supply, which consumes energy and in this way taxes health. Although, schooling does not directly enter the utility function specified in (3.2.1) it is an input into the health production function and contained in HI. The levels of health produced in (3.2.2), (3.2.3), and, (3.2.4), are also likely to vary with socio-demographic characteristics, A, such as age and gender and the disease environment, D (Figure B2).

There are two sources of unobserved heterogeneity that influence the health production formulation in (3.2.2), (3.2.3), and, (3.2.4). The first deals with factors that are known to the individual but unobserved by the researcher such as the inherent healthiness of the individual, μ. The other unobservable is that which is unknown to the individual and the researcher, e, which includes measurement error. The issue of unobserved heterogeneity is discussed in detail in chapter four.

Now, suppose that the household can either make investment in the education of the child (S_K^1) in period one, and reap the fruit of this investment in period two in the form of human capital ($\tilde{S}_K$) or the household can choose not to make this investment and send the child into the labor market (part-time or full-time) in period one. The level of accumulated human capital in period two can be defined as:

$$\tilde{S}_K = \tilde{S}_K(S_K^1, H_K^1; A, SI, \mu_K, e_{S_K}) \tag{3.2.5}$$

The level of education attained in period two is a function of previous years of schooling, S_K^1, health status in period one, H_K^1, and uncontrollable factors such as school infrastructure and the quality of teachers, SI[4]. Schooling in period one (S_K^1) is expected to positively influence the level of human capital in the future. Health status in the current period can either increase or decrease the value of $\tilde{S}_K$ in the future. If the household in (3.2.1) earns labor income, w, and owns assets or non-labor income V, utility in (3.2.1) is bounded by the resource constraint:

$$P_c(C_H^1 + C_H^2) + P_{HI}(HI_{H-K}^1 + HI_K^1 + HI_H^2) + P_s S_K^1 = w_{H-K}(L_{H-K}^1) + V + w_K^1(L_K^1) + w_H^2(L_H^2, \tilde{S}_K) \qquad (3.2.6)$$

Equation (3.2.6) states that expenditures on consumption goods, health inputs, and education cannot exceed total income. The right-hand side of this two-period resource constraint, is a result of the earlier assumption that the household can choose to fully invest in education, in which case the child would supply zero labor and $w_K^1(L_K^1) = 0$ or decide to send the child into the labor market in period one part-time or full-time ($w_K^1(L_K^1) \neq 0$). If the household chooses to educated the child in period one and succeeds, then w_H^2 is expected to be higher since w_H^2 is a function of the human capital acquired, $\tilde{S}_K$.

Unlike the traditional labor supply models for developed countries, it is assumed that the household is better able to earn labor income when it is healthy. Therefore, the earnings of every member of the household are a function of health (H_{H-K}^1, H_K^1, and H_H^2). With that in mind, the earnings functions for each member can be written as:

$$w_{H-K}^1 = w_{H-K}^1(H_{H-K}, L_{H-K}; A, E, IN, \alpha, e_{w_{H-K}}) \qquad (3.2.7)$$

[4] Note that (3.2.5) can also depend on the health status of the parent(s) within the household, H_{H-K}^1 .

$$w_K^1 = w_K^1(H_k^1, L_K^1, S_K^1; A, E, IN, \alpha, e_{w_K}^1) \qquad (3.2.8)$$

$$w_H^2 = w_H^2(H_H^2, L_H^2, \tilde{S}_K; A, E, IN, \alpha, e_{w_K}^2) \qquad (3.2.9)$$

The earnings function is also influenced by uncontrollable factors such as local community infrastructure, IN, socio-demographic characteristics, A, and education, E. Notice that earnings in period two also depend on the gains from investing in human capital in period one, $\tilde{S}_K$. The term $\tilde{S}_K$ is expected to positively impact overall earnings for the household given that it chose to invest in education in the previous period. Earnings will also be affected by unobservable factors, α, such as ability, school quality, and random fluctuations in wages and measurement error captured by e_w.

The full extent of health on utility can be derived by substituting (3.2.2), (3.2.3), and (3.2.4) into the utility function in (3.2.1). The utility function in (3.2.1) can then be rewritten as:

$$\begin{aligned} U = {} & U^1(C_H^1, H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1), H_K^1(HI_K^1, L_K^1), L^1; A, E, \xi) \\ & + \beta U^2(C_H^2, H_H^2(HI_H^2, L_H^2), L_H^2; A, E, \xi) \end{aligned} \qquad (3.2.1')$$

The household can now choose C_H^1, HI_{H-K}^1, HI_K^1, L_{H-K}^1, L_K^1, S_K^1, C_H^2, HI_H^2, and L_H^2 subject to the resource constraint in (3.2.6) (see Appendix C).

2.2.1 Utility Maximization: The Health and Human Capital Accumulation Model

To determine the optimal investment in human capital for the household represented in (3.2.1′), the first order necessary condition (FONC) is solved for S_K^1. This yields the demand equation for education in period one.

$$S_K^* = S(H_K^1; P_S, A, SI, \mu_K, e_{S_K}) \qquad (3.3.1)$$

The FONC for S_K^1 indicates that the household must weigh the price of schooling in the current period against the returns to schooling in the next period (i.e. the opportunity cost of schooling) in its decision to invest in human capital. Assuming that the second order sufficient condition (SOSC) for S_K^1 is satisfied, the demand curve for schooling in period one is illustrated below, assuming a constant price for education in that period.

Figure C1 illustrates that the household maximizes utility in period one by investing in S_K^* amount of human capital. However, the optimal investment is determined by the child's health in period one, and other factors, which are out of the control of the household (as indicated by (3.3.1)). Good health/poor health in period one is expected to increase/decrease schooling (i.e. through the number of healthy school days and academic performance) and shift the demand curve rightwards/leftwards.

Notice that health as a choice variable in the demand for schooling poses an estimation problem because H_K^1 is endogenous in the human capital accumulation model. H_K^1 is determined by immeasurable factors such as the inherent healthiness of the individual therefore; estimating (3.3.1) with ordinary least squares (OLS) will produce biased estimates of the causal effect of health on schooling. The likelihood of endogeneity of H_K^1 as well as the possibility of measurement error using instrumental variables can be mitigated. This can be done by using equation (3.2.3) as the first stage regression in our econometric analysis of the causal effects of health on schooling, where P_{HI} and D in (3.2.3) are valid instruments for H_K^1. A detailed description of these instruments is presented in chapter four and five.

2.2.2 Utility Maximization: The Health and Labor Supply Model

As mentioned earlier, the causal effect of health on labor supply is to be determined in period two when the entire household supplies labor. In period two, all gains from investments in human capital are realized and the household supplies an amount of labor equal to L_H^2. In this period, the study evaluates the household's labor supply decision and solves for the optimal amount of labor supply. Two different cases are considered in solving for the optimal amount of L_H^2. In the first case, health is treated as an endogenous variable while in the second case health is considered to be exogenous.

2.2.2.1 Case 1: Health is Endogenous

A maximization of (3.2.1′) with respect to C, HI, and L subject to (3.2.6) and (3.2.9) will produce the following reduced form equations for C, HI, and L (Appendix C).

$$C = C(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.1}$$

$$HI = HI(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.2}$$

$$L_H^* = L(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.3}$$

Although, (3.4.1) and (3.4.2) are informative, the problem with treating health as a choice variable is that H_H^2 disappears from the reduced form equation for L in (3.4.3). The reduced form equation for labor supply, L in case one is only an assessment of the total effect of prices and the disease environment on labor supply, which is not very helpful in understanding the causal effect of health on labor supply. To determine how health affects labor supply, a structural labor supply function that is conditional on health is needed. To derive this important link between health and labor supply, consider the following case.

2.2.2.2 Case 2: Health is Exogenous

In this case, the individual's utility maximization problem remains the same except that H_H^2 is now a conditioning variable. Case two yields the structural equation for labor supply listed below which now depends on health, H_H^2 (Appendix C).

$$L_H^* = L(H_H^2; P_C, w_H^2, V, A, E, IN, \mu, \alpha) \tag{3.4.4}$$

Unlike (3.4.3), the structural equation for labor supply (3.4.4) provides the opportunity to separately identify the causal effects of health on labor supply. First, once the model conditions on H_H^2, the prices of health care inputs, P_{HI} and the disease environment, D, do not affect labor supply. Second, equation (3.2.4) can be used as the first stage regression in our econometric analysis of labor supply, where P_{HI} and D are valid instruments for health (H_H^2). A detailed description of these instruments is presented in chapter four, section 4.4.2.

Chapter 3

Review of Relevant Literature

Many empirical studies of the effects of health on economic welfare have been published over the past 3 decades. Initially, most research in this field sought to document health status of households in developed economies as a part of a larger inventory of human capital components influencing wages, labor supply and earnings. This literature has branched in recent times into evaluating the implications of health on the economic welfare of households in developing countries, particularly in Africa.

The following review documents evidence on health and economic welfare based on developed and developing country experience. The first section analyzes relevant literature on the implications of disease on labor supply and human capital accumulation in Africa. The last section of this review is a brief assessment of the empirical evidence of health on economic welfare in the developed country context.

3.1 Malaria, Labor Supply, and Schooling in Sub-Saharan Africa

This section reviews the existing literature on the role of malaria in the supply of labor and child education in Sub-Saharan Africa. Very few empirical studies have analyzed the role of malaria in labor supply and human capital accumulation in Africa. Studies by Audibert (1986) and Wang'Ombe and Mwabu (1993) are the only known empirical studies on malaria and labor supply in Sub-Saharan Africa. Another study by

Nur (1993) also analyzes the role of malaria in the supply of labor in Sub-Saharan Africa but only through a descriptive analysis of the data.

Compared to the literature on the role of malaria in labor supply in Sub-Saharan Africa, less work has been done on malaria and schooling in the region. A study by Leighton and Foster (1993) attempted to evaluate the economic impact of malaria on child schooling in Sub-Saharan Africa using a case study method. However, there is no real empirical study evaluating the causal effects of malaria on child schooling in Sub-Saharan Africa.

3.1.1 Malaria and Agricultural Labor Supply in Sub-Saharan Africa

Most studies of health and labor supply in Sub-Saharan Africa have focused predominantly on agricultural employment. Because most farmers (especially self-employed farmers) enjoy the advantage of labor substitution, the true cost of ill health on labor supply is often minimized. Labor substitution[5] minimizes economic loss from illness by insuring workers against future losses in income, which arises from inability to work due to illness. Therefore, when analyzing the impact of disease on the agricultural sector the observed cost of ill health may be lower than the true cost. More importantly, the prevalence of diseases such as malaria, HIV/AIDS, and tuberculosis in Africa makes analyzing the impact of health on the labor supply patterns of working adults equally important for all sectors of employment, not just farming.

Studies evaluating the link between health and labor supply have focused on patterns of agricultural labor supply in the event of a prime-age adult illness and/or death.

[5] Labor substitution in this context refers to the process whereby the hours lost in labor supply are compensated for by family members

Studies by Audibert (1986), Nur (1993), and Wang'Ombe and Mwabu (1993) all address the issue of malaria and labor supply in terms of farm occupation and generate similar results.

Audibert (1986) was the first major longitudinal study to measure the effect of malaria on agricultural productivity in Sub-Saharan Africa. Audibert estimated the impact of a variety of illnesses, including malaria, on agricultural non-wage peasant production using data on rice farmers in Cameroon. The study measured loss in productivity using a single production function with factors such as land properties and quality of labor as exogenous regressors affecting the production of rice. In contrast to Audibert, which uses an output approach, the present study evaluates the marginal impact of health status on labor using the number of hours of labor supplied as a measure of outcome. In addition, our goal is not to measure the impact of health status on the marginal loss in output but instead to measure the impact of health status on the quantity of labor supplied.

Audibert's study found malaria to be insignificant in explaining variations in rice output. Nur (1993) studied the effects of malaria on labor inputs into agriculture by examining the extent of family labor substitution in the event of a male adult being incapacitated by malaria. In addition to family labor substitution, he also examined the traditional system of mutual aid in which, other farmers provided assistance on a reciprocal basis when an adult male was ill. Nur's sample was based on 250 randomly selected tenant farmers in Gezira, Sudan.

Like Audibert, Nur's main findings on the number of hours of labor supply lost due to illness were counter-intuitive. The results indicate that farmers in Sudan were more likely to be incapacitated by malaria in the dry season (low prevalence season), but less likely to be incapacitated in the peak prevalence season. However, the study stated that high prevalence detected in the sample coincides with the peak season (the land preparation and harvest seasons). The results in the study revealed that the sum of reported hours lost in agriculture due to malaria in the land preparation and harvest seasons exceed those lost in the low prevalence months by almost 2200 hours.

In terms of family labor substitution and mutual aid, Nur determined that other family members compensated for all the hours lost in agricultural labor supply due to malaria. The family members in the sample tend to contribute more than the actual number of hours lost. This is possibly because the time contributed by labor substitutes was not as productive as each hour lost. According to the study, women in the family are far more likely to be labor substitutes than children and non-family labor substitutes (mutual aids).

Unlike Audibert, Nur failed to control for seasonal variation in labor use. Since the peak malaria season coincides with the harvest season, it is reasonable to expect that these months generate fewer hours lost due to illness than otherwise reported in the months of low prevalence (dry season). One reason is because labor substitution is less likely in the dry season than in the rainy or harvest season. Also, the opportunity cost of being sick is higher in the harvest season than it is in non-harvest seasons. These factors are not accounted for by Nur's study since the results are drawn from a simple review of frequency tables that are limited in interpretability and lend no econometric legitimacy to the study.

Wang'Ombe and Mwabu (1993) also examined the direct effects of malaria on labor productivity. Wang'Ombe and Mwabu use a cross-section of 302 households in Western Kenya. The study estimated a production function for cassava farming to evaluate the extent that malaria affects productivity and household income, similar to equation (3.4.4). Because the production equation estimated in the study only controlled for family size and the number of malaria cases per household as exogenous regressors, the study found no statistically significant direct effects of malaria on productivity. The authors cite possible model misspecification for the lack of significant results.

The studies reviewed above all fail to control for factors that could have prolonged the duration of sickness and are not directly related to malaria. For instance, the pre-existence of other chronic conditions such as HIV/AIDS and Sickle Cell Anemia was not accounted for. Failure to control for the existence of pre-existing conditions can create endogeneity in the estimation process.

3.1.2 Malaria and Schooling in Africa

In the context of health and child schooling in Sub-Saharan Africa, much effort has been geared towards evaluating the impact of chronic illness and adult mortality and morbidity on schooling. One exception to the existing literature is the study by Leighton and Foster (1993) that attempted to measure the economic impact of malaria on schoolchildren in Kenya and Nigeria. This study was based solely on focus group interviews and a spreadsheet method of estimation, which failed to control for micro-level factors that influence the decision to attend school. The study determined that between 4% and 40% percent of all school absences annually can be attributed to malaria in Kenya and Nigeria, respectively. The findings in the study are questionable for several reasons. First, the study is based on an unusually long recall period of one year. Second, it used a very small focus group sample. Third, as mentioned earlier it is based on a questionable estimation strategy.

3.2. Health and Labor Supply and Schooling in Sub-Saharan Africa

This section reviews the existing literature on illness and labor supply and schooling in Sub-Saharan Africa. Unlike the published literature on malaria in the region, more studies have been conducted on the effects of chronic illness in Sub-Saharan Africa. These studies have mostly focused on the effect of HIV/AIDS on the Sub-Sahara population. This part of the review discusses studies on health and economic outcomes in Sub-Saharan Africa.

3.2.1 HIV/AIDS and Labor Supply in Sub-Saharan Africa

Studies in the area of health and the labor supply in Sub-Saharan Africa have focused on the impact of adult morbidity and/or death on household decision to work. Beegle (2003) and Thirumurthy, et al. (2006) studied the impact of HIV/AIDS on labor supply in Tanzania and Kenya, respectively. Beegle evaluated the impact of the HIV/AIDS epidemic on Sub-Saharan household’s farm labor supply before and after the death of a prime-age adult. This study used data from the World Bank and the University of Dar Es Salaam for the Kagera region of Tanzania and studied over 800 households between 1991 and 1994. These data are now published as part of the Living Standard Measurement Survey (LSMS).

Like other studies, Beegle ignored seasonal variation in agricultural labor use and found insignificant changes in labor supply for individuals in households that experienced a prime-age adult death from HIV/AIDS. Another study by Thirumurthy, et al. analyzed how antiretroviral therapy (ARV) influences the labor supply of treated patients and their family members. Results from Thirumurthy, et al. indicate that individuals just beginning ARV treatment are five times more likely to increase participation and four times more likely to increase hours worked than those who have not undergone treatment. Like Beegle, Thirumurthy, et al. findings suggest that neither girls nor older boys experienced any significant spillover effects in terms of increased labor participation rates when another household member suffered from or died due to HIV/AIDS.

3.2.2 HIV/AIDS and Schooling in Sub-Saharan Africa

Martha Ainsworth, et al. (2005) and Anne Case, et al. (2002) analyzed the impact of orphan status and the death of a prime-aged adult in the household on school hours for children in Sub-Saharan Africa. Ainsworth, et al. estimated the number of school hours lost per week for orphaned girls to be between five and 13 hours conditional on school attendance. In addition, the study found the impact of adult mortality to be greater on school attendance for younger children than for older children. In particular, children in poor households with recent adult death had a 10 percentage point lower attendance rate than poor children in households without an adult death. Case, et al. also examined the effect of orphan status on school enrollment and found results similar to Ainsworth, et al. for school enrollment rates of orphans in poorer households.

3.2.3 Hookworm and Schooling in Sub-Saharan Africa

Edward Miguel and Michael Kremer (2004) conducted a school-based mass randomized treatment experiment with deworming drugs in Kenya. The purpose of the field experiment was to identify the impacts of deworming on education and health while controlling for the likelihood of treatment externalities. In order to determine the extent of externalities resulting from treatment, seventy-five primary schools participating in the experiment were phased into deworming treatment in a randomized order.

Miguel and Kremer found significant gains in school attendance due to treatment for both treatment and control groups. Findings in the study suggest that the program reduced overall school absenteeism by seven percentage points, a one-quarter reduction in total school absenteeism. In terms of externalities, the study found that deworming creates positive externalities both within and across schools. In particular, implementing treatment to a fraction of pupils in a school led to a 6.2 percentage points gain in school participation for children in the control group at the same school.

3.3 A Brief Review of Developed Country literature

This section provides a brief review of empirical evidence on health, labor supply and human capital accumulation in the United States.

3.3.1 Health and Labor Supply: A Brief Review of Developed Country literature

Early works by Michael Grossman and L. Benham (1974), Ann Bartel and Paul Taubman (1979), and H.S. Luft (1975) emphasize the interrelationship between health and labor force participation. They all conclude that poor health in preceding periods reduce labor supply in following periods. Grossman and Benham in particular, emphasize the importance of a model that views wage rates, hours of work and health as a set of interrelated household decisions.

In the Grossman and Benham framework, work-time and wages are recognized as interdependent. Using this framework, Robert Haveman, et al. (1994) adopted a 3-equation simultaneous equation model to capture the interrelationship between labor force participation and health. The study utilized annual data on white males with a history of significant labor force attachment from the Michigan Panel of Income Dynamics (PSID). The findings by Haveman, et al. support that of Grossman and Benham in concluding that prior health limitations have a significant negative effect on work time and wages.

Another study examining the impact of health status on work hours is Susan Ettner, et al. (1997). This study examined the impact of psychiatric disorders on employment and conditional work hours and income using the National Comorbidity Survey (NCS). Ettner, et al. found small and sometimes insignificant effects of health on work hours depending on whether two-stage instrumental variables (IV) estimation or ordinary least squares (OLS) was used. The slight reduction in conditional work hours was only observed for men.

Ronald Kessler and Richard Frank (1997) also examined the relationship between psychiatric disorders and work impairment by using the same data as Ettner, et al. Kessler and Frank examined eight job conditions using cluster analysis. Unlike Ettner, et al., Kessler and Frank found significant reduction in workdays. This result also varies significantly depending on the constellation of disorders and the occupation in which the worker is employed. Other works on the economic effects of poor health include Thomas Chirikos and Gilbert Nestel (1985), Jean Mitchell and Richard Burkhauser (1990) and Matthew Kahn (1998).

The study by Chirikos and Nestel was based on a tobit regression of retrospective history of self-reported health appraisals on hours of work and wages over a ten-year period. The authors found enough evidence to support the significance of prior health status on labor hours and wages, but determined the strongest effect to be for hours worked, not wages. Mitchell and Burkhauser estimated the extent to which arthritis limits the ability of workers to fully function in the workplace. The study is also based on a simultaneous tobit model similar to Chirikos and Nestel. Mitchell and Burkhauser found that a history of ill health (in this case, Arthritis) has a greater effect on hours than on wages.

Kahn (1998) investigated the effect of diabetic status on health on a more general scale than Chirikos and Nestel and Mitchell and Burkhauser. The findings of this study suggest that the effects of diabetes duration on employment rates have lessened over time. The study determined that even though people with diabetes still earn less than their non-diabetic counterparts, diabetic income and labor force participation rates are far higher than in earlier years. Overall, improvements in labor market participation for diabetics can be attributed to changes in the Social Security Disability Insurance System and technological advances that improve the quality of life for diabetics.

3.3.2 Health and Schooling: A Brief Review of Developed Country literature

Barbara Shapiro, et al. (1995) documents the effects of chronic illness on school attendance of children in the United States. The study collected data on children and adolescents with sickle cell disease in order to gain information on the natural history of pain and its impact on school attendance and sleep. The data was based on a home diary experiment from a self-report system used for research on sleep and circadian rhythms. Shapiro, et al. found that sickle cell patients reported missing 41% of school days on which they reported pain versus an average of 9% on days without pain. On average, 2.7 consecutive school days over a period of 10 months were associated with clinic visits or other medical problems related to sickle cell disease.

Chapter 4

Research Design

This chapter describes the data and methods used in this study. The first section explains the objectives and hypotheses underlying this dissertation. The second section details the source of data used in the analysis. The third section describes the variables used in the estimation of the models presented in the study. The fourth and fifth sections focus on the methodologies and findings of the study.

4.1 Objectives and Hypotheses

The objective of the study is to evaluate the causal effects of malaria on labor supply and schooling outcomes. To arrive at this goal, the results of the two-period household utility model developed in chapter two are applied to a sample of household in the Sub-Saharan Africa region. The first goal is to estimate the equations for the demand for education (3.3.1) and labor supply (3.4.4).

Assuming a linear specification for both schooling and labor supply, equations (3.3.1) and (3.4.4) can be written as:

$$S = \beta_0 + \beta_1 H + \beta_2 P_S + B_3 A + B_4 SI + \mu + e \qquad (4.1)$$

$$L = \beta_0 + \beta_1 H + \beta_2 P_C + \beta_3 V + \beta_4 w + \beta_5 A + \beta_6 E + \beta_8 IN + \mu + \alpha \qquad (4.2)$$

In equation (4.1), the demand for schooling (S) is conditioned on health status (H), price of schooling (P_S), socio-demographic characteristics (A), and school infrastructure (SI). Health status, which is characterized in this study as the incident of illness is defined as the presence of malaria. Holding constant all other factors that may influence the demand for human capital, malaria is expected to decrease the demand for schooling; therefore, a negative sign is expected on β_1. Improvements in school amenities and the quality of teachers denoted by SI are other factors influencing schooling outcomes in equation (4.1). These variables are expected to positively impact the demand for schooling. On the other hand, increases in the level of price indicators such as the price of education, P_S, will decrease the demand for school.

In equation (4.2), labor supply is conditioned on health (H), price of consumption goods (P_C), wages, non-labor income (V), socio-demographic characteristics (A), education (E), and community infrastructure (IN). Holding all other factors constant in equation (4.2), the partial effect of health on labor supply (β_1) is expected to yield a negative sign for those who are sick with malaria. On the contrary, increases in the price of non-health consumption goods, wages, the level of education and community infrastructure are expected to increase the labor supply in equation (4.2). Furthermore, an increase in non-labor income (V), is more likely to reduce labor supply. As indicated in chapter two, current health status poses an estimation problem in the labor supply and schooling models. Current health status (H) in both models is determined by immeasurable factors such as the inherent healthiness of the individual.

The second goal of this study is to determine the extent to which improvements in preconditioning factors of the socio-economic and physical environment have impacted the incidence of malaria in Sub-Saharan Africa. As mentioned earlier in the chapter two, the framework suggested by Rosenfield (1984) highlights the relevance of socio economic factors that act as baseline inputs into the health production function (Figure B2). According to Rosenfield (1984), these factors predetermine health outcomes in tropical environments. In this study, Rosenfield's framework is used to evaluate the effect of changes in preconditioning factors of the socio-economic environment on health status over time. The goal is to determine how the disease environment has been altered by changes in baseline health production variables such as income and health care availability.

The unit of analysis for both the labor supply and schooling models is the household. In the analysis of the disease environment, the unit of analysis is at the household and community levels. Income measures for the analysis of the disease environment are aggregated at the household level while data on health facilities are derived at the community level.

4.2 Description of Data

The data used in this study are part of the Living Standard Measurement Surveys (LSMS). The World Bank began the surveys in the 1980s for the purpose of developing new methods of monitoring levels of living, identifying the effects of government policies, and advancing communications between those who collect and use data as well as policymakers around the world. To date, the LSMS has been conducted for the following five Sub-Saharan Africa countries - Cote d'Ivoire, Ghana, Malawi, South Africa and Tanzania.

The Cote d'Ivoire LSMS is the oldest of the five living standard surveys, which began in 1985 and ended in 1988. The living standard surveys for Ghana, Malawi, South Africa, and Tanzania are more current. However, LSMS for Malawi and South Africa are each limited to one wave of information. Ghana and Tanzania contain the most recent and detailed panel information of the five countries surveyed in Sub-Saharan Africa. LSMS Ghana however, lacks detailed information on household health compared to LSMS Tanzania.

4.2.1 The Kagera Region of Tanzania

The analyses in this study are based on data from the Tanzania Living Standard Measurement Survey. The dataset is arguably the most detailed of all five living standard surveys for measuring the effect of health status of the Sub-Saharan Africa population. The living standard survey for Tanzania was conducted in the Kagera region of the country. Kagera is located on the western shore of Lake Victoria adjacent to Uganda, Rwanda, and Burundi. Kagera is the 15th largest region in Tanzania and its regional capital Bukoba Town is about 1,500 kilometers from the country's capital, Dar Es Salaam. The region covers a total of 40,838 square kilometers and lies at 3,750 feet above sea level[6].

[6] Location information is derived from the Tanzania Chamber of Commerce Industry and Agriculture website at www.kagera.org.

Figure 2: Map of Kagera Tanzania

Source: Tanzania Chamber of Commerce Industry and Agriculture

The Kagera region comprises of five districts namely, Bukoba (the regional capital), Muleba, Karagwe, Biharamulo and Ngara. The entire region has a population of approximately 2 million people (Table 1). The most populous region Bukoba, has a population of over 470,000. Total population for the Kagera region is expected to increase by 400,000 for the year 2007.

Table 1: Kagera Population

Kagera Population			
District	Population	No. of Households	Average Household Size
Bukoba	476,351	64,510	4.3
Karagwe	425,476	89,047	4.8
Muleba	386,328	79,107	4.9
Biharamulo	410,794	67,131	6.1
Ngara	334,939	49,082	6.8
Total Population Kagera District	2,033,888	394,128	5.2

Projected population for 2007: 2,417,000

4.2.2 Current State of Malaria and Other Illnesses in Kagera

Although malaria is generally prevalent throughout Tanzania, it is a big public health concern in Kagera where malaria is the leading cause of death[7]. On July 10 2006, the Tanzania Red Cross National Society (TRCNS) reported a rise in the number of reported malaria cases in Kagera. The rise in reported cases had resulted in a drastic increase in malaria related mortality in the region. The two districts most affected by the outbreak were Karagwe and Muleba. From January to May 2006, the number of deaths among children under the age of five rose to 3,944 in Karagwe and 3,542 in Muleba from about 300 deaths in January of 2006.

Coupled with the threat of malaria in the region is the prevalence of HIV/AIDS in Tanzania. In 1983, the first reported cases of HIV/AIDS in Tanzania were from the Kagera region. The first cases of HIV/AIDS in the area were reportedly imported from neighboring countries as a result of Kagera having the largest common border with other countries in east Africa. Like most regions in Africa, there are too few hospitals in Kagera to sustain the extent of disease prevalence in the region. In 2006, the TRCNS estimated that there are 13 hospitals, 13 clinics, and 202 dispensaries providing health care serving a population of over 2 million in the Kagera region.

The Living Standard Measurement Survey (LSMS) of the Tanzanian region of Kagera began in 1991 with the goal of measuring the impact of adult mortality (predominantly due to AIDS) on households and evaluating the effectiveness of policies geared toward preventing the disease in this Sub-Sahara section. The desire for a LSMS

[7] According to the Tanzania Chamber of Commerce Industry and Agriculture malaria is the leading cause of death in Kagera, Tanzania.

style survey in Kagera was prompted by the disturbing rate of HIV infection and AIDS death of the adult population in Kagera.

4.2.3 LSMS Tanzania - Kagera

The LSMS in Tanzania-Kagera now consists of five waves of household and community level data from 1991-2004. Over 800 households were surveyed in the first four rounds (1991-94). The final round (2004) consists of over 2700 households from the original baseline, which were re-contacted 10 to 14 years later. The sections of LSMS Tanzania-Kagera contain responses to questions on a vast array of topics. Of relevance to this study, are questions on current illness, school and work hours, and other socio-demographic information.

4.2.4 Attrition in LSMS Tanzania-Kagera

The Tanzania-Kagera LSMS had a low attrition rate of 10% from wave one to four. The main reason for attrition in the LSMS of Tanzania-Kagera is death in the household, which led to the relocation of the household. However, attrition in the fifth wave is a bigger issue since wave five of the survey was conducted a decade later. Section 4.6 presents the estimation problems posed by attrition in surveys such as the LSMS Tanzania – Kagera.

4.2.5 Sample Construction: Schooling

In Kagera, it is expected that children be enrolled in school at the age of seven (Ainsworth, et al. 2005). School enrollment in this study is defined as students who reported enrollment in formal schooling in all 5 waves of the survey. Children who were being home schooled as well as those only attending Koranic schools were excluded from the sample. The study also focuses on primary education in Kagera as most children interviewed in the survey fall within this category.

Primary schools in the Kagera survey offer a maximum of seven grade levels. The analysis in this study focuses on children aged seven to 20 who are enrolled in primary school at the time of survey. The mean age observed for the Kagera school sample is 12 years. Chapter five presents additional characteristics of the Kagera school sample.

4.2.6 Sample Construction: Labor Supply

A sample of adults between the age of 18 and 65 was chosen as the likely work force for LSMS Tanzania - Kagera. Majority of the Kagera work force reported working as a farmer from wave one to wave five. The second largest work category in Kagera at the time of the survey was those working for someone other than himself or herself in form of formal employment (wage employment). Self-employed individuals are the least of the Kagera work force.

4.2.7 Definition: Prevalence of Illness

Estimates of malaria prevalence used in this study are based on self-reported measures. John Bound (1991) argued that self-reported measures of health may be more reliable than other objective measures of health. It is expected that a self-reported measure of health will be appropriate for identifying malaria in Africa given the commonality of the illness. In addition, the weakness of health systems in Sub-Saharan Africa may contribute self-treatment versus formal care for common ailments. Self-treatment for malaria is common in Sub-Saharan Africa where most people live on less than $1 a day and are unable to afford physician care (S.C. McCombie 1996). Studies that are more recent however, indicate a huge increase in formal treatment seeking in private health clinics and community health centers (Wakgari Deressa 2007).

The analysis to follow utilizes both doctor diagnosed and self-diagnosed self-reported episodes of malaria. A vast majority of reported malaria cases in the LSMS for the Kagera region are self-diagnosed. Doctor diagnosed episodes only constitute a small fraction of the total (Table 2a). Self-diagnosed episodes of malaria in the Kagera sample, was lowest in wave four of the survey. Wave five of the survey had the highest number of reported episodes (doctor and self-diagnosed)[8].

Table 2a Episodes of Illness by Diagnosis Type

Malaria Diagnosis in LSMS - Kagera					
Diagnosis Type	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Self-Diagnosed	418	590	538	393	1969
Doctor-Diagnosed	190	229	163	165	998

4.2.8 Definition: Malaria

The malaria status in the LSMS is based solely on self-reported episodes of doctor diagnosed and self-diagnosed cases. Respondents in the LSMS were asked to report all types of acute illness within the past four weeks and all types of chronic illness within the past 6 months. The malaria sample was then defined as inclusive of all individuals who

[8] Wave five of the LSMS for Kagera was administered 10 years after the wave four survey.

reported a lack of school or work attendance due to the malaria illness in the past seven days. To minimize measurement error with this definition of malaria status, the estimation process also controls for pre-existing conditions that can also result in low attendance.

4.3 Description of Variables

The following sections describe the dependent and independent variables selected for the models in equation (4.1) and equation (4.2) for schooling and labor supply.

4.3.1 Dependent Variables

The outcome variable in the labor force participation equation (4.2) is the number of *hours worked* per week. This variable is defined as the number of hours worked within a seven-day workweek for adults aged 18-65. The LSMS labor force sample is made up of respondents between the ages of 18 and 65 who reported work hours in the past seven days as an employee, a farmer, or a self-employed businessman. The outcome variable in the schooling equation (4.1) is the *number of school hours attended per week.* This variable is defined as the number of hours present in school during the school week for those between the ages of 7 and 20.

4.3.2 Explanatory Variables

Factors that determine labor supply in equation (4.2) are health status, (H), prices, (P), non-labor income, (V), socio-demographic indicators, (A), education, (E), and community infrastructure, (IN).

Health Status, H

In our conceptual framework, health status is approximated using self-reported measures of acute and chronic illness. This definition of health status differs from actual health by a measurement error, e:

Actual Health = H + e

The measurement error, e is sometimes treated as random and uncorrelated with other determinants of health. A problem with such treatment is that 'e' is unlikely to be random. For instance, in the labor supply case, individuals who have reduced their hours of work, are more likely to report poor health status, functional limitations, and other conditions (Chirikos and Nestel 1984). This can be attributed to the existence of factors that are known to the individual but unobserved by the researcher; described in the theoretical framework as μ. In this case, ignoring μ would lead to biased estimates of the measure of health status.

In addition, there are studies, which suggest that the concerns about non-random 'e' do not necessarily induce a bias on self-reported measures of health. One case is when an instrumental variables approach is used to capture the other part of H that is not addressed by self reported H (Ettner 1997). Although, self-reported health may bias the estimated coefficients downwards, the endogeneity of self-reported health may also bias the estimated effect upwards. Therefore, the two effects may cancel out unlike in the case of more objective measures of health that are biased towards zero only.

The determinants of health status (H) are evaluated as the existence of acute and chronic illness. Acute illness is measured as the incident of malaria. Acute illness in the

LSMS is defined as illness that limits the ability to work, attend school or perform regular chores. The incident of malaria in the survey is recorded by wave for individuals who reported being incapacitated by malaria within the survey period. Chronic conditions in the LSMS are addressed as pre-existing health conditions (such as HIV, Asthma and Kwashiorkor, and Malnutrition) that prevent the individual from working or attending school.

Prices (P), Education (E), and Socio-demographic Indicators (A)

The LSMS contains information on a collection of household expenditures on food and non-food items including the price of food, pharmaceutical products, and school fees. The LSMS price questionnaire consisted of thirty food items, thirteen non-food items, and six pharmaceutical products collected at different markets in Kagera. With the exception of a few items, food items in the survey were weighted to the nearest 50 grams. Pharmaceutical products were not weighted in the survey but measured in tablets. Complete pharmaceutical price information is only available for three of the six items in waves. A list of food and pharmaceutical items relevant to this study can be found on Table A4.

In addition to questions on household expenditures, the LSMS asks questions about the educational attainment of each respondent every wave. The survey also asks detailed questions on socio-demographic characteristics such as age, sex, marital status, and income. Household income in the LSMS is defined as the sum of six components: employment income from self-employment in agriculture; non-farm self-employment income; income from rent; transfer income from individuals and organizations; and other non-labor income. Because no information on wages is available in the LSMS dataset, household income is used as a proxy for the wages in the analyses.

Community and School Infrastructure (SI and IN)

In each wave, community leaders are interviewed by the LSMS. They are asked to identify the local schools and hospitals in the district. Representatives from each school and hospital facility are then re-interviewed in each subsequent wave on the changes in the demand and supply of these facilities. Three community infrastructure variables were chosen for the labor supply equations. These variables measure access to motorable roads, electricity and pipe borne water. For the schooling equation, the number of teachers per school was chosen as the measure of school infrastructure.

4.4 Econometric Strategy

Equation (4.1) and (4.2) can be estimated using standard econometric methods like OLS. OLS is unbiased in the absence of measurement error and unobserved heterogeneity. However, as mentioned in the theoretical framework, H_H^2 is endogenous to the labor supply model while, and H_K^1 is endogenous to the human capital accumulation model. Given the possibility of endogeneity in both models and the likelihood of measurement error in the health status variable, OLS is biased.

There are several solutions when dealing with unobserved heterogeneity. This study considers three identification strategies using the fixed effects, two stage least squares (2SLS), and the limited information maximum likelihood (LIML) estimators. The fixed effects (FE) estimator captures all unobserved, time in-variant factors that affect labor force participation and schooling. FE by no means solves the endogeneity problem and may also impose additional problems of attrition bias and insufficient within variation in panel data estimations. Given that the endogeneity problem posed by H_H^2 in the labor supply equation and, H_K^1 and in schooling equation will remain unsolved using FE, a more appropriate method of estimating (4.1) and (4.2) is by instrumental variables. Another problem with the FE estimator is that it imposes a strict exogeneity assumption on the regressors within the model.

Estimation by instrumental variables will involve estimation of equation (4.1) and equation (4.2) using factors such as the disease environment, D, and prices, as valid instruments for health. The instrumental variables regression can be implemented using the 2SLS and LIML methods. Having more than one instrument in the instrumental variables estimation also allows us to test overidentifying restrictions. Consistency of the instrumental variables estimates rely on the validity of the instruments. Invalid instruments can produce 2SLS and LIML parameter estimates that much more inconsistent than OLS.

4.4.1 2SLS Estimation Methods: Two Stage Least Squares (2SLS)

Estimation of the effect of malaria on schooling and labor supply is based on the structural equation

$$y_{it} = \alpha_0 + \beta h_{it} + \delta X_{it} + u_{it} \qquad (1)$$

where y_{it} is the outcome (hours of school and work) in the schooling and labor supply equations for individual i at time t; h_{it} is a vector of malaria and pre-existing health status indicators for an individual i at time t; X_{it} is a vector of socio-demographic, school and community level infrastructure, prices, and health status indictors. As discussed earlier, health status in (1) is correlated with u_{it}. Estimating the above structural model through OLS is biased. To mitigate the endogeneity problem in (1), the analysis employs a 2SLS regression of the schooling and labor supply models where the reduced form equation is given by:

$$h_{it} = \pi_0 + \pi_1 Z_{it} + \pi_2 X_{it} + v_{it} \qquad (2)$$

and Z_{it} is a vector of indicators of the disease environment serving as additional instruments in the reduced form estimation, and v_{it} is the reduced form error term. Assuming that Z_{it} is uncorrelated with the structural u_{it} and provided that the disease environment is sufficiently correlated with health status then,

$$Cov(h_{it}, Z_{it}) \neq 0$$

Finally, the study estimates a panel 2SLS model of schooling and labor supply where (2) is the first stage regression of health status on a vector of instruments in (2) and (1) is the second stage regression. The second stage regression uses hours of school and hours of work as the two outcome measures as depicted in equation (4.1) and equation (4.2) and a vector of health status indicators as well as a vector of socio-demographic, school and community level infrastructure, and price indicators as exogenous regressors.

4.4.2 Instruments

As discussed earlier, the likelihood of endogeneity in both equations can be alleviated using instrumental variables. It was also mentioned in chapter three that, P_{HI}

and D, are valid instruments for mitigating the endogeneity of h_{it} in (1). Two instruments of the disease environment available in the LSMS data for the estimation of the reduced form model in (2) are a rainfall season indicator and a measure of rainfall amounts in the Kagera region.

A good instrument satisfies the following assumptions – It must be uncorrelated with the error term and is correlated with the endogenous explanatory variable. As mentioned earlier, having more than one instrument in the 2SLS estimation facilitates the test for overidentification. The overidentification test can be implemented by first estimating the structural equations in (1) to obtain the 2SLS residuals $\hat{\mu}_1$. Then $\hat{\mu}_1$ is regressed on all exogenous explanatory variables in (1) to obtain the R-squared, R_1^2. Under the null hypothesis that all instruments are uncorrelated with μ_1, $nR_1^2 \sim \chi_q^2$ where q is the number of instruments minus the total number of endogenous explanatory variables. If the null hypothesis is rejected - nR_1^2 > the critical value in the chi-Square distribution (χ_q^2), it is safe to conclude that at least one of the instruments is not exogenous.

Considering the known correlation between malaria and high amounts of rainfall, monthly rainfall estimates and a rainfall season indicator can be considered as measures of the disease environment (Chris Drakeley, et al. 2005)[9]. The LSMS keeps monthly record of the rainfall amounts in Tanzania-Kagera. In the LSMS, monthly rainfall records are measured in millimeters over a period of 60 months. This record is only available for four of the five waves of the LSMS for Tanzania. To complete the rainfall

[9] Drakeley et al. (2005) found that plasmodium falciparum prevalence showed a negative relationship with altitude and rainfall amounts in Tanzania

data, the LSMS rainfall estimates are merged with additional rainfall data from the NOAA National Data Centers for wave five.

High prevalence for malaria usually falls in the rainy season months, which varies between regions in the Sub-Sahara. The rainy seasons in the Kagera district of Tanzania are the short 'Vuli' rainy season, which falls between October and January, and the long 'Msimu' rainy season from March to May. In addition to instrumenting for the likely endogeneity of malaria, controlling for the rainfall patterns using a season indicator ensures that seasonal variations in labor force participation and schooling are reasonably captured.

4.5 Analyzing Improvements in the Disease Environment

The panel structure of LSMS also allows for a direct test of whether changes in pre-conditioning factors have translated into improvements in the disease environment over time in Kagera. Factors preconditioning exposure to malaria used in this study are access to community health facilities and changes in household income levels over time. Using a difference-in-difference (DID) estimator, the estimated change in a preconditioning factor from wave t-1 to wave t can be obtained as:

$$\hat{\delta} = \gamma_t - \gamma_{t-1}$$

γ_t and γ_{t-1} represent levels of measurable preconditioning factors over time and $\hat{\delta}$ is the difference between predicted values of these factors over time. If the parameter $\hat{\delta}$ is significantly different from zero, it is safe to conclude that the preconditioning factors have improved the disease environment in Kagera over time.

Using the first and last waves of the LSMS Kagera, improvements in the disease environments are determined using the following DID estimation of malaria on two sets of preconditioning factors. Two preconditioning factors were chosen for the DID estimation: a count of health facilities in Kagera over time and changes in the levels of household income in Kagera over time.

$$malaria = \beta_0 + \delta_0 year + \beta_1 income + \beta_2 healthfacilities + \beta_3 year * income + \beta_4 year * healthfacilities + u$$

In the above DID equation, the parameters of interest, β_3 and β_4 measure the effects of changes in income and changes in the number of health facilities on the incidence of malaria. The parameter δ_0 captures the changes in the reported cases of malaria within the sample between 1991 and 2004.

4.6 Other Identification Strategies

In addition to 2SLS and LIML, another identification strategy explored in the estimation of the labor supply and schooling equations is the Average Treatment Effects on the Treated (ATET) estimator. Unlike the 2SLS and LIML the ATET takes into consideration the different states of the health status indicator (H). A thorough evaluation of both equations using ATET is presented in Appendix D.

Chapter 5

Research Results

This chapter presents the estimation results for the schooling and labor supply models. Section 5.1 presents the sample evidence on episodes of malaria for the school and work samples. Section 5.2 documents the school enrollment characteristics of school age children in Kagera. Section 5.3 details the summary statistics of the Kagera school sample. Section 5.4 and 5.5 review the summary statistics of the Kagera labor force sample. Estimation results for the school and work sample using OLS, FE, 2SLS, and LIML are presented in section 5.6 and 5.7. Lastly, section 5.8 presents difference in difference estimates (DID) of changes in the disease environment in Kagera.

5.1 Episodes of Malaria in the Study

The rates of reported malaria episodes by survey period for the Kagera sample ranges between a low of 7% and a high of 23% for the selected school and work samples. The percentage of reported malaria cases for children enrolled in school ranged between 7% and 13% from wave one to wave five of the LSMS (Table 2b). Unlike the school sample, workers in Kagera reported a wider range of malaria episodes by wave. The malaria rate for self-employed farmers ranged between 11% and 20% across waves. Non-farm self-employed workers and individuals in wage employment also reported a similar range of malaria cases (between 9% and 23%) across waves. Wave five had the highest number of reported malaria cases in the LSMS for children enrolled in school and for most types of employment. Overall, fewer people reported being sick with malaria in earlier waves than they did in more recent waves of the LSMS; a finding consistent with current trends in reported malaria cases for Tanzania (Table A3).

Table 2b: Malaria Cases

Reported Malaria Cases By Wave					
Sample	**Wave 1**	**Wave 2**	**Wave 3**	**Wave 4**	**Wave 5**
School Sample					
% Reporting a Malaria Episode	7%	11%	12%	11%	13%
Work Sample					
% of Self Employed (Farm) Reporting a Malaria Episode	11%	14%	13%	11%	20%
% of Self Employed (Non-Farm) Reporting a Malaria Episode	13%	9%	11%	15%	23%
% of Wage Employed Reporting a Malaria Episode	9%	17%	15%	10%	12%

5.2 School Enrollment in Kagera

There were over 11,000 children between the age of 7 and 20 in the Kagera sample from 1991 to 2004. Majority of these children were introduced into the survey in 2004 (wave five). Approximately 6400 children of school age were enrolled in school in at least one wave of the Kagera survey. In wave one, the school enrollment rate for children in the survey was 49%. Between wave two and wave three, the overall enrollment rate remained steady at 56%. By wave five, school enrollment for children of school-going age had increased to 62%.

In terms of school enrollment rates by gender composition, males were more likely to be to be enrolled in school than females. However, enrollment for both groups increased steadily in each wave with the exception of waves two and three when female school enrollment stalled at 54%. In addition to the overall low school enrollment in the sample, the school starting age was well over seven years. In fact, the lowest grade level, P1, consisted of students as old as age 16. This finding is consistent with that of Ainsworth, et al. (2005), which found that school enrollment was delayed for orphans in Tanzania and that children tend to drop out of school due to orphan status or death of an adult in the family.

Table 3: School Enrollment Rate

Children Ages 7 to 20: School Enrollment					
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Total	2232	2112	1996	1821	3114
Number Currently Enrolled in School	1104	1186	1125	1048	1936
Percentage of Total Currently Enrolled in School	49%	56%	56%	58%	62%
Percentage of Males Currently Enrolled in School	51%	58%	59%	60%	67%
Percentage of Females Currently Enrolled in School	48%	54%	54%	55%	58%

5.3 Summary Statistics Kagera School Sample

Table 4 presents the summary statistics for the entire Kagera school sample for all five waves of the LSMS data. The final primary school sample for all five waves of the survey consisted of 4,189 students. This sample size was conditioned on having reported current school enrollment. Therefore, the school sample excludes students who were at home on vacation or holiday at the time of the survey.

The summary statistics for primary school students in the Kagera showed that school children in Kagera reported attending approximately 28 hours of school per week between wave one and five of the study. The average school enrollment age in the schooling sample was 12 years. Overall, there were more students in lower grade levels in the Kagera schools than there were in upper primary grade levels. In particular, there were 70% more students enrolled in the P1 (the lowest grade level) than there were in P7 (the highest grade level). Again, this finding is consistent with that of Ainsworth, et al. (2005).

The average annual household income for parents with children in primary school in the Kagera survey was well below $200. Most parents with children in primary school in Kagera reported having less than an elementary school education. However, fathers in the sample were more likely to have completed primary school than mothers. In terms of malaria status, at least 11% of primary school children reported an episode of malaria in all five waves of the survey. However, the rate of chronic illness in the sample was much lower than the malaria rate. Merely 3% of primary school children reported a chronic illness in all five waves.

Other variables relevant to the Kagera school sample include school level infrastructure variables and measures of the cost of education. Approximately 10 teachers were staffed in a typical primary school in Kagera during the survey period. However, each primary school in the Kagera sample was staffed by a minimum of one teacher and a maximum of 29 teachers. In terms of costs, school students in Kagera live relatively close to home and on average commute about three miles to school. Real expenditures on tuition for primary school in Kagera were negligible as the average household spent about $3[10] a year on school fees in the survey.

[10] School fees in the analysis exclude expenditures such as money spent on books and supplies, school uniform and transportation as most of these variables contained missing observations in the sample.

Table 4: Summary Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Kagera School Characteristics (1991-2004)						
School Hours						
Hours	School Hours (Past 7 Days)	4189	28.311	12.686	0.000	55.000
Grade Level						
p1	First Grade Level	4189	0.136	0.342	0.000	1.000
p2	Second Grade Level	4189	0.138	0.345	0.000	1.000
p3	Third Grade Level	4189	0.128	0.334	0.000	1.000
p4	Fourth Grade Level	4189	0.121	0.326	0.000	1.000
p5	Fifth Grade Level	4189	0.104	0.306	0.000	1.000
p6	Sixth Grade Level	4189	0.077	0.266	0.000	1.000
p7	Seventh Grade Level	4189	0.031	0.172	0.000	1.000
Socio-Demographic Characteristics						
Age	Age in Years: Range 7-20	4189	12.604	2.940	7.000	20.000
Gender	Percent Male: 1 = male; 0 female	4189	0.527	0.499	0.000	1.000
Mother Education	Mother Education: 1 = primary education; 0 = no education	4189	0.136	0.343	0.000	1.000
Father Education	Father Education: 1 = primary education; 0 = no education	4189	0.178	0.383	0.000	1.000
Household Income	Annual family income in dollars	4189	199.531	473.660	0.000	11880.640
School Characteristics						
Teachers	Number of teachers at school	4189	9.671	4.106	1.000	29.000
Distance	Distance to school from home in miles	4189	3.276	26.109	0.000	500.000
Fees	Annual school fees in dollars	4189	0.851	7.571	0.000	300
Health Status						
Malaria	Malaria Cases	4189	0.113	0.317	0.000	1.000
Chronic Illness	Cases of Chronic Illness	4189	0.033	0.180	0.000	1.000
Disease Environment						
Malaria Season	Rainfall Season	4189	0.279	0.449	0.000	1.000
Rainfall	Total Monthly Rainfall (mm)[a]	4189	388.826	287.313	2.700	872.300

[a]This variable is constructed from the Living Standard Measurement Survey for Tanzania (1991-994) and NOAA National Data Centers (2004)

5.4 Summary Statistics Kagera Work Sample

Three employment categories where analyzed in the labor supply estimation. These employment groups are self-employment in farm occupations, self-employment in non-farm occupations, and wage employment (Table A5). The largest category of the Kagera work sample was farm employment, which accounted for 71% of the total workforce. Workers in wage employment made up 18% of the total work sample. The smallest employment category in the survey was the self-employment non-farm group, which accounted for 15% of the Kagera work sample.

The typical worker in the Kagera sample was age 34 years old, female and married. Precisely, the gender composition of the Kagera work sample was 45% male and 55% female. In addition, 56% of these workers were married at the time of the survey and the average worker reported having at least a primary school education. Workers in the Kagera sample earned an average of $189.93 dollars a year and worked approximately 20 hours per week.

In terms of community level indicators, access to community infrastructure was measured as the availability of to electric power, pipe borne water, and motorable roads in each community. With the exception of access to a motorable road, most Kagera workers reported living in communities without access to either of the other two community infrastructures. Specifically, 80% of workers in Kagera lived in areas without pipe borne water, while 59% resided in areas without electrical power. However, 96% of Kagera workers reported living near a motorable road at the time of the survey.

5.4.1 Farm Self-employment Characteristics

Two main types of farm activities were chosen for the definition of the farm employment work category. The first type of farm activity refers to adults who reported work hours on a farm or garden belonging to themselves or their household at the time of the survey. The second farm activity pertains to those who served as caretaker for animals or transformed animal products belonging to themselves or their households during the survey period. These two types of farm activity make up the single definition of farm related self-employment used in the analysis.

Table 5a presents the summary statistics for the Kagera farm sample. The average age for a typical Kagera farmer in the farm employment category was 35 years. Compared to other workers in non-farm self-employment, self-employed farmers are less likely to be educated. Most farmers in the survey had less than a secondary education. Majority of the farmers were female and married. Self-employed farmers in the Kagera survey earned on average $182.54 annually during the survey period. Farmers in Kagera were also less likely to have access to community level infrastructures such as motorable roads, pipe borne water and electric power supply compared to workers in non-farm self employment and wage employment. Ninety five percent of self-employed farmers lived in communities near a motorable road, while 36% had access to electric power and only 17% had access to pipe borne water.

Table 5a: Summary Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Kagera Self-Employment (Farm) Characteristics (1991-2004)						
Work Hours						
Hours	Number of hours worked per week	7,822	24.585	15.576	0.000	130.000
Socio-demographic Characteristics						
Age	Age in Years: Range 18 - 65	7,822	35.038	13.929	18.000	65.000
Gender	Percent Male: 1 = Male; 0 Otherwise	7,822	0.412	0.492	0.000	1.000
Education	Highest Level of Education: 1 = Primary; 2 = Secondary; 3 = College	7,822	1.893	1.300	0.000	6.000
Married	Marital Status : 1 = Married; 0 Otherwise	7,822	0.580	0.494	0.000	1.000
Income	Annual Income: Employment income in dollars	7,822	182.540	377.082	0.000	7401.588
Health Status						
Malaria	Malaria Cases = 1 if reported malaria in at least 1 wave; 0 otherwise	7,822	0.154	0.361	0.000	1.000
Chronic Illness	Chronic Illness = 1 if reported chronic illness in at least 1 wave; 0 otherwise	7,822	0.107	0.309	0.000	1.000
Disease Environment						
Malaria Season	Rainfall Season: 1 = Interviewed during rainy season; 0 otherwise	7,822	0.885	0.319	0.000	1.000
Rainfall	Total Monthly Rainfall[a:] Total Monthly Rainfall measured in millimeters	7,077	358.799	278.346	1.000	872.300
Community Infrastructure						
Electric Power	Electric Power: 1= Access; 0 otherwise	6,886	0.364	0.481	0.000	1.000
Pipe Water	Pipe Borne Water: 1 = Access; 0 otherwise	6,886	0.174	0.379	0.000	1.000
Road	Motorable Road: 1 = Access; 0 otherwise	6,886	0.951	0.215	0.000	1.000
Prices[b]						
Food	Average food price in dollars	6,393	0.156	0.196	0.003	3.800
Pharmaceuticals	Average pharmaceutical price in dollars	5,445	0.032	0.106	0.001	0.700

[a]This variable is constructed using estimates from the Living Standard Measurement Survey for Tanzania (1991-994) and NOAA National Data Centers (2004); Food items are weighted to the nearest 50 grams and pharmaceutical items are measured in tablets

In terms of weekly work hours, self-employed farmers in the Kagera survey reported working on average 25 hours per week. The number of hours worked per week for self-employed farmers were closer to the work hours of those in non-farm self-employment but lower than hours reported by workers in wage employment. When asked about malaria status, 15% of self-employed farmers in the survey reported being incapacitated with malaria. This was the lowest percentage of all types of employment in the Kagera work sample. Although, farmers reported lower malaria rates on average, they were also more likely to report having a chronic illness. Eleven percent of surveyed farm workers in Kagera reported a chronic illness in at least one wave of the survey.

5.4.2 Non-farm Self Employment Characteristics

A non-farm self employed worker in the analysis is defined as an individual who owns his/her own business or is employed by his/her family in a non-farm sector. This is the smallest employment category in all five waves of the survey with (Table 5b). The average age for workers in non-farm self-employment is 34 years. On average, non-farm self-employed workers reported a higher income than those in the self-employment farm category. Self-employed workers in the Kagera sample were mostly male and better educated than those in farm employment, but less educated than workers in wage employment.

Table 5b: Summary Statistics

Kagera Self-Employment (Non-farm) Characteristics (1991-2004)						
Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Work Hours						
Hours	Number of hours worked per week	1,659	24.940	21.842	0.000	130.000
Socio-demographic Characteristics						
Age	Age in Years: Range 18 - 65	1,659	33.474	12.015	18.000	65.000
Gender	Percent Male: 1 = Male; 0 Otherwise	1,659	0.592	0.492	0.000	1.000
Education	Highest Level of Education: 1 = Primary; 2 = Secondary; 3 = College	1,659	2.001	1.151	0.000	5.000
Married	Marital Status : 1 = Married; 0 Otherwise	1,659	0.586	0.493	0.000	1.000
Income	Annual Income: Employment income in dollars	1,659	221.565	591.714	0.000	7401.588
Health Status						
Malaria	Malaria Cases = 1 if reported malaria in at least 1 wave; 0 otherwise	1,659	0.165	0.371	0.000	1.000
Chronic Illness	Chronic Illness = 1 if reported chronic illness in at least 1 wave; 0 otherwise	1,659	0.096	0.295	0.000	1.000
Disease Environment						
Malaria Season	Rainfall Season: 1 = Interviewed during rainy season; 0 otherwise	1,659	0.820	0.384	0.000	1.000
Rainfall[a]	Total Monthly Rainfall[a:] Total Monthly Rainfall measured in millimeters	1,413	308.379	273.839	1.000	872.300
Community Infrastructure						
Electric Power	Electric Power: 1= Access; 0 otherwise	1,414	0.526	0.499	0.000	1.000
Pipe Water	Pipe Borne Water: 1 = Access; 0 otherwise	1,414	0.212	0.409	0.000	1.000
Road	Motorable Road: 1 = Access; 0 otherwise	1,414	0.967	0.177	0.000	1.000
Prices[b]						
Food	Average food price in dollars	1,260	0.199	0.260	0.003	1.500
Pharmaceuticals	Average pharmaceutical price in dollars	1,031	0.044	0.124	0.002	0.700

[a]This variable is constructed using estimates from the Living Standard Measurement Survey for Tanzania (1991-994) and NOAA National Data Centers (2004); Food items are weighted to the nearest 50 grams and pharmaceutical items are measured in tablets

In addition, work hours for non-farm self-employment were closer to those reported in farm employment. Specifically, the average worker in non-self employment in the Kagera sample worked approximately 25 hours a week. These workers were more likely to reside in communities with access to motorable roads, pipe borne water and electric power. Ninety seven percent of workers in non-farm self employment reported living in a community with access to a motorable road; while 53% reported having access to electric power and 21% had access to pipe borne water.

In terms of health status, non-farm self-employed workers were more likely to report episodes of malaria than workers in other employment categories. The malaria rate for workers in this group was 17% for all five waves of the survey. On the contrary, 10% of self-employed workers reported a chronic illness in all five waves of the survey.

5.4.3 Wage Employment Characteristics

Wage employment in the analyses is defined as employment by an employer who is not a member of the employee's household (i.e. a firm or the government). The employment characteristics for this work group are presented below on Table 5c. Workers in the wage employment category reported working an average of 40 work hours per week. This was the highest weekly work hours for all three work categories.

A typical worker in this cohort was 32 years of age, male, and better educated than workers in other work categories. Unlike those in farm employment, workers in wage employment had better access to community level infrastructures such as electric power, pipe-water, and motorable roads. Approximately, 96% of wage-employed workers had access to a motorable road, 25% had access to pipe borne water, while 50% had electric power supply at the time of the survey. Workers in wage employment also earned more income than workers in the other two categories of employment over the sample period. Particularly, a worker in wage employment had twice the earning power of a farm worker and earned 39% more income than a self-employed individual in Kagera. In terms of health status, 16% of workers in wage employment reported being sick with malaria during the survey period. This rate of malaria infection is high relative to that reported in farm employment but slightly lower than the malaria rate for those in self-employed non-farm occupations. On the contrary, wage workers were significantly less likely to report a chronic illness when compared to their counterparts. Only 7% of workers in wage employment suffered from a chronic condition in at least one wave of the survey.

5.5 Other Kagera Sample Characteristics

Food and pharmaceutical price data were used in addition to community infrastructure as indicators of the Kagera work environment. These variables are derived directly from the LSMS dataset. The calculation of food prices in the dataset is based on an average of a 30 food items priced at different points in time and weighted to the nearest 50 grams. Pharmaceutical prices are measured in tablets in the survey. Table A4 provides a list of the food and pharmaceutical items relevant to this study. The prices were collected for each item at the cluster level, where one cluster consists of 16 households. Food and pharmaceutical prices were only used in the labor supply estimation. Over the survey period for the Kagera work sample, food prices for all 30 items averaged at of 0.2 cents and peaked at 3.80 dollars. Pharmaceutical prices for selected items averaged at 0.04 and peaked at 0.70 cents.

In addition, the estimated labor supply and schooling equations also include an indicator of rainfall amounts and a measure of seasonal rainfall variations. Rainfall amounts in the LSMS were measured in millimeters over a 60-month survey period from wave one to four. Rainfall estimates for wave five was derived from the NOAA national data centers. Monthly rainfall estimates ranged between one and 872.30 millimeters over the sample period (Table 4). For the school sample, monthly rainfall amounts averaged at 384.706 millimeters over the sample period (Table 4). Average rainfall amounts for the total work sample was 341.27 millimeters for all waves.

A malaria season indicator was also included in the analyses. The malaria season indicator is a dummy variable capturing rainy and non-rainy months in Kagera. Additional coverage of seasonal rainfall patterns in Kagera can be found in chapter four.

Table 5c: Summary Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Kagera Wage Employment Characteristics (1991-2004)						
Work Hours						
Hours	Number of hours worked per week	1,987	40.650	21.095	0.000	130.000
Socio-demographic Characteristics						
Age	Age in Years: Range 18 - 65	1,987	32.389	11.637	18.000	65.000
Gender	Percent Male: 1 = Male; 0 Otherwise	1,987	0.721	0.449	0.000	1.000
Education	Highest Level of Education: 1 = Primary; 2 = Secondary; 3 = College	1,987	2.121	1.280	0.000	6.000
Married	Marital Status : 1 = Married; 0 Otherwise	1,987	0.508	0.500	0.000	1.000
Income	Annual Income: Employment income in dollars	1,987	365.769	567.481	0.002	7300.002
Health Status						
Malaria	Malaria Cases = 1 if reported malaria in at least 1 wave; 0 otherwise	1,987	0.161	0.368	0.000	1.000
Chronic Illness	Chronic Illness = 1 if reported chronic illness in at least 1 wave; 0 otherwise	1,987	0.074	0.262	0.000	1.000
Disease Environment						
Malaria Season	Rainfall Season: 1 = Interviewed during rainy season; 0 otherwise	1,987	0.831	0.374	0.000	1.000
Rainfall[a]	Total Monthly Rainfall[a:] Total Monthly Rainfall measured in millimeters	1,718	335.444	281.118	1.000	872.300
Community Infrastructure						
Electric Power	Electric Power: 1= Access; 0 otherwise	1,679	0.496	0.500	0.000	1.000
Pipe Water	Pipe Borne Water: 1 = Access; 0 otherwise	1,679	0.245	0.430	0.000	1.000
Road	Motorable Road: 1 = Access; 0 otherwise	1,679	0.958	0.200	0.000	1.000
Prices[b]						
Food	Average food price in dollars	1,458	0.183	0.204	0.005	1.500
Pharmaceuticals	Average pharmaceutical price in dollars	1,188	0.042	0.122	0.002	0.700

[a]This variable is constructed using estimates from the Living Standard Measurement Survey for Tanzania (1991-994) and NOAA National Data Centers (2004); Food items are weighted to the nearest 50 grams and pharmaceutical items are measured in tablets

5.6 Main Estimation Results

The estimation results of the causal effects of malaria on school and work hours are presented in this section. Section 5.6.1 summarizes the estimated results for the school sample beginning with the first stage 2SLS estimates and the overidentification test. Section 5.6.2 presents the 2SLS, OLS and FE results for the entire school sample. As indicated earlier, the Kagera labor supply sample is categorized into three main types of employment (namely, self-employed farmers, self-employed non-farm workers and wage employed workers) in order to separately identify the impact of malaria on each group. A separate binary indicator is included in the analyses identify each work group. Section 5.6.3 and 5.6.4 presents the first and second stage estimation results for the three employment groups in the sample.

5.6.1 First Stage IV Estimation Results - Kagera School Sample

First stage estimation results for the Kagera sample are presented in Table 6a. The three instruments used in the first stage as proxies for the disease environment are one binary indicator of the seasonal variations in rainfall and a level and squared indicator of monthly rainfall amounts[11]. The first stage results show a significant correlation between malaria and seasonal variations in rainfall for the Kagera region. The coefficient on the monthly rainfall indicator is small magnitude. Even so, the first stage results show a positive correlation of the three instruments with the incidence of malaria. A joint significance test of the instruments indicates the statistical significance of the

[11] Analysis of the reported cases of malaria in the overall Kagera data showed a non-linear relationship between malaria episodes and rainfall amounts (Figure A2).

three instruments in the first stage. Table 6a shows that the partial F-test on the three instruments is statistically significant at the 1% level.

The n*R-Sq test was used to test overidentifying restrictions in the first stage IV estimation process. The overidentification test is based on the null hypothesis that the instruments are uncorrelated with the error term in the first stage IV regression. The n*R-Sq estimate in the first stage estimation is 1.028 with a p-value of 0.5982. Therefore, the null hypothesis is rejected and it is safe to conclude that at least one of the IVs is not exogenous in the first stage estimation. Although the overidentification test and n*R-Sq test confirm the validity of the instruments in the first stage estimates, they do not guarantee the exogeneity of the health status variables in the second stage.

In addition to the instruments evaluated at the first stage, the first stage estimates also provide a meaningful insight into the determinants of malaria in the Kagera school sample. Two variables found to be significantly correlated with the incidence of malaria in the school sample are age and the household spending on tuition. As presented in the first stage results on Table 4a, age decreases the probability of reporting an incidence of malaria in the Kagera school sample by 0.003%. On the other hand, a unit increase in household spending on school fees increases the probability of reporting an incidence of malaria by 0.002%. An explanation of the significant positive correlation between the probability of reporting an episode of malaria and household spending on education lies in basic economic theory. The positive correlation indicates a likely tradeoff between household spending on education and household spending on health. In other words, the more the household allocates resources to education the less it has left to allocate to health.

Table 6a: First Stage Results

Kagera School Sample				
Dependent Variable: Malaria (N = 4189)				
1st Stage Explanatory Variables	**Coefficient**	**Standard Error**	**t**	**P>t**
Age	-0.003917*	0.002121	-1.850	0.065
Grade 1	-0.005497	0.016421	-0.330	0.738
Grade 2	0.026833	0.016353	1.640	0.101
Grade 3	0.008775	0.017239	0.510	0.611
Grade 4	0.026348	0.018167	1.450	0.147
Grade 5	0.036911*	0.019930	1.850	0.064
Grade 6	0.033975	0.022608	1.500	0.133
Gender	-0.014960	0.009957	-1.500	0.133
Mother Education	0.003289	0.015023	0.220	0.827
Father Education	0.006277	0.011492	0.550	0.585
household Income	0.000002	0.000011	0.180	0.858
Distance to School	0.000229	0.000205	1.120	0.264
School Fees	0.002353***	0.000744	3.160	0.002
Number of Teachers	-0.001640	0.001219	-1.350	0.179
Chronic Illness	-0.014631	0.028712	-0.510	0.61
Instruments				
Rainfall Season	0.036441***	0.013810	2.630	0.008
Total Rainfall	-0.000132	0.000084	-1.590	0.112
Total Rainfall Squared	1.10E-07	0.0000001	1.170	0.243
F-Statistic (P-value)	2.73 (0.000)			
Joint Significance - IVs Only (P-value)	7.11 (0.000)			
Overidentification Test - N*R-Sq (P-value)	1.028 (0.5982)			

***1% Significance Level **5% Significance Level *10% Significance Level
Overidentification Test (N*R-Sq): Reject $H_{0;}$ at least some IVs are not exogenous

The malaria season indicator is positively related to malaria incidence and statistically significant even though the magnitude of the coefficient is small. The coefficients on the rainfall amount variables are not individually statistically significant. However, an overall test reveals the joint statistical significance of all the instruments in the schooling estimation at the 1% level.

5.6.2 2SLS, OLS and FE Estimation Results - Kagera School Sample

Table 6b presents the 2SLS, OLS, and FE estimation results for the dependent variable school hours for the Kagera school sample. The table shows a statistically significant causal relationship between malaria and school hours. The incidence of malaria in the Kagera school sample is associated with a 23.67 hour decrease in weekly school attendance. This estimate is statistically significant at the 1% level. OLS and FE estimates of the relationship between malaria and school hours also show a significant causal effect although the magnitudes of the coefficients for both regressors are smaller than the estimated 2SLS effect. Unlike 2SLS malaria estimates, the OLS and FE results are likely to be biased downwards due to the endogeneity of malaria.

Beside the malaria variable, another health status indicator in the school sample estimations is chronic illness. This variable is insignificant in explaining school hours for the Kagera school sample. This lack of significance can be attributed to the relatively small number of students in the Kagera sample reporting a known form of chronic illness (see Table 4). Another indicator of the number of school hours attended per week is age. 2SLS estimates show that being a year older increases weekly school attendance by 0.82 hours.

Grade level was also a strong indicator of school attendance. The omitted grade level in the analysis grade seven. With the exception of students in the 1st grade, hourly school attendance per week is positively correlated with grade level for all grades. In particular, students who had completed the 4th grade have the highest hourly school attendance rate compared to those in grade seven at the time of survey.

In addition to age in the schooling equation gender also proved significant in predicting school hours. Males in the Kagera school sample are less likely to attend school on a weekly basis compared to females. In particular, male students in the Kagera sample attended 1.076 fewer weekly school hours compared to female students. Furthermore, 2SLS, OLS, and FE estimates all show that family income increases weekly school attendance by 0.002 hours a week. However, parental education proved insignificant in explaining school attendance.

In terms of the indicators of school infrastructure and the cost of schooling, school fees and the number of teachers in a school are both significant predictors of school attendance. School fees, an indicator of the cost of schooling is significant in both 2SLS and FE estimations at the 10% level or better. 2SLS estimates show that a one unit increase in school fees increases weekly school attendance by 0.065 hours, while OLS and FE estimates predict a much lower effect. The predicted positive relationship between weekly school attendance and school fees is counterintuitive.

Table 6b: Second Stage 2SLS Results

Kagera School Sample						
Dependent Variable: Hours of School Attended Per Week (N = 4,189)						
	Two Stage Least Squares		**Ordinary Least Squares**		**Fixed Effects**	
Variables	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Malaria	-23.674***	9.018	-2.918***	0.534	-2.064***	0.802
Age	0.820***	0.097	0.900***	0.067	0.682***	0.198
Grade 1	-0.563	0.681	-0.735	0.541	-0.031	0.983
Grade 2	8.902***	0.732	7.698***	0.540	8.829***	1.231
Grade 3	9.012***	0.723	8.573***	0.569	10.883***	1.410
Grade 4	9.424***	0.795	8.573***	0.607	9.903***	1.573
Grade 5	6.441***	0.907	5.881***	0.673	6.289***	1.740
Grade 6	8.365***	1.009	7.694***	0.753	7.512***	1.937
Gender	-1.076**	0.433	-0.881***	0.335		
Mother Education	-0.407	0.624	-0.689	0.490	-0.145	0.913
Father Education	0.107	0.484	0.067	0.362	-1.755***	0.681
Income	0.002***	0.000	0.002***	0.000	0.002***	0.001
Distance to School	-0.012	0.009	-0.016**	0.007	-0.011	0.008
School Fees	0.065*	0.039	0.024***	0.006	0.013	0.022
Number of Teachers	0.057	0.052	0.080**	0.041	-0.149	0.193
Chronic Illness	-1.462	1.205	-1.413	1.010	2.453	1.442
Constant	15.675***	1.958	12.304***	0.934	16.177***	2.886
Hausman Statistic (p-value)	31.42 (0.009)					

***1% Significance Level **5% Significance Level *10% Significance Level

Note: Malaria = 1 if reported at least one episode of malaria in all 5 waves; 0 otherwise. Age = age in years; Grade 1-7 = Completed Primary School Grade Levels; the omitted group is grade 7. Gender = 1 if male, 0 otherwise; Mother/Father Education = 1 if mother/father has at least a primary school education; 0 otherwise. Household Income = income in dollars; Chronic illness = 1 if reported at least one chronic illness; 0 otherwise Distance to School = Estimated distance traveled from home to school in miles. Number of teachers (this is a community level variable) = Number of teachers per school.

Higher school fees might reflect higher school quality, which is not controlled for in the estimation. In addition, parents who spend more on school related expenses may be inclined to encourage higher school attendance compared to those who spend less. The positive relationship between school attendance and tuition can also be attributed to school quality and perhaps a public versus private school quality phenomenon since the majority of students in the survey were enrolled in public schools. In fact, there are no records of child enrollment in private schools for the first 4 waves of the survey and the fraction of reported enrollments in public school in wave five is small. In addition to school quality, other unobserved factors such as family characteristics or family values regarding education could be driving the observed effect.

As described in chapter four, the main proxy for school infrastructure in the estimation is number of teachers per school. 2SLS and FE estimates of the relationship between the number of teachers per school and school attendance are statistically insignificant. However, the OLS coefficient for school infrastructure is statistically significant and suggests that an additional teacher staffed in a Kagera school moderately increases school attendance by 0.08 hours per week.

5.6.3 First Stage IV Estimation Results - Kagera Work Sample

The first stage results for the Kagera work sample are presented in Table 7a. The instruments in the first stage 2SLS regression for the work sample are identical to those used in section 5.6.1 for the school sample. All three instruments for the disease environment for the labor supply sample are jointly statistically significant in the first stage. The n*R-Sq test was done to ensure the endogeneity of the instruments in the first stage. With a χ^2 statistic of 3.479 and a p-value of 0.1756, it is safe to conclude that at least one of the three instruments is not exogenous in the first stage.

As indicated earlier, studies evaluating the causal effects of malaria on labor supply have thus far focused on farm employment. The motivation behind disaggregating the Kagera work sample by employment type was to determine the extent to which type of employment influences the incidence of malaria, with the expectation that farm employment increases the likelihood of contracting malaria. To clarify this, the analyses begin with evaluating the employment type indicators in the first stage 2SLS estimation.

According the first stage results, workers in wage employment have a higher probability of reporting an episode malaria compared to those in self-employment. There is no statistically significant difference in the incidence of malaria between workers in farm employment and those in self-employment. The effect of malaria on workers by type of employment is further investigated in the second stage through interacting malaria with employment type variables.

As seen on Table 7a, age is a significant indicator of malaria incidence for the work sample. However, the magnitude of the age effect on the adult work sample is rather small. Chronic illness and food prices also proved significant in the first stage work sample estimation. Chronic illness for the entire work sample increases the incidence of malaria for working adults by 2 percentage points. On the other hand, food prices have the strongest effect on malaria incidence in terms of magnitude. With the exception of the age, chronic illness and food prices, other regressors in the labor supply estimation proved insignificant in explaining the incidence of malaria; a finding similar to that obtained the first stage estimation for the Kagera school sample. An overall significance of the malaria first stage regression was evaluated using a joint F-test. The joint F-test confirmed the statistical significance of the first stage labor supply 2SLS estimates at the 1% level.

Table 7a: First Stage Results (without interaction terms)

Kagera Work Sample				
Dependent Variable: Malaria (N = 5,876)				
1st Stage Explanatory Variables	**Coefficient**	**Std. Err.**	**t**	**P>t**
Farm Employment	-0.016	0.011	-1.510	0.131
Wage Employment	-0.0009***	0.013	-0.670	0.502
Age in Years	0.001	0.000	2.820	0.005
Gender	0.003	0.009	0.370	0.715
Education	-0.001	0.004	-0.200	0.845
Married	-0.008	0.009	-0.830	0.408
Income	-8.60E-06	0.000	-0.800	0.422
Chronic Ill	0.021*	0.013	1.670	0.096
Electric Power	-0.006	0.013	-0.450	0.653
Pipe Water	-0.010	0.014	-0.710	0.480
Motorable Road	-0.027	0.029	-0.950	0.344
Pharmaceutical Price	0.099	0.094	1.050	0.295
Food Price	0.051*	0.029	1.760	0.079
Instruments				
Malaria Season	-0.110**	0.044	-2.510	0.012
Total Rainfall	-0.0002***	0.000	-3.940	0.000
Total Rainfall Squared	2.86E-07***	7.08E-08	5.110	0.000
F-Statistic (P-value)	3.26 (0.000)			
Joint Significance - IVs Only (P-value)	9.07 (0.000)			
Overidentification Test - N*R-Sq (P-value)	3.479 (0.1756)			

Overidentification Test (N*R-Sq): Reject $H_{0;}$ at least some IVs are not exogenous

5.6.4 2SLS, OLS and FE Estimation Results - Kagera Work Sample

This section presents the main estimation results for the Kagera work sample. First, the 2SLS[12] estimates for the work sample are presented on Table 7b without interaction terms. In Table 7c, the complete work sample results are presented with interaction terms to determine the extent to which malaria limits work hours for each employment type.

2SLS estimates of work hours without interaction of malaria with type of employment (Table 7b) show that a typical Kagera worker who reported being sick with malaria lost 13.9 hours of work per week over the sample period. This estimate is however insignificant at the 10% level. OLS and FE estimates of the causal effects of malaria on labor outcomes also indicate a loss in weekly work hours. Although, the coefficients derived by the OLS and FE estimators are significantly smaller in magnitude than the 2SLS result, the OLS and coefficients on malaria are statistically significant at the 5% and 10% levels.

Table 7b also shows a negative casual effect of chronic illness on the Kagera work sample. 2SLS estimates of the effect of chronic illness for all types of employment indicates that being chronically ill decreases the number of work hours by approximately 2.5 hours a week. OLS and FE effects of chronic illness on work hours are similar to those obtained by 2SLS and range between -2.68 and -1.45 hours per week. Estimates of the casual effect of chronic illness on weekly work hours are significant at the 1% level for 2SLS and OLS and at the 5% level for FE.

---

[12] OLS and FE results are presented solely for the purpose of comparison. Given the endogeneity of malaria status in the estimation of the schooling and labor supply models, OLS and FE estimates of the causal effect are biased.

Table 7b also shows the effect of occupation choice on work hours. 2SLS estimates suggest that workers in farm employment and those in wage employment work more hours on average than those in self-employment. All three estimators confirm that workers in farm employment worked 15 hours more per week than those in self-employment; whereas, those in wage employment supplied 16-21 more hours of work per week compared to those in self-employment. All estimates on work hours by employment type are significant on the 1% level.

Other statistically significant indicators of labor supply for the Kagera adult sample are, age and gender, marital status, income, and community infrastructure. In terms of age and gender, 2SLS estimates show that being a year older increases work hours by approximately 0.040 hours per week, while being male increases work hours by 4.46 hours per week. 2SLS estimates also show that married workers work on average 1.77 hours more than their unmarried counterparts. Also, household income slightly increases the number of weekly hours by 0.006 hours in the 2SLS estimation.

In terms of community infrastructure, access to a motorable road, electric power, and pipe borne water are all statistically significant indicators of work hours using 2SLS. Access to a motorable road increases the number of hours worked per week by 4 hours while, electric power supply increases work hours by 1.5 hours per week. Accessibility of pipe borne water also increases weekly work hours by 3.3 hours. Pharmaceutical and food prices proved insignificant in predicting work hours in the 2SLS, OLS, and FE estimates. In addition, education had no statistically significant effect on the number of hours worked per week for the Kagera work sample.

Table 7b: Second Stage 2SLS Results (without interaction terms)

Kagera Work Sample						
Dependent Variable: Number of Hours Worked Per Week (N = 5,876)						
	Two Stage Least Squares		**Ordinary Least Squares**		**Fixed Effects**	
Variables	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Malaria	-13.914	8.730	-1.368**	0.544	-1.184*	0.619
Farm Employment	15.306***	0.502	15.737***	0.443	15.893***	0.530
Wage Employment	20.783***	0.576	21.093***	0.534	16.572***	0.723
Age in Years	0.040**	0.017	0.024*	0.014	-0.215**	0.087
Gender	4.459***	0.432	4.573***	0.401		
Education	0.238	0.170	0.217	0.156	-0.592*	0.349
Marital Status	1.786***	0.416	1.848***	0.386	2.558**	1.048
Income	0.006***	0.000	0.006***	0.000	0.004***	0.001
Chronic Illness	-2.468***	0.608	-2.688***	0.553	-1.452**	0.705
Electrical Power	1.529***	0.572	1.043**	0.497		
Pipe Borne Water	3.312***	0.648	3.868***	0.560		
Motorable Road	4.114***	1.310	4.723***	1.234		
Pharmaceutical Price	1.680	4.411	-1.199	3.170	-7.098	7.048
Food Price	-1.025	1.410	-1.028	1.186	1.742	1.456
Constant	-1.770	1.906	-3.767	1.452	14.276	3.406
Hausman Statistic (p-value)	24.43 (0.0274)					

***1% Significance Level **5% Significance Level *10% Significance Level

Note: Malaria = 1 if reported at least one episode of malaria in all 5 waves; 0 otherwise Farm employment = 1 if worker is a self employed farmer; 0 otherwise. Wage Employment = 1 if worker is employed by a firm or government; 0 otherwise (base group is self employment non-farm). Age = age in years; Gender = 1 if male, 0 otherwise; Education = 1 if worker has at least a primary school education; 0 otherwise; Marital status =1 if worker is married; 0 otherwise. Income = income in dollars; Chronic illness = 1 if reported at least one chronic illness; 0 otherwise Electrical power (this is a household level indicator) = 1 if worker resides in a residence with electrical power; 0 otherwise. Pipe Borne water (this is a community level indicator) = 1 if worker lives in area with access to pipe borne water; 0 otherwise. Motorable road (this is a community level indicator) = 1 if worker lives in area with motorable road; 0 otherwise. Pharmaceutical prices = average price of selected pharmaceutical products priced at the community level (see chapter four for further description). Food prices = average price of selected food items priced at the community level (see chapter four for further description)

Table 7c presents a comprehensive evaluation of the causal relationship between the malaria incidence and work hours by type of employment. This particular estimation differs from the parsimonious 2SLS estimation presented in Table 7b as interaction terms between malaria and the employment type indicators are introduced. Interacting malaria with employment variables facilitates the assessment of the causal effect of malaria from one sector to the other. The first stage estimates of this model are presented in the appendix on Table A7. The comparison group in this estimation is the self-employment group.

The overidentification test for the comprehensive labor supply model is presented on Table A7. The χ^2 statistic and a p-value for the N*R-Sq test are 11.168 and 0.0833, respectively. Therefore, the comprehensive model fails the test overidentification test suggesting that all the instruments in the first stage are exogenous.

The previous work sample results presented on Table 7b suggested a statistically insignificant causal effect of malaria on weekly work hours. Once malaria incidence is defined as a function of employment type, the casual effect of malaria on labor supply (in terms of work hours) is now centered on farm employment. In essence, associating malaria with type of employment concentrates the malaria effect into the farm employment sector where farmers lose approximately 35 hours of work per week compared to workers in self-employment. This estimate is statistically significant at the 5% level in the 2SLS regression. In terms of the effect of malaria on wage employment, the comprehensive model produced statistically insignificant results indicating no decipherable causal relationship between malaria and work hours for workers in wage employment compared to those in self-employment.

Although the 2SLS results in the comprehensive model, confirm a negative causal relationship between malaria incidence and work hours, the concern over the validity of the rainfall instruments in the first stage remains. The failed overidentification test on Table A7 arises from the estimation of 2SLS in instrumental variable settings with weak instruments. This particular problem was worsened by the inclusion of additional instruments to in the comprehensive model[13].

Recently in empirical economic literature, concerns have been raised over the reliability of inference based on conventional instrumental variables settings using 2SLS when the instruments are only weakly correlated with the endogenous regressors. Older studies including that of John Bound, et al. (1995) emphasize the potential pitfalls of estimating 2SLS in the case of weak instruments. The most citied example in these studies is that of Joshua Angrist and Alan Krueger (1991) (AK-91), which estimated wage equations using quarter of birth as an instrument for educational attainment. Paul Bekker (1994), Bound, et al. (1995), Gary Chamberlain and Guido Imbens (2004) and Jeffery Wooldridge and Guido Imbens (2007), to name a few, contend that the results obtained by AK-91 are influenced by weak instruments. Although, the standard errors reported in AK-91 are reasonable and the specified model passed the overidentification test, the coverage properties are still very poor suggesting that 2SLS estimates may be misleading especially in the case of many weak instruments Wooldridge and Imbens (2007).

[13] Introducing interaction terms to the labor supply model to create a comprehensive model (Table A7) necessitated instrumenting for the additional endogenous regressors (i.e. the interaction terms between malaria and employment type).

Table 7c: Second Stage 2SLS Results (with interaction terms)

Kagera Work Sample						
Dependent Variable: Number of Hours Worked Per Week (N = 5,876)						
	Two Stage Least Squares		Ordinary Least Squares		Fixed Effects	
Variables	Coefficient	Std. Err.	Coefficient	Std. Err.	Coefficient	Std. Err.
Malaria	4.909	11.851	-1.035	1.105	-2.184*	1.236
Malaria * Farm Employment	-35.100**	16.302	-0.207	1.237	1.727	1.369
Malaria * Wage Employment	14.946	21.015	-1.218	1.536	-2.135	1.738
Farm Employment	19.961***	2.200	15.773***	0.478	15.647***	0.569
Wage Employment	19.263***	2.624	21.248***	0.568	16.802***	0.749
Age in Years	0.041**	0.019	0.024*	0.014	-0.215**	0.087
Gender	4.384***	0.450	4.579***	0.401		
Education	0.181	0.193	0.214	0.156	-0.572*	0.349
Marital Status	1.760***	0.483	1.854***	0.386	2.586**	1.048
Income	0.006***	0.001	0.006***	0.000	0.004***	0.001
Chronic Illness	-2.165***	0.699	-2.680***	0.553	-1.449**	0.705
Electrical Power	1.753***	0.642	1.043**	0.498		
Pipe Borne Water	3.090***	0.731	3.875***	0.560		
Motorable Road	3.978***	1.453	4.714***	1.234		
Pharmaceutical Price	-5.343	5.304	-1.155	3.171	-6.999	7.047
Food Price	-1.099	1.560	-1.030	1.186	1.728	1.456
Constant	-4.125	2.298	-3.818	1.458	14.379	3.408
Hausman Statistic (p-value)	29.30 (0.0147)					

***1% Significance Level **5% Significance Level *10% Significance Level
Malaria*Farm = malaria interacted with farm employment indicator. Farm = 1 if worker is a self-employed farmer; 0 otherwise. Malaria*Wage = malaria interacted with wage employment indicator. Wage = 1 if worker is employed by a firm or the government; 0 otherwise. Non-farm self employment is the omitted group.

5.7 Limited Information Maximum Likelihood (LIML) Estimation

The main concern with estimating 2SLS in cases with weak or irrelevant instruments involves the construction of confidence intervals with good coverage properties. Wooldridge and Imbens (Summer, 2007) state that although, conventional methods such as 2SLS and LIML are rarely misleading when there are many weak instruments, LIML is generally much better than 2SLS. In particular, LIML has better properties compared to 2SLS when proportional adjustments to the LIML standard errors are made. A simple adjustment to the LIML inference statistics for the endogenous explanatory variable can be made as proposed in Marcelo Moreira (2001). Moreira developed reliable statistical inference tests with better coverage for structural parameters based on instrumental variable settings with weak instruments. Moreira showed that confidence regions based on the likelihood ratio (LR) statistics have coverage probabilities close to their nominal levels no matter how weak the instruments. Others, such as Jean-Marie Dufour (1997), have also showed that LR-type tests behave more smoothly in the presence of identification problems compared to Wald-type tests[14].

The instruments in the labor supply estimation are weak as indicated by the overidentification test (Table A7). The result of estimation under weak instruments is that the obtained 2SLS inference statistics are potentially misleading. A viable alternative to 2SLS under weak instruments is LIML. However, conventional LIML standard errors are invalid, but proportional adjustments can be made. The labor supply and schooling equations are re-estimated in section 5.7.1 and 5.7.2 using the LIML estimator based on the conditional likelihood ratio (CLR) approach developed in Moreira.

[14] Dufour (1997) shows that LR-type tests do not require the indentifiability; a property not shared by Wald-type statistics.

In addition, the Anderson-Rubin statistic, the Lagrange Multiplier score, and the Wald statistic are reported alongside the Moreira CLR statistic for comparison.

5.7.1 Limited Information Maximum Likelihood (LIML) Estimation Results - Kagera School Sample

Table 8a provides the LIML estimates for the Kagera school sample. The objective of the LIML estimation is to provide potentially better inference tests given the possibility of weak instruments. Although the weak instrument problem is a bigger concern in the labor supply estimates, the LIML results provide a sensitivity check for the schooling estimates in section 5.6.2.

The causal effect of malaria on the Kagera school sample obtained using the LIML estimator is similar to that obtained by the 2SLS estimator. According to the LIML estimates, malaria is associated with a 24 hour loss in school attendance per week. The CLR statistic for malaria indicates that the malaria coefficient is statistically different from zero at the 1% level. The Anderson-Rubin statistic, the Wald statistic, the Lagrange Multiplier score also indicate the statistical significance of the causal effect.

Table 8a also displays the estimated LIML coefficients for the other variables in the Kagera schooling model. Standard errors are invalid for the LIML estimation therefore, additional inference on other variables in the model is invalid. Nonetheless, the estimated LIML coefficients are similar in magnitudes to those obtained in the 2SLS estimation of the Kagera schooling model. The direction of the malaria effect on the Kagera school sample generated by the LIML estimator is also comparable that obtained via OLS and FE.

Table 8a: LIML Results

Main Estimation Results: Kagera School Sample				
Dependent Variable: Hours of School Attended Per Week (N = 4,189)				
	LIML		**2SLS**	
Variables	Coefficient	Std. Error	Coefficient	Std. Error
Malaria	-24.064	8.951	-23.674***	9.018
Age	0.8501	0.098	0.820***	0.097
Grade 1	-0.562	0.683	-0.563	0.681
Grade 2	8.925	0.734	8.902***	0.732
Grade 3	9.019	0.725	9.012***	0.723
Grade 4	9.433	0.797	9.423***	0.795
Grade 5	6.45	0.908	6.441***	0.907
Grade 6	8.385	1.01	8.364***	1.009
Gender	-1.079	0.434	-1.076**	0.433
Mother Education	-0.68	0.726	-0.407	0.624
Father Education	-0.102	0.667	0.107	0.484
Income	0.002	0.0004	0.002***	0.000
Distance to School	-0.011	0.009	-0.012	0.009
School Fees	0.067	0.039	0.065	0.039
Number of Teachers	0.059	0.052	0.057	0.052
Chronic Illness	-1.578	1.173	-1.462	1.205
Constant			15.675***	1.958
Hausman Statistic (p-value)			31.42 (0.009)	

***1% Significance Level **5% Significance Level *10% Significance Level

CLR Test: H_0: b[malaria] = 0.0000

Likelihood Ratio Statistic: 9.7279; Critical Value: 3.8415; Reject H_0

Anderson-Rubin Statistic: 10.7741; 95% Critical Value: 7.8147; Reject H_0

Wald Statistic: 7.1322; Critical Value: 3.8415; Reject H_0

Lagrange Multiplier Score: 9.2363; 95% Critical Value: 3.8415; Reject H_0

5.7.2 Limited Information Maximum Likelihood (LIML) Estimation Results - Kagera Work Sample

Table 8b presents the LIML estimation results for the Kagera work sample. The LIML estimates on Table 8b account for the weak instrument problem. As indicated earlier, LIML estimates have improved coverage properties over 2SLS in settings with weak instruments. Therefore, a gain in statistical inference power is expected with LIML over 2SLS. Table 8a presents LIML results alongside 2SLS, OLS, and FE estimates for comparison.

The LIML estimates of the causal effect of malaria on work hours are similar to those obtained using 2SLS with the exception of an improved significance of the coefficient. The 2SLS results described in section 5.6.4 revealed a negative casual effect of malaria on work hours, which was statistically significant at the 15% level. LIML estimates of the same specification using rainfall amounts and a rainfall season indicator as instruments produced similar results. In particular, the LIML coefficient on malaria suggests that workers in Kagera lose approximately 15 hours to malaria per week.

Table 8b also presents the test CLR, Anderson-Rubin, Wald, and LM statistics for the work sample. The CLR test verifies the extent of the weak instruments problem and confirms that the malaria effect for the Kagera work sample is not different from zero. Anderson-Rubin, Wald, and LM tests also suggest that the malaria effect is not different from zero. In all, it is reasonable to conclude that the rainfall instruments do not properly mitigate the endogeneity of malaria in the labor supply equation.

Table 8b: LIML Results

Main Estimation Results: Kagera Work Sample				
Dependent Variable: Number of Hours Worked Per Week (N = 5,876)				
	LIML		**2SLS**	
Variables	Coefficient	Standard Error	Coefficient	Standard Error
Malaria	-15.746	9.434	-13.914	8.730
Farm Employment	15.275	0.512	15.306***	0.502
Wage Employment	20.771	0.585	20.783***	0.576
Age in Years	0.041	0.017	0.040**	0.017
Gender	4.452	0.436	4.459***	0.432
Education	0.234	0.172	0.238	0.170
Marital Status	1.777	0.42	1.786***	0.416
Income	0.006	0.001	0.006***	0.000
Chronic Illness	-2.431	0.618	-2.468***	0.608
Electrical Power	1.539	0.579	1.529***	0.572
Pipe Borne Water	3.292	0.655	3.312***	0.648
Motorable Road	4.082	1.325	4.114***	1.310
Pharmaceutical Price	-1.443	4.478	1.680	4.411
Food Price	-0.923	1.437	-1.025	1.410
Constant	-1.54	1.973	-1.770	1.906
Hausman Statistic (p-value)	24.43 (0.0274)			

***1% Significance Level **5% Significance Level *10% Significance Level

CLR Test: H_0: b[malaria] = 0.0000

Likelihood Ratio Statistic: 3.4720; Critical Value: 3.8415; Fail to Reject H_0

Anderson-Rubin Statistic: 6.9141; 95% Critical Value: 7.8147; Fail to Reject H_0

Wald Statistic: 2.9395; Critical Value: 3.8415; Fail to Reject H_0

Lagrange Multiplier Score: 3.0182; 95% Critical Value: 3.8415; Fail to Reject H_0

5.8 Difference in Difference Estimates of Changes in the Disease Environment

The results of the DID estimation of changes in the disease environment due to increases in income levels and the number of health facilities in Kagera are presented below on Table 10. The 'year' variable captures changes in malaria prevalence over time in Kagera from 1991-2004. As indicated below on Table 10, an increase in reported malaria cases occurred between 1991 and 2004. Precisely, the DID estimates show that malaria prevalence in the region increased by 43% percent over the sample period and is statistically significant on at the 1% level.

In terms of improvements in the disease environment over time due to rising income levels, the derived DID estimates are counterintuitive. The coefficient on the income interaction term is positive suggesting that malaria prevalence has increased over time with changes in income levels. Although, the magnitude of the coefficient on the income interaction is negligible, it is statistically significant at the 10% level. One reason for the observed relationship is that even though income levels have risen over time, the disease environment may have worsened at a faster pace.

On the other hand, the effect of a rise in the number of health facilities is also suspect. The model failed to confirm a statistically significant improvement in the disease environment due to changes in the number of health facilities. The coefficient on the health facilities interaction with time is nonetheless negative although statistically insignificant.

Table 10: Improvements in the Disease Environment

DID Estimates of Improvements in the Disease Environment				
Dependent Variable: malaria (N = 14,647)				
Variable	**Coefficient**	**Std. Err.**	**z**	**P>|z|**
income	-0.0094	0.0056	-1.69	0.090
health facility	-0.0254	0.1089	-0.23	0.816
year	0.4348	0.0361	11.41	0.000
year * income	0.0098	0.0056	1.75	0.081
yr * health facility	-0.1394	0.1132	-1.23	0.218
_cons	-1.3159	0.0299	-43.89	0.000
Chi-Sq Statistic = 197.15				
P>Chi-Sq = 0.000				

A joint significance test of the estimates in Table 10 reveals the statistical significance of the income and health facilities indicators in predicting changes in the disease environment over time. A consistent conclusion about the disease environment cannot be reached from Table 10. However, the results generated by the DID estimator make intuitive sense when looking at the reported cases of malaria in Tanzania over time and the current state of the disease in the Kagera region.

The upward trend in reported cases of malaria for Tanzania on Table A3 suggests an increased persistence of the disease irrespective of the changes in the socio-economic environment. In addition, the recent report by the TRCNS of a rise in the number of malaria cases in Kagera for the year 2006 lends some validity to the results obtained by the DID estimator. A 43% increase in malaria prevalence in the LSMS sample over time and the recent report by the TRNS in Kagera suggest a possible adaptation of the malaria virus over time to the disease environment. Evidence of such adaptation of the malaria virus to treatment in Sub-Saharan Africa has been documented in recent epidemiological literature (Bloland 2001).

5.9 Other Findings

This study also made additional attempts to better identify the labor supply equation using Average Treatment Effects on the Treated (ATET). The Average Treatment effects estimator has been mostly utilized in the labor supply literature to measure the impact of specific interventions on an outcome of interest. In health economics literature, studies using the ATET estimator analyze the effect of specific health interventions on health related outcomes (an example is Gabriel Picone, et al. 2006). As an extension to the analyses in section 5.6, the ATET estimator was also applied to both the labor supply and schooling models. A description of the ATET estimator and how it was applied to the data as well as the results derived from the analyses are presented in Appendix D.

As part of the identification strategy, a separate analysis carried out in the study involves substituting household income for a measure of the household wage rate in the labor supply model. The structural labor supply model presented in chapter two defines labor supply as a function of household health, prices, wages, non-labor income, socio-demographics, and community level infrastructure. However, the analyses so far have relied on household income as a proxy for the wage rate. As an extension to the labor supply estimations in section 5.6, the model is re-estimated using a proxy for the wage rate, which was defined in terms of the number of hours worked per member of the household. This particular specification yielded causal effects identical to those obtained in previous regressions, but the proxy wage variable was statistically insignificant. A formal table of the results for the wage rate regression will not be presented in this study.

Chapter 6

Conclusion

This chapter summarizes the main findings of the effects of malaria on schooling and labor supply. Section 6.2 discusses the policy implications of the findings in the study. Additionally, the limitations of this research and opportunities for future research are discussed in section 6.3 and 6.4, respectively.

6.1 Summary of Findings

This dissertation is the first attempt to empirically measure the causal effect of malaria on schooling in Sub-Saharan Africa. It is also the first attempt to investigate the effect of malaria on different sectors of employment in the region. The causal effects of malaria on economic outcomes are examined for the young and adult segments of the population in the Kagera region of Tanzania using panel data from 1991-2004. The main findings drawn from this study are summarized below.

Health status as described in chapter four is mainly measured as the incidence of malaria. Given that malaria status is endogenous in equations (4.1) and (4.2), a seasonal indicator of the variations in rainfall patterns in Tanzania and a rainfall amount variable are used as instruments to mitigate the endogeneity problem, while also controlling in the estimation for the existence of chronic illness.

This dissertation finds a negative causal relationship between malaria and school attendance for children of schooling age in Kagera. Following the same definition of health status, the effect of malaria on adults of working age in Kagera is inconclusive. In terms of the school sample, this dissertation finds that children of schooling age in Kagera may lose up to 24 hours of school per week due to malaria. The largest causal effects generated for the school sample was derived using the 2SLS and LIML estimator at 23-24 hours lost in school attendance per week. Standard regression methods such as OLS and FE generate modest estimates suggesting a 2-3 hour decrease in weekly school attendance attributable to malaria.

The attempt to mitigate the endogeneity of malaria in order to obtain unbiased estimates of the causal effect of malaria on labor supply proved insufficient. Although similar instruments were employed in the labor supply estimations and the results point to a negative causal effect, identification problems within the model could not be resolved. Therefore, inference on the malaria effect is invalid and the overall evidence on the causal effect of malaria on the Kagera work sample is inconclusive.

Even though this study finds that the effect of malaria on healthy work days is inconclusive; it is important to note that this study does not measure the impact of malaria on labor productivity. Malaria may very well impact labor productivity more than it influences labor supply. In addition, the short-term effects of missing work may also be mitigated by labor substitution; a phenomenon not studied in this dissertation. It is likely that the extent of economic loss from illness may be minimized through family labor substitution, even if the time contributed by a labor substitute is not as productive as the time lost. Therefore, in some cases the tradeoff is between the number of hours worked and labor productivity.

Another element of health status in equation (4.1) and (4.2) is chronic illness. This study finds a negative effect of chronic illness on labor supply. In particular, chronic illness can be related to a loss of 1-2 hours in weekly school attendance for the school sample and similarly for the labor supply sample. The effect of chronic illness on school attendance is statistically insignificant.

6.2 Policy Implications

This dissertation has shown that malaria limits the number of healthy school days in Sub-Saharan Africa. The adverse effects of malaria on schooling are likely to exceed school absenteeism, as absenteeism may increase failure and school dropout rates for children. Therefore, policies geared towards malaria eradication in Sub-Saharan Africa will not only benefit the health of the population, but also enhance the quality of human capital in the region. Malaria eradication in Sub-Saharan Africa is feasible. However, some methods of eradication of the malaria virus have been strongly opposed by policy makers.

Successful efforts to eradicate malaria in the industrialized world involved house spraying of residual insecticides (such as DDT) and anti-malarial drug treatment (CDC – Eradication of Malaria in the U.S.). While eradication was a success for industrialized countries, developing countries face a bigger obstacle in their efforts to wipe out the disease. DDT spraying is by far the cheapest method of eradicating malaria. However, environmental concerns have led to its ban in most countries. Since many Sub-Saharan

African countries depend on international aid to fund many disease prevention and treatment programs, they face the likelihood of donor sanctions if DDT is used for eradicating malaria.

Beyond eradication, other costs of malaria treatment and prevention include the costs of insecticide treated nests for households in Sub-Saharan Africa and the emergence of drug resistant plasmodium falciparum. The direct costs of malaria prevention and treatment also eat into the meager incomes of households in Sub-Saharan Africa. Guyatt Brooker, et al. (2000) showed that the actual costs of prevention and treatment for households in Sub-Saharan Africa are unrealistically expensive in a region where most people live on less than $1 a day. Ettling, et al. (1994) also evaluated the financial burden of malaria on households in Sub-Saharan Africa and found that over 25% of earnings in low-income households in Malawi is spent on malaria treatments.

The findings in this dissertation suggest that malaria does limit the number of healthy school days for children in Sub-Saharan Africa. Cumulative effects of absenteeism can indeed translate in to indirect costs to society in terms of human capital accumulation. Effective malaria control policies may salvage such losses and improve overall economic welfare in the region.

6.3 Limitations

A major limitation of this research is the lack of strong instruments for mitigating the endogeneity of malaria, especially for the labor supply model. Finding a better measure of the disease environment that is correlated with malaria incidence should improve the indentifiability of the causal effect for the labor supply sample. Other potential instruments for malaria in the labor supply and schooling models are the cost of anti-malaria drugs and geographic elevation. Although, the LSMS collected data on prices for Chloroquine, a malaria treatment drug, no information is available in the survey on anti-malaria drugs. Drakeley, et al. (2005) determined that malaria is negatively correlated with altitude in Tanzania. Although data on geographic elevation lacks significant variation, this problem can be overcome by interacting geographic elevation with other variables such as regional identifiers in the model.

In addition to the problem of weak instruments, another problem faced in this study is the unavailability of panel type data for other countries in Sub-Saharan Africa. There are relatively few panel data collected on households in Sub-Saharan Africa. Although LSMS datasets are available for 5 countries in Sub-Saharan Africa to date, most of these datasets are limited to one wave while others lack detailed information on household health.

6.4 Future Research

Future research should consider the cumulative impact of malaria on labor supply and schooling. This can be done by estimating the probability of repeating a grade due to repeated episodes of malaria for the children of school age. The cumulative impact of malaria on labor supply can be analyzed through estimating the extent of job loss stemming from multiple episodes of malaria. Future research should also address the costs and benefits of malaria eradication and anti-malaria treatment initiatives possibly through field experiments.

References

Ainsworth, Martha; Beegle, Kathleen and Koda, Godlike. 2005. "The Impact of Adult Mortality and Parental Deaths on Primary Schooling in North-Western Tanzania." *The Journal of Development Studies*, 41:3, pp. 412-39.

Amexo, Mark; Tolhurst, Rachel and Barnish, Guy, et al. 2004. "Malaria Misdiagnosis: Effects on the Poor and Vulnerable." *Lancet*, 364, pp. 1896-98.

Audibert, Martine. 1986. "Agricultural Non-Wage Production and Health Status." *Journal of Development Economics*, 24, pp. 275-91.

Angrist, Joshua D. and Krueger Alan B. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, CVI: 4, pp. 979-1014

Bartel, Ann and Taubman, Paul. 1979. "Health and Labor Market Success: The Role of Various Diseases." *The Review of Economics and Statistics*, 61:1, pp. 1-8.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *The Economic Journal*, 75, pp. 493-517.

Becker, Sascha O. and Ichino, Andrea. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *The Stata Journal*, 2:4, pp. 358-77.

Beegle, Kathleen. 2003. "Labor Effects of Adult Mortality in Tanzanian Households." *World Bank Policy Research Working Paper 3062*.

Bekker, Paul A. 1994. "Alternative Approximations of the Distributions of Instrumental Variable Estimators." *Econometrica*, 62:3, pp. 657-81.

Bloland, Peter B. 2001. "Drug Resistance in Malaria." *World Health Organization Communicable Disease Surveillance & Response (CSR)*, 4.

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

Bound, John; David, Jaeger A. and Baker, Regina M. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the

Endogeneous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90:430, pp. 443-50.

Brooker, S; Guyatt, J. and Omumbo, R., et al. 2000. "Situation Analysis of Malaria in School-aged Children in Kenya - What Can Be Done?" *Parasitology Today*, 16:5, pp. 183-86.

Case, Anne; Paxson, Christina and Ableidinger, Joseph. 2002. "Orphans in Africa." *NBER Working Paper 9213*.

Center for Disease Control and Prevention. 2006. Malaria. http://www.cdc.gov/malaria/disease.htm. (accessed September 22, 2006)

Chamberlain, Gary and Imbens, Guido. 2004. Random effects estimators with many instrumental variables." *Econometrica*, 72:1, pp. 295-306.

Chirikos, Thomas N. and Nestel, Gilbert. 1984. "Economic determinants and consequences of self-reported work disability." *Journal of Health Economics*, 3, pp. 117-36.

______. 1985. "Further Evidence on the Economic Effects of Poor Health." *The Review of Economics and Statistics*, 67:1, pp. 61-69.

Currie, Janet and Madrian, Brigitte C. 1999. "Health, Health Insurance and the Labor market." *Handbook of Labor Economics,* 3, pp. 3309-415.

Dehejia, Rajeev H. and Wahba, Sadek. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics*, 84:1, pp. 151-61.

Deressa, Wakgari. 2007. "Treatment-seeking behavior for febrile illness in an area of seasonal malaria transmission in rural Ethiopia." *Malaria Journal*, 6:49.

Drakeley, Chris; Carneiro, Ilona and Reyburn, Hugh, et al. 2005. "Altitude Dependent and -Independent Variations in *Plasmodium falciparum* Prevalence in Northeastern Tanzania." *The Journal of Infectious Diseases*, 191, pp. 1589-98.

Dufour, Jean-Marie. 1997. "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models." *Econometrica*, 65:6, pp. 1365-87.

Ettling, M, McFarland, DA and Schultz, LJ, et al. 1994. "Economic Impact of Malaria in Malawian Households." *Tropical Medicine and Parasitology*, 45:1, pp. 74-9.

Ettner, Susan L.; Frank, Richard G. and Kessler, Ronald C. 1997. "The Impacts of Psychiatric Disorders on Labor Market Outcomes." *Industrial and Labor Relations Review*, 51:1, pp. 64-81.

Fox, Matthew P.; Rosen, Sydney and MacLeod, William B., et al. 2004. "The Impact of HIV/AIDS on Labor Productivity in Kenya." *Tropical Medicine and International Health*, 9:3, pp. 318-24.

Grossman, Michael. 1973. "The Correlation Between Health and Schooling." *NBER Working Paper 22*.

Grossman, Michael and Benham, L. 1974. "Health, hours, and wages. *In The Economics of Health and Medical Care* (Mark Perlman, ed.). London.

Haveman, Robert; Wolfe, Barbara and Kreider, Brent, et al. 1994. "Market work, wages, and men's health." *Journal of Health Economics*, 13:1994, pp. 163-82.

Imbens, Guido and Wooldridge, Jeffrey. Summer 2007. "Estimation of Average Treatment Effects Under Unconfoundedness." *NBER Lecture Notes*.

_____. 2007. "Weak Instruments and Many Instruments." *NBER Lecture Notes*.

Kahn, Matthew E. 1998. "Health and Labor Market Performance: The Case of Diabetes." *Labor Economics*, 16:4, pp. 878-99.

Kessler, Ronald C. and Frank, Richard G. 1997. "The imapact of psychiatric disorders on work loss days." *Psychological Medicine*, 27, pp. 861-73.

Leighton, Charlotte and Foster, Rebecca. 1993. "Economic Impacts of Malaria in Kenya and Nigeria." *A Health Financing and Sustainability Project (Abt Associates).*

Lloyd, Cynthia B. and Blanc, Ann K. 1996. "Children's Schooling in Sub-Saharan Africa: The Role of Fathers, Mothers, and Others." *Population and Development Review*, 22:2, pp. 265-98.

Luft, H.S. 1975. "The Impact of Poor Health on Earnings." *Review of Economics and Statistics*, 57 (February), pp. 43-57.

McCombie, S.C. 1996. "Treatment Seeking for Malaria: A Review of Recent Research." *Social Science and Medicine*, 43:6, pp. 933-45.

Miguel, Edward and Kremer, Michael. 2004. "Worms: Identifying impacts on education and health in the presence of treatment externalities." *Econometrica*, 72:1, pp. 159-217.

Mitchell, Jean M. and Burkhauser, Richard V. 1990. "Disentangling the effect of arthritis on earnings: a simultaneous estimate of wage rates and hours worked." *Applied Economics*, 22, pp. 1291-309.

Moreira, Marcelo J. 2001. "Tests With Correct Size When Instruments Can be Arbitrarily Weak." *Center for Labor Economics Working Paper Series, 37, UC Berkeley*.

Murphy, Sean C. and Breman Joel G. 2001. "The Childhood Malaria Burden in Africa: Cerebral Malaria, Neurological Sequelae, Anemia, Respiratory Distress, Hypoglycemia, and Complications of Pregnancy." *American Journal of Tropical Medicine and Hygiene,* 64:(1,2)S, pp. 57-67.

Nicholson, Walter. 2002. Microeconomic Theory: Basic Principles and Extensions. Mason, Ohio: Southern-Western (Thomas Learning)

Nur, El Tahir Mohamed. 1993. "The Impact of Malaria on Labor Use and Efficiency in the Sudan." *Social Science and Medicine*, 37:9, pp. 1115-19.

Picone, Gabriel; Khwaja, Ahmed and Salm, Martin. 2006. "A Comparison of Treatment Effect Estimators Using Severity of Illness Information from Hospital Charts." *Forthcoming, Journal of Applied Econometrics*.

Roll Back Malaria (RBM). 2006. Malaria in Africa. http://www.rollbackmalaria.org/. (accessed August 25, 2006).

Rosenfield, P.L; Golladay, F. and Davidson, R.K. 1984. "The Economics of Parasitic Diseases: Research Priorities." *Social Science and Medicine*, 19:10, pp. 1117-26.

Sachs, Jeffery and Malaney, Pia. 2002. "The economic and social burden of malaria." *Nature*, 415, pp. 680-85.

Sascha, Becker O. and Ichino, Andrea. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal*, 2:4, pp. 358-77.

Shapiro, Barbara S.; Dinges, David F. and Orne, Emily C., et al. 1995. "Home management of sickle cell-related pain in Children and adolescents: natural history and impact on school attendance." *Pain*, 61, pp. 139-44.

Shchultz, Paul T. and Tansel, Aysit. 1983. "Measurement of Returns to Adult Health: Morbidity Effects on Wage Rates in Cote d'Ivoire and Ghana." *Living Standards Measurements Study Working Paper No. 95*.

Societies, International Federation of Red Cross and Red Crescent. 27 February 2007. "Tanzania: Malaria Outbreak in Kagera." *DRF Bulletin*.

Strauss, John and Thomas, Duncan. 1998. "Health, Nutrition, and Economic Development." *Economic Literature*, 36:2, pp. 766-817.

The Global Fund. 2006. Fighting Malaria. http://www.theglobalfund.org/en/about/malaria/. (accessed August 25, 2006).

_____. 2006. Fighting AIDS. http://www.theglobalfund.org/en/about/aids/. (accessed September 22, 2006).

_____. 2006. Fighting Tuberculosis. http://www.theglobalfund.org/en/about/tuberculosis/. (accessed September 22, 2006).

Thirumurthy, Harsha; Zivin, Joshua G. and Goldstein, Markus. 2006. "The economic Impact of AIDS Treatment: Labor Supply in Western Kenya." *Job Market Paper*.

United Nations Children's Fund (UNICEF). 2006. Health – Malaria. http://www.unicef.org/health/index_malaria.html. (accessed September 22, 2006).

United States Agency for International Development (USAID). 2006. Accelerating the Fight Against Malaria. http://www.usaid.gov/press/factsheets/2005/fs050915.html. (accessed August 2, 2006).

Wang'Ombe, Joseph K and Mwabu, Germano M. 1993. "Agricultural Land Use Patterns and Malaria Conditions in Kenya." *Social Science and Medicine*, 37:9, pp. 1121-30.

Wooldridge, Jeffery M. 2003. Introductory Econometrics: A Modern Approach. Mason, Ohio: South-Western (Thomson Learning).

World Health Organization (WHO) Health System Statistics. 2006. http://www.who.int/whosis/database/core/core_select_process.cfm?countries=all&indicators=nha. (accessed August 25, 2006).

_____. 2006. International Travel and Health. http://www.who.int/ith/en/. (accessed June 23, 2006).

World Health Organization (WHO) – Roll Back Malaria (RBM). 2006. The Africa Malaria Report 2003. http://www.rbm.who.int/amd2003/amr2003/amr_toc.htm. (accessed November 13, 2006).

_____. 2006. World Malaria Report 2005. http://www.rbm.who.int/wmr2005/html/toc.htm. (accessed June 23, 2006).

Appendix

Appendix A: Tables

Table A1: Number of reported cases of HIV, TB and Malaria

Country	*HIV Prevalence (2005)*	*TB Prevalence (2004)*	*Malaria Prevalence (2002)*
Global	**38 600 000**	**14 602 353**	**408,388,001**
Sub-Saharan Africa	**24 500 000**	**3 843 167**	**43 614 238**
Angola	320 000	48,006	1409328
Benin	87 000	9,779	…
Botswana	270 000	46,815	28858
Burkina Faso	150 000	41,049	1451125
Burundi	150 000	36,473	1808588
Cameroon	510 000	1,556	…
Central African Republic	250 000	21,882	…
Chad	180 000	53,448	…
Comoros	500	740	…
Congo	120 000	18,035	2640168
Côte d'Ivoire	750 000	116,349	…
Democratic Republic of Congo	…	307,554	…
Djibouti	15 000	1,585	5021
Equatorial Guinea	8900	18,498	…
Eritrea	59 000	403,098	75386
Ethiopia	…	4,619	427831
Gabon	60 000	4,858	…
Gambia	20 000	81,480	…
Ghana	320 000	37,739	2830784
Guinea	85 000	4,713	…
Guinea-Bissau	32 000	297,330	194976
Kenya	1 300 000	9,782	124197
Lesotho	270 000	14,493	…
Liberia	…	63,606	…
Madagascar	49 000	63,159	1543130
Malawi	940 000	75,845	2853317
Mali	130 000	14,975	723077
Mauritania	12 000	1,670	167423
Mauritius	4100	123,360	22
Mozambique	1 800 000	11,767	4458589
Namibia	230 000	38,850	442527
Niger	79 000	683,847	681707
Nigeria	2 900 000	58,658	2605381
Rwanda	190 000	387	…
Senegal	61 000	51,383	…
Sierra Leone	48 000	66	…
Somalia	44 000	45,215	96922
South Africa	5 500 000	316,260	15649
Sudan	350 000	131,543	3056400
Swaziland	220 000	11,580	14863
Togo	110 000	43,012	…
Uganda	1 000 000	179,843	7216411
United Republic of Tanzania	1 400 000	180,069	7489890
Zambia	1 100 000	81,187	…
Zimbabwe	1 700 000	87,006	1252668
North Africa	**440 000**	**78 978**	**424**
Algeria	19 000	17,432	307
Egypt	5300	25,364	10
Libyan Arab Jamahiriya	…	1,146	107
Morocco	19 000	32,640	…
Tunisia	8700	2,396	…
	WHO Estimates 2005	*WHO Estimates 2004*	*WHO Estimates 2002*

Table A2: Per Capita Health Expenditures in Sub-Saharan Africa

Per capita total expenditure on health at average exchange rate (US$)					
Country	**2000**	**2001**	**2002**	**2003**	**2004**
Angola	15.5	21.7	18	23.9	25.5
Benin	15.1	16.7	16.8	20.8	24.2
Botswana	129.8	129.8	143.9	231.6	328.6
Burkina Faso	11.1	11.3	13.7	18.9	24.2
Burundi	3.4	3.1	3	2.7	3
Cameroon	30.7	31.3	34.8	44.3	50.7
Cape Verde	55.5	61.3	66.2	77.7	97.8
Central African Republic	9.8	9.7	10.5	12.1	13.2
Chad	10.7	11.8	12.3	16.3	19.6
Comoros	8	7	9.6	12.7	13.2
Congo	19.8	19	20.1	24.1	27.6
Côte d'Ivoire	29.5	24.1	25.6	30.8	33
Democratic Republic of the Congo	9.8	4.3	3.3	4.1	4.7
Equatorial Guinea	42.9	64.9	166.6	111.2	168.2
Eritrea	8.9	8.4	7.4	8.7	9.9
Ethiopia	5.1	5	4.9	5	5.6
Gabon	164.4	190.7	190.3	206.3	231.3
Gambia	21	20	16.9	17.4	18.5
Ghana	17.4	18.6	18.8	24.2	27.2
Guinea	17.2	18	20	22.1	21.8
Guinea-Bissau	7.4	6.2	7	7.6	8.7
Kenya	18.1	17.7	18.6	19.9	20.1
Lesotho	28	30.8	25.8	39.7	49.4
Liberia	5.8	6.1	5.1	5.4	8.6
Madagascar	6	10	10.3	10.9	7.3
Malawi	9.2	11.3	16	18.3	19.3
Mali	14.3	16.1	16.3	20.7	23.8
Mauritania	11.3	9.1	13.3	13.9	14.5
Mauritius	142.2	142.2	167	180.3	222.3
Mozambique	11.3	9.3	10.3	10.6	12.3
Namibia	126.1	109.9	97.4	147.4	189.8
Niger	5.3	6.2	6.4	7.6	8.6
Nigeria	18	19.1	18.8	20.8	23
Rwanda	9	8	8.3	13.4	15.5
Sao Tome and Principe	20.8	30.4	31.4	45.2	47.8
Senegal	19	20.3	24.2	31.6	39.4
Seychelles	403.5	406.5	463.1	515.4	534.4
Sierra Leone	5.4	5.8	6.8	6.8	6.6
South Africa	235.6	216	198.4	300.7	390.2
Swaziland	82.7	67.3	63	111.4	145.8
Togo	11	12.1	12.9	15.6	17.9
Uganda	15.6	16.9	17.7	17.5	19
United Republic of Tanzania	10.8	10.9	10.4	10.6	12
Zambia	17.1	18.7	21.2	23.9	29.6
Zimbabwe	44.4	65.1	150.5	39.9	27.2

Source: World Health Statistics

Table A3: Annual Reported Malaria Cases – Sub-Sahara Countries (WHO Estimates)

Annual Reported Malaria Cases - Sub-Sahara Countries											
Country	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993
Angola	..	1,409,328	1,385,597	1,635,884	1,471,993	1,169,028	893,232	..	156,603	667,376	722,981
Benin	..	..	779,041	707,408	709,348	650,025	670,857	623,396	579,300	546,827	403,327
Botswana	22,418	28,858	48,237	71,403	72,640	59,696	101,887	80,004	17,599	29,591	55,331
Burkina Faso	..	1,451,125	1,203,640	1,032,886	867,866	721,480	672,752	582,658	501,020	472,355	502,275
Burundi	..	1,808,588	2,855,868	3,057,239	1,936,584	687,301	670,857	974,226	932,794	831,481	828,429
Cameroon	..	..	..	..	..	664,413	787,796	931,311	784,321	189,066	478,693
Cote d'Ivoire	..	..	400,402	1,491,943	..	..	983,089	1,109,011	755,812	..	421,043
Central African Rep	95,644	..	140,742	89,614	127,964	105,664	99,718	95,259	100,962	82,057	82,072
Chad	..	..	386,197	369,263	392,815	395,205	343,186	278,048	293,564	278,225	234,869
Comoros	..	..	3,718	9,618	9,793	3,844	..	15,509	15,707	13,860	12,012
Congo	..	..	..	..	..	17,122	9,491	14,000	28,008	35,957	15,504
Djibouti	5,036	5,021	4,312	4,667	6,140	5,920	4,314	6,105	5,982	6,140	4,166
Equatorial Guinea	..	..	..	..	..	..	..	..	12,530	14,827	17,867
Eritrea	72,023	75,386	125,746	119,155	147,062	255,150	..	129,908	81,183	..	..
Ethiopia	565,273	427,831	400,371	383,382	647,919	604,960	509,804	478,411	412,609	358,469	305,616
Gabon	..	..	..	..	..	80,247	57,450	74,310	54,849	82,245	70,928
Gambia	..	..	..	..	127,899	..	325,555	266,189	135,909	299,824	..
Ghana	3,552,869	2,830,784	3,383,025	3,349,528	2,895,079	1,745,214	2,227,762	2,189,860	1,928,316	1,672,709	1,697,109
Guinea	..	..	..	889,089	807,895	817,949	802,210	772,731	600,317	607,560	..
Guinea-Bissau	..	194,976	202,379	246,316	197,454	2,113	10,632	6,457	197,386	..	158,748
Kenya	..	124,197	132,590	74,194	122,792	80,718	..	3,777,022	4,343,190	6,103,447	..
Liberia	..	..	..	..	..	777,754	826,151	239,998	..	..	..
Madagascar	2,114,400	1,543,130	1,429,491	1,383,239	1,141,474	..	..	..	196,358	..	..
Malawi	..	2,853,317	2,955,627	3,774,982	4,193,145	2,985,659	2,761,269	6,183,290	..	4,736,974	4,686,201
Mali	809,428	723,077	612,895	546,634	530,197	12,234	384,907	29,818	95,357	263,100	295,737
Mauritania	..	167,423	243,942	259,093	253,513	168,131	189,571	181,204	214,478	156,080	43,892
Mauritius	..	22	61	62	73	52	65	82	46	65	54
Mozambique	5,087,865	4,458,589	3,978,397	3,278,525	2,336,640	194,024	..	12,794	..	..	..
Namibia	444,081	442,527	537,115	519,113	429,571	353,110	390,601	345,177	275,442	401,519	380,530
Niger	..	681,707	606,802	646,757	815,895	872,925	978,855	1,162,824	778,175	806,204	726,666
Nigeria	2,608,479	2,605,381	2,253,519	2,476,608	1,965,486	2,122,663	1,148,542	1,149,435	1,133,926	1,175,004	981,943

Table A3: Annual Reported Malaria Cases – Sub-Sahara Countries (WHO Estimates)

Annual Reported Malaria Cases - Sub-Sahara Countries (Contd.)											
Country	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993
Rwanda	856,233	..	..	915,916	906,552	1,279,581	1,331,494	1,145,759	1,391,931	371,550	733,203
Senegal	..	..	..	1,120,094	1,145,112	948,823	861,276	..	628,773	450,071	..
Sierra Leone	..	..	..	..	409,670	249,744	209,312	7,192	..	..	..
Somalia	23,349	96,922	10,364	10,364	9,055	..	..	..	..	..	3,049
South Africa	13,446	15,649	26,506	64,622	51,444	26,445	23,121	27,035	8,750	10,289	13,285
Sudan	3,084,320	3,056,400	3,985,702	4,332,827	4,215,308	5,062,000	4,065,460	4,595,092	6,347,143	8,562,205	9,867,778
Swaziland	36,664	14,863	19,799	45,581	30,420	4,410	23,754	38,875	..	..	..
Tanzania	10,712,526	7,489,890	..	..	423,967	30,504,654	1,131,655	4,969,273	2,438,040	7,976,590	8,777,340
Togo	..	..	431,826	398,103	412,619	368,472	366,672	352,334	..	328,488	561,328
Uganda	12,343,411	7,216,411	5,622,934	3,552,859	3,070,800	2,845,811	2,317,840	..	1,431,068	2,191,277	1,470,662
Zambia	..	..	2,010,185	1,139,489	2,992,203	3,399,630	..	3,215,866	2,742,118	3,514,000	3,514,000
Zimbabwe	..	1,252,668	1,609,296	1,533,960	1,804,479	1,719,960	1,849,383	1,696,192	761,791	324,188	877,734

Source: WHO Reported Malaria Cases

Table A4: Construction of Price Variables

Food Items	**Measurement**	**Pharmaceutical Items**	**Measurement**
Raw Cassava	Grams	Aspirin	Tablets
Dry Cassava	Grams	Paracetamol/Panadol	Tablets
Cassava Flour	Grams	Nivaquine/Cloroquine	Tablets
Maize Flour	Grams		
Sorghum	Grams		
Finger Millet	Grams		
White Bread	Grams		
Rice	Grams		
Sugar	Grams		
Sweet Potato	Grams		
Irish Potato	Grams		
Kidney Beans	Grams		
Salt	Grams		
Groundnuts	Grams		
Tomato	Grams		
Cooking Banana	Grams		
Sweet Banana	Grams		
Orange	Grams		
Cooking Oil	Grams		
Local Brew	Grams		
Tea Leaves	Grams		
Onions	Grams		
Chicken Eggs	Grams		
Chicken	Grams		
Beef	Grams		
Goat's Meat	Grams		
Fish	Grams		
Peas	Grams		
Fresh Milk	Grams		
Milk Powder	Grams		

Table A5: Summary Statistics – Total Work Sample

Kagera Work Characteristics (1991-2004)						
Variable	**Description**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Work Hours						
Hours	Number of hours worked per week	11,955	19.89	19.14	0	130.00
Employment Type						
Farm Employed	Self Employment in the Agricultural Sector	11,050	0.71	0.45	0	1.00
Self Employed	Self Employment in Non-Agricultural Sector	11,040	0.15	0.36	0	1.00
Wage Employed	Wage Employment in Private or Government Sector	11,042	0.18	0.38	0	1.00
Socio-demographic Characteristics						
Age	Age in Years: Range 18 - 65	11,955	33.94	13.52	18	65.00
Gender	Percent Male: 1 = Male; 0 Otherwise	11,955	0.45	0.50	0	1.00
Education	Highest Level of Education: 1 = Primary; 2 = Secondary; 3 = College	11,955	1.84	1.29	0	6.00
Married	Marital Status : 1 = Married; 0 Otherwise	11,955	0.56	0.50	0	1.00
Income	Annual Income: Employment income in dollars	11,955	189.93	409.61	0	7401.59
Health Status						
Malaria	Malaria Cases = 1 if reported malaria in at least 1 wave; 0 otherwise	11,955	0.15	0.36	0	1.00
Chronic Illness	Cases of Chronic Illness = 1 if reported chronic illness in at least 1 wave; 0 otherwise	11,955	0.10	0.30	0	1.00
Disease Environment						
Malaria Season	Rainfall Season: 1 = Interviewed during rainy season; 0 otherwise	11,955	0.81	0.39	0	1.00
Rainfall	Total Monthly Rainfall[a:] Total Monthly Rainfall measured in millimeters	10,204	341.27	280.68	1	872.30
Community Infrastructure						
Electric Power	Electric Power: 1= Access; 0 otherwise	12,009	0.41	0.49	0	1.00
Pipe Water	Pipe Borne Water: 1 = Access; 0 otherwise	12,009	0.20	0.40	0	1.00
Road	Motorable Road: 1 = Access; 0 otherwise	12,009	0.96	0.21	0	1.00
Prices[b]						
Food	Average Food Price in dollars	9,394	0.17	0.21	0.002	3.80
Pharmaceuticals	Average Pharmaceutical Price in dollars	7,707	0.04	0.12	0.001	0.70

[a]This variable is constructed using estimates from the Living Standard Measurement Survey for Tanzania (1991-994) and NOAA National Data Centers (2004); Food items are weighted to the nearest 50 grams and pharmaceutical items are measured in tablets

Table A6 First Stage 2SLS Results (with interaction terms) – Kagera Work Sample

	2SLS First Stage					
	Dependent Variable: Malaria		Dependent Variable: Malaria in Farm Employment		Dependent Variable: Malaria in Wage Employment	
1st Stage Explanatory Variables	**Coefficient**	**Std. Err.**	**Coefficient**	**Std. Err.**	**Coefficient**	**Std. Err.**
Farm Employment	-0.0796	0.0872	0.2497***	0.0751	0.0067	0.0334
Wage Employment	0.0015	0.1099	0.0325	0.0947	0.2946***	0.0421
Age in Years	0.0009***	0.0003	0.0006**	0.0003	0.0002	0.0001
Gender	-0.0034	0.0095	-0.0027	0.0082	0.0032	0.0036
Education	-0.0007	0.0037	-0.0032	0.0032	-0.0017	0.0014
Married	-0.0077	0.0091	-0.0029	0.0078	0.0035	0.0035
Income	-9.00E-06	0.0000	-5.10E-06	9.32E-06	6.88E-08*	4.14E-16
Chronic Ill	0.0214*	0.0128	0.0228**	0.0110	0.0069	0.0049
Electric Power	0.0053	0.0127	0.0111	0.0110	0.0017	0.0049
Pipe Water	-0.0094	0.0143	-0.0121	0.0123	0.0025	0.0055
Motorable Road	-0.0264	0.0288	-0.0255	0.0248	-0.0105	0.0110
Pharmaceutical Price	0.0940	0.0948	0.0073	0.0816	0.0960***	0.0363
Food Price	0.0503*	0.0291	0.0281	0.0211	-0.0017	0.0112
Instruments						
Rainfall Season	-0.1169	0.0741	0.0199	0.0638	0.0112	0.0284
Total Rainfall	-0.0395***	0.0138	0.0011	0.0119	0.0022	0.0053
Total Rainfall Squared	4.28E-07***	1.58E-07	-3.46E-08	1.36E-07	-6.66E-08	6.05E-08
Season*Farm Employment	0.0352	0.0909	-0.1015	0.0783	-0.0113	0.0348
Season*Wage Employment	-0.2905	0.1139	-0.0452	0.9810	-0.1548***	0.0436
Total Rainfall*Farm Employment	0.0173	0.0150	-0.0220*	0.0130	-0.0028	0.0058
Total Rainfall*Wage Employment	0.0089	0.0174	-0.0020	0.0149	-0.0213	0.0066
Total Rainfall Sq*Farm Employment	-1.71E-07	1.71E-07	2.69E-07	1.47E-07	8.47E-07	6.56E-08
Total Rainfall Sq*Wage Employment	-7.46E-08	1.94E-07	1.01E-07	1.67E-07	2.69E-07***	7.45E-08
Constant	0.3046	.0767	-0.0099	0.0661	-0.0075	0.0294
F-Statistic (P-value)	2.49 (0.0001)		11.06 (0.0000)		34.97 (0.0000)	
Joint Significance - IVs Only (P-value)	3.29 (0.0005)		2.38 (0.0109)		5.40 (0.0000)	
Overidentification Test - N*R-Sq (P-value) 11.168 (0.0833)						

Omitted group in first stage is Non-farm Self Employment group; Overidentification Test (N*R-Sq): Reject $H_{0;}$ IVs are exogenous

Appendix B: Figures

Figure B1: Estimated incidence of clinical malaria episodes – caused by any species – resulting from local transmission, country level Averages, (2004)

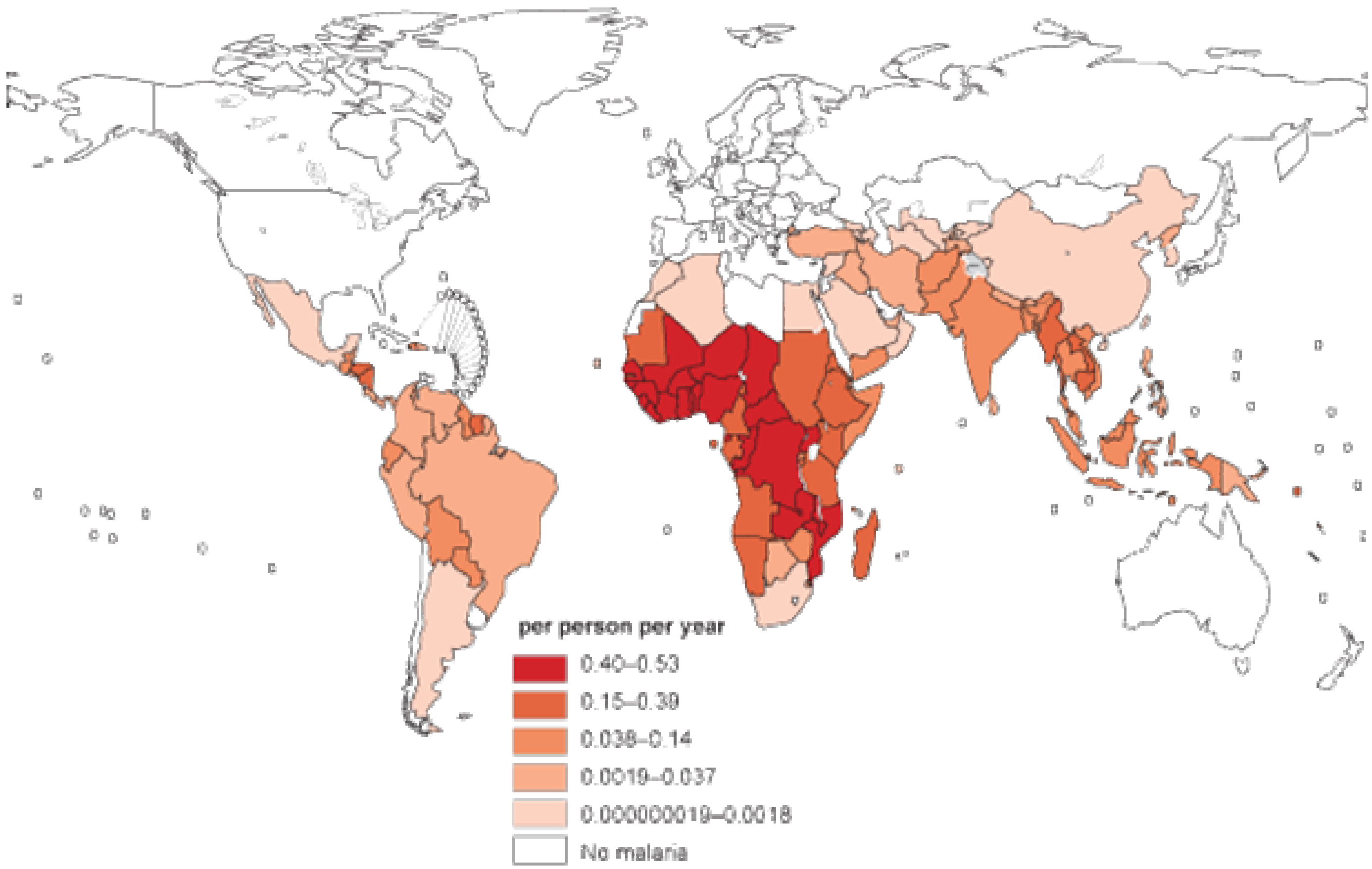


Figure B1 is derived from: WHO – World Malaria Report 2005

Figure B2: Rosenfield's Conceptual Framework for Tropical Illness

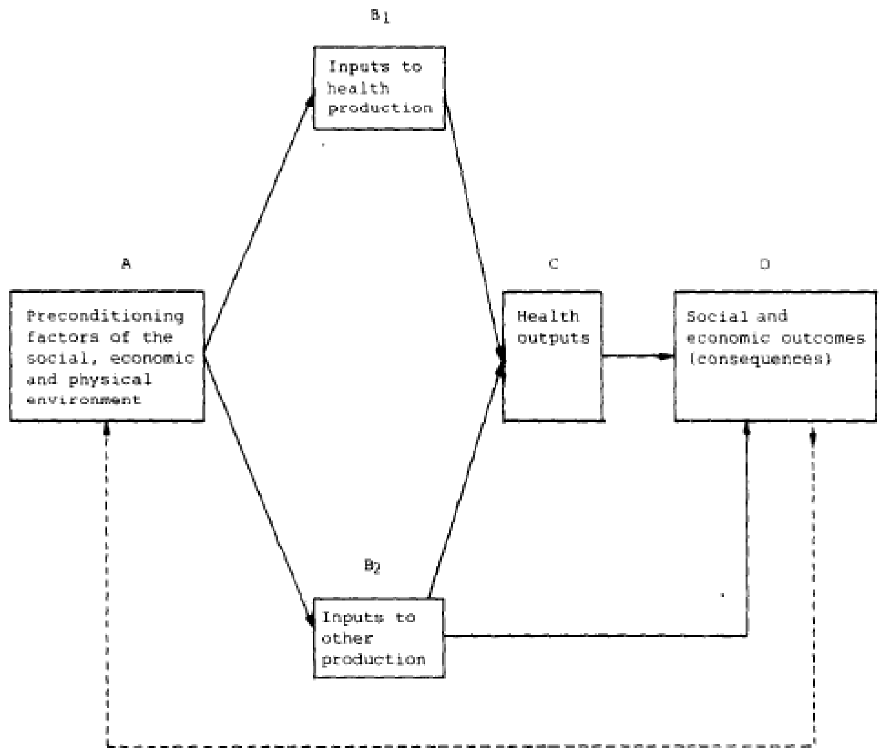
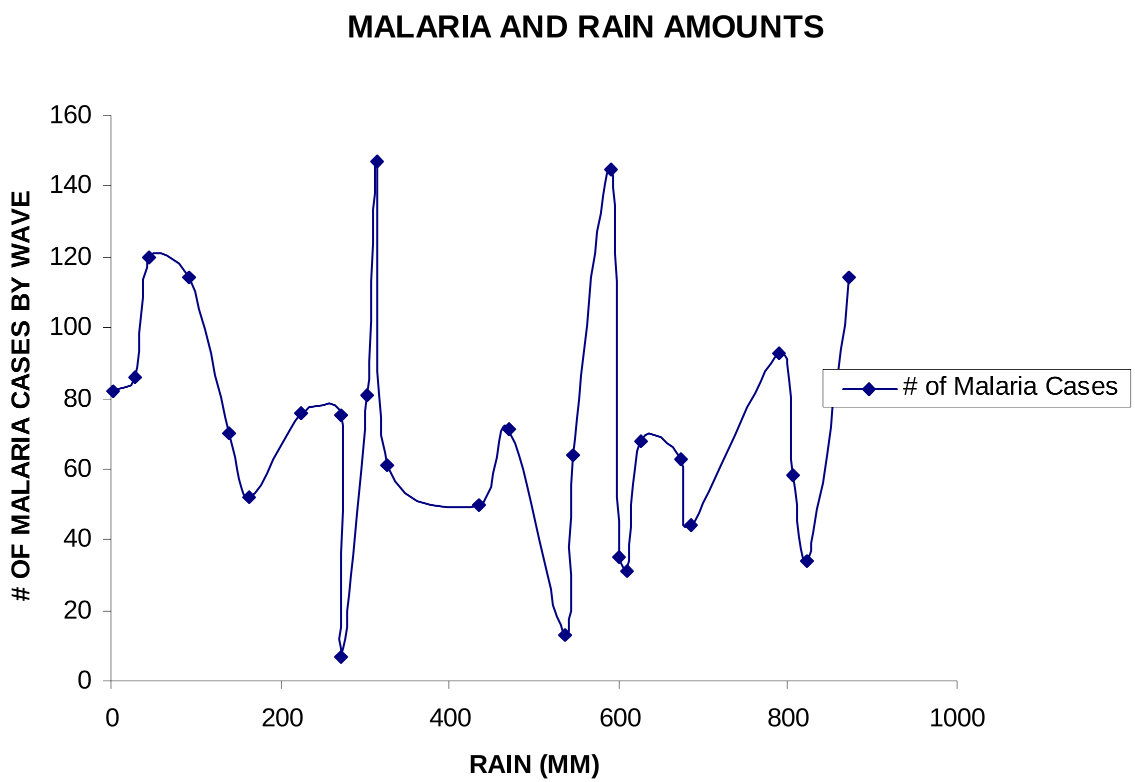


Fig. 1. Conceptual framework for assessing the social and economic consequences of the tropical diseases [1, p. 8].

Figure B2 is derived from: Rosenfield (1984)

Figure B3: Relationship between Episodes of Malaria and Rainfall Amounts



* Figure B3 shows monthly rainfall estimates plotted against malaria cases per month for the year 1991-1994 and 2004.

Appendix C: Two Period Household Health, Human Capital, and labor Supply Model

Health and Education Model

$$U^1(C_H^1, H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1), H_K^1(HI_K^1, L_K^1), L_K^1) + \beta U^2(C_H^2, H_H^2(HI_H^2, L_H^2), L_H^2) \quad (3.2.1')$$

$$P_c(C_H^1 + C_H^2) + P_{HI}(HI_{H-K}^1 + HI_K^1 + HI_H^2) + P_s S_K^1 = w_{H-K}(L_{H-K}^1) + V + w_K^1(L_K^1) + w_H^2(L_H^2, \tilde{S}_K) \quad (3.2.6)$$

$$H_{H-K}^1 = H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1; A, D, \mu_{H-K}^1, e_{H-K}^1) \quad (3.2.2)$$

$$H_K^1 = H_K^1(HI_K^1, L_K^1; A, D, \mu_K^1, e_K^1) \quad (3.2.3)$$

$$H_H^2 = H_H^2(HI_H^2, L_H^2; H_{H-K}^1, H_K^1, A, D, \mu_H^2, e_H^2) \quad (3.2.4)$$

$$\tilde{S}_K = \tilde{S}_K(S_K^1, H_K^1; A, SI, \mu_K, e_{S_K}) \quad (3.2.5)$$

$$w_{H-K}^1 = w_{H-K}(H_{H-K}, L_{H-K}^1; A, E, IN, \alpha, e_{w_{H-K}}) \quad (3.2.7)$$

$$w_K^1 = w_K^1(H_k^1, L_K^1, S_K^1; A, E, IN, \alpha, e_{w_K}^1) \quad (3.2.8)$$

$$w_H^2 = w_H^2(H_H^2, L_H^2, \tilde{S}_K; A, E, IN, \alpha, e_{w_K}^2) \quad (3.2.9)$$

Utility Maximization Problem: Period 1

$$\begin{aligned} L = \max U &= U^1(C_H^1, H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1), H_K^1(HI_K^1, L_K^1), L^1; A, E, \xi) \\ &+ \beta U^2(C_H^2, H_H^2(HI_H^2, L_H^2), L_H^2; A, E, \xi) \\ &+ \lambda[w_{H-K}^1(L_{H-K}^1) + V + w_K^1(L_K^1) + w_H^2(L_H^2, \tilde{S}_K) \\ &- P_c(C_H^1 + C_H^2) - P_{HI}(HI_{H-K}^1 + HI_K^1 + HI_H^2) - P_s S_K^1] \end{aligned}$$

$$C_H^1 : \frac{\partial U^1}{\partial C_H^1} - \lambda P_C = 0$$

$$HI_{H-K}^1 : \left(\frac{\partial U^1}{\partial H_{H-K}^1}\right)\left(\frac{\partial H_{H-K}^1}{\partial HI_{H-K}^1}\right) - \lambda P_{HI} = 0$$

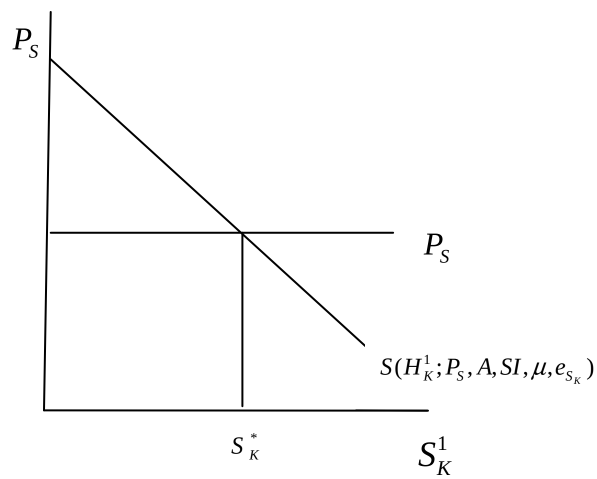
$$HI_K^1 : \left(\frac{\partial U^1}{\partial H_K^1}\right)\left(\frac{\partial H_K^1}{\partial HI_K^1}\right) - \lambda P_{HI} + \left(\frac{\partial w_H^2}{\partial \tilde{S}_K}\right)\left(\frac{\partial \tilde{S}_K}{\partial HI_K^1}\right)] = 0$$

$$S_K^1 : -\lambda P_S + \lambda \left(\frac{\partial w_K^2}{\partial \tilde{S}_K} \right) \left(\frac{\partial \tilde{S}_K}{\partial S_K^1} \right) = 0 \quad \text{Rewritten, } \left(\frac{\partial w_K^2}{\partial \tilde{S}_K} \right) \left(\frac{\partial \tilde{S}_K}{\partial S_K^1} \right) = P_S$$

Demand for Education in period one

$$S_K^1 = S(H_K^1; P_S, A, SI, \mu_K, e_{S_K}) \tag{3.3.1}$$

Figure C1



Health and Labor Productivity Model

Case 1: H* is endogenous; Individual chooses C, HI, and L

$$
\begin{aligned}
L = \max U &= U^1(C_H^1, H_{H-K}^1(HI_{H-K}^1, L_{H-K}^1), H_K^1(HI_K^1, L_K^1), L^1; A, E, \xi) \\
&+ \beta U^2(C_H^2, H_H^2(HI_H^2, L_H^2), L_H^2; A, E, \xi) \\
&+ \lambda[w_{H-K}^1(L_{H-K}^1) + V + w_K^1(L_K^1) + w_H^2(L_H^2 + \tilde{S}_K) \\
&- P_c(C_H^1 + C_H^2) - P_{HI}(HI_{H-K}^1 + HI_K^1 + HI_H^2) - P_s S_K^1]
\end{aligned}
$$

$$C_H^2 : \frac{\partial U^2}{\partial C_H^2} - \lambda P_C = 0$$

$$HI_H^2 : \left(\frac{\partial U^2}{\partial H_H^2}\right)\left(\frac{\partial H_H^2}{\partial HI_H^2}\right) - \lambda P_{HI} = 0$$

$$L_H^2 : \left(\frac{\partial U^2}{\partial L_H^2}\right)\left(\frac{\partial H_H^2}{\partial L_H^2}\right) + \lambda w_H^2 = 0$$

Reduced Form Equations for C, HI, and L

$$C = C(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.1}$$

$$HI = HI(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.2}$$

$$L_H^2 = L(P_C, P_{HI}, V, A, E, D, IN, \mu, \alpha) \tag{3.4.3}$$

Case 2: H is given; Individual chooses C, and L

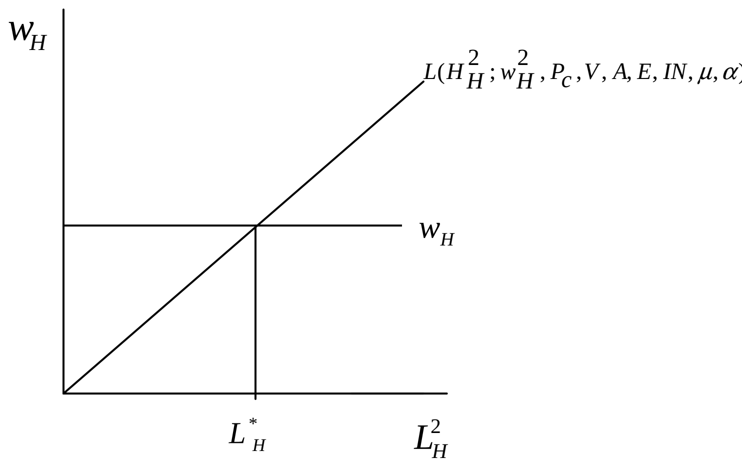
$$C_H^2 : \frac{\partial U^2}{\partial C_H^2} - \lambda P_C = 0$$

$$L_H^2 : \left(\frac{\partial U^2}{\partial L_H^2}\right) + \lambda w_H^2 H_H^2 = 0$$

Structural Equation for L

$$L_H^2 = L(H_H^2; w_H^2, P_C, V, A, E, IN, \mu, \alpha) \tag{3.4.4}$$

<u>Figure C2</u>



Dictionary of Variables in Theoretical Framework

A = Socio-demographic Characteristics (age, gender, and marital status)

C_H^1 = Household Consumption of Non-Health Goods in Period One

C_H^2 = Household Consumption of Non-Health Goods in Period Two

D = Disease Environment

E = Education

H_{H-K}^1 = Household Health less that of the Child in Period One

H_K^1 = Child Health in Period One

H_H^2 = Overall Household Health in Period Two

HI = Health Inputs

HI_{H-K}^1 = Household Health Inputs less that of the Child in Period 1

HI_K^1 = Child Health Inputs in Period 1

HI_H^2 = Overall Household Health Inputs in Period 2

IN = Community Infrastructure

L^1 = Labor Supply in Period One

L^2 = Labor Supply in Period Two

L_{H-K}^1 = Labor Supply (Household minus Child) in period one

L_K^1 = Labor Supply (Child) in period one

L_H^2 = Household Labor Supply in period two

P_{HI} = Price of Health Inputs

P_S = School Fees

S_K^1 = Child Schooling in Period 1

$\tilde{S}_K$ = Accumulated Human Capital by Period 2

SI = School Infrastructure

V = Non-Labor Income

w_{H-K}^1 = earnings (Household minus Child) period one

w_K^1 = earnings (Child) period one

w_H^2 = Household earnings period two

α, ξ, μ, e = error terms

Appendix D: Analysis of Malaria, Labor Supply and Schooling using Average Treatment Effects on the Treated (ATET)

OLS, FE and 2SLS, and LIML estimates of the causal effect of health status on schooling and labor supply outcomes were presented in the previous sections. This section explores a different estimation procedure for further identification of the parameters in (1) of chapter four. The approach chosen for further identification of the schooling and labor supply equations is based on the method of average treatment effects on the treated (ATET) with instrumental variables. Unlike the estimation procedure detailed in section 4.4.1 of chapter four, which ignores the different states of treatment within the sample, estimating ATET considers different states of the health status indicator.

In general, estimating equation (1) of chapter four by 2SLS assumes that the predicted person specific gain from treatment is zero given a vector of observables. This is equivalent to the ATET case of estimating equation (3) below:

$$E(y \mid h, X) = \mu_0 + \alpha h + g_0(X) \qquad (3)$$

Where h is the endogenous binary treatment indicator with $h = 1$ denoting treatment and $h = 0$ otherwise, X denotes a vector of observable variables, and $g_0(X)$ is the expected value of the person specific effects of treatment conditional on the observables. In previous chapters, it was assumed $E(v_1 - v_0 \mid X) = 0$; where v_1 and v_0 are the person specific effects of treatment. If this assumption holds true, the gain from treatment is the same for everyone in the treatment group in (3) and estimating the effect of health status on the schooling and labor supply models by the standard 2SLS procedure is sufficient.

If the $E(v_1 - v_0 \mid X) \neq 0$, the effect of malaria is not the same for everyone in the sample and 2SLS is insufficient. Clearly, making such a restrictive assumption as $E(v_1 - v_0 \mid X) = 0$ is highly unrealistic in this case, as one may expect that the effect of malaria on school and work hours vary from one individual to another. In other words, to allow for varying effects of malaria on the school and work outcomes, the assumption of zero person-specific gains from treatment is relaxed to derive the ATET estimator:

$$E(y \mid h, X) = \mu_0 + \alpha h + h[g_1(X) - g_0(X)] \qquad (4)$$

Where $g_0(X) - g_1(X)$ measures the average gain of treatment to a participant with characteristics X. To identify the parameters in (4), two assumptions are required about treatment assignment. The first assumption is referred to as conditional independence or selection on observables or unconfoundedness.

Assumption (1a): Unconfoundedness or Conditional Independence

$$(y_0, y_1) \perp h \mid X$$

Conditional independence states that conditional on X, the outcomes are independent of treatment[15]. A weaker conditional independence assumption can be specified if one is only interested in the treatment effect on the treated.

Assumption (1b): Weak Conditional Independence

$$y_0 \perp h \mid X$$

Assumption (1b) implies that there is no omitted variable bias once X is controlled for. Alternatively, this implies that malaria status in our sample is exogenous. Assumption (1b) is not sufficient in our case knowing the likely endogeneity of malaria

---

[15] y_1 and y_0 are the outcomes for the treated and non-treated groups, respectively.

status; therefore, the analysis relies on assumption (1a). Assumption (1a) is not directly testable however; there are indirect ways to assess it. The test of conditional independence relies on estimating the causal effect that is known to equal zero. If the null hypothesis is rejected, then the likelihood of conditional independence is less plausible (Imbens and Wooldridge 2007). In addition to (1a), equation (4) also relies on the overlap or matching assumption.

Assumption (2): Overlap

$0 < \Pr[h = 1 \mid X] < 1$

Under assumption (2), there exist treated and non-treated cases for each value of X , in the sense that there is overlap between the treated and untreated sub-samples. Although, this assumption is not required to identify the treatment parameters for the treated group, it is needed for identifying the treatment parameter for a randomly selected individual within the sample. The lack of overlap can be detected by plotting the distribution of the covariates by treatment groups. Because this becomes cumbersome in cases with more than two covariates, a direct method of determining overlap is to examine the distribution of propensity score in both treatment groups.

ATET Method: Propensity Score Matching Methods

Matching by propensity score is recommended as a way to control for confounding factors that may bias the effect of treatment in observational studies (Sascha Becker and Andrea Ichino 2002)[16]. Propensity score matching estimates are superior to

[16] According to Becker and Ichino, "propensity score matching method is a way to correct the estimation of treatment effects controlling for the existence of these confounding factors based on the idea that the bias is reduced when the

other matching estimators when the dimensionality of observable characteristics (X) is high (Rajeev Dehejia and Sadek Wahba 2002). When X is of a small dimension, matching is straightforward and non-propensity score matching methods are sufficient for deriving unbiased estimates of the treatment effect. This study, explores 4 matching methods namely, stratification matching, nearest-neighbor matching, radius matching, and kernel matching methods, all of which can be defined in terms of propensity scores.

The propensity score $p(X)$ is the probability of an individual receiving treatment conditional on observables, X. If $0 < p(X) < 1$ and $p(X)$ is a consistent estimate of $p(X)$, then a consistent estimator of the ATET can be obtained using a parametric probit or logit model as:

$$\overbrace{ATET} = \left(N^{-1} \sum_{i=1}^{N} h_i \right)^{-1} \left\{ N^{-1} \sum_{i=1}^{N} [h_i - \hat{p}(x_i)] Y_i / (1 - \hat{p}(x_i)) \right\}$$

In the analyses to follow, consistent $\overbrace{ATET}$ estimates are obtained using 4 propensity score matching methods with the logit model.

Matching on propensity score is based on pairing comparable treated and control units in terms of observable characteristics, which are assumed independent of treatment. Due to the unobserved heterogeneity problem with the health status indicator, malaria, an instrumental variables approach to ATET will be used to mitigate this problem. This is arrived at in the estimation process by adding the instruments to the list of covariates when estimating the propensity scores which will be used in calculating the treatment effects.

comparison of outcomes is performed using treated and control subjects who are as similar as possible" (Becker and Ichino 2002).

The stratification, nearest neighbor, radius, and kernel matching methods are the four most widely used propensity score matching methods. The reason for using all four methods in this study is mainly for comparison. Each one of these propensity score matching methods differ by the process of selection of the comparison units (matches) and selection of one method over another boils down to a tradeoff between the quality and quantity of matches (Becker and Ichino 2002).

The stratification matching method provides ATET estimates which are based on dividing the range of variation of the propensity scores into intervals (blocks) such that within each block, treated and untreated units have on average the same propensity score. After matching, the ATET estimates are then obtained as an average of the ATET of each block with weights given by the distribution of the treated units across blocks. One disadvantage of this method is that it ignores observations in blocks where either treated or untreated units are missing. On the other hand, the nearest neighbor method makes up for this deficiency by taking each treated unit and locating the untreated unit with the closest propensity score as a match (i.e. the nearest neighbor to each treated unit). After matching is complete, the ATET estimates are derived as the average of the difference between the outcome of the treated units and that of the control/untreated units. While this method ensures that all treated units find a match, it does not necessarily improve the quality of the matches for those, which would have been disregarded using the stratification method.

The radius method offers a solution to the above deficiency. The radius matching method pairs the treated units only with the untreated units whose propensity score fall in a predefined radius (neighborhood) of those of the treated units. The downside to this method lies in the specified radius in that, if the radius is set too small, it is possible that no matches exits in the neighborhood because the neighborhood contains no untreated units. Alternatively, specifying a small radius also improves the quality of the matches. Another matching approach is the kernel method, which is based on matching treatment units using a weighted average of all control/untreated units with weights that are proportional to the distance between the propensity scores of the treated and untreated units. All ATET estimates for the schooling and labor supply equations are presented below.

ATET Assumptions

Although the ATET assumption of unconfoundedness is not directly testable, Imbens and Wooldridge (2007) propose an indirect test (described under assumption (1)). When the indirect test is applied to the labor supply and schooling model, neither model passes the conditional mean independence test. In addition to the test of conditional independence, Imbens and Wooldridge (2007) suggest a direct way of assessing overlap in ATET estimations (described under assumption (2)). Two histogram plots of the propensity scores for the labor supply and schooling models are used to access the overlap assumption (Figure D1). The histogram plots show considerable evidence of overlap in the distribution of the propensity scores for both samples.

ATET Estimation Results – Comparison of Means (School Sample)

Table D1 presents the sample means for the two comparison groups of the Kagera schooling sample. Of the 4,951 observations from the Kagera school sample, 549 fall into the treatment group for the ATET analyses. A comparison of means by the number of school hours attended per week suggests that those without treatment attend on average 3 hours more school per week than those with treatment (i.e. children reporting an episode of malaria). However, children in the treated group were of similar age to those in the control group. In terms of the gender composition of both groups, females were just as likely to be in the treated group as males. Surprisingly, parents of children in the treatment category are more likely to be educated than those without treatment even though those in treatment are more likely to come from households with lower income. Furthermore, the likelihood of falling into treatment does not seem to vary by the pre-existence of chronic illness since fewer children in the treated group have a chronic illness.

Table D1: ATET Comparison of Means – Kagera School Sample

Sample Means of Characteristics for the Kagera School Sample (N = 4,189)										
	N	Age	Gender	Mother Education	Father Education	Household Income	Chronic Illness	Malaria Season	Total Rainfall	School Hours
Treated Group										
Mean	549	12.51	0.5	0.27	0.29	157.48	0.02	0.32	349.66	25.58
Std. Dev.		12.51	0.50	0.44	0.45	355.13	0.14	0.47	294.79	13.97
Control Group										
Mean	4402	12.52	0.53	0.21	0.26	171.31	0.03	0.22	381.6	28.26
Std. Dev.		3.00	0.50	0.41	0.44	452.02	0.17	0.42	287.74	12.76

Note: School hours = number of hours of school attended in a week; Age = age in years; Gender = 1 if male, 0 otherwise; Mother/Father education = if mother/father has at least a primary school education, 0 otherwise; Household Income = income in dollars; Chronic illness = 1 if reported at least one chronic illness, 0 otherwise; Malaria season = 1 if month of interview falls within malaria season, 0 otherwise; Total Rainfall = rain amounts in mm.

ATET Estimation Results – Treatment Effects (School Sample)

Four types of matching estimators in the estimation of the treatment effects were explored in the ATET analyses. All estimates of the treatment effect use rainfall amounts and a rainfall season indicator as instruments. The results of all four matching estimators are presented on Table D2 for the purpose of comparison. The radius matching method by definition matches treated units only with control units whose propensity score fall in a predefined radius of those of the treated units. The analyses using radius matching first compares neighbors within a (0.001) radius; then, the radius is gradually expanded for a more realistic comparison of both groups given the tight range of the estimated propensity scores (Table D2)[17].

Radius matching estimates of the treatment effect range between 2.397 and 3.865 school hours per week lost due to malaria illness. Using the nearest neighbor, stratification, and kernel matching methods all generate the same treated effects as the radius method. All estimates of the treatment effect are statistically significant but far smaller in magnitude than the 2SLS estimates derived in section 5.6.2. The estimates produced by the propensity score matching methods are however very close to those derived with the OLS and FE estimators. Findings of treatment effects similar to OLS are not uncommon in practice. Picone, et al. (2006) estimated treatment effects on

[17] Note that the tradeoff of expanding the radius to match pairs with very close propensity scores is a decrease in the number of possible matches. As shown in Table D2, the number of close matches falls drastically when a smaller radius is defined.

individuals with Acute Myocardial Infraction (AMI) using Medicare claim data and found no advantages over using a standard least squares regression.

Table D2: ATET Results – Kagera School Sample

Malaria Impact: Estimates of treatment effect				
Propensity Score Range [.04736965, .40351615]				
Matching Method	Number Treated	Number in Control	ATET	Standard Error
Radius				
r = 0.001	524	4171	-2.397	0.650[a]
r = 0.0001	507	3140	2.793	0.673[a]
r = 0.00001	331	609	-3.865	0.932[a]
Nearest Neighbor	525	485	-2.811	0.866[a]
Stratification	461	3715	-2.706	0.692[a]
Kernel	525	4231	-2.639	0.617[a]

a Analytical standard

b Bootstrapped standard errors with 250 replications

5.7.4 ATET Estimation Results – Comparison of Means (Work Sample)

A comparison of means by work category is presented in Table D3. Comparison of means for the farm employment category yielded 1208 treated units and 6637 control units. In terms of non-farm self-employment, 275 workers fall into the treated group while, 1369 make up the control group. On the other hand, wage employment generated 316 treated units and 1668 control units. Although, non-farm self-employment had the least number of workers in both treated and control groups this employment category also had the highest percentage of workers in treatment as a percentage of total workers in each group. Seventeen percent of non-farm self-employed workers reported having malaria compared to 15% of self-employed farmers and 16% of workers in wage employment.

As shown on Table D3, the largest difference in hours worked per week between workers in the treated group and those in the control group is also for the non-farm self employment category. Workers in the treated group for non-farm self employment worked 10 fewer hours per week than those in the treated farm employment group and 23 hours less per week than those in treated wage employment category. In terms of wages, treated workers in self-employment earn less income than their other treated counterparts. In particular, treated units in wage employment earn on average 65% more income than treated units in self-employment while, treated units in farm employment earn 7% more than treated units in self-employment.

Also, a majority of treated units in farm employment are more likely to be female and less educated than the treatment units in other employment classifications. In terms of pre-existing health conditions, treated and untreated farmers are more likely to have an existing chronic condition than workers in other groups. In addition, the treated units are in every work classification on average more likely to be chronically ill compared to workers in the control units.

On the contrary, comparisons of means for the rainfall and malaria season variables yielded counter-intuitive results of both treated and control units for each work sample. For all three classification of employment, rainfall amounts for the survey period are greater for workers in the treated units compared to those in the control groups. Likewise, those in treatment are less likely to report being in malaria season compared to those in the control group. This calls to question the validity of the two instruments in the ATET estimations to follow. However, as shown earlier in figure B3, the overall data suggests a non-linear relationship between episodes of malaria by wave and rainfall amounts, which may in part explain the observation on Table D3.

ATET Estimation Results – Treatment Effects (Work Sample)

Panels A - D of Table D4 present the treatment effects for the dependent variable 'work hours' using the radius, nearest neighbor, stratification, and kernel matching methods. As indicated earlier, the farm employment group is the largest of the 3 work categories. The analysis begins with the treatment effect with the farm employment work group (Panel A; Table D4).

According to the ATET estimates using 4 types of matching methods, self employed farmers in Kagera who reported being sick during the sample period lost between 1.712 and 2.640 hours of weekly work hours. The lowest treatment effect for this work group was derived by the 0.001 radius matching method. Although, the 0.00001 radius generated the largest treatment effect (-2.640 hours), it also paired the least number of workers in both treatment and control groups (Panel A; Table D4). All estimates of the treatment effects for the farm employment sample are statistically significant at the 1% level.

Table D3: ATET Comparison of Means – Kagera Work Sample

Sample Means of Characteristics for the Kagera Work Sample															
		Age	**Sex**	**Education**	**Marital Status**	**Income**	**Chronic Illness**	**Electric Power**	**Pipe Borne Water**	**Motorable Road**	**Drug Prices**	**Food Prices**	**Malaria Season**	**Total Rainfall (mm)**	**Work Hours**
Self Employment (Farm)															
Treated Group (N = 1208)	Mean	35.52	0.38	1.64	0.63	157.27	0.12	0.41	0.18	0.94	0.03	0.19	0.79	322.85	18.60
	Std. Dev.	13.74	0.49	1.23	0.48	335.45	0.32	0.49	0.38	0.23	0.10	0.26	0.41	284.36	12.02
Control Group (N = 6637)	Mean	34.93	0.42	1.93	0.57	187.54	0.10	0.37	0.17	0.95	0.03	0.16	0.90	364.53	19.79
	Std. Dev.	13.97	0.49	1.30	0.49	384.93	0.31	0.48	0.38	0.21	0.08	0.21	0.30	276.68	12.27
Self Employment (Non-Farm)															
Treated Group (N = 275)	Mean	34.01	0.57	1.80	0.61	146.78	0.11	0.61	0.24	0.96	0.04	0.23	0.70	261.06	8.30
	Std. Dev.	12.38	0.50	1.08	0.49	419.07	0.32	0.49	0.43	0.20	0.12	0.31	0.46	269.47	14.24
Control Group (N = 1369)	Mean	33.37	0.60	2.06	0.58	244.70	0.09	0.62	0.20	0.97	0.04	0.19	0.84	319.68	12.49
	Std. Dev.	11.93	0.49	1.16	0.49	634.92	0.29	0.49	0.40	0.18	0.11	0.27	0.37	275.04	18.10
Wage Employment															
Treated Group (N = 316)	Mean	32.66	0.68	1.81	0.55	423.78	0.09	0.57	0.27	0.95	0.04	0.21	0.72	290.57	31.25
	Std. Dev.	11.20	0.47	1.21	0.50	610.39	0.28	0.50	0.44	0.22	0.10	0.24	0.45	279.74	23.01
Control Group (N = 1668)	Mean	32.36	0.73	2.18	0.50	360.78	0.07	0.56	0.26	0.96	0.03	0.19	0.86	343.11	33.01
	Std. Dev.	11.72	0.44	1.28	0.50	568.36	0.26	0.50	0.44	0.19	0.10	0.27	0.35	280.68	22.68

Note: work hours = number of hours worked per week. Age = age in years; Gender = 1 if male, 0 otherwise; Education = 1 if worker has at least a primary school education; 0 otherwise; Marital status =1 if worker is married; 0 otherwise. Income = income in dollars; Chronic illness = 1 if reported at least one chronic illness; 0 otherwise Electrical power (this is a household level indicator) = 1 if worker resides in a residence with electrical power; 0 otherwise. Pipe Borne water (this is a community level indicator) = 1 if worker lives in area with access to pipe borne water; 0 otherwise. Motorable road (this is a community level indicator) = 1 if worker lives in area with motorable road; 0 otherwise. Pharmaceutical prices = average price of selected pharmaceutical products priced at the community level Food prices = average price of selected food items priced at the community level Malaria season = 1 if month of interview falls within malaria season, 0 otherwise; Total Rainfall = rain amounts in mm.

Panel C details the results of the ATET estimations for the wage employment sample. The results of in Table D4 show that workers in wage employment lost -11.05 hours weekly work hours to malaria during the survey period. This estimate is significant at the 1% level using the 0.0001 radius matching method. No other treatment estimates of radius matching method or other matching algorithms are significant for the wage employment sample and will not be discussed.

Results for the total sample using all four matching methods are presented in panel D. For the entire work sample, a 0.001 radius matching and kernel matching methods produce statistically significant estimates of the ATET at the 5% level between -1.052 and -1.537 hours. Similar results were generated using a 0.0001 radius matching, which was significant at the 10% level.

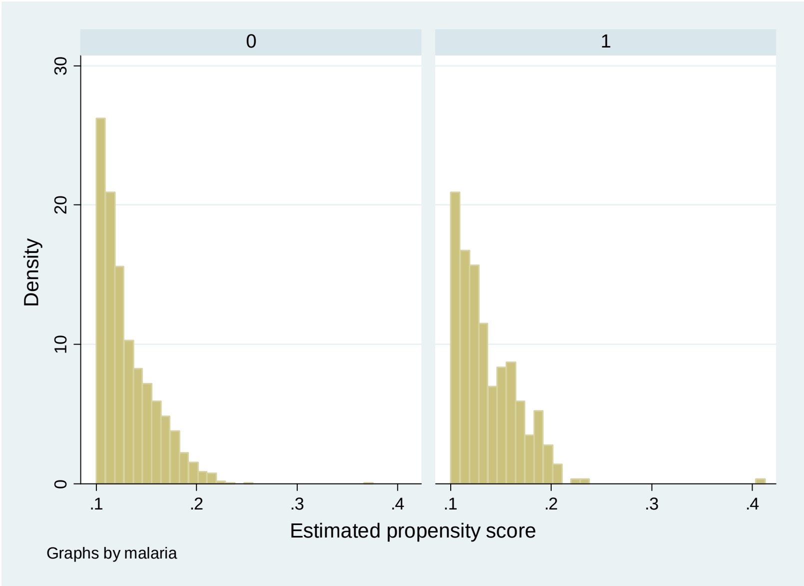
As seen previously with the ATET results for the school sample, propensity score estimates may produce results parallel to OLS. With the exception of the ATET results derived for the self-employment and wage employment samples the estimated treatment effects mirror the causal effects generated using standard OLS. In addition, propensity score methods appear to mirror OLS estimates the closer the range of the propensity score estimates (Table D4).

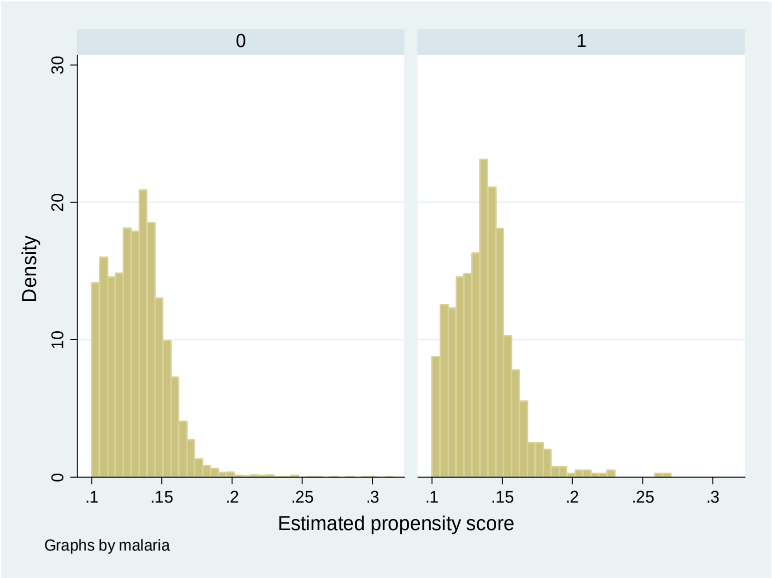
Table D4: ATET Results – Kagera Work Sample

Panel A - Self Employment (Farm) Malaria Impact: Estimates of treatment effect				
Propensity Score Range [.06469756, .44094733]				
Matching Method	Number Treated	Number in Control	ATET	Standard Error
Radius				
r = 0.001	561	3495	-1.712	0.549[a]
r = 0.0001	536	2468	-2.029	0.586[a]
r = 0.00001	310	442	-2.640	0.882[a]
Nearest Neighbor	564	505	-2.246	0.73[a]
Stratification	510	3445	-1.860	0.562[a]
Kernel	564	3558	-1.785	0.532[b]
Panel B - Self Employment (Non-Farm) Malaria Impact: Estimates of treatment effect				
Propensity Score Range [.0134235, .78899154]				
Matching Method	Number Treated	Number in Control	ATET	Standard Error
Radius				
r = 0.001	58	142	-6.178	2.724[a]
r = 0.0001	15	15	-8.767	8.034[a]
r = 0.00001	2	2	11.75	11.750[a]
Nearest Neighbor	78	65	-4.513	3.485[a]
Stratification	70	318	-3.925	2.180[a]
Kernel	78	347	-4.814	2.119[b]
Panel C- Wage Employment Malaria Impact: Estimates of treatment effect				
Propensity Score Range [.05082005, .56096773]				
Matching Method	Number Treated	Number in Control	ATET	Standard Error
Radius				
r = 0.001	86	397	-2.763	2.739[a]
r = 0.0001	38	53	-11.051	3.709[a]
r = 0.00001	5	5	-6.400	10.668[a]
Nearest Neighbor	98	84	1.250	3.793[a]
Stratification	79	529	-4.029	2.823[a]
Kernel	98	580	-1.131	2.543[b]
Panel D - Malaria Impact (Total): Estimates of treatment effect				
Propensity Score Range [.06072613, .22507454]				
Matching Method	Number Treated	Number in Control	ATET	Standard Error
Radius				
r = 0.001	842	5360	-1.537	0.694[a]
r = 0.0001	821	4554	-1.318	0.713[a]
r = 0.00001	641	1240	0.650	0.923[a]
Nearest Neighbor	843	757	0.723	0.944[a]
Stratification	843	5428	-1.052	0.682[a]
Kernel	843	5428	-1.425	0.630[b]
a Analytical standard				
b Bootstrapped standard errors with 250 replications				

Figure D1: ATET Assumptions – Overlap

Test of Overlap – Total Kagera School Sample Test of Overlap – Total Kagera Work Sample





About the Author

Born in Nigeria, Taiwo Abimbola relocated to the United States of America in 1998 to obtain her undergraduate degree from Lincoln University of Pennsylvania. She graduated summa cum laude with a B.A. in Economics from Lincoln University in 2002. In 2004, she earned an M.A. in Economics at the University of South Florida (USF). Thereafter, Taiwo was accepted into the PhD program at the USF College of Business, where her main areas of research were Health Economics and Development Economics. While attending USF, Ms. Abimbola was an instructor for both Economics of Health Care and Principles of Microeconomics courses.