An analysis of factors determining malaria incidence in India

with particular reference to Uttar Pradesh

Lubina F Qureshy

Doctor of Philosophy

(Economics)

University of Sussex

2010

UNIVERSITY OF SUSSEX

Lubina F Qureshy

Degree of Doctor of Philosophy

An analysis of factors determining malaria incidence in India with particular reference to Uttar Pradesh

Summary

This thesis identifies, *inter alia*, the socio-economic factors that affect malaria incidence at both the household and district levels and investigates how these differ across rural and urban settlement-types. In addition, state level data for India are used to examine the effect of aggregate income relative to that of public health expenditure on malaria incidence. The household and district-level analysis focuses on the state of Uttar Pradesh and exploits the National Family Health Survey, which is the Demographic Health Survey (DHS) for India, for two time periods - 1992-93 and 1998-99 - and combines these data with the district-level census data for 1991 and 2001.

A key theme of the micro-level analyses is whether household wealth exerts a negative impact on malaria incidence. Wealth is measured using the DHS data by constructing a consumer durable asset-index by Principal Components Analysis and malaria incidence was modelled using a probability model. The household-level analysis reveals that the relationship between socio-economic status and malaria incidence is not always negative. For example, owning a water pump, indicative of a higher socio-economic status, has a positive impact on malaria incidence and being of a lower caste has a negative impact. Variables that support the negative socio-economic status and health relationship include having an electricity connection in the house, having access to a protected public drinking water supply rather than an open source, and living farther away from open water sources.

The aggregate (or panel data) analysis was undertaken using data for 15 states in India covering the time period 1978 to 2000. The aggregate analysis reveals that income has a

negative impact on malaria incidence but direct expenditure on health is more effective in bringing about a decline in malaria incidence—an increase of a rupee in aggregate income per person reduces malaria incidence by 0.1 percent whereas an equivalent increase in real health expenditure per capita results in a 0.4 percent decline in malaria incidence.

The research undertaken for this thesis is unique in using the DHS to identify the factors affecting malaria incidence and shows that these data are very useful in exploring the relationship between malaria incidence and a host of socio-economic factors in order to identify areas for effective policy intervention. Such a holistic approach is critical in controlling and, eventually, eradicating malaria rather than relying primarily on more direct treatment strategies based on insecticide-treated bed nets and drug therapy. The areas where public spending could be directed to attack malaria identified by the empirical analysis include education, particularly raising awareness on prophylactic measures through adult literacy centres, controlling the breeding of mosquitoes in open water collection sites such as public taps and around water pumps and improving water flow in agricultural fields to prevent stagnant water collection.

Acknowledgements

DPhils are meant to be long journeys, but this one seems to have taken longer than most. Needless to say, I had the support of numerous individuals in helping me through this thesis. Above all, I am very grateful to my supervisors, Dr Barry Reilly, Dr Julie Litchfield and Prof Mike Sumner. I am indebted to Dr Reilly for his impeccable mentoring, endless patience, and refreshing sense of humour. He is no doubt the best teacher I have ever had and is a wonderful person. He patiently guided me through numerous drafts, enhanced my econometric skills, provided very good comments, and took the time to discuss the smallest detail even on the phone once I was based in Indonesia. Dr Litchfield gave me a lot of encouragement at the beginning of the DPhil and with enrolling for the programme. She continued to be warm, friendly, and approachable and gave very useful comments. She also helped me financially by finding bits and pieces of work for me at the Poverty Research Unit and in teaching. Prof Sumner, along with Dr Litchfield, saw me through the research proposal phase before he retired. I benefited from excellent comments and the many discussions we had at this stage. I cannot thank my supervisors enough for their support when I had a prolonged back problem and could not work on the DPhil for a long time.

I thank Dr Howard White for his comments on the rural analysis of this thesis. I thank the Richard Stapley Trust for giving me small grants for two consecutive years. I am very grateful to Dr Sonia Bhalotra for her rather quick response to my emails and guiding me to the source for the panel data I have used in this work. I must also mention the fantastic collection of the National Medical Library at Bethesda. It is perhaps the best organised library I have been to. Dr Rastogi at the WHO, New Delhi office, also provided me very promptly with the malaria incidence data I needed. I also thank my friend Asim for sending me all the census data I needed for this DPhil via email just a few hours after I requested him for it and Ritu for bringing over some of the data I needed on her visit to the US. Thanks Leonardo for introducing me to Latex and giving me several lessons, making it very easy to compile the thesis. I also thank Rafael for helping me with one of Stata's programmes. I am grateful for the calm company of one of my best friends, Yoko, in the initial years at Sussex and then in Washington DC, for her encouragement, and for her comments. My friends, Kaushik, Chiranjeev, and Mari also commented on some parts and on rather short notice. Thanks also to Yamini, who found the first piece of literature on socio-economic factors affecting malaria for me in one of Delhi's medical libraries.

My mother, my sisters, Asbah and Yasmin, and my brother, Aamir, gave me constant support and encouragement. It would not have been possible for me to enroll for the programme without the financial support of my brother who also continued to support me later. Thank you Ana for your encouragement and Meenakshi, Mudita, Kirsty, Tashi, Annika, and Elaine for your support. Yunib, Shalu, the three little Zs, my Aunt and Uncle, Pedro, Caitlin, Pooja, Jakob, Laia, Estrella, and Gonzalo accommodated me in my many visits to England. Thanks so much for your hospitality.

This note would be incomplete without mentioning TERI (The Energy and Resources Institute) where I spent six years of my work life before enrolling for the DPhil at Sussex. I owe my interest in development and in pursuing the DPhil to the work I did there. At TERI I also found some of my best and lasting friends, many of whom are mentioned here.

Thanks, Kiko, for bearing with this seemingly never-ending piece of work on a day to day basis, commenting on various parts of the thesis, supporting me, and motivating me. And thank you Minah for your help with my children, Izan and Alba.

I dedicate this thesis to my father who believed in education for the sake of education and whose strong sense of ideals and a principled, courageous stand on numerous subjects has left an indelible mark on us.

Contents

List of Tables

List of Figures

Abbreviations

ABER	Annual Blood Examination Rate
API	Annual Parasite Incidence
CPIAL	Consumer Price Index for Agricultural Labourers
CPIIW	Consumer Price Index for Industrial Workers
DDT	Dichlorodiphenyltrichloroethane
DHS	Demographic Health Survey
EMCP	Enhanced Malaria Control Project
FSM	The Federated States of Malaya
GDP	Gross Domestic Product
G6PD	Glucose-6-phosphate dehydrogenase
IRS	Indoor Residual Spraying
ITNs	Insecticide-treated bed nets
MPW	Multipurpose health worker
NFHS	National Family Health Survey
NMCP	National Malaria Control Programme
NMEP	National Malaria Eradication Programme
NSDP	Net State Domestic Product
NSS	National Sample Survey
NTPC	National Thermal Power Corporation
NVBDCP	National Vector Borne Diseases Control Programme
OBC	Other 'backward' castes
PSU	Primary Sampling Unit
RBCs	Red Blood Cells
RBM	Roll Back Malaria
SC	Scheduled caste
ST	Scheduled tribe
SEAR	Southeast Asia Regional Office
SPR	Slide Positivity Rate
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
WHO	World Health Organization

Chapter 1

Introduction

> *(M)alaria's probably the all-time biggest killer among diseases. Next to the common cold it's just about the most prevalent disease on the planet. We're not talking about a disease which shoots off the charts suddenly some century like the plague or smallpox or syphilis. (T)here's no place on earth that's off the malaria map: Arctic circle, freezing mountaintop, burning desert, you name it, malaria's been there. We're not talking millions of cases here; more like hundreds of millions. We don't even know how many, because malaria's so widespread it doesn't always get on the charts. And besides, it is a master of disguise; it can mimic the symptoms of more diseases than you can begin to count—lumbago, the flu, cerebral hemorrhage, yellow fever. And even when it's properly diagnosed it's not like quinine is always going to get you home safe.*

The character, Murugan, in Amitav Ghosh's novel
"The Calcutta chromosome: a novel of fevers, delirium and discovery": 55-56

1.1 The problem

The quote from Amitav Ghosh's book above is a simple, succinct, and effective description of malaria (Ghosh 1995). Indeed, about 3.2 billion people are estimated to be at risk of contracting malaria because they live in areas considered endemic to the disease.[1] The

[1] WHO (2005a) defines an area as endemic to malaria if there is "constant measurable incidence both of cases and of natural transmission" in that area over a succession of years. In other words, the disease occurs there on a consistent basis (CDC 2004). Natural transmission refers to occurrence of malaria due to climate and other factors considered generally to be the cause of malaria as opposed to cases of malaria

annual estimates of clinical episodes range from 350 to 500 million and the disease claims over a million lives each year.

There is evidence in the medical writings found in China and India indicating that malaria has existed at least since 2700 BC. The Roman involvement in drainage works especially of marshy land around 26 BC was probably motivated by malarial concerns as well (Heggenhougen et al. 2003, Rich et al. 1998). But, despite centuries of awareness about the disease and breakthroughs in understanding what causes malaria and how it can be transmitted among humans in the late 19th Century we have not been able to tame either the insect or the parasite that it carries (Najera 2001).

Part of the reason for the inability to bring malaria under control has to do with the ability of the malaria-causing *Anopheles* mosquito to breed in even a small amount of stagnant water. Preventing the formation of such breeding grounds in poor agrarian societies lacking the infrastructure to enable drainage of excess water is difficult. Further, the *Anopheles* mosquitoes tend to become resistant to insecticides such as DDT (Dicholro-diphenyltrichloroethane) and the parasite to drugs like chloroquine (Sharma et al. 1993). Yet another problem that has emerged in recent years is that of climate change bringing with it warm and humid climate conditions, conducive to the breeding of mosquitoes and the survival of the malaria parasite, *Plasmodium*, particularly the more deadly strain, *Plasmodium falciparum* (Basher and Cane 2002). These impediments make the task of malaria control complex and challenging, constantly needing new and varied control methods.

Poor countries seem to bear the burden of malaria. Gwatkin and Guillot (2000) estimate three-fifths of global malaria deaths to occur among the poorest fifth of the world's population. They arrive at this estimate by comparing total malaria deaths occurring in the richest (in terms of per capita income adjusted for purchasing power) 20 percent of the countries against the poorest 20 percent. These are rough estimates given that the set of high income countries would also include poor population groups and, similarly, the poorest countries would have some people with above average per capita incomes. Another study using a data source covering 150 countries from 1950 to 1995 by Gallup and Sachs (2001) obtain similar results. They find that 44 of the countries that had severe malaria[2] fall at the bottom when ranked by income per capita. This negative association

occurring in areas where such conditions don't exist, but where malaria can be imported such as through travelers returning from malarial areas.

[2]The degree of malaria was derived using historical maps of the geographical extent of malaria risk. These were combined with data on world population to get the proportion of population living in high risk

was also found between economic growth and malaria prevalence. While the growth of income per capita for countries with severe malaria over 1965-90 was 0.4 percent per year, the average for other countries was 2.3 percent. McCarthy et al.'s (2000) cross-country study covering 187 malarial countries over three five-year periods of 1983-88, 1988-93 and 1993-94 and controlling for reverse causation and climate also found a significant negative association between high malaria morbidity and GDP growth per capita. Their study attributed this to such factors as labour days lost due to illness and human capital loss from school absenteeism.

As the study above finds, malaria is an economic and social burden that causes loss in productivity and impedes human capital formation. Sometimes the only source of income that the poor may have is their (potential) labour and poor nutrition and ill-health can prevent them from using it effectively, thus perpetuating their poverty (Dasgupta 1997). Not surprisingly, the poor themselves identify illness as one of the most important causes of poverty in Participatory Poverty Assessment studies in various developing countries (Narayan 2001, Njie 2001). Sen (1987) avers that poverty is ultimately a matter of 'capability deprivation' and the inability to lead a healthy life is the deprivation of the most basic capability.

The economic costs of malaria are mostly measured in terms of days lost from work (Najera and Hempel 1996). As argued by Grossman (1972), the amount of time people spend working is a function of their 'stock of health' thus differentiating health capital from other forms of human capital, such as education where the 'stock of education' determines earnings directly rather than through the number of hours worked. On average, a malaria episode is estimated to result in the loss of four working days aside from the incremental days lost due to reduced capacity to work (McCarthy et al. 2000). Early estimates in Italy dating back to 1880 also attributed labour loss due to malaria at 4.5 days per worker (Snowden 2006). The total economic loss to India in 1991 is estimated at half a billion US dollars (Sharma 1996).

Aside from the output and productivity loss, households suffer a direct cost in terms of treatment and prevention that impacts on their disposable income (Sachs and Malaney 2002). In a study of malaria incidence in Nadiad *taluka* (sub-district) of Kheda district in Gujarat covering 100 respondents in seven villages and 50 in one town, Bhati et al. (1996)

---

areas in each country. Finally, the index of malaria intensity was obtained as the proportion of population at risk of malaria times the proportion of cases of *falciparum* malaria.

calculate the mean monetary loss due to malaria to be Rs. 394 (approximately \$9) in urban and Rs. 158 (\$3.5) in rural areas over the period October 1993 to December 1994.[3] The estimates are based on reported direct expenditure (including costs on treatment, transportation to commute to a health centre for treatment, and on a special diet, if any) and indirect expenses (that is, wage loss due to days lost at work) resulting from a malaria episode in the household.

In addition to the individual and immediate costs, long-term economic losses attributed to malaria incidence can accrue due to its negative impact on human capital formation. Anaemia caused by malaria in pregnancy is related to low birth weight that may affect cognitive and behavioural development of children and has also been found to be associated with low academic achievement in later years (Sachs and Malaney 2002, Guyatt and Snow 2004, Breslau et al. 2004). Older children affected with chronic malaria can also suffer from impairment and disability (Holding et al. 2004). Further, malaria morbidity tends to increase school absenteeism as evidenced in a rural community in Sri Lanka where Konradsen et al. (1997) attribute one-tenth of school days lost to malaria. In southwest Cameroon, Kimbi et al. (2005) estimate school days lost per child every month due to malaria at one and a half days.

1.2 Combating the problem

The above paragraphs provide an idea of the burden that malaria imposes on individuals and society. So, what are the measures that can be undertaken to combat the problem? Given evidence that poor countries bear the burden of malaria, the obvious deduction is that income growth and better socio-economic status would bring down incidence. This leads us into the wider debate on whether growth in aggregate income does, in fact, deliver better health outcomes. This has been examined by a number of authors. Pritchett and Summers (1996), for example, using a panel of 58 developing countries spanning a period of 31 years from 1960 to 1990 find a negative relation between infant and child mortality and per capita income. Their findings suggest that a one percent rise in income in the developing countries can prevent as many as 33,000 infant and 53,000 child deaths annually.

Anand and Ravallion (1993) argue that while it may be true that higher income countries will have better health outcomes, this is not merely a result of economic growth but

[3] Official exchange rate: 1 US Dollar = Rs. 44.65 Indian Rupees on 8 April 2006 (Oanda 2006)

of how that growth has been channelled, for example, to reducing poverty or providing better public health services that in turn improve health outcomes. This would mean that if one controls for such factors as absolute poverty and public health services, the relationship between income and health would at least weaken if not disappear. In another study on Sri Lanka, Anand and Kanbur (1995) find that health spending per capita had a significant negative effect on the infant mortality rate for two sets of time periods they considered (1960-78 and 1952-81) but income per capita had a significant impact only for the longer time period indicating that the benefits from income growth don't accrue as quickly as direct spending on health. Further, they find that an increase of one Sri Lankan rupee spent on health per person had the same impact in reducing infant mortality rate as 33 Sri Lankan rupees spent on aggregate income or GDP per person, thus highlighting the greater impact of public health spending in bringing down infant mortality.

Lending support to the Sri Lankan study, Haddad et al. (2003) also find that direct interventions in health improvements bring quicker returns than waiting for the benefits of income growth to accrue. They tested the effect of income growth on reducing malnutrition both at the household level using data from 12 countries and at the country level covering 61 developing countries. Their aim was to investigate if the millennium development goal of halving malnutrition by 2015 would be met by relying on income growth alone. They find that if income growth per capita is projected to be as high as 2.5 percent annually (which is not likely to be the case for most developing countries) it would result in reducing the malnutrition rate by 27 percent over 20 years by 2015. However, if improvements in household infrastructure such as sanitation are allowed the malnutrition rate would fall by 34 percent.

The above literature provides evidence that simply relying on income growth to deliver an impact on health outcomes is not sufficient or at least much slower than introducing direct interventions. Thus, while income growth may generate benefits over time, benefits from direct expenditure on public health measures accrue more quickly. This issue for malaria is under-studied and this thesis serves to fill this gap both at the micro and the aggregate level. It argues that public health expenditure is likely to have a larger negative impact on malaria incidence at the aggregate level as compared to income. It also argues that income and socio-economic status may not always be negatively correlated with malaria incidence at the household level and, hence, increases in income growth that lead to an increase in household disposable income may not necessarily bring down malaria incidence. This is likely because individuals or households may not use even a part of an increase in their incomes in taking such simple prophylactic measures as using a mosquito net at night or using mosquito repellents as their priorities may lie elsewhere such as in improving nutrition or in spending more on education, especially for those households that are poor.

Households may also not have the willingness to invest in improving their private infrastructure such as repairing leaking pipes or taps that may form stagnant water pools ideal for mosquito breeding. Such repair works would also have a positive externality for other households in the neighbourhood by preventing the breeding of mosquitoes. Moreover, malaria control is very often likely to fall prey to the problem of the 'tragedy of the commons' (Hardin 1968): if a cluster of houses falls next to a a public water body such as a river, canal, pond or stream that breeds *Anopheles*, households in the neighbourhood will be unwilling or even unable to invest in using measures to control the breeding. Government intervention is the answer for investment in control measures like spraying insecticides or introducing guppy fishes in the water bodies to feed on mosquito larvae or lining canals (see, for example, Bhati et al. (1995))

The next question then is where should public spending be directed? Some areas for policy intervention are alluded to above. But, what is the general stance on this subject in the current global initiative for controlling and subsequently eradicating malaria? This global initiative is called the Roll Back Malaria (RBM) partnership and was launched by the World Bank, the World Health Organization (WHO), the United Nations Children's Fund (UNICEF), and the United Nations Development Programme (UNDP) in November 1998. The three priority areas marked by the RBM are to provide artemisinin-based[4] combination therapy, to provide insecticide-treated bed nets (ITNs), and to develop a vaccine (WHO 2005d). All these strategies are aimed at protecting people from contracting malaria but none of these deal with the control of breeding areas of the malaria-causing *Anopheles* vector, and hence aiming to kill the problem at its root rather than devising end-of-the-pipe solutions. This is not to say that such solutions are not essential or useful,

[4]Artemisinin or *qinghaosu* is a compound derived from an ancient herb used in traditional Chinese medicine. It is currently considered the choice for anti-malarial treatment especially in places where the malaria parasite has developed resistance to chloroquine. However, in order to avoid the recurrence of malaria with the use of artemisinin alone, it is advisable to use combination therapies involving its use with another drug, mefloquine, also used for the treatment and prevention of chloroquine-resistant malaria strains (WHO 2008).

but these need to be complemented with other strategies to make the control of malaria a successful enterprise.

The development of a vaccine is likely to be one of the best solutions to the problem, but it is not clear when it will be ready and available for use in child immunization programmes. Further, it is not clear how effective and sustainable the vaccine will be given that previous attempts to develop a vaccine have failed because of the ability of the parasite to mutate (Chauhan 2008, MVI 2008). The use of ITNs at night has brought down malaria incidence and reduce deaths in many African countries (WHO 2006). But, reliance on this alone is not a long-term solution for ITNs are not affordable by the poor, even when available at a subsidised rate. They are only used at night while sleeping when the peak biting activities of the vector can start as early as dusk. They would also be ineffective in cases where the vector exhibits outdoor feeding patterns. Moreover, even when ITNs are provided free of cost to poor people, they need to be re-impregnated with insecticide at regular intervals. Since this is often at a cost to the user, people who are poor are not likely to be able to comply (Heggenhougen et al. 2003). Similarly, drug use is not a guarantee for long-term success as there continues to be fear that the parasite will become resistant to the drug. Therefore, along with these strategies one needs better infrastructure and engineering works that can prevent the formation of breeding areas and an understanding of the socio-economic context to enable more effective policy intervention. Such a multi-pronged strategy using drug therapy, ITNs and environmental control strategies is also supported by the Macroeconomic Commission on Health, established by the WHO in the year 2000. Malaria control would only be feasible with such a combined approach as Jeffery Sachs, who chaired the Commission in 2000 and 2001, argues (Sachs 2002). In fact, malaria was conquered in wealthier countries like Italy, United States, Greece, and Spain not simply because of socio-economic development but through an attack on various fronts such as using drainage of marshy areas, insecticide spraying and improvement of living conditions of the affected people (Snowden 2006, Sachs and Malaney 2002).

Perhaps the lack of adequate attention to the socio-economic context including improvement in infrastructure in the global initiative to fight the disease comes from lack of research connecting the influence of specific socio-economic variables to malaria incidence. This is the second gap that this thesis seeks to address.

There are scattered studies in the field of biological, environmental, and social sciences that relate malaria incidence to such factors as caste, occupation, education, type of house construction (e.g., whether the roof has eaves that allow entry of mosquitoes), irrigation, and proximity to a river, canal or stream (Guthmann et al. 2001, Guthmann et al. 2002, Konradsen et al. 2003, Sharma 1998). These studies enable one to deduce where policy interventions may be effective: for example, in lining canals to prevent breeding of mosquitoes, to invest in improving house construction, to invest in better drainage facilities in irrigated areas, or to educate and promote awareness so that people take prophylactic measures.

Most studies focus on rural areas as malaria is primarily seen as a rural phenomenon but there has been a rise in urban malaria in recent years. The factors determining malaria in the two settlement types differ on account of differences in the habitats of the prominent vectors – *An. culicifacies* in rural and *An. stephensi* in urban areas – and intrinsic differences in the infrastructure in urban and rural areas. Thus, for example, in rural areas the vector can be found breeding along rivers, ponds and streams, and in agricultural fields whereas in urban areas it has been found resting in public toilets and different kinds of water containers, used extensively in areas of water shortage such as in the city of Delhi in India (Sharma et al. 1993, Hati 1997, Karn and Harada 2002).

The scattered studies mentioned above examine the influence of one or two variables on malaria incidence rather than a host of factors. This thesis investigates the relationship with a host of socio-economic factors separately for rural and urban settlement types.

1.3 The research questions

Recapitulating from the discussion above, the research questions addressed in this thesis are:

1. What are the socio-economic factors that affect the incidence of malaria at the household and district level? How do these differ by settlement types and are the effects of these factors robust over time?

2. Does household wealth and other socio-economic status variables have a negative impact on malaria incidence?

   These two questions are interrogated using micro-level analysis and the following question is addressed using aggregate level data.

3. Does aggregate income have a negative impact on malaria incidence? How does it compare with the effect of public health expenditure?

The first two research questions have been examined using household and district level data. The two sources of information are (1) the household Demographic Health Survey (DHS) for India, called the National Family Health Survey (NFHS) for two time periods, 1992-93 and 1998-99 and (2) the district level census for 1991 and 2001. The NFHS is rich in data on such socio-economic variables as education, occupation, type of house, and availability of amenities like sanitation. The census provides data on such public infrastructure variables as extent of area irrigated by canals and distance to public drinking water sources. Two important contributions here are (a) the use of the DHS to study malaria, which has not been done to date for any country, and (b) combining the DHS with the census data to enrich the analysis.

In order to explore the second research question we need data on income or another welfare metric. The NFHS does not collect data on conventional welfare metrics like income, consumption, and expenditure. This drawback was overcome by collapsing the information available on the ownership of different consumer durable assets such as a television or a fan into an asset index, using Principal Components Analysis (e.g., Filmer and Pritchett (2001)), to serve as a proxy for permanent household income.

The third research question is investigated using a panel for 15 states in India covering the time period 1978 to 2000 prepared by Ozler et al. (1996) and updated by Burgess and Pande (2005).[5] This dataset contains information on such variables as aggregate income, health expenditure, and poverty. The data on malaria incidence were added to these data for this study.

1.4 The geographical area of the study

The country chosen to address the above questions is India since in Asia it has the largest number of malaria cases accounting for at least two-fifths of the total reported cases annually since 1992 (Table 1.1).

Approximately ninety five percent of the Indian population resides in areas considered susceptible to malaria. While in the early 1950s, nearly 75 million cases and 0.8 million

[5]I am very grateful to Dr Sonia Bhalotra for pointing out these sources to me and for providing me with her updated dataset (Bhalotra 2006).

Table 1.1: Malaria cases in India: number and percentage in Asia (1992 to 2003)

Year	*No. of cases*	*Percentage in Asia*[a]
1992	2,125,826	46
1993	2,207,431	50
1994	2,511,453	51
1995	2,988,231	55
1996	3,035,588	54
1997	2,660,057	45
1998	2,222,748	54
1999	2,284,713	36
2000	2,031,790	40
2001	2,085,484	51
2002	1,842,019	44
2003	1,781,336	46

a.The data coverage consists of a total of 30 countries.
Source: Calculations based on data from the Global Atlas of Infectious Diseases (WHO 2003).

deaths from malaria were reported every year, with the launch of the National Malaria Control Programme (NMCP) in 1953 and the National Malaria Eradication Programme (NMEP) in 1958 cases fell to 100,000 and by 1965-66 no deaths were reported. The programmes relied on insecticide spraying primarily DDT. However, there was a resurgence of the disease in the 1970s and malaria morbidity in 1976 rose to over six million cases (WHO 2005b) attributed to the development of resistance of the malaria parasite, *Plasmodium falciparum*, to the drug chloroquine and to the development of resistance of the malaria-causing *Anopheles* vector especially *Anopheles culicifacies* to insecticides, mainly DDT. In 1992 the number of reported cases reduced to a little over two million (see Table 1.1). But the 1990s saw another increased phase of malaria incidence with the total number of episodes touching three million in 1996. Sharma (1998) attributes this to chloroquine resistance in yet another *Plasmodium* species, *P. vivax*[6] as well as to problems in administering the malaria programme. The upsurge in malaria cases after getting close to eradicating the disease is disappointing. While it warns us about the complex nature of malaria and the likelihood of a resurgence of the disease if caution is not exercised in keeping incidence low, it also provides encouragement since it has been possible in the past to bring down incidence to levels as low as 100,000 from 75 million in a span of just eight years.

With the aid of malaria control programmes the number of cases declined to nearly two million in 2001 and finally close to 1.8 million in 2003 (WHO 2005c). In recent years there have also been concerns regarding the rising trend in *P. falciparum* cases in some

[6]This is a standard norm for writing '*Plasmodium*' when it appears as a part of the full name such as *Plasmodium vivax* or *Plasmodium falciparum* and, hence, abbreviated as *P.vivax* or *P. falciparum*. The norm for writing *Anopheles* is '*An*' when written with the name of a species. For example, *Anopheles culicifacies* is written as *An. culicifacies*.

areas of India. For example, in Mandla district in the state of Madhya Pradesh there is evidence to show that *P. falciparum* is replacing *P. vivax*. In the state of Chattisgarh[7] in 1975-76 more than half of the reported cases were *P. vivax*, but by 2002 these were reduced to a third with *P. falciparum* showing an upward trend (Singh et al. 2004a, Singh et al. 2004b). There are also fears that climate change may introduce malaria into areas that are currently not endemic but could become prone to the disease with changing weather conditions such as the northern Indian state of Himachal Pradesh (Bhattacharya et al. 2006).

While the country of study is India and an inter-state analysis covering 15 major states has been undertaken to address the third research question above, the other research questions have been addressed by focusing on one state, Uttar Pradesh, using the NFHS.[8] Focus on one state is important since malaria transmission can vary depending on geography. Therefore, assessing the malaria relationship at the all-India level would be too general to arrive at informative policy conclusions. Uttar Pradesh had the highest incidence in 1992-93 at 7.19 percent. It is also the most populous state,[9] accounting for nearly one-sixth of the country's population in 1991 and is one of the poorest states with a poverty head count ratio of 31.7 in 1999-2000 (Deaton and Dreze 2002). The level of literacy in the state is also quite low – 57 percent in 2001 (GoI 2001). A high proportion of its land area falls under irrigation at nearly two-fifths (GoI 1979-2001). These are important factors that can influence malaria incidence. In recent years there has been an increased focus on a number of peninsular states under the Enhanced Malaria Control Project (EMCP) initiated in 1997. These states include Gujarat, Madhya Pradesh, Maharashtra, and Orissa. There are fears that these are witnessing increasing *P.falciparum* cases mostly in tribal areas. However, the study of malaria in these states is effectively a study of tribal malaria. Since we are interested in also exploring other factors that influence malaria, Uttar Pradesh provides a much better case study. It is also worth noting that the attention given to the peninsular states has also come at the cost of poor surveillance in Uttar Pradesh in recent years. Sixteen deaths due to malaria were recorded in the city of Kanpur in Uttar Pradesh recently attributed to poor surveillance so that early detection

[7]The state of Chattisgarh was formed from Madhya Pradesh in November 2000. The 1975-76 reference is to that part of Madhya Pradesh that is now Chattisgarh (GoI 2005).

[8]At the time of the survey India was divided into 25 states and currently India is divided into 29 states (GoI 2005).

[9]Population of India in 1991 was 846 million and that of Uttar Pradesh was 139 million (GoI 2000).

was not possible (The Hindu 2008). While the WHO recommends a target of 10 percent for the annual blood examination rate to test malaria cases, in Uttar Pradesh this rate has been as low as three percent in some districts (GoI 2006c, Thankappan et al. 2006).

1.5 The structure

Chapter 2 discusses the literature covering transmission and factors affecting malaria and control measures. This forms the theoretical structure for the empirical analysis and the basis for selecting variables to be employed in the malaria incidence model.

Chapter 3 describes the three datasets – the NFHS and census used for the Uttar Pradesh analysis and the panel data employed for the inter-state analysis. It details the sampling methodology, how the variables were cleaned and constructed (where relevant), and their summary statistics. It also discusses the use of wealth as a welfare metric in the absence of other conventional metrics like consumption and income and describes how Principal Components Analysis can be used to collapse asset wealth variables into a single index.

Chapter 4 details the empirical methodology. It first discusses the probit methodology used for the rural household analysis and the method to combine the NFHS and district census data to study district level factors affecting malaria incidence. It then explains the methodology with specific reference to the urban sample followed by an explanation of how to decompose the malaria difference between rural and urban areas. Finally, it explains the panel methodology for the inter-state analysis.

Chapter 5 deals with an analysis of rural malaria incidence at the household level investigating the influence of income, socio-economic, and other factors on malaria incidence and if the effects of these factors are robust for the two time periods studied here. Chapter 6 investigates this for urban Uttar Pradesh. The effect of district level factors on malaria incidence is only explored for the rural sample because of data constraints. The decomposition analysis to examine the factors that determine differences in malaria incidence across settlement types is also undertaken in Chapter 6.

Chapter 7 explores the impact of aggregate income and public health expenditure on malaria incidence, in line with the approach of other authors studying the impact of income and public expenditure on health outcomes like infant mortality, for example.

The last chapter draws out the conclusions with particular reference to the policy relevance of the current study.

Chapter 2

Factors influencing malaria incidence

The epidemiology[1] of malaria relies on an interdependence of three types of environments (Heggenhougen et al. 2003, Flessa 1999):

1. The biological environment comprising the parasite, *Plasmodium*, and the breeding, resting and feeding habits of the malaria-causing vector, the female *Anopheles* mosquito.

2. The physical environment constituting the ecological and climatological factors such as humidity, temperature, and topography.

3. The host environment representing socio-economic factors or living conditions of the human population as well as the level of immunity.

The strategies to control malaria are designed to attack the biological and the host environments. The first section of this chapter begins by explaining how malaria is transmitted and then focuses on the factors that affect transmission. This is discussed in turn with respect to the biological, physical, and the host environment and with specific reference to how the factors differ between rural and urban areas. The second section details the current and past strategies that have been used to control malaria. Both the strategies and the transmission factors together influence malaria and form the basis for selecting the explanatory variables used in the malaria incidence model in this thesis.

[1]Epidemiology refers to the factors that affect the presence or absence of a disease (MedlinePlus 2003).

2.1 Transmission and factors that influence malaria

Malaria is believed to have become widespread among humans with settled agriculture that made possible the concentration of large numbers of people with animals, facilitating the transfer of malaria from animals to humans (Heggenhougen et al. 2003, Rich et al. 1998, Poolsuwan 1995).[2] The malaria parasite enters the human body when an infected female *Anopheles* mosquito takes a blood meal. It then goes through various life stages evading the human immune system and infects the liver, multiplies, ruptures the liver cells, enters the blood stream infecting the Red Blood Cells (RBCs), feeds on the haemoglobin and develops finally into a form that can infect another female *Anopheles* mosquito when it bites the human host. Inside the mosquito the parasite matures, reaches the sexual stage (this takes nine to 14 days) and when this mosquito takes its next blood meal it infects another human host through its saliva (WHO 2005d).[3] This is how malaria is transmitted between mosquitoes and humans. We now discuss the different environments that aid transmission starting with the biological environment.

2.1.1 The biological environment

Plasmodium, often referred to as the "beast within" the female *Anopheles* mosquito, is a one-celled parasite that has four species: *falciparum*, *vivax*, *ovale*, and *malariae*. In India, *vivax* dominates the malaria cases at 65 percent and *falciparum* occurs in the remaining cases (Prakash et al. 2003). *Ovale* is not found in India and malariae occurs in small numbers in the foothills of the state of Orissa. All malaria mortality in India is due to *P. falciparum*. Malaria caused by this species, if left untreated, can cause cerebral malaria. *P. vivax* may cause relapsing malaria but seldom death (WHO 2005b).

The *Anopheles* genus has 422 species though in India only 57 have been reported, of which only six are major malaria vectors. The most prominent vector is *An. culicifacies*, responsible for transmitting two-thirds of the malaria in India. The other two vectors of

[2]Later, by 1903, the separation of domestic animals from humans by introducing stables was recognised in helping to establish a clear preference of mosquitoes for feeding on animals (Najera 2001).

[3]Symptoms of malaria (e.g., fever, chills, and nausea) in an infected human host normally appear after nine to 14 days but in case of *vivax* and *ovale* it can even take months. However, the infection can be passed on to another mosquito before this time once the form of the parasite that can infect mosquitoes called gametocyte is in the blood. For *P. vivax* and *P.ovale* gametocytes can appear in the blood even within three days but for *P. falciparum* it takes 10 days. The parasite then needs a period to mature within the mosquito to be able to infect another human host when this mosquito takes a blood meal. The length of this period is called the sporogonic cycle. In highly endemic areas individuals develop immunity over time and can act as passive carriers of gametocytes, that is, they may not show symptoms of malaria but carry gametocytes in their blood that can infect a mosquito taking a blood meal (WHO 2004).

importance are *An. stephensi*, accounting for about 12 percent of the total malaria cases in the country, followed by *An. fluviatilis*. Uttar Pradesh, the state selected for this study, is host mainly to *An. culicifacies* in rural areas and *An. stephensi* in urban areas. *An. fluviatilis*, usually active in hills and foothills up to an altitude of 2,500 metres, is found in the *terai*[4] area of district Nainital in Uttar Pradesh.

The efficiency with which malaria is transmitted depends on how closely associated the *Anopheles* species are with the host environment. This is determined by where the different species breed, rest, and feed. The feeding habits can be zoophilic (feeding on animals) or anthropophilic (feeding on humans) as well as endophagic (feed indoors) or exophagic (feed outdoors). The resting habits can be endophilic (rest indoors after blood meals) or exophilic (rest outdoors after blood meals). Although a clear preference may be shown for one or the other of the resting and feeding characteristics, some species may show all the traits. The feeding hours of the different species also vary though generally most species feed from dusk to very early hours of the morning. Knowledge of all these factors influences vector control programmes. For example, knowing that a vector is endophagic and feeds at night would imply focus on use of ITNs, whereas awareness of resting habits would indicate where an insecticide spraying programme would work best – indoors or outdoors. Similarly, knowledge of breeding areas can help in determining which water bodies can be introduced with guppy fishes to control mosquito larvae (CDC 2006, Sharma 2002).

An. culicifacies is primarily endophagic, indoor resting, and mainly zoophilic, but readily feeds on humans. Its typical breeding sites are stagnant, unpolluted, sunlit surface water bodies such as rainwater puddles, irrigated fields, forest pools, lakes, hoof prints, borrow pits, disused wells, and drains. This species shows varied biting behaviour but usually feeds from dusk until before midnight. Rural settings are more conducive to its breeding (MRC 2003, Srivastava et al. 2001, WHO 2005b, Sharma 2002).

The next vector of importance, *An. stephensi*, found in urban areas, is also zoophilic but has increasingly shown a preference for human blood meals. It is primarily endophagic and feeds from late night until very early hours of the morning. The day time resting habits of the vector are ambiguous making it hard to attack. In Hati's (1997) study undertaken in Calcutta, however, the species was found resting in dark, humid places

[4] *Terai* (or *Tarai*) means literally moist land and is a belt of marshland that runs parallel to the lower Himalayan ranges in southern Nepal and northern India and stretches from the river Yamuna in the west to the river Brahmaputra in the east. It is composed of alternate layers of clay and sand, with high water tables that create springs and wetlands (Britannica 2006, Wikipedia 2006).

including temporary human shelters as well as in a brick-built public lavatory. Karn and Harada (2002) also found a high density of *An. stephensi* in toilets in slums and squatter settlements of Mumbai. It can breed in the smallest of containers to store water and it is this feature of *An. stephensi* that has led to a rise in urban malaria. It also shares other breeding habitats with *An. culicifacies*. The breeding of both *An. culicifacies* and *An. stephensi* is associated with the development of irrigation systems and urbanisation. The third vector of importance, *An. fluviatilis* is primarily anthropophilic and breeds in fresh water springs (Adak et al. 2005, Ansari and Razdan 2004).

The habitat of a species may be different from the areas where it feeds. For example, the species *An. dirus* primarily resides in forest fringes in northeast India but can fly up to 1.5 km for feeding in inhabited villages. *An. stephensi* can fly up to three kilometres and *An. fluviatilis* has a flight range of up to 1.5 km. It is also important to know the longevity of the vector so that a control programme can spread over the entire time frame that the vector is likely to survive. *An. stephensi*, for example, can survive from 26 up to 80 days. Vector control programmes, therefore, need to focus on a fairly large area taking account of the flight range of the species as well as its longevity (Srivastava et al. 2001, Sharma 2002).

2.1.2 The physical environment

Climate and topography play an important role in supporting the habitat of the *Anopheles* vector, thus influencing the potential for malaria transmission. Temperature, precipitation and relative humidity are the three main factors that affect malaria transmission (Pampana 1969). Piyaratne et al. (2005) found that *An. culicifacies* occurrence was higher in the temperature range of 28-32 degrees Celsius as this provides the optimum egg, larval, and pupal development conditions for this species. However, for most mosquito species the ideal breeding temperature is between 25 and 27 degrees Celsius and the maximum survival temperature for the vector as well as the parasite is 40 degrees Celsius. The minimum temperature for mosquito development is as low as 8-10 degrees Celsius and that for parasite development ranges between 14 and 19 degrees Celsius. The transmission of *P. vivax* requires a minimum temperature of 15 degrees Celsius whereas that of *P. falciparum* needs at least 19 degrees Celsius. Furthermore, the temperatures need to prevail over a certain period of time in order to complete sporogony (Patz et al. 1998, Sharma 2002).

The malaria vector can be found at altitudes between 1,600 m and 3,300 m, but these altitudes are generally not very conducive to the survival of the *Plasmodium* parasite because of lower temperatures at these elevations—Drakeley et al. (2005) found a negative relation between altitude and *Plasmodium falciparum* incidence in children in Kenya. However, microclimatic influences in the form of indoor heating can promote the spread of malaria even at high altitudes (Sharma 2002, Malakooti et al. 1998).

Humidity is essential for promoting the breeding of mosquitoes, but extreme conditions of rain can restrict mosquito proliferation. Srivastava et al. (2001) maintain that while the number of breeding sites rise in proportion to the amount of rainfall, high rainfall can flush mosquito eggs and low rainfall prevents the formation of breeding sites. Furthermore, average monthly relative humidity below 55 percent and above 80 percent also shortens the life of the mosquito and, hence, that of malaria transmission (Pampana 1969). Not surprisingly, therefore, Singh and Sharma (2002) as well as Akhtar and McMichael (1996) did not find a clear relationship between rainfall and malaria incidence in Madhya Pradesh (central India) and west Rajasthan in western India, respectively. However, Bhattacharya et al. (2006) observe that even in the dry months of the year in India when average rainfall recorded is low malaria incidence can persist.

Topography such as type of forest cover and terrain are also important in determining whether a location is likely to be exposed to malaria. For example, evergreen forests of north-east India and moist deciduous forests such as those found in some districts of Kerala, Karnataka, and Tamil Nadu in India support the species, *An. dirus* (Srivastava et al. 2001). A malaria study conducted in the villages of Mandla district in Madhya Pradesh reports that the hilly terrain coupled with perennial tributaries of the Narmada river formed ideal breeding grounds for the *Anopheles* vector in the form of rock pools, pits, and seepages (Singh et al. 2004b).

2.1.3 The host environment

In the host environment, malaria incidence is affected by socio-economic factors and living conditions as well as the level of immunity of the human host. We first discuss how immunity to malaria develops and which groups of the population are likely to be immune.

Immunity

In highly endemic areas individuals develop immunity over time and can act as passive carriers of gametocytes,[5] that is, they may not show symptoms of malaria but carry gametocytes in their blood that can infect a mosquito taking a blood meal (WHO 2004). In the human host the level of immunity can affect a person's susceptibility to malaria. In areas endemic to malaria the host population tends to develop immunity to the malaria parasite. In certain ethnic groups in sub-Saharan African countries, parts of India, and in the middle east this has manifested itself in the form of the inherited blood disorder called the sickle-cell trait.[6] This disorder is developed over generations of being subjected to malaria. While it provides protection against malaria if the gene is inherited only from one parent, it is fatal for children who inherit it from both parents. In India the sickle-cell trait is found among those belonging to the lowest castes and in tribal populations. The Glucose-6-phosphate dehydrogenase (G6PD) deficiency is also a hereditary enzyme defect that tends to break down red blood cells (haemolysis) when a person is exposed to infection and has been found in tribal populations in India. Ethnic groups carrying this trait are partly protected against malaria. Thus, ethnicity, by being associated with such characteristics, can be an important factor influencing the incidence of malaria across populations (Sachs and Malaney 2002, Pant et al. 1992, MedlinePlus 2006).

Certain blood group types can also carry immunity to malaria. On the basis of data collected from the relevant literature Arthreya and Coriell (1967) find a correlation between blood group B and malaria endemicity. They found that persons with blood group B develop immunity over time and have better chances of resisting the malaria parasite. Other blood groups such as A are more susceptible to contracting malaria because of sharing of a common antigen-blood group A substance between the tissues of the host and the parasite so that the body does not identify it as a foreign substance and an immune response is not triggered. Thus, if an infecting organism contains blood group A activity then individuals with A and AB groups would be more susceptible to this organism and those with blood groups O and B will not as these blood groups will contain the anti-A

[5] Gametocytes are a form of the malaria parasite that can infect mosquitoes (WHO 2004).

[6] The red blood cells (RBCs) of a person carrying the sickle-cell trait are hard and crescent-shaped because they carry abnormal haemoglobin (S) that reduces the amount of oxygen distorting the shape of the cells. Normal RBCs, carrying normal haemoglobin, are disc shaped and can move freely through the body. The presence of the sickle-cell trait can cause blood vessels to get blocked and prevent oxygen flow to tissues and organs because the RBCs are sticky and hard preventing free movement (NIH 2008, MedlinePlus 2006).

antibody increasing the chances of resisting the infection. Blood groups A and AB, on the other hand, would be at a disadvantage because of the absence of anti-A antibodies.[7] The other commonly used aspect to classify blood types is the *Rhesus* factor, which can be negative or positive, depending on whether a person has this antigen (positive) on the surface of its RBCs or not (negative). A number of other antigens can exist along with these two main types. One of these is the Duffy negative antigen mainly found in Africans. This is a blood type system classified according to whether it contains the Duffy antigen on the surface of the red blood cells in addition to the blood types ABO and *Rhesus*. *Plasmodium vivax* uses the Duffy antigen to enter the blood cells. Thus, persons who do not contain the Duffy antigen (Duffy negative) are able to resist the parasite (MedlinePlus 2003).

Socio-economic factors and living conditions

Malaria incidence is affected by such factors as the proximity of the house to breeding sites, type of house (i.e., material used for construction and whether the house has eaves that allow entry of mosquitoes or has crevices in which mosquitoes can hide) living conditions such as failure to clean or dry residual water after using it for domestic purposes, whether household members work in irrigated fields, and if people have poor knowledge and understanding about the disease. Sharma (1998) points out that *An. culicifacies* thrives in freshly dug pits. These are not uncommon in villages in India since earth is often used for house construction, plastering, and maintenance. Field-based analysis in the northern coastal area of Peru supports the view that type of house, its location and the house environment can influence the transmission of malaria. Thus, such factors as living more than 100 metres from a canal or residing far from agricultural fields and having a level of education equal to primary school or above serve as a protection against the risk of contracting malaria. On the other hand, living in a house that is more than four years old or sleeping in a bedroom that has spaces between the walls and in the roof can statistically significantly increase the risk of developing malaria (Guthmann et al. 2001, Guthmann et al. 2002). Konradsen et al.'s (2003) study for Sri Lanka support

[7] An antigen is a substance that the body's immune system identifies as foreign and triggers the release of antibodies as a part of the immune response mechanism. Red blood cells (RBCs) can contain A or B antigens. Blood group A contains the A antigen and will produce antibodies against antigen B. Group B has the B antigen and produces antibodies against A. AB group has both antigens so it does not produce antibodies against either and O has neither A or B antigens and produces antibodies against both antigen types A and B (MedlinePlus 2003).

these findings. They find that houses located near streams had more *An. culicifacies* than those that were farther way. House construction was also observed as a significant risk, though less strong than living close to a stream, for harbouring *An. culicifacies*, known to hide inside houses. Proximity to breeding sites may also be determined by type of work a person is involved in. Snowden (2006), for example, referring to Italy in the late 19th and early 20th centuries, points to the high rate of malaria incidence among agricultural labourers and mine workers. The incidence of malaria was also higher in the younger age groups and in the male populace since this segment of the population very often migrated to work in areas infested with malaria-causing mosquitoes.

The fact whether animals are kept inside or outside the house at night as well as the cattle-human ratio may also affect exposure to mosquito bites. For example, *An. dirus*, one of the most efficient malaria vectors in north-east India enters both cattle sheds as well as houses for feeding (Srivastava et al. 2001). Some agricultural communities, such as in Kenya, even use animals as decoys to prevent nuisance biting in case the mosquitoes have a preference for feeding on cattle (zoophilic) rather than on humans (anthropophilic) (Macintyre et al. 2002, CDC 2006).

Incidence may also be affected by hydrological changes or developmental work. In the Mandla villages in Madhya Pradesh there was a sudden rise in malaria cases from 1986 to 1987. This was on account of the construction of a major irrigation project on the river Narmada, the Bargi dam, leading to the formation of new breeding sites (Singh et al. 2004b). Ratnagiri district in the state of Maharashtra on the west coast of India was classified as a non-problem area according to the annual parasite incidence (positive cases of malaria), rainfall, topography, and vector species. But the inception of the Konkan railway line disturbed the ecology of this area and resulted in the creation of new breeding areas for the malaria vector. The absence of livestock in this area further facilitated the human-mosquito contact (Jotkar et al. 1997).

The above paragraphs refer primarily to conditions prevailing in rural areas and indicate that rural environments are more conducive to the spread of malaria. Aside from the fact that rural environments (as compared to urban locations) generally provide the habitats congenial to mosquito breeding, higher population density in urban areas can also be associated with lower incidence by allowing for a higher human-mosquito ratio, which reduces the number of infective bites per person. These can vary from less than one bite per year in areas with low transmission to nearly as many as 3000 in high transmission areas (Robert et al. 2003). Rickman et al. (1990) found that one or two infective bites per person affected only five out of 10 volunteers without immunity to malaria. Thus, a reduction in the number of bites per person, especially in endemic areas where people acquire anti-malarial immunity over time could be associated with lower malaria incidence (Gupta et al. 1999).

It is believed that urban malaria has risen in recent years. However, the existence of malaria in urban areas in India was known as early as 1911 when the British moved their capital from Calcutta to Delhi and built the city of New Delhi outside the old 'walled' city because of the high incidence of malaria there. This incidence was attributed mainly to the use of open wells where the main vector of urban malaria mosquitoes, *An. stephensi*, thrived. Once these wells were closed the breeding of the vector was minimised bringing down the malaria incidence rate. However, there was a re-emergence of urban malaria especially in the 1960s because *An. stephensi* found new areas for breeding, particularly in overhead water tanks, as well as in all kinds of fresh water collections in different types of containers (discarded cans, tin drums, or even coconut shells) just about large enough to allow the mosquito sufficient space for oviposition. The overhead tanks are generally installed in inaccessible places to prevent vandalism, with the result that they are often not cleaned or are left inadequately covered, allowing breeding of mosquitoes. *An. stephensi* has also been found to breed around plants with residual water collection and in different types of pots in urban gardens. Various other random water collections like small water pools from leaking pipes, rain water and other drains as well as water containers of air coolers[8] also serve as suitable breeding areas for the vector. A study in Gurgaon in the outskirts of Delhi found that the breeding in random water collections was particularly intense because these pools of water are generally missed out in larvicide control treatments under the Urban Malaria Scheme operational in areas with a population of more than 40,000 people. In addition to these sites, the vector continues to breed in both used and disused wells. Neglected drains in houses as well as air conditioning equipment in relatively well-off households also add to the many breeding areas, leaving no income group free of a potential breeding site in its vicinity (Sharma et al. 1993, Sharma 1999, Hyma et

[8]Air coolers throw cool air into a room with the aid of a fan beneath which is placed a container of fresh water and behind the fan there is a thatched structure around which water is thrown up with the help of a pump. It is a cheap and effective way of cooling air used extensively in the hot season in India.

al. 1983, Mariappan et al. 1992, Hati 1997, Chakraborty et al. 1998).

Interestingly most studies found negligible breeding in sewage pools because of the presence of too much organic matter that seems to inhibit breeding. This is particularly true of the secondary malaria vector of urban areas, *An. culicifacies*. Since a lot of peri-urban areas share rural features, it is not surprising to find the co-existence of this vector with *An. stephensi*. *An. culicifacies* is found in canals and rivers and in open drains, sites that are not off-limits for the principal vector either. A study in Delhi found that this species had abandoned its more traditional riverine breeding areas because of the high pollution content of the river Yamuna and preferred other clear water sources. However, it is not found to breed in such devices as containers and tanks. This feature is reserved for the flexible adaptive behaviour of *An. stephensi*, known to adapt to nearly any habitat, making the control of urban malaria rather difficult along with the fact that it has developed resistance to DDT and other insecticides in many parts of India, along with *An. culicifacies* (Sharma et al. 1993, Hati 1997, Batra et al. 2001). The resistance of the *Anopheles* sub-species to DDT was primarily a result of intense use of DDT in agriculture and cotton production in particular. The resistance built gradually over time since the 1960s. In the early 1970s DDT was first banned in industrialized countries and there has been pressure to discontinue its use and exercise a complete ban throughout the world (Robert et al. 2003). However, DDT might still be effective for malaria control in some countries such as Ecuador where following the increase in DDT use since 1993 there has been a substantial reduction in malaria incidence (Roberts et al. 1997, Roberts et al. 2000).[9]

Bang (1985) argues that in the WHO South East Asia region (SEAR) and especially so in India, urban malaria is mainly a result of unplanned urbanisation associated with in-migration and the inability of the existing infrastructure to cope with, for example, the supply of water or repair of leaking water pipes, which would prevent breeding. There is a proliferation of water tanks in urban areas as it is common to have regulated water supply at certain times of the day, encouraging residents to store water for use at other times. This view is also supported by Yadav et al. (2003) based on their study of Ahmedabad city in Gujarat. In Calcutta, the construction of the underground metro rail favoured breeding

[9]Sri Lanka also had a dramatic fall in malaria cases from 2.8 million cases to merely 17 from 1945 to 1963 as a result of DDT spraying. The programme was terminated in 1964, following which mortality and morbidity levels rose again (Pinikahana and Dixon 1993).

of *An. stephensi*, once again because of its ability to use new habitats for breeding.

Along with construction activities, a large number of illegal settlements sprang up to host rural migrant labourers who arrived in the city to work on these sites and generally lived in poor environmental conditions (such as living in temporary shelters or huts and having inadequate water supply) that are highly conducive to mosquito breeding. Pattanayak et al. (1994) point out that epidemics around construction sites are not uncommon because of the collection of residual water and the use of cement tanks for storing water for construction work. In a South Delhi construction site a number of rocky pits were found as a result of the quarrying work, conducive to the breeding of *An. stephensi* (Sethi et al. 1990). In Goa, Sumodan et al.'s (2004) study also collected the maximum number of *An. stephensi* from construction sites and inside huts of construction workers.

Most industrial townships have been associated with malaria incidence during their construction phase. The Mirzapur Thermal Power Project in Uttar Pradesh reported 2,421 malaria cases in 1979 at the initiation of the project. The number of cases were reported at 11,455 in 1980, within one year of the start of the project. The Mathura Oil Refinery (in Uttar Pradesh) also had high malaria incidence during its construction phase, which continued even after the completion of the project. The National Thermal Power Corporation (NTPC) unit in Dadri (in Uttar Pradesh) was also associated with a rise in malaria incidence. One of the main reasons for malaria incidence in industrial locations is water logging. NTPC was, in fact, located in the water logged area of the Upper Ganga Canal. Industries are established near water bodies to allow ease in the use of water for cooling and waste disposal. The Mathura refinery, too, was built near a river for this purpose. Moreover, most industries do not take account of health measures or undertake appropriate engineering works to allow good drainage aiding the proliferation of mosquito vectors.

Aside from the above reasons, new construction sites or industries attract labour from rural areas as mentioned above. Many of these workers may be endemic to malaria and hence act as potential carriers of *Plasmodium* enabling transmission in new areas. In a study undertaken in a South Delhi locality by Sethi et al. (1990), migrant construction workers were compared with local inhabitants (living in separate settlements) for malaria incidence. The migrants, generally from a young age group, were found to have higher malaria prevalence than the locals. The study found that 60 percent of the migrants went back once a year to their original place of residence with a length of stay of approximately five weeks. The authors suggest that such movements of population keep up the endemicity of malaria in cities.

Finally, the development of transport networks have also been found to aid in transmitting malaria. ‘Airport’ malaria has been reported in Belgium, France, and Switzerland among residents living near the airport or among employees, even though these countries are not endemic to malaria. A possible reason for these malaria cases (mainly *P. falciparum*) was attributed to the transportation of infected mosquitoes sheltering in hand luggage, cargo holds, animal compartments or wheel bays on in-bound flights from sub-Saharan Africa countries of Cote de Ivoire, Benin, Cameroon, and Senegal (Van den Ende et al. 1998, Guillet et al. 1998).

The above discussion highlights the fact that the factors affecting malaria incidence are varied and cannot be reduced to one single cause or source. There are complex factors,[10] and complex interactions among these factors, at work requiring an attack on the vector as well as the host environment, thus making the task of malaria control challenging. The next section looks briefly at various measures used to combat malaria since the 1900s.

2.2 Malaria control measures

Globally, malaria control measures have gone through three phases. The first phase from 1900 to 1950 consisted of attacking the malaria vector and using drainage techniques. The next phase, until the early 1960s, used insecticide (mainly DDT) spraying to kill mosquitoes. The third and current phase aims at providing such devices as insecticide-treated bed nets to prevent mosquito bites and also recognises that environmental management and understanding of socio-economic factors are important to enable successful eradication (Heggenhougen et al. 2003).[11]

A number of breakthroughs define the history of malaria. One of the first of these came

[10]Yet another example of the complexity can be gauged from the possibility of malaria transmission by such issues as whether household members visit forest areas (that may harbour certain malaria causing *Anopheles* species) for logging and firewood collection even if they otherwise don’t reside within these forests. This is demonstrated by the occurrence of a case of malaria caused by the species *P. ovale* in Jorhat district of Assam. Since *P. ovale* is not found in India the team studying malaria in this district investigated the issue in greater detail and concluded that the person, in all likelihood, contracted malaria when he was collecting firewood or was engaged in log cutting in the nearby Nagaland hills that border Myanmar where *P. ovale* is found (Prakash et al. 2003).

[11]This phase also aims at developing a malaria vaccine but it is not available for use yet. As mentioned in the first chapter, the development of a vaccine is very challenging because the parasite tends to mutate.

about in 1880 when the French physician, pathologist, and parasitologist, Charles-Louis-Alphose Laveran found the malaria parasite, *Plasmodium*. In 1898 the British bacteriologist Sir Ronald Ross discovered that the vector that causes malaria is *Anopheles* and that malaria can be transmitted between humans through the bites of the female *Anopheles* mosquito confirming the British physician Patrick Manson's deduction that malaria is transmitted by mosquitoes four years prior to Ross. Around the same time the Italian scientists Giovanni Battista Grassi, Amico Bignami and Giuseppe Bastianelli showed (by infecting humans with malaria) that human malaria could only be transmitted by the *Anopheles* mosquito that carries *Plasmodium* in its digestive tract. Following these discoveries considerable attention was focused in some countries on the control of those species of *Anopheles* that were carriers of *Plasmodium*. In Europe mosquitoes were studied to find out which species were harmful, identifying their haunts, and requiring a concentrated attack on these species to eradicate the disease. Watson (1937) used this approach in Malaysia by studying the micro-epidemiology of malaria and identifying the main vector species and their breeding and feeding habits to devise malaria control measures. Focus on only exterminating those species of mosquitoes that caused malaria was referred to as 'species sanitation' by the Dutch malariologist, Nico Swellengrebel, who used this technique in Indonesia around the 1910s and 1920s. The Dutch school of malariology following epidemics over 1919-21 in North Holland followed larviciding (killing larvae before they hatched into adults) around main cities and adulticiding (attacking adult mosquitoes) in cattle sheds. Concomitant to use of these measures they also encouraged the use of nets, took regular blood samples from those who were suspected to have malaria and used quinine and increased awareness in the general public to encourage them to use protective measures (Najera 2001, Snowden 2006, Britannica 1997, Hamoudi and Sachs 1999).

The importance of drainage techniques in malaria control was recognised in the early 1900s—water management through drainage was key to the effective control of malaria in the British colony of Ismailia in Egypt in 1903 (Watson 1937). While Panama met with success in introducing efficient drainage to control malaria over the period 1904-08, the Mian Mir project in India (1902-04) could not meet this objective. The failure is partly attributed to terminating the programme too soon for fears of excessive expenditure although Ross denied that such programmes could be any more expensive in the long run than the use of end-of-the-pipe curative methods such as chemoprophylaxis[12] through the intake of quinine (Najera 2001). In the Tennessee Valley river basin of USA, too, local water management and drainage techniques were key to the success of malaria eradication. Other examples from India include the construction of the Sarda Canal (Clyde 1931 in Sharma and Sharma 1989) and Irwin canal (Rao and Nasiruddin, 1945 in Sharma and Sharma 1989) and the Cauveri-Mettur project in south India (Russell and Knipe 1942 in Sharma and Sharma 1989). During World War II Indian troops in Assam were protected from malaria through organised drainage (Sharma 2002). However, current policy on malaria control in India does not emphasize these methods. Instead, it focuses on distribution of ITNs, insecticide spraying and using drug therapy (GoI 2006c). Konradsen et al. (2002) criticize the sidelining of such techniques in the 1950s when too much emphasis was given to the environment-unfriendly practice of using insecticide sprays, mainly DDT. This method had to be abandoned in the 1960s when it was realized that the malaria-causing *Anopheles* vector was becoming resistant to DDT.

A pointer to how environment-friendly agriculture can help counteract malaria can be seen from the example of rice fields in Malaysia (known as the Federated States of Malaya (FSM) at the time) provided by Watson (1937). Rice grows in marshy areas and these can be ideal breeding grounds for mosquitoes. The very term malaria (or bad air) is derived from the original reasoning for its existence, which is the fetid smell associated with marshes. Yet the rice fields of FSM were malaria free back in 1909. This was because the forest acted as a protective cover against malaria and prevented the breeding of those mosquitoes that carry *Plasmodium*—the streams that fed the rice fields ran under dense shade and kept malaria away. Similarly, a sub-species of *An. minimus*, Theobald, was brought under control in the tea gardens in India by using plants that can grow in shades in the drains (Covell 1955 in Sharma and Sharma 1989).

Current malaria control practices include the provision of such counteractive measures as insecticide treated bed nets (ITNs) and indoor residual spraying (IRS).[13] However, the success of such measures depends on the co-operation of the local communities. For example, bed nets need to be re-impregnated with insecticide at regular intervals often at a cost to the user. In the Gambia between 1992-94 a programme that introduced bed nets

[12]Chemoprophylaxis is a term used for the prevention of infectious diseases through the use of chemical agents (MedlinePlus 2003)

[13]Another vector control strategy includes the use of larvivorous fish that feed on mosquito eggs (WHO 2005a).

and re-impregnation was successful, but after this period once a re-impregnation cost was introduced at $0.50 per net the programme failed (D'Alessandro et al. 1995).

Further, as Ehiri and Anyanwu (2004) warn, all pesticides are toxic by nature with adverse health implications. Thus, the widespread use of ITNs in malaria endemic countries could be risky especially where the treatment and re-treatment of nets is carried out by untrained people and in uncontrolled settings. Moreover, in case the species found in an area is exophilic with outdoor biting activity, ITNs would not be useful in preventing bites. *An. culicifacies* has, in fact, shown a departure from previous resting and feeding behaviour from being endophilic and endophagic to becoming exophilic and exophagic (Saxena et al. 1992).

There has been increased emphasis on conducting studies on the socio-economic aspects of communities in order to better address their problems and engage them in malaria control programmes. Such studies also help in discovering other means of controlling malaria that local communities may already be employing. To give an example, a socio-cultural study conducted in eight villages of the coastal plain of Chiapa, Mexico, revealed that women had greater participation as family health promoters and were in charge of the application of self-protection preventive measures. An implication of this would be that a campaign aimed at influencing households to undertake prophylactic measures should be aimed at the women. The study also found that more than 50 percent of the respondents preferred to use mosquito coils and smoke (produced by burning flammable material, which varied from weeds to dry cattle dung) to prevent mosquito bites rather than insecticide spraying or bed nets. Yet another device used in the Chiapa households to prevent mosquito bites was an electric fan, limited to respondents who had completed elementary education and to houses that had a concrete floor, indicating that it was a function of wealth. It must be noted, however, that the possibility of owning a fan is not simply a function of household wealth but also depends on whether a particular house has access to electricity or not. This very often depends on the closeness of the village to the grid (Rodriguez et al. 2003).

Research and awareness regarding the use of environment friendly herbal remedies against mosquito bites such as the use of *Neem* oil may also help in propagating the use of these as a part of a malaria control programme. An experiment conducted in Mandla district of Madhya Pradesh revealed that the application of *Neem* oil deters *An.*

culicifacies from taking blood meals from the host for five to six hours. Since tribals customarily apply oil to their body parts a malaria control programme that encourages the use of *Neem* oil would be in consonance with the socio-cultural practices of the tribals (Mishra et al. 1995).

Bhati et al. (1995) also argue for community involvement and raising awareness based on a control-target approach carried out in Nadiad *taluka* in Kheda district in Gujarat. Community seminars held in the target group of people raised awareness and sense of responsibility regarding malaria control in the community. The target group, unlike the control group, believed that the task of malaria control was not just a responsibility of the government but ought to be a joint effort involving the public as well.

The literature discussed above provides some insights into the different kinds of explanatory variables that can be used to model the incidence of malaria including socio-economic and climatological factors. Since the epidemiology of malaria differs according to local conditions it is not advisable to use a general model for estimating incidence. However, Ross did develop a mathematical model that was later refined by Macdonald (1957) with the aim of obtaining the number of malaria cases in a particular locale (Hamoudi and Sachs 1999).

Macdonald's model was stated in terms of a basic reproduction number which was the number of secondary infections resulting from a single infection. The higher this number, the harder it was considered to control malaria. Using Macdonald's notation, the basic re-production number, z_0, is given by the equation:

$$z_o = \frac{ma^2bp^n}{-rlnp} \tag{2.1}$$

where m stands for the density of *anophelines* relative to the human population, which would depend on number of breeding sites available and hence climate, topography and other such factors as discussed above. Drainage techniques to remove excess water that can serve as a potential breeding area and larval control can bring down the mosquito density. The parameter a indicates the human-biting propensity of the vector (so it gives the average number of persons bitten by one mosquito in one day) dependent on whether the vector is zoophilic or anthropophilic. Thus, the presence of animals could deter human biting in species that are zoophilic. Following the above equation this is an important factor as the reproduction rate, z_o, is directly proportional to the square of a. The parameter b is the proportion of infective mosquito bites or the probability whether a bite by an infected mosquito will transmit malaria. This depends on such factors as the number of infected sporozites present in the saliva of the mosquito. The number of infective bites could be controlled by the use of such prophylactic measures as mosquito repellents and insecticide treated bed nets.

The probability of mosquito survival is captured by p and was defined by Macdonald as the 'probability that a mosquito will survive a whole day' and p^n is the probability of a mosquito surviving for n days. The parameter p normally ranges between 0.6 and 0.95, which can be shown to be equal to $Ee^{E/n}$, where E = *-1/lnp* is defined as the life expectancy of the mosquito and n is the number of days a mosquito survives. *Ceteris paribus*, the reproduction number is proportional to p^n/lnp. Thus, p, as Hamoudi and Sachs (1999) suggest, is an essential factor in controlling malaria and explains why, historically, immense efforts were concentrated on using insecticide spraying particularly on adult populations of mosquitoes. The number of days it takes the parasite to complete its life cycle, n, within the mosquito also enters the above equation and this parameter determines when the mosquito is infectious. If the mosquito dies before this period the infection will not be passed on. The survival of the mosquito and of parasite development depends on such conditions as temperature and other factors discussed in the literature. Finally, the parameter, r, is the recovery rate that depends on adaptations and the level of immunity of the host human population and is defined by Macdonald as "the proportion of affected people who receive one infective innoculation only, who revert to the unaffected state in one day". Given the long history and presence of malaria among the human population there may have been a number of evolutionary adaptations to malaria in the human populace. Some of these adaptations include the sickle cell trait, the Duffy blood group negativity, and the G6PD deficiency. Adaptations that reduce susceptibility to infections affect the parameter b and those that increase immune responsiveness affect r. Similarly, prophylaxis affects b and therapy affects r. The model proposed by Ross and Macdonald looks at the number of infections resulting from one single infection or the reproduction rate of malaria whereas in the current study we are examining the factors responsible for why a particular person contracts malaria and not the total number of malaria cases contracted in the population. However, through the use of this equation it is possible to infer the kind of effect an attack on the different parameters entering the model can have on malaria incidence in a particular area.

2.3 Summary

Malaria is transmitted through a complex interaction of the biological, physical, and host environment. Malaria control strategies are aimed at the biological and host environment. Understanding the breeding, resting, and feeding habits of the malaria vector, which falls under the study of the biological environment, can affect decisions on whether to use indoor insecticide spraying when the vector is mainly indoor resting or encouraging the use of bed nets in case the vector also exhibits indoor biting behaviour. Knowledge of breeding areas of the malaria causing *Anopheles* species, on the other hand, is important in directing methods of mosquito larval control to those habitats. The use of larvivorous fish in ponds and canals could be one such strategy. In the host environment, an understanding of the living conditions and socio-economic factors that are likely to be associated with malaria incidence also enables effective implementation of control strategies. For example, proximity of a house to a water body serving as a breeding area for *Anopheles* is often associated with higher malaria incidence. On the other hand, level of education or awareness can also prevent people from taking prophylactic measures, calling for programmes that increase awareness about taking preventive measures and being aware of vector habits. In India, indoor residual spraying, distribution of insecticide treated nets, and use of larvivorous fish are the main control strategies used along with distribution of anti-malarial drugs (GoI 2006c). However, the environmental control strategy of draining fields with excess water to prevent breeding are not so common. This would be particularly relevant for the state of Uttar Pradesh with nearly two-fifths of its land area under irrigation (GoI 1979-2001). Prior to implementation of a control strategy, however, it is important to study which factors actually influence malaria incidence. On the basis of the literature covered above, such factors as type of house, proximity to a water body, extent of irrigated area, and level of education have been identified as potential explanatory variables to be used to study malaria incidence in Uttar Pradesh in this thesis.

Chapter 3

Data

The analysis in this thesis is conducted at three levels – household, district, and state – requiring three different sources of data. The household analysis uses the National Family Health Survey (NFHS) for two time periods, 1992-93 and 1998-99. The data for the third phase of this series were available in 2008 but these did not cover malaria incidence. An explanation for why this was not undertaken is not available in the reports (IIPS 2008). The district analysis combines the district census for 1991 and 2001 with the household data, and the state level analysis uses a panel based on 15 states in India covering the period 1978 to 2000. The NFHS and the district census have been used to analyse the effect of income, socio-economic, and other factors on malaria incidence by focusing on one state, Uttar Pradesh. The panel data is employed to undertake an inter-state analysis of the impact of income, public health expenditure, and other variables on malaria incidence.

The dependent variable used throughout is malaria incidence. For the household and district analysis for Uttar Pradesh these data are drawn from the NFHS where malaria incidence data are collected by reference to whether a household member had malaria over the past three months. The inter-state analysis uses the malaria estimates available from the National Malaria Eradication Programme (NMEP) of the Government of India based on testing blood smears of febrile patients for the malaria parasite. These sources report different estimates. The first section of this chapter attempts to reconcile these differences. There might be a tendency to trust the reliability of the NMEP data more than that of the NFHS because the former is based on a blood examination. One could speculate this as one of the reasons why the NFHS did not collect any malaria incidence data in its third phase. However, this section argues that the NMEP data also suffer from inaccuracies and underestimation and when we take account of these factors, the difference in estimates from the two sources is attenuated considerably.

Section 3.2 begins by explaining the sampling methodology used in the NFHS. It then identifies the explanatory variables employed for the household and district level analysis based on the literature reviewed in the previous chapter. The cleaning procedures for the household variables drawn from the NFHS are elaborated next, followed by the summary statistics of these variables presented separately for the rural and urban areas and for the two phases of the NFHS. As mentioned in the introduction to this thesis, a measure of current income, consumption, or expenditure is not available in the NFHS. Instead, the information on the ownership of different consumer durable assets was collapsed into an asset wealth index to serve as a proxy for permanent income. The subsequent section first justifies the use of wealth as a welfare metric in the absence of more conventional metrics like current income, consumption, or expenditure. It then explains the data reduction technique (i.e., Principal Components Analysis) to create an asset wealth index and presents the results of this exercise.

Section 3.4 begins by explaining how the district level census variables were complied in order to combine these with the household survey data. Cleaning procedures and summary statistics for the variables selected are then presented.

Finally, Section 3.5 elaborates on the variables used in the inter-state analysis, the different sources of data used to compile the panel data, and the summary statistics. In this part of the analysis, use of current income does not pose a problem as the data on current state domestic product has been used.

3.1 Malaria incidence data

As noted above, the household and district level analysis uses the NFHS malaria incidence data and the inter-state analysis employs the NMEP data. There are differences in these estimates and this section attempts to reconcile these differences.

The NMEP data are based on actual blood smear examinations and are presented as the annual parasite incidence (API), which is defined as the number of (blood) slides found positive for malaria per thousand population. The NMEP also reports the slide positivity rate (SPR), defined as the number of slides positive per hundred slides examined (Sharma

1999, Srivastava et al 2003). The blood samples are collected from fever cases. The NFHS, on the other hand, does not examine blood for malaria and the incidence data reported here are collected by asking the question "Did any member of the household suffer from malaria in the last three months" with reference to the symptoms of malaria mainly fever with chills recurring on alternate days. Thus, the number of reported malaria cases in the NFHS dataset set may be compared to the number of fever cases reported by the NMEP.

Malaria incidence in Uttar Pradesh was 73 per thousand population in 1992-93 according to the NFHS data. The annual parasite incidence from NMEP, on the other hand, was 0.840 per 1000.[1] However, as mentioned above, these figures are not strictly comparable as the number of malaria cases when using the NFHS are potentially fever cases suspected of malaria since these are not based on a blood examination. Moreover, one needs to be circumspect about the malaria cases reported by the NMEP. As per the in-depth evaluation committee of the NMEP (1985), underreporting is common and malaria cases are "grossly" underestimated. Sharma (1999), on the basis of epidemiological data analysis conducted by other researchers, field visits to endemic districts, evaluations of the NMEP and discussions with health officials suggests that the actual number of malaria cases are at least 10-15 times more than those reported by the NMEP. The WHO Southeast Asia Regional Office (WHO SEAR) estimates of malaria incidence are also six times higher than those of the NMEP—while NMEP reported 2.5 to three million cases over the 1990s, the WHO SEAR office reported 15 million over the same period (Yadav et al. 2003). One major reason for this is that NMEP only covers parasitologically-confirmed cases, based on the examination of blood smears collected at regular intervals at the village level. Moreover, NMEP was intended to provide treatment even to the most isolated cases in far flung villages rather than monitoring the actual incidence. Thus, the data provided by the NMEP should ideally be considered for looking at the trend in malaria rather than the actual number of cases at a particular point of time. Attention now turns to the problems with blood smear collection and reasons why the reported cases may be an underestimate.

Blood smears are collected as a part of the malaria surveillance programme initiated under the National Vector Borne Diseases Control Programme (NVBDCP) of the Ministry of Health and Family Welfare later called the National Malaria Eradication Programme

[1] This is calculated as the average of the annual parasite incidence reported for 1992 and 1993 by the NMEP (GoI 1986-2001).

(NMEP). The data are collected through active and passive surveillance. While both methods are used in rural areas, urban areas are covered only through passive surveillance.

Active surveillance or active case detection requires fortnightly visits to different hamlets in each village in different states by multipurpose health workers (MPWs). The MPWs are responsible for identifying fever cases in the village between visits and collect a blood smear from those who had fever. Thus, the samples collected are not random. The blood smears collected by the MPWs are then forwarded to laboratory technicians at the primary health centres and sub-centres. Once the slides have been examined the data on the number of malaria cases and slides examined are sent to the state health departments for collating at the state level. These results are then sent to the NMEP at the Centre for national level estimates (Sharma 1996).

Under passive case detection fever cases are expected to report voluntarily to malaria clinics or primary health centres or sub-centres. While the problem of shortage of supply of MPWs may be countered to some extent by the passive surveillance scheme under which people reporting to any clinic are screened for malaria, this may miss out on the numerous individuals who have recourse to informal institutions or do not report to a medical practitioner at all if the symptoms are not severe. In fact, a majority of the cases report to private clinics. These malaria cases and even those reporting to well-known government hospitals like the All India Institute of Medical Sciences and the Ram Manohar Lohia Hospital are not included when arriving at final estimates. Fever cases visiting such alternative-medicine centres as homeopathic, *ayurvedic* and *yunani* are not considered either (Sharma 1999). The main aim of the NMEP was to provide treatment especially when there was a resurgence of malaria in the late 1960s. As a part of this a number of drug distribution centres were opened under the Modified Plan of Operation of the NMEP in 1977 that provide drugs to any fever case without taking a blood smear and are not counted in the total number of malaria cases. Morbidity figures of people working in defence, paramilitary forces, police, border road organisations, tea estates and coffee plantations as well as other autonomous institutions and private hospitals are also excluded (Sharma 1996).

The malaria surveillance report of the NVBDCP points to the shortage of MPWs and, hence, an inability to collect the requisite number of blood samples needed to arrive at authentic API rates. According to Sharma (1999), there are about 40-50 percent vacancies for MPWs in rural areas. Moreover, MPWs are overburdened. They are responsible for eight programmes, which includes family planning that takes priority, especially because it incorporates an incentive scheme. On the other hand, there is no accountability for poor malaria surveillance. Further, there is no provision for replacement in case an MPW is on leave.

Not surprisingly the annual blood examination rate (ABER) (number of blood smears collected in the year as a proportion of population covered under surveillance) falls short of the WHO-recommended target of 10 percent. As mentioned in the introduction, in states like Uttar Pradesh, which are not covered under the Enhanced Malaria Control Programme (EMCP), the ABER has been as low as three percent in some districts (GoI 2006c, Thankappan et al. 2006). Calculations from the data collected from WHO show a mean ABER of seven percent for Uttar Pradesh as a whole with as low as 0.57 percent in some districts. Mathur et al. (1992) also report that in Barmer district in Rajasthan over the period 1982-1999 the ABER was consistently low because of irregular surveillances.

Aside from the problem of inadequate coverage, there is the additional problem of inaccurate results. The blood samples are taken back to the laboratories that may be located a considerable distance from the hamlets visited, a delay that can often render the slides unusable. Often, especially when malaria transmission is at its peak, there is a backlog of slides because of shortage of technicians to examine these. This may take six to eight weeks to clear and the unexamined blood slides are not counted in calculating malaria incidence. Yet another source of error comes from the fact that at the time of blood collection a person who had malaria two weeks ago may not test positive for malaria—the microscopic examination may not show a positive case of parasitaemia in the peripheral blood. On several occasions, people who did not have fever may have had malaria, but this sample is not covered at all. Coupled with these problems there may also be poor training of the technicians in examining the slides. In a study undertaken in Orissa, it was observed that the staff in laboratories at the primary health centres often failed to detect malaria parasites in blood smears that contained them (Yadav et al. 2003).

The malaria estimates of the NMEP can be cross-checked on the basis of consumption of anti-malarial drugs, mainly chloroquine. The yearly production of chloroquine phosphate base in India is 600 metric tonnes. Of this 150 metric tonnes is exported to neighbouring countries. One metric tonne of the chemical yields four million tablets of 150 mg each or 2,400 million tablets on average per year. Of these, 300 million are set aside for the drug distribution centres, primary health centres, and other bodies that come under the NMEP. The balance of 2,100 million tablets is available for treatment in the private sector alone. On average, the private sector uses seven tablets per malaria episode. This would mean that 300 million fever cases are treated outside the NMEP alone. Since these are fever cases, to calculate actual incidence the slide positivity rate is used. Assuming each of the fever cases was tested for malaria through a blood smear examination and using the SPR of 2.82 reported by the NMEP in 1991 there would be 8.46 million malaria cases. Adding to this number the 2.1 million reported by the NMEP generates a total of 10.57 million malaria cases. Another 50 percent may be attributed to those not recorded due to poor surveillance (and other factors) under the NMEP and the total adds up to approximately 15 million (Sharma 1996). Using similar calculations Pattanayak et al. (1994) also arrive at 10-15 million cases. In 1975 and 1980 the NMEP reported 5.2 and 2.9 million cases of malaria respectively, whereas the consumption of anti-malaria drugs alone in these years indicates at least 12 million malaria cases in 1975, or twice as high, and 20 million in 1980, or seven times higher, than that reported (Annon 1981 in Yadav et al. 2003). Isolated evidence in rural areas have, in fact, confirmed actual incidence to be much higher than that reported by the NMEP. Das et al. (2005) for western Uttar Pradesh and Lal et al. (1996) for Rohtak in Haryana have also reported much higher incidence than that reported by NMEP.

Table 3.1: Slide Positivity Rates (SPR) reported by different studies compared with PHC[a] data

Author	SPR reported by study	SPR reported using PHC data
Sharma et al (1983)	21.2	4.7
Sharma et al (1983)	43.2	12.6
Malhotra et al (1985)	58.7	5.3
Ghosh et al (1985)	26.3	7.6
Choudhury et al (1987)	5.5	1.5

Source: Sharma (1996); a. PHC: Primary Health Centre

The estimates calculated above may be regarded as conservative since several studies have contested the SPR reported by the NMEP using the primary health centre data. Some of these are reported in Table 3.2. Using a higher estimate of SPR would give a much higher incidence of malaria.

The malaria incidence reported by the NFHS may be considered comparable to the number of fever cases and not to the actual number of malaria cases, as mentioned above. Using the total Uttar Pradesh incidence rate (rural and urban combined) of 73 per thousand population and comparing this with the API of 0.840 per thousand population for the state indicates the former to be about 80 times higher than the API. However, the NFHS incidence ought to be adjusted using the SPR to arrive at the actual number of cases. The total number of households (rural and urban) surveyed by the NFHS in Uttar Pradesh is 10,110. Using the average SPR of 2.82 (for 100 slides) reported at the all-India level by the NMEP would then generate malaria incidence of 28.5 cases per thousand population. The API calculated using the NMEP data is 0.840 per thousand population indicating that the NFHS incidence is about 30 times higher and not a factor of 80 or higher. As discussed above, the SPR itself is debatable and cautious estimates would also approximate actual API to be at least 10 to 15 times higher than that reported by the NMEP.

A summary of the main points in the above discussion on the measurement of malaria incidence is provided below:

1. The estimates of malaria from the two sources - NFHS and NMEP - are not directly comparable because NFHS uses malarial fever cases whereas NMEP uses blood tests on actual fever cases. As such, the Uttar Pradesh incidence in 1992-93 was 73 per 1000 under the former and 0.890 per 1000 under the NMEP. The appropriate criterion for comparison of the two estimates should be malaria like fever cases.

2. The NMEP reporting is considered inaccurate. Sharma (1999) suggests that the actual malaria incidence is 15 times higher.

3. The reasons why NMEP malaria cases are underreported:

    (a) Number of slides collected for a blood exam are not sufficient and do not meet the WHO requirement of a 10 percent annual blood exam rate (ABER). In Uttar Pradesh it has been reported to be as low as 0.57 percent in some districts.

    (b) Missing out on malaria cases when collecting blood smears.

    (c) Inaccurate testing so that those who had malaria are not tested positive resulting in underreporting. (It is also possible for those who had malaria a long time back to also test positive because of some parasitaemia in the blood though this is not likely to be a high number of cases.)

(d) Many slides are rendered unusable because they are not sent to testing centres in time and in peak seasons there is a backlog of slides many of which are wasted.

(e) Both the above factors contribute to a low SPR. As noted in Table 3.1, the SPRs reported in independent evaluations can differ by wide margins from those reported by the NMEP. In Malhotra et al's (1985) study the SPR reported is as high as 58.7 whereas the corresponding NMEP SPR was only 5.3.

An attempt was made to calculate a comparable estimate of malaria cases for the rural Uttar Pradesh NFHS sample and is repeated in the table below. There is no doubt that a large difference in the malaria incidence still remains to be explained but this exercise is only meant to draw attention to the fact that the margin of error falls down considerably and does not expect to result in an accurate assessment of the error.[2]

Table 3.2: Discrepancy in NFHS and NMEP reported malaria cases

NMEP malaria cases	*NFHS malaria cases*	*Discrepancy*
0.890 per 1000 population (reported based on blood smears)	73 per 1000 population (report based on fever cases)	NFHS 80 times higher
	28.5 per 1000 population (calculated based on SPR=2.82)	NFHS 30 times higher

Thus, the Uttar Pradesh incidence of 73 per thousand seems high especially when (erroneously) compared with the API. But, on the basis of the above-discussed calculations and the very high likelihood of underreporting by the NMEP, the cases reported in the NFHS data are comparable with the official figures with a considerable margin of error in the NMEP reporting. The NFHS data gives us the number of fever cases suspected of malaria since it uses the method of recall asking a household head whether he/she had

[2]The implications in terms of estimation for the state-level analysis in Chapter 7, which uses the NMEP data of malaria incidence as the dependent variable, is that estimated standard errors will be inflated upwards, affecting the inferences there. Thus, the measurement error would affect the efficiency of the estimates. On the other hand, for binary dependent variable models, i.e., the probit in our case (Chapters 5 and 6), there could be a mis-classification of the dependent variable (the NFHS data). The resultant maximum likelihood estimates may be subject to bias, with even a very small incidence of mis-classification yielding sizable bias in the estimated effects. However, it is not possible to address these issues given the data we have exhibit such a problem as it is hard to measure malaria incidence accurately.

malaria in the last three months. It has been standard practice to use malaria-like fever cases to arrive at estimates of malaria in the past such as in Italy (Snowden 2006). Further, the advantage of using the NFHS data is that they contain information on socio-economic variables at the household level enabling us to analyse how such factors influence malaria incidence.

We now turn to describing the sampling method followed in the NFHS and the summary statistics for relevant variables of the Uttar Pradesh rural and urban samples for 1992-93 and 1998-99. Malaria incidence by settlement types and time periods in Uttar Pradesh are also presented here.

3.2 Household data

Uttar Pradesh, the state selected for the current analysis, is divided into five geographic regions – Hill, Western, Central, Eastern, and Bundelkhand. Each of these had two domains, urban and rural, so that the total number of sampling domains were 10. Within these domains, the IIPS (1995) used a systematic, multi-stage stratified sampling procedure. In rural areas the first stage involved selecting the primary sampling units (PSUs), which are villages or clusters of villages. The sampling frame was the 1991 census list of villages. The list was stratified by various variables. The first level involved assigning districts to the five regions. The next level of stratification was by village size, percentage of population from scheduled castes or tribes, percentage of males in non-agricultural activities, and female literacy rate. From this ordering the villages were then selected systematically with probability proportional to the population size of the village. At least 50 households comprised a village or a group of villages. In the second stage, the households were selected using systematic sampling within the PSU. In urban areas, in the first stage, wards were selected with probability proportional to population size from the 1991 census list arranged by district and within the district by the level of female literacy. Within each ward, one census enumeration block (with 150-200 households) was selected with probability proportional to size and, finally, households were selected using systematic sampling within this block. An average of 30 households were selected per village or block.

Nearly 80 percent of the population of Uttar Pradesh resides in rural areas as reflected in the rural bias of the NFHS sample—the number of rural households surveyed was

7,795 whereas the urban sample was 2,315 households.[3] The analysis has been conducted separately for the rural and urban areas because of reasons discussed in the first two chapters. The data description also deals separately with the rural and urban samples.

It may be noted that the focus of the analysis is on heads of households and not all members in the household as the quality of the malaria incidence data for other members of the households is suspect. As IIPS (2000) notes, the questions were addressed to the heads of households or another informed member in the absence of the head. In many instances, the head was not aware if another household member had suffered from malaria. Support for this discrepancy also comes from the malaria incidence reported by the NFHS based on incidence for the entire household at 73 per thousand population for Uttar Pradesh in 1992-93, very close to that calculated for heads of households only for this study at 71 per thousand.

The NFHS data were examined to select variables relevant to the study of malaria incidence (see Table 3.3).[4] On the basis of the summary statistics for these variables, it was decided which variables should be used as continuous measures and which as categorical. All relevant variables were assessed for missing values and, when encountered, corresponding observations were dropped. The cleaning procedure described here for the different variables selected for analysis was compatible for both 1992-93 and 1998-99. The terms in parentheses in bold typeface indicate the mnemonics used for the variables. For some variables preliminary regressions were run to decide on further categorisations or cleaning and this is indicated below.

Gender

The code '1' was used for male and '2' for female in the original data. The latter was converted to '0' to make a dummy variable for whether the head of household was male or not (**genderh**).

[3]The population of India in 1991 was 846 million and that of Uttar Pradesh was 139 million. The rural population of Uttar Pradesh was 111.5 million (GoI 2000).

[4]This table also presents the district level variables selected and will be referred to later in this chapter when discussing the district statistics. The variables listed in this table are those that were used in the final regressions.

Table 3.3: Explanatory variables used to model malaria incidence

Explanatory variables	*Variables from NFHS*	*Variables from Census data*
	Physical (natural) environment	
Climate, ecology	Not available; district dummies to capture these 'unobservables'	District-level rainfall;[a] percentage of district area under forest cover
	Biological environment	
Resting and breeding areas of the malaria vector		
Rainwater puddles, forest pools, and other water bodies	Not available; district dummies used;[b] also see 'proximity to breeding sites' below	Percentage of district area under forest; average rainfall; percentage of district area under different irrigation systems
Irrigated fields	Ownership of irrigated and non-irrigated land	
Hoof prints	Ownership of livestock	
	Host (human) environment	
Developmental work such as irrigation works or dam construction; level of socio-economic development	Not available; captured in district dummies	Percentage of district area irrigated by different irrigation systems; population density, percentage of villages in district with primary, middle, high school, and adult literacy centre;[c] commuting facility and power supply for agriculture and domestic use
Proximity to breeding sites	Source of drinking water and fuel; ownership of irrigated and non-irrigated land	Average distance to drinking water source; percentage of district area irrigated by different irrigation types; percentage of district area under forest
Socio-economic status, household characteristics, living conditions		
i. Income	Asset-wealth variables[d]	
ii. Occupation	Occupation of head	
iii. Education	Level of education of head	
iv. Type and features of house	Type of house head owns; whether house has separate kitchen[e]	
v. Living conditions	Whether household has electricity or not; whether household has access to sanitation	
vi. Social status, ethnicity and other characteristics	Whether head belongs to a scheduled caste or tribe; gender and age of household head	

a. Temperature was not used as district level temperature is available on a monthly basis whereas the NFHS covers malaria incidence in the last three months and the documentation does not indicate specific months in which different districts were surveyed. Altitude may be considered a proxy for temperature, but the coefficient on altitude was not well determined. b. Overlaps with climate and ecology. c. Data on percentage of villages with a health centre was used in the preliminary regressions, but the effect was not well determined. d. Education, occupation, and the ownership of irrigated and non-irrigated land and livestock (included under biological environment) also capture income, but these have been entered separately because the literature on malaria mentions the influence of these variables directly on malaria incidence. e. The variables whether animals are kept inside the house at night, number of rooms and number of members also indicate living conditions but the coefficients on these variables were not significant.

Caste

The caste variable had three categorisations: scheduled caste, scheduled tribe, and other.[5] Preliminary regressions with dummy variables for each of these categories did not generate well determined coefficients. Therefore, the dummy for whether or not a person is a scheduled caste (SC) or scheduled tribe (ST) was used with the code '1' when the head belonged to an SC or ST and '0' otherwise (**caste**).

Age

This variable was available as a continuous measure. However, an examination of the data revealed spikes at multiples of five and 10. Therefore, the data were regrouped into class intervals chosen such that the multiples of five and 10 were included in each interval. The less than 18 years age group was dropped given our focus on the 18-plus age-group. It should be stressed that these constituted less than one percent of the sample. The categories constructed were: 18 to 27 years (**age18-27**); 28 to 37 years (**age28-37**); 38 to 47 years (**age38-47**); 48 to 57 years (**age48-57**); 58 to 67 years (**age58-67**); and more than 67 years (**age-ab67**). Each of these was treated as mutually exclusive dummy variables with a dummy for whether the head belonged to that age group or not.

Education

Education level of the household head is available as a categorical variable in the NFHS dataset. The education variable chosen for the analysis from the NFHS is available as a 're-coded' education variable with four codes for four levels of education chosen: (1) illiterate (**eduh-ill**); (2) literate or completed primary or lower level (**eduh-pri**); (3) completed middle school (**eduh-mid**); and (4) completed high school or higher level (**eduh-hi**). These were converted into four mutually exclusive dummy variables.

Occupation

Occupation of household head originally had 16 categories, which were: high level professional; low-level professional; in administration/executive/managerial duties; clerk; sales worker; service worker; farming, fishing, hunting and gathering; production and trans-

[5]The 'other' category included all other castes and people belonging to those religions that do not follow the caste system – Muslims, Christians, and Sikhs.

portation; household duties; student; retired; beggar; inmate of an institute; disabled; unemployed; other; and 'don't know'. From these categories the following were dropped: inmate of an institute, beggar, and 'don't know', which constituted less than one percent of the sample. The remaining categories were combined into the following four groups: (1) head works in the agriculture sector (farming, fishing, hunting and gathering) (**OCagr**); (2) wage earner (sales, clerical, other services) (**OCwage**); (3) production and transportation (**OCptran**); (4) unemployed,[6] student, in house duties (or domestic chores), retired, disabled, and 'other' (**OCother**).

House-type

The NFHS categorised house-type into *kachcha* (**houskach**), semi-*pucca* (**houssemi**), and *pucca* (**housepuc**). These were converted to three dummy variables for each category. A *kachcha* house is a hut made of temporary materials, a semi-*pucca* house is one with a temporary wall, ceiling or floor, and a pucca house is a permanent structure constructed of such materials as bricks (IIPS 1995).

Separate kitchen

This variable was already available as a dummy variable for whether a household had a separate kitchen or not (**kichsep**).

Electricity

This variable was also available as a dummy for whether a household had electricity ('1') or not ('0') (**electr**) and did not require any cleaning.

Sanitation

Sanitation or type of toilet facility as specified in the NFHS had eight categories. These were collapsed into a dummy for whether or not a household had a sanitation facility (**sanitdum**). The following were included under 'had a sanitation facility': private flush toilet; shared flush toilet; public flush toilet; private pit toilet/latrine; shared pit toilet/latrine; public pit toilet/latrine; and 'other'. The category 'no facility, use open areas' was included under 'no sanitation facility'.

[6] An unemployed person is defined as one who is not working but is seeking or is available for work.

Cooking fuel

The NFHS reports nine fuel categories. These were combined to obtain three mutually exclusive dummy variables for: (1) miscellaneous (coal, charcoal, kerosene, electricity, liquefied petroleum gas (LPG) , biogas, other) (**fmisc**), (2) wood (**fwood**), and (3) dung (**fdung**).

Drinking water source

The drinking water source variable had 13 categories: piped into residence; public tap; well in residence; public well; hand pump in yard/plot; public hand pump; spring; river or stream; pond or lake; dam; rainwater; tanker truck; and other. These categories were re-grouped into four: private protected or **wpvtprot** (piped into residence); open private source or **wpvtopen** (well or hand pump in residence); protected public source or **wpubprot** (public well or public tap);[7] and open public source or **wpubopen** (i.e., rivers, springs, ponds and dams, tanker truck, and other). Corresponding to these, four dummy variables were constructed. The groups made here are based on the logic that an open source is more likely to breed mosquitoes as opposed to a 'closed' or protected source. With that logic 'tanker truck'[8] is also included in the open public category since mosquitoes can breed inside tanks.

Irrigated and non-irrigated area

The relevant agricultural land variables in the NFHS are: whether a household owns agricultural land or not (**ownagr**), size of irrigated area (**szirr**), and size of non-irrigated area (**szunirr**).

In order to ensure that there are no contradictions, that is, even though a person does not own land, there is a figure reported under either or both of **szirr** and **szunirr**, these were replaced with the value zero whenever a household did not own any agricultural land. Finally, any further observations with missing values were excluded from the analysis.

Rather than use irrigated and non-irrigated area as continuous variables these were categorised because the category less than one acre of land is not reported precisely and it

[7] A public well is a government-owned tube well provided for drinking water. (Source: Personal communication with Asim Mirza, Field Analyst, The Energy and Resources Institute, New Delhi.)

[8] This is a large tank of water carried on a truck from which water is delivered to areas lacking water supply. It is usually supplied by the government. (Source: Personal communication with Asim Mirza, Field Analyst, The Energy and Resources Institute, New Delhi.)

takes care of extreme values since for irrigated area these were included under the category above five acres and for non-irrigated area above 10 acres. The following categorisations have been used:

For irrigated area: No irrigated area (**irr-0**); less than one acre (**irr-L1**); one or two acres (**irr-1t2**); three, four, or five acres (**irr-3t5**); more than five and less than or equal to 10 acres (**irr-5t10**); more than 10 acres (**irr-ab10**)

For non-irrigated area: No non-irrigated area (**unir-0**); Less than one acre (**unir-L1**); one acre (**unir-1**); two acres (**unir-2**); three, four, or five acres (**unir-3t5**); more than five acres (**unir-ab5**).

Additionally, dummy variables for ownership of both irrigated and non-irrigated area were also constructed. These are: does not own any irrigated or non-irrigated land (**ir0unir0**); owns less than one, equal to one or equal to two acres irrigated with less than one or one acre non-irrigated land (**irt2niL2**); owns less than or equal to one or two acres irrigated with more than or equal to two acres non-irrigated (**irt2ni2a**); owns more than or equal to three acres irrigated with owns less than or equal to one acre non-irrigated (**ira2niL2**); owns more than or equal to three acres irrigated with more than or equal to two acres non-irrigated (**ira2ni2a**) land.

For the 1998-99 dataset, instead of using a split category for irrigated and non-irrigated area, as explained above for the 1992-93 data, a dummy for the ownership of agricultural area was used (one for when a household owned agricultural land and zero when a household did not own any agricultural land). A dummy was used instead of a more disaggregated variable because in this dataset there seems to be a mistake in reporting the less than one acre category of land with a negligible number of households reported under this heading. Given that there have not been any radical land reforms over the period 1992-93 and 1998-99 this is not possible and can only be explained in terms of misreporting and it is likely that these households have been reported in another land category.[9]

Agricultural assets

The descriptive statistics of whether a household owns a tractor, thresher, bullock cart or water pump (**owntract**, **ownthres**, **ownbullc**, **ownwpump**, respectively) showed that

[9]Communication with the Indian Institute of Population Studies, which administers the survey, did not throw any light on this issue either.

many households that did not own agricultural land did not give a response to whether they owned any of these assets or not. These appear as missing values and were replaced with ‘0’ to indicate that these households do not own these assets. It is possible that a person who does not own agricultural land owns these for renting out purposes. But this outcome is unlikely and it was assumed that those heads of households who did not own agricultural land did not own these assets either. While all these assets were included in the preliminary regressions, the final regression used only the variable **ownwpump**.

Non-agricultural consumer durable asset wealth variables

These include the ownership of the following assets: radio, bicycle (**bicyc**), clock, fan, sewing machine (**sewmach**, television (**TV**), VCR or VCP (**VCR**), sofa, motorbike (**mobike**), car, and refrigerator (**refrig**). No cleaning was required in the case of these assets except to exclude observations with missing values. All of these assets are dummy variables for whether a household owns a particular asset or not. These variables were collapsed into an asset wealth index using Principal Components Analysis, which is discussed in a subsequent section.

The variables number of members and rooms, ownership of livestock, and whether animals are kept inside the house were also included in preliminary regressions. However, since these did not have well determined effects, they were not included in the final regressions.

We now present the summary statistics for these variables. The next two subsections present these for the rural samples for 1992-93 and 1998-99 and the subsequent subsections for the corresponding urban samples. To test whether the differences in variable means across the two time periods and the two settlement types are statistically significant or not, the following test statistics were used: (1) χ^2 for categorical variables (2) z-test for proportions for binary variables, and (3) t-test for continuous variables.

3.2.1 Rural Uttar Pradesh: 1992-93

Five districts (Chamoli, Pithorgarh, Hardwar, Kheri, and Lucknow) were dropped due to perfect classification[10] in terms of malaria incidence (an issue explained in the methodology chapter) and the final usable sample comprised of 7,287 households. Table 3.4 presents

[10]This was found on running preliminary regressions using the relevant variables.

the summary statistics for the relevant variables. Some of these pertaining to the main characteristics of the rural households are discussed here.

The malaria incidence for the cleaned 1992-93 rural sample is estimated at 8.8 percent.[11] The household heads were mostly male with only one-tenth female heads. Twenty-three percent of the heads of households belonged to the age-category 28 to 37 years and the percentage of those who fell in the 38 to 47 years bracket was higher by two percentage points. Half the heads were illiterate and the largest proportion (0.60) worked in farming and related activities followed by production and transport with less than 20 percent.

The proportion of scheduled caste (SC) or scheduled tribe (ST)[12] households was 0.22. The scheduled castes (also referred to as *dalits* or 'outcastes') have traditionally not been included in the Hindu caste system even though their culture is essentially that of the Hindu community. The caste status is inherited at birth and is, thus, imposed on them. They were traditionally considered 'untouchable' being regarded as causing pollution on physical contact because of the nature of their work (performing menial tasks like sweeping). A community is listed as belonging to the scheduled caste if it suffers extreme social, education and economic deprivation because of the practice of 'untouchability'. SCs were traditionally not allowed to raise their status such as through prohibition on land ownership or on wearing clothes that could make them appear to be of a higher status. STs, on the other hand, are those groups of people who are considered to exhibit 'primitive'[13] traits, a distinctive culture, geographic isolation, and shyness of contact with the community at large. They are also referred to as *adivasis* (literally, original inhabitants) and are non-agricultural communities that dwell mainly in forest areas away from caste settlements and live on wild fruits, nuts, and roots and also resort to fishing and hunting for food. SC and ST communities are given special status under the Constitution of India as they are considered to be among the weaker sections of society. The Constitution of India forbids caste/tribe discrimination and given the low traditional status accorded to SCs and STs, it directs the State, under Article 46 and Articles 338 and 338A, to promote the educational and economic interests as well as the social development

[11] This estimate is different from the one mentioned in Section 3.1 at 7.1 percent because of differences in sample size resulting from the cleaning procedures discussed in the preceding section. The earlier estimate is based on the original NFHS sample size of 10,110. The estimates presented here are for 7,287 households.

[12] Those heads of households who are not SC/ST could be Hindu belonging to other castes or could belong to any of the other religious groups (such as Muslim, Sikh, or Christian) found in India and that do not follow the Hindu caste system.

[13] This word is conventionally used by the Ministry of Law and Justice to define STs (GoI 2006b).

Table 3.4: Summary statistics[a] of variables from NFHS, rural Uttar Pradesh, 1992-93

Variable	*Variable description*	*Mean (SD)*[b]
malhead	Head suffered from malaria in last 3 months (0: No; 1: Yes)	0.0878
Social characteristics		
genderh	HH[c] head is male (0: No; 1: Yes)	0.9271
caste	Head belongs to SC/ST (0: No; 1: Yes)	0.223
age18–27	Age of head >=18 and <=27 years (0: No; 1: Yes)	0.0937
age28–37	Age of head >=28 and <=37 years (0: No; 1: Yes)	0.229
age38–47	Age of head >38 and <=47 years (0: No; 1: Yes)	0.2411
age48–57	Age of head >=48 and <=57 years (0: No; 1: Yes)	0.1846
age58–67	Age of head >=58 and <=67 years (0: No; 1: Yes)	0.1614
age–ab67	Age of head >67 years (0: No; 1: Yes)	0.0902
Education categories		
eduh–hi	Head completed high school or higher level (0: No; 1: Yes)	0.1402
eduh–mid	Head completed middle school (0: No; 1: Yes)	0.1009
eduh–pri	Head completed primary school (0: No; 1: Yes)	0.2097
eduh–ill	Head is illiterate (0: No; 1: Yes)	0.5492
Occupation categories		
OCagr	Head works in the agricultural sector (farming) (0: No; 1: Yes)	0.5964
OCwage	Head is a wage-earner (s.a. sales worker, clerical and other services) (0: No; 1: Yes)	0.1169
OCother	Head is unemployed, a student, in house duties, retired, disabled, or 'other' (0: No; 1: Yes)	0.1025
OCptran	Head works in production and transportation (0: No; 1: Yes)	0.1842
House type categories		
housepuc	Lives in a pucca house (0: No; 1: Yes)	0.0829
houssemi	Lives in a semi-pucca house (0: No; 1: Yes)	0.2937
houskach	Lives in a kachcha house (0: No; 1: Yes)	0.6234
Living conditions		
no–mem	Household size	6.6027 (3.51)
rooms	No. of rooms in the house	2.8503 (2.14)
kichsep	HH uses separate room as kitchen (0: No; 1: Yes)	0.3766
electr	HH has electricity (0: No; 1: Yes)	0.1967
sanitdum	House has access to sanitation (0: No; 1: Yes)	0.0755
Cooking fuel categories		
fmisc	Main fuel source is 'misc' (0: No; 1: Yes)	0.0231
fwood	Main fuel source is wood (0: No; 1: Yes)	0.7937
fdung	Main fuel source is dung (0: No; 1: Yes)	0.1832
Drinking water source categories		
wpvtopen	HH uses open private source for drinking water (0: No; 1: Yes)	0.0464
wpvtprot	HH uses protected private source for drinking water (0: No; 1: Yes)	0.3901
wpubopen	HH uses open public source for drinking water (0: No; 1: Yes)	0.0531
wpubprot	HH uses protected public water source (0: No; 1: Yes)	0.5104
Irrigated area categories		
irr–0	HH does not own any irrigated land (0: No; 1: Yes)	0.3174
irr–L1	HH owns less than 1 acre irrigated area (0: No; 1: Yes)	0.2389
irr–1t2	HH owns 1 or 2 acres irrigated area (0: No; 1: Yes)	0.2321
irr–3t5	HH owns 3, 4, or 5 acres irrigated area (0: No; 1: Yes)	0.1149
irr–5t10	HH owns more than 5 and less than or equal to 10 acres of irrigated area (0: No; 1: Yes)	0.0622
irr–ab10	HH owns more than 10 acres irrigated area (0: No; 1: Yes)	0.0346

Continued on next page...

Table 3.4 continued

Variable	*Variable description*	*Mean (SD)*[b]
Non–irrigated area categories		
unir–0	HH does not own any non–irrigated land (0: No; 1: Yes)	0.7642
unir–L1	HH owns less than 1 acre non–irrigated area (0: No; 1: Yes)	0.104
unir–1	HH owns 1 acre non–irrigated area (0: No; 1: Yes)	0.0467
unir–2	HH owns 2 acres non–irrigated area (0: No; 1: Yes)	0.0263
unir–3t5	HH owns 3, 4, or 5 acres non–irrigated area (0: No; 1: Yes)	0.036
unir–ab5	HH owns more than 5 acres non–irrigated area (0: No; 1: Yes)	0.0228
Irrigated and non–irrigated area categories		
ir0unir0	Does not own any irrigated or non–irrigated land (0: No; 1: Yes)	0.2054
irt2niL2	Owns less than 1, equal to 1 or or equal to 2 acres irrigated with less than 1 or 1 acre non–irrigated (0: No; 1: Yes)	0.0696
irt2ni2a	Owns less than or equal to 1 or 2 acres irrigated with more than or equal to 2 acres non–irrigated (0: No; 1: Yes)	0.0222
ira2niL2	Owns more than or equal to 3 acres irrigated with owns less than or equal to 1 acre non–irrigated (0: No; 1: Yes)	0.0107
ira2ni2a	Owns more than or equal to 3 acres irrigated with more than or equal to 2 acres non–irrigated (0: No; 1: Yes)	0.0213
Agricultural assets		
ownwpump	HH owns a water pump (0: No; 1: Yes)	0.1029
ownbullc	HH owns a bullock cart (0: No; 1: Yes)	0.0999
owntract	HH owns a tractor (0: No; 1: Yes)	0.0247
ownthres	HH owns a thresher (0: No; 1: Yes)	0.0386
ownlive	HH owns livestock (0: No; 1: Yes)	0.8297
Non agricultural assets		
radio	HH owns a radio (0: No; 1: Yes)	0.2758
bicyc	HH owns a bicycle (0: No; 1: Yes)	0.5389
clock	HH owns a clock or a watch (0: No; 1: Yes)	0.457
fan	HH owns a fan (0: No; 1: Yes)	0.0988
sewmach	HH owns a sewing machine (0: No; 1: Yes)	0.1482
TV	HH owns a TV (0: No; 1: Yes)	0.0689
VCR	HH owns a VCR or VCP (0: No; 1: Yes)	0.0102
sofa	HH owns a sofa set (0: No; 1: Yes)	0.0298
mobike	HH owns a motor bike (0: No; 1: Yes)	0.0306
car	HH owns a car (0: No; 1: Yes)	0.0022
refrig	HH owns a refrigerator (0: No; 1: Yes)	0.0078

a. Statistics based on a total number of observations of 7287 rural Uttar Pradesh households; b. SD: Standard deviation (presented in parentheses for continuous variables); c. HH: Household.

of the SCs and STs (GoI 2006a, GoI 2006b, Ambedkar Unpublished).

The main fuel the households used was firewood with 80 percent of the households having reported using wood, nearly a fifth used dung and only two percent used 'miscellaneous' sources that include gas and electricity. Most of the houses (a little over 90 percent) did not have access to a toilet/sanitation facility. Only about a fifth of the respondents had electricity in their house, 62 percent lived in a *kachcha* house, about a third in a *semi-pucca* house and a little under 10 percent in a pucca house.

Among the non-agricultural consumer durable assets, more than half the households owned a bicycle and about 50 percent also owned a clock. The next most-owned asset was a radio at about a third of the sample. A tenth of the households also owned a fan, 15 percent owned a sewing machine and the ownership of TVs stood at seven percent. Cars, refrigerators, and VCRs/VCPs were owned by less than one percent of the sample and sofas and motorbikes by about three percent. An index was constructed using these assets and this is discussed in the next section as noted earlier.

3.2.2 Rural Uttar Pradesh: 1998-99

The incidence of malaria for comparable rural samples is 5.1 percent for 1998-99 and 9.6 percent for 1992-93 (see Table 3.5). Six districts were not covered in this dataset—Garhwal, Almora, Mathura, Lucknow, Jaswantnagar, Sidhdharthanagar, and Gorakhpur. Of these, Lucknow was dropped due to perfect classification in the 1992-93 data. The districts Chamoli, Pithorgarh and Hardwar were also dropped due to perfect classification (predicted failure perfectly) in both datasets. The 1998-99 data also had seven new districts not covered in 1992-93. These were dropped from the 1998-99 specification. The comparable specification included 42 districts in all (rather than 48 as was the case for the earlier specification for 1992-93). The remaining sample consisted of 6,071 households (rather than 7,287). For 1998-99 the usable sample consisted of 4,659 households. Comparable rural specifications were then obtained for the two years and we now discuss the summary statistics for these.

The summary statistics for comparable specifications of the rural sample are presented in Table 3.5. The percentage of illiterates fell by four percentage points and all other education categories show a slight increase. The percentage of heads in agriculture fell by eight percentage points and that in production and transport as well as the ‘other’ category increased by five. The proportion living in *kachcha* houses fell and that in *semi-pucca* rose. A slightly smaller proportion have a separate kitchen. There is a four percentage point increase in access to sanitation and an eight percentage point increase in having protected private water supply. The percentage owning irrigated area rose by six percentage points although the overall ownership of agricultural land fell. A larger proportion of households owned a thresher in 1998-99 at 0.06 compared to 0.04 for 1992-93. For this set of variables (excluding the consumer durable assets), we reject the null of no difference in means

Table 3.5: Summary statistics for comparable rural samples of 1992-93 and 1998-99

Variable	*1992-93 Mean (SD)*[a]	*1998-99 Mean (SD)*[a]	χ^2*, z and t-stat*[b]	*Critical values*[c]
malhead	0.0955	0.0514	8.48***	±1.96
Social characteristics				
caste	0.2143	0.2408	-3.24***	±1.96
genderh	0.9308	0.9197	2.17***	±1.96
Age categories				
agehead	45.83	45.84	-0.03	±1.96
	(14.71)	(15.11)		
age18–27	0.0957	0.0976	26.96***(5)	11.07
age28–37	0.2336	0.2437		
age38–47	0.2461	0.2435		
age48–57	0.1792	0.1628		
age58–67	0.1588	0.1489		
age–ab67	0.0867	0.1035		
Education categories				
eduh–hi	0.1436	0.1563	37.43***(3)	7.82
eduh–mid	0.1034	0.1212		
eduh–pri	0.2102	0.2221		
eduh–ill	0.5427	0.5004		
Occupation categories				
OCagr	0.5828	0.5019	198.17***(3)	7.82
OCwage	0.1204	0.1088		
OCother	0.1005	0.1434		
OCptran	0.1963	0.2459		
House-type categories				
housepuc	0.0937	0.0946	41.18***(2)	5.99
houssemi	0.3146	0.3572		
houskach	0.5917	0.5482		
Cooking fuel categories				
fmisc	0.0267	0.0972	887.40***(2)	5.99
fwood	0.7776	0.7029		
fdung	0.1957	0.1999		
Drinking water source categories				
wpvtprot	0.4149	0.5002	222.84***(2)	5.99
wopen	0.0458	0.0081		
wpubprot	0.4910	0.4917		
Other characteristics				
kichsep	0.3770	0.2918	9.19***	±1.96
electr	0.2003	0.2095	-1.17	±1.96
sanitdum	0.0782	0.1197	-7.20***	±1.96
Agricultural land and agricultural assets				
ownagr	0.7926	0.7576	4.30***	±1.96
ownsirr[d]	0.6923	0.7572	-7.39***	±1.96
ownwpump	0.1141	0.1057	1.37	±1.96
ownbullc	0.0908	0.1000	-1.61	±1.96
owntract	0.0280	0.0379	-2.86***	±1.96
ownthres	0.0387	0.0636	-5.87***	±1.96
ownlive	0.8208	0.7705	6.42***	±1.96
Non-agricultural consumer-durable assets				
radio	0.2784	0.2650	1.53	±1.96
TV	0.0768	0.1463	-11.51***	±1.96
bicyc	0.5462	0.6245	-8.10***	±1.96
sewmach	0.1586	0.1580	0.08	±1.96
clock	0.4640	0.5211	-5.84***	±1.96
fan	0.1071	0.1881	-11.87***	±1.96
refrig	0.0091	0.0146	-2.67***	±1.96
mobike	0.0338	0.0423	-7.48***	±1.96
car	0.0023	0.0048	-0.67	±1.96
No. of households	6071	4588		

a. SD: Standard deviation (presented in parentheses for continuous variables); b. This column presents the χ^2-statistic for categorical, z-statistic for proportions, and t-statistic for continuous variables; the degrees of freedom for the χ^2 statistics are reported in parentheses; c. All critical values are reported for the 5 percent level of significance; d. ownsirr: whether household owns irrigated land or not. For all other abbreviations, refer to Table 3.4. * significant at 10%; ** significant at 5%; *** significant at 1%.

across the two time periods at the five percent level of significance for all variables except **agehead**, **ownwpump**, and **electr**.

More than 50 percent of the households owned bicycles and there was an increase of eight percentage points over the time period considered from 54 to 62 percent. The other asset that was owned by many rural households in both time periods was a clock – a little less than half the households in 1992-93 and just over 50 percent in 1998-99. The ownership of televisions doubled over this six-year period from about seven to 14 percent. Fans were also owned by a larger number of households in the second phase of the NFHS – an increase of eight percentage points compared to 1992-93. The ownership of motorbikes did not show much of a rise and cars and refrigerators continued to be owned by less than one percent of the sample. For all these asset wealth variables except radio, **sewmach**, and car, we reject the null of no difference in means. The asset wealth index is discussed in the next section. We now turn to a discussion of the urban statistics.

3.2.3 Urban Uttar Pradesh: 1992-93

The sample used to compare rural against urban Uttar Pradesh comprised of 6,582 households after making the changes necessary to match variables in the two samples (see Table 3.6). The urban summary statistics for 1992-93 are presented in Table 3.7. The rural statistics are also presented in order to refer to the differences across the two settlement-types.

The malaria incidence in urban Uttar Pradesh was 2.39 percent (six percentage points lower than rural malaria) in 1992-93. Nine percent of the heads of households in urban areas belonged to SC/ST as compared to a fifth in the rural settlement type. The heads of households were mostly male. The distribution across age categories showed a similar pattern for urban and rural with a slightly higher percentage in the above 57 years group for rural (25 percent) as compared to urban areas (18 percent). The occupation status was dominated by heads working as wage earners in the urban settlement type—nearly half the total. In the rural settlement type, as expected, heads worked predominantly in farming. Half of the urban heads of households had completed a high school level of education as compared to only 12 percent in the rural areas and a much smaller proportion at 0.27 were illiterate as compared to rural areas where nearly 60 percent were not literate. Most urban heads (67 percent) lived in *pucca* houses whereas most rural houses (about three-fifths)

Table 3.6: Differences in rural and urban specifications (1992-93)

Variable	*Variables adjusted to enable rural and urban comparison based on the urban regression*
Age	In the urban regression, the 18 to 27 years age-group was dropped due to perfect prediction of failure and combined with the 28 to 37 years group for both samples. The age group 58 to 67 was conflated with the above-67 group because there were only six percent of heads in the latter category in urban areas.
Drinking water source	In the rural specification the categories were: public protected and public open, private protected and private open. In urban UP since only 0.81 percent of the households owned a private well this was combined with public open and private open to read as one category: open source. Thus, the comparable specifications included three categories: public protected, private protected and open (to include both private and pubic sources).
Land	Ownership of agricultural land rather than the split up variables of ownership of irrigated and non-irrigated area were used. The main reason for this was that in urban areas almost all the land owned is irrigated with a negligible proportion of households of 0.005 owning non-irrigated area.[a]
Agricultural assets	Ownership of a water pump,[b] thresher, and bullock cart were dropped due to perfect prediction of failure and the assets ownership of tractor and livestock were dropped due to collinearity. The mean value of these assets was less than one percent that is, less than one percent of the households owned each of these assets with the exception of water pump and livestock with mean values of 1.5 percent and 17 percent,[c] respectively.
Districts	Seven of the districts were perfectly classified (perfectly predict failure) and several districts could not be used in the analysis as they did not report any malaria incidence at all. Of 54 districts, 38 did not have any malaria. It was, therefore, decided to use regions rather than districts. UP has been classified into five regions: Central, Bundelkhand, East, West, and Hill. Central was used as the base in the regressions.

a. In rural areas the percentage of households owning non-irrigated area is 25 percent. b. Combining water pump with ownership of irrigated area did not resolve the problem as every household that owned a water pump also owned irrigated land. c. Livestock was not included in the rural specification either because it was found to perform poorly. Before dropping the other agricultural assets altogether different combinations of assets were collapsed into one dummy variable to indicate ‘whether the household owns any of these assets’. But these variables also performed poorly. The combinations were: livestock, water pump, thresher, tractor, and bullock cart (agricultural assets other than land); and irrigated area, non-irrigated area, livestock, and water pump. A third attempt included collapsing the non-agricultural assets into an index using Principal Components Analysis, which also performed poorly. In fact, the coefficient on the index of any combination of agricultural assets was not well determined.

Table 3.7: Characteristics of rural and urban Uttar Pradesh households (comparable specifications, 1992-93)

Variable	*Rural*	*Urban*	χ^2 *and z-stat*[a]	*Critical values*[b]
malhead	0.0845	0.0239	6.30***	±1.96
Social characteristics				
caste	0.222	0.0919	13.54***	±1.96
genderh	0.9193	0.9356	-2.50***	±1.96
agehead	45.9829	44.6487	3.94***	±1.96
	(14.79)	(13.46)		
Age categories			56.12***(3)	7.81
age18–37	0.3263	0.3293		
age38–47	0.2446	0.2955		
age48–57	0.1819	0.1878		
age–ab57	0.2472	0.1874		
Occupation categories			3604.18***(3)	7.81
OCagr	0.5767	0.0626		
OCwage	0.1194	0.4784		
OCptran	0.1966	0.3212		
OCother	0.1073	0.1351		
Education categories			2237.65***(3)	7.81
eduh–hi	0.1284	0.4581		
eduh–mid	0.0981	0.1077		
eduh–pri	5.9912	0.1604		
eduh–ill	0.5623	0.2739		
House-type categories			11514.46***(2)	5.99
housepuc	0.0744	0.6716		
houssemi	0.2905	0.1500		
houskach	0.6351	0.1784		
Cooking fuel categories			29071.41***(2)	5.99
fmisc	0.0232	0.5685		
fwood	0.7944	0.3703		
fdung	0.1823	0.0613		
Drinking water source categories			2025.47***(2)	5.99
wpubprot	0.5415	0.173		
wpvtprot	0.3555	0.8117		
wopen	0.103	0.0144		
Other characteristics				
sanitdum	0.0672	0.8041	-69.06***	±1.96
kichsep	0.3555	0.6126	-21.22***	±1.96
electr	0.1826	0.814	-54.24***	±1.96
ownagr	0.7777	0.1647	50.00***	±1.96
Regions				±1.96
Hill	0.1551	0.2122	-6.20***	±1.96
East	0.2607	0.1730	18.67***	±1.96
West	0.2788	0.4387	-13.29***	±1.96
Central	0.1808	0.1279	5.77***	±1.96
Bundelkhand	0.0295	0.0387	-1.76*	±1.96
No. of households	6582	2220		

a. This column presents the χ^2-statistic for categorical, z-statistic for proportions, and t-statistic for continuous variables; b. All critical values are presented for the 5 percent level of significance; the degrees of freedom for the χ^2 statistics are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

were *kachcha*. The main fuel source in the urban settlement type was the miscellaneous category. Little under three-fifths of the households used this category, which covers such fuels as liquefied petroleum gas and electric stoves. A large proportion (0.37) also used wood as their main fuel source but this was not the predominant fuel as is the case in the rural settlement type. The majority of urban households met their drinking water needs mostly from private protected sources, that is, through piped water supply in the house. In rural areas this category covered 35 percent of the households. Four-fifths of the urban households also had access to sanitation and electricity in contrast to less than one-tenth and one-fifth, respectively, in the rural settlement type. A very small proportion of households (0.15) owned land and nearly all of this was irrigated. In the rural settlement type ownership of land was a dominant feature with nearly 90 percent of the households owning land. It may be noted that for all variables we reject the null of no difference in means across the two settlement types at the five percent level of significance (see Table 3.7).

3.2.4 Urban Uttar Pradesh: 1998-99

The 1998-99 urban samples summary statistics are discussed with reference to a comparable specification used for 1992-93. Table 3.8 presents the summary statistics for comparable urban specifications and Table 3.9 for rural and urban specifications for 1998-99.

Comparing urban 1992-93 and 1998-99 samples

Malaria incidence was 2.4 percent in 1992-93 and dropped by one percentage point by 1998-99 (the difference in means is statistically significant at the five percent level). The percentage of heads of households belonging to SC/ST increased from nine percent to 15.5 percent from 1992-93 to 1998-99 and the percentage of female heads rose from six percent in 1992-93 to 11 in 1998-99. The distribution of heads across different age categories did not show much change. The proportion of heads working in farming reduced by nearly half over the six-year period from 0.06 to 0.03. Wage earners registered a slight fall of three percentage points and settled at 45 percent in 1998-99 and the percentage of heads in production and transportation fell by one percentage point. The disabled, household activities, unemployed, and the other category had the largest increase of a little over six percentage points and stood at 20 percent by 1998-99. The test statistics

Table 3.8: Summary statistics for comparable urban samples of 1992-93 and 1998-99

Variable	*1992-93 Mean (SD)*[a]	*1998-99 Mean (SD)*[a]	χ^2 *and z-stat*[b]	*Critical values*[c]
malhead	0.0239	0.0143	2.14**	±1.96
Social characteristics				
caste	0.0919	0.1506	-5.64***	±1.96
genderh	0.9356	0.8935	4.73***	±1.96
agehead	44.65	45.58	-2.11**	±1.96
	(13.46)	(13.86)		
Age categories				
age–37	0.3293	0.3137	2.65 (3)	7.82
age38–47	0.2955	0.2940		
age48–57	0.1878	0.1940		
age–ab57	0.1874	0.1982		
Occupation categories				
OCagr	0.0626	0.0351	85.04***(3)	7.82
OCwage	0.4784	0.4560		
OCptran	0.3212	0.3036		
OCother	0.1351	0.2054		
Education categories				
eduh–hi	0.4581	0.4631	5.95 (3)	7.82
eduh–mid	0.1077	0.1054		
eduh–pri	0.1604	0.1780		
eduh–ill	0.2739	0.2536		
House-type categories				
housepuc	0.6716	0.7369	75.89***(2)	5.99
houssemi	0.1500	0.1661		
houskach	0.1784	0.0970		
Living conditions				
sanitdum	0.8041	0.8232	-1.51	±1.96
kichsep	0.6126	0.5363	4.78***	±1.96
electr	0.8140	0.8649	-4.25***	±1.96
Cooking fuel categories				
fmisc	0.5685	0.6512	51.99***(2)	5.99
fwood	0.3703	0.3119		
fdung	0.0613	0.0369		
Drinking water source categories				
wpubprot	0.1541	0.2208	59.68***(2)	5.99
wpvtprot	0.8117	0.7774		
wopen	0.0333	0.0202		
Land				
ownagr	0.1649	0.1685	-0.29	±1.96
radio	0.5536	0.4792	4.61***	±1.96
TV	0.5477	0.6821	-8.50***	±1.96
bicyc	0.5860	0.6333	-2.99***	±1.96
sewmach	0.5626	0.5500	0.79	±1.96
clock	0.7878	0.8839	-7.90***	±1.96
fan	0.7005	0.8036	-7.32***	±1.96
refrig	0.2405	0.2988	-4.08***	±1.96
mobike	0.2221	0.2214	0.05	±1.96
car	0.0319	0.0292	0.50	±1.96
No. of households	2220	1680		

a. SD: Standard deviations (presented in parentheses for continuous variables); b. This column presents the χ^2-statistic for categorical, z-statistic for proportions, and t-statistic for continuous variables; the degrees of freedom for the χ^2 statistics are reported in parentheses; c. All critical values are reported for the 5 percent level of significance. * significant at 10%; ** significant at 5%; *** significant at 1%.

for all these variables reveal a rejection of the null of no difference in means over the time period considered. The level of education categories did not show much change and the only category that exhibited some movement was heads who had studied up to the primary school level, which rose by two percentage points while the illiterate category fell by one percentage point. In house-type the proportion of heads living in a semi-*pucca* house was the same but those residing in a *kachcha* house fell by nearly eight percentage points. In the *pucca* house category, on the other hand, there was a rise of six percentage points. A larger proportion of households used the miscellaneous fuel category in 1998-99 at 0.64 as compared to 0.57 in 1992-93 and, at the same time, there was a fall of two percentage points in households using dung and nearly five percentage points in those using wood as fuel. Among drinking water source categories, there is a rise of about seven percentage points in households using a protected public water source and a small fall of four percentage points in protected private sources. There was a negligible change in households with access to sanitation (the z-test fails to reject the null of no difference in means), a five percentage point increase in those with access to electricity, statistically significant at the five percent level of significance and a fall of nearly nine percentage points in households with a separate kitchen over this period, which is statistically significant at the five percent level – we reject the null hypothesis of no significant difference in the means over the two years.[14] This is contrary to what one would expect. However, it is not implausible especially for poor households that may not regard having a separate kitchen as an important improvement in house construction. Finally, the ownership of agricultural land showed a small decline of less than a percentage point – the z-test fails to reject the null of no difference.

Among the consumer durable assets, a large decrease of eight percentage points was registered for radio ownership, while there was an increase of 13 percentage points in the ownership of TV sets. Bicycle ownership also registered an increase of four percentage points. The ownership of clocks and fans rose by about 10 percentage points and the ownership of refrigerators also increased by as much as five percentage points. For all these variables, we reject the null of no difference in means across the two years. However, we fail to reject the null for **sewmach**, **mobike**, and **car**.

[14]Using sample weights does not make a difference to this result either. In 1992-93, the percentage of households with a separate kitchen is 57 percent whereas in 1998-99 it reduces to 52 percent. The z-test fails to reject the null of no difference.

Table 3.9: Summary statistics for comparable rural and urban samples for 1998-99

Variable	*Rural*	*Urban*	*χ^2 and z-stat*[a]	*Critical values*[b]
malhead	0.0487	0.0143	6.17***	±1.96
Social characteristics				
caste	0.2579	0.1506	8.84***	±1.96
genderh	0.9152	0.8935	2.61***	±1.96
agehead	45.3956	45.58	-0.45	±1.96
	[14.98]	[13.86]		
Age categories			46.69***(3)	7.82
age18–37	0.3515	0.3137		
age38–47	0.2454	0.2940		
age48–57	0.1618	0.1940		
age–ab57	0.2413	0.1982		
Occupation categories			2584.31***(3)	7.82
OCagr	0.4757	0.0351		
OCwage	0.1092	0.4560		
OCptran	0.2696	0.3036		
OCother	0.1455	0.2054		
Education categories			1460.65***(3)	7.82
eduh–hi	0.1423	0.4631		
eduh–mid	0.1177	0.1054		
eduh–pri	0.2181	0.1780		
eduh–ill	0.5218	0.2536		
House-type categories			8822.86***(2)	5.99
housepuc	0.0882	0.7369		
houssemi	0.3434	0.1661		
houskach	0.5684	0.0970		
Fuel categories			6115.88***(2)	5.99
fmisc	0.0943	0.6512		
fwood	0.7061	0.3119		
fdung	0.1996	0.0369		
Water source categories			32562.89***(2)	5.99
wpubprot	0.5113	0.2208		
wpvtprot	0.0297	0.7774		
wopen	0.4582	0.0202		
Other characteristics				
sanitdum	0.1141	0.8232	-52.48***	±1.96
kichsep	0.2747	0.5363	-18.93***	±1.96
electr	0.2042	0.8649	-46.63***	±1.96
ownagr	0.7290	0.1685	39.05***	±1.96
Regions				
Hill	0.0399	0.1244	-11.88***	±1.96
East	0.3254	0.1131	16.63***	±1.96
West	0.3612	0.3804	-1.38	±1.96
Central	0.1672	0.2119	-4.02***	±1.96
Bundelkhand	0.1063	0.0815	2.87***	±1.96
No. of households	4103	1680		

a. This column presents the χ^2-statistic for categorical, z-statistic for proportions, and t-statistic for continuous variables; b. All critical values are presented for the 5 percent level of significance; the degrees of freedom for the χ^2 statistics are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Comparing urban 1998-99 with rural 1998-99

The specification used for the rural and urban comparison here was the same as that used for 1992-93. After ensuring comparable specifications, the size for the rural sample was 4,103 households and that for urban households was 1,680 households (Table 3.9).

There is a 3.4 percentage point difference in malaria over the time period considered – a drop from 4.8 to 1.4 – which is statistically significant. The decline of nearly 10 percentage points in the percentage of SC/ST households is also significant. The decline in male-headed households is much smaller at about two percentage points but this difference attains statistical significance. The large difference in heads in agriculture at 48 percent in rural and nearly four percent in urban areas is not surprising. Similarly, a much larger proportion are wage earners in urban areas, not surprisingly. The summary statistics for the education categories show that almost the same proportion of heads attained middle-school and primary education but the percentage of high-school educated is much higher in urban areas. There is a substantial difference in the *pucca* house category as well. Only nine percent of rural houses fall into this category whereas the urban sample has 74 percent *pucca* houses. The dominant cooking fuel in urban areas is the miscellaneous category, which includes LPG whereas in rural areas firewood dominates. More than three quarters of the urban households rely on a private drinking water source whereas 50 percent of the rural households depend on a public protected source of drinking water. The difference between rural and urban areas for all categorical variables (age, education, occupation, fuel type, water source, and regions) is statistically significant. As expected, a high proportion of urban households (four-fifths) have access to sanitation as compared to those in the rural areas (a tenth). Similarly, the number of urban households that have a separate kitchen is twice as high as that in rural areas. Access to electricity is as high as 80 percent in urban and as low as 20 percent in rural areas. In rural areas 73 percent of the households own agricultural land whereas only a fifth own agricultural land in the urban settlement type in 1998-99. All these variables also reject the null of no difference in proportions. The z-statistics calculated for differences in proportions across regions also reject the null of no difference for all regions except the West, implying that there is a statistically significant difference in the proportion of households sampled across the two settlement types, which is likely because some regions may be more rural than the others.

3.3 Using wealth as a welfare metric

Income can be proxied using measures of welfare like occupation, education, and consumer durables, which may be regarded as reflecting permanent income or wealth of the household. It is argued that these factors proxy for long-run welfare or permanent income, which serves to better fulfill the goal of poverty elimination than a focus on current income alone (Fouarge and Muffels 2000).

Permanent income variables (or wealth) may also have certain practical advantages over the use of consumer expenditure or current income. While it might be possible to get a more accurate estimate of wealth from a household survey, good estimates of a flow variable like income require repeated interviewing over a period of time (Swaminathan 1989). Moreover, differences in recall periods, changes in survey instruments, and the nature of interviewer training can have large systematic effects on the measurement of household expenditures and can induce coefficient bias if measured with error and used as an explanatory variable. Further, consumer price indices may not be reliable. Data on wealth variables, on the other hand, are more likely to be recalled accurately—people are more likely to recall with greater accuracy whether they own a clock as compared to recalling what they ate a week back, let alone a month back.[15]

In addition to these issues, the practical motivation for using permanent income variables is the absence of data on malaria incidence in the National Sample Survey (NSS), the standard source for obtaining poverty estimates in India.

Aside from the above two main arguments in favour of using wealth as a welfare-metric, yet another motivation comes from the problem of endogeneity inherent in using consumer expenditure and other such flow variables in analysing the relationship between health and socio-economic variables, as explained now.

The association between poor socio-economic status and poor health is an empirical regularity. Several studies have examined this relation with the causality running in both directions. Poorer countries generally show higher infant mortality and lower life expec-

[15] The successive rounds of the National Sample Survey (NSS) have been the main data source for measuring welfare in India. The poverty line in India is defined in terms of the expenditure required to meet a per capita calorie consumption norm of 2,400 calories for rural and 2,100 calories for urban areas. The equivalent monthly per capita expenditure at 1973-74 prices is Rs 49 for rural and Rs 57 for urban areas (GoI 1979). Thus, to calculate the standard of living, data is collected on amount spent on various food items consumed by the household over the last month. The questionnaire is canvassed to the head of the household who needs to recall what was consumed over the last month except for the last NSS round of 1999-2000, where weekly recall was used.

tancy as compared to richer countries. In addition, the poorest among the population of the world experience higher mortality than the richest. On the other hand, illness may have a negative impact on household income or wealth. Poor health can also lower living standards due to increased expenditure on health. Gertler and Gruber (2002) find evidence for Indonesian households of reduced consumption because of the incidence of household illness due to an inability to smooth consumption against the economic costs of illness. In fact, the poor may pay proportionately more from their income than the more privileged groups (Pritchett and Summers 1996, Hueveline et al. 2002, Wagstaff 2002, Pannarunothoi and Mills 1997).

Thus, given that the health-income causality runs in both directions there are issues of endogeneity inherent in estimating such a relation. Strauss and Thomas (1998) outline the problem of simultaneity in health modelling with reference to the likelihood of feedbacks between health and income. Thus, better health is likely to increase productivity or number of hours worked, in turn contributing to higher income. This would feed back to enabling better health care, nutrition and other such 'health inputs', thus improving health outcomes.

If malaria incidence (the dependent variable in this study) is modelled as a function of income or current consumption or expenditure the problem of endogeneity lies in the possibility of income, in turn, being a function of ill-health (in this study having malaria) since illness can lead to a fall in earnings because of days lost at work or a decline in productivity because of physical inability to cope with work following the illness. Modelling malaria as a function of asset wealth, on the other hand, provides a better option as asset wealth could be assumed to be at least weakly exogenous, that is, current levels of the asset variables are not likely to explain the dependent variable of malaria incidence. This is because asset wealth is less likely to be influenced contemporaneously by income or consumption changes.[16] Thus, malaria incidence is not likely to contemporaneously affect the holdings of assets through the channel of a decline in income or consumption.

This argument also gets support from Russell's (2005) study on the coping sequence of a household faced with illness costs (with specific reference to the head of household

[16]The consumption-wealth channel is examined quite extensively in the literature on monetary policy and economic stabilisation and argues that non-durable consumer spending is affected by changes in asset values, brought about through changes in monetary policy. The literature also mentions the reverse link of wealth (a stock variable) not being affected contemporaneously by consumption, a flow variable (Ludvigson et al. 2002).

falling sick). The response may consist of both mobilising additional resources as well as adjusting spending depending on options available to them and their initial endowments. The latter includes such tactics as delaying payments on bills or repayment of loans, and delaying redemption of pawned jewellery; cutting social spending; cutting spending on relatively expensive food items and finally curtailing food consumption to one meal a day. The strategies for resource mobilisation can be sequenced as: resorting to obtaining credit from local shops for essential fuel and food items; seeking or accepting financial help through social/kinship networks; using savings; borrowing money from no or low-interest sources (i.e., family or NGOs/micro-credit institutions); borrowing from moneylenders at higher interest rates; renting a room; pawning jewellery; other members of the family (e.g., spouse/eldest child) seeking work; and only as a last resort, selling productive assets such as land and livestock. Thus, selling assets of 'core' value, that is those that contribute to current and future income, occurs only when other alternatives are exhausted. Moreover, such assets are also harder to dispose off because they are not very liquid (Wilkes et al. 1997). It may be argued that holdings of what Wilkes terms 'surplus' assets like fans and clocks that do not contribute to income are not likely to be used as a last-level of response but earlier on. However, these assets are not liquid either and may not be good stores of value over time because of depreciation. Therefore, selling them may not be a preferred strategy. Thus, the use of wealth also addresses the potential problem of endogeneity in looking at the relationship between malaria incidence and socio-economic factors.

The main drawback of using wealth as a welfare metric is the problem of aggregating it into an index because of problems in defining what to include in wealth. Different studies use different definitions of wealth. Swaminathan (1989), in her south Indian village study, uses all types of land, other agricultural assets such as machinery; livestock and draught animals; assets used in business activity; gold, houses and other buildings and consumer durables. Financial assets were also included, although these were only owned by a few of the households. Sahn and Stifel (2000) use household characteristics (water source, toilet facilities, construction materials), durables (ownership of radio, TV, refrigerator, bicycle, motorcycle and/or car), and the education of the head of the household to represent the stock of human capital. Hentschel et al. (1998) use basic needs indicators like access to public services: water source, sanitation, waste disposal; education and occupation of each family member; housing quality; household size, crowding and age/sex composition; principal language spoken and location of the residence. Filmer and Pritchett (2001) use 21 asset variables (collapsed into a Principal Components Index, also employed in this thesis). The data are from the NFHS for India (1992-93) and include household ownership of consumer durables like clock/watch, bicycle, radio, television, sewing machine, refrigerator, and car; characteristics of the household's dwelling, such as toilet facilities, drinking water source, rooms in the dwelling, house-type and main source of lighting and cooking; and whether or not a household owned more than six hectares of land. They obtained a cut-off value to then classify a household into the poor, middle, or rich category using this index. The bottom 40 percent were considered poor. They also assessed if the asset index identifies similar households as poor as compared to the per capita consumption expenditure criterion conventionally used to obtain poverty head count ratios in India. To do this they arrived at the percentage contribution of each state in the total percentage of the asset index poor for India and compared this against the contribution of each of these states to the all-India head count ratio (36 percent in 1992-93) measured using per capita consumption expenditure. The rank correlation coefficient between these two measures is 0.794, statistically significant at the one percent level.

Thus, wealth variables cover a wide range from basic needs indicators to household assets and characteristics. These could be categorised in different ways to develop indices. Based on the literature on malaria discussed above and the variables available in the NFHS, for the purpose of the current analysis the following classification was considered (monetary assets are not included in the list below as these are not covered by the NFHS):

1. Living conditions: Type of house, type of fuel used, demographic indicators (number of members), number of rooms, whether house has a separate kitchen, drinking water source, whether household has sanitation, whether animals are kept inside the house.

2. Agricultural assets: Ownership of agricultural land (irrigated and non-irrigated), livestock, and agricultural machinery: tractor, thresher, water pump, and bullock cart.

3. Non-agricultural assets/consumer durables: Ownership of radio, TV, fan, clock, bicycle, sewing machine, refrigerator, sofa, and motorbike.

4. Human assets/capital/income sources: Education of household head, age, gender,

caste, and occupation.

.

Indices corresponding to these different categories using Principal Components Analysis were experimented with. However, the constructed measures were found to perform poorly in the malaria incidence model except for the non-agricultural consumer durable wealth index. Moreover, these assets did not perform well individually. In any event, this study takes the view that it is better to treat the other variables individually and not use indices as this approach could fulfill the purpose of better informing policy on the role of specific socio-economic factors in determining malaria incidence—a number of the variables, as the literature review sections of this chapter suggest, bear directly on the incidence of malaria and collapsing them into an index leads to a loss in information.

3.3.1 Principal Components Analysis

Principal Components Analysis[17] is a data reducing technique used, for example, when a set of variables such as non-agricultural consumer durable wealth assets in this study, need to be collapsed into an index value that can be considered to represent the data fairly well. In this method this original set of, say, p variables $(x_1, x_2, ...x_p)$ common to each household, i, are formed into p new variables $(\epsilon_1, \epsilon_2, ...\epsilon_p)$ that are orthogonal linear combinations of the original ones. This new set of variables capture information common to the original variables and may be represented as:

$$\begin{aligned} \epsilon_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \ldots + w_{1p}x_p \\ \epsilon_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3 + \ldots + w_{2p}x_p \\ &\vdots \\ \epsilon_p &= w_{p1}x_1 + w_{p2}x_2 + w_{p3}x_3 + \ldots + w_{pp}x_p \end{aligned} \tag{3.1}$$

where $(\epsilon_1, \epsilon_2, ...\epsilon_p)$ are the p principal components and w_{ij} are the weights assigned to the jth variable for the ith principal component. The weights are estimated such that the first principal component, ϵ_1, contains the maximum information common to all the

[17] The reference for this section is Sharma (1996).

variables by accounting for the maximum variance in the data, the second component accounts for the variance not accounted for by the first component and so on.

In the above system of equations only the left hand side is known and, thus, the system is indeterminate. This problem is resolved by finding the first principal component followed by the second that is orthogonal to the first and so on. This is done by solving the equations:

$$(\sum -\lambda I)\gamma = 0 \tag{3.2}$$

where $\sum$ is the covariance matrix, which is equal to $E(X'X)$. X is a p-component random vector and p is the number of variables. $\gamma' = (\gamma_1\gamma_2 \dots \gamma_p)$ is a vector of weights to form the linear combinations and $\epsilon = \gamma' X$ is the new variable, a linear combination of the original ones and its variance is given as $var(\epsilon) = E(\epsilon\epsilon') = E(\gamma' X X' \gamma)$ or $\gamma' \sum \gamma$. Thus, we need to find the weight vector γ' such that $\gamma \sum \gamma$ is maximum for the different linear combinations and $\gamma'\gamma = 1$. This is a maximisation problem formulated using the Lagrange multiplier, λ:

$$z = \gamma' \sum \gamma - \lambda(\gamma'\gamma - 1) \tag{3.3}$$

Then,

$$\frac{\delta z}{\delta \gamma} = 2 \sum \gamma - 2\lambda\gamma = 0 \tag{3.4}$$

is given as:

$$(\sum -\lambda I)\gamma = 0 \tag{3.5}$$

In order for this system to have a non-trivial solution, the determinant of $(\sum -\lambda I)$ should be zero or $|\sum -\lambda I| = 0$. This determinant is then a polynomial in λ of order p. Solving this equation yields p roots $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p)$ that are called eigenvalues or roots of $\sum$. These have eigenvectors associated with them, which is the vector of weights, γ_i obtained by solving the following two equations:

$$(\sum -\lambda I)\gamma = 0 \tag{3.6}$$

and

$$\gamma'\gamma = 1 \tag{3.7}$$

The sum of the weights is set to one in order to fix the scale of the principal components since the weights assigned to the variables to obtain the principal components are affected by the relative variance of the original variables. Thus, these are standardised to a variance of one such that each variable then accounts for the same variance. Thus,

$$\sum_{ij} w_{ij}^2 = 1 \; where \; i = 1 \ldots p; j = 1 \ldots p \tag{3.8}$$

and the covariance between w_i and w_j is set to zero:

$$\sum_{ij} w_i w_j = 0 \; \forall \; i = j \tag{3.9}$$

In this study the original variables were standardised to a mean of zero and variance of one in order to convert the data available in the form of dichotomous variables into continuous ones and the principal components method was applied to this new set of variables.

The values for each of the above equations, given the weights, corresponding to each observation on the original variable are called scores. For each equation in the above system of equations or for each principal component there are as many scores as the original number of observations on the variables. Each principal component is associated with a variance. The first principal component accounts for the maximum variability in the original data followed by the second component and so on. The total variance of all the principal components is the same as that of the original data. The variance of each principal component is called the eigenvalue.

3.3.2 Summary statistics

Rural Uttar Pradesh

The summary statistics (column headed mean) for the non-agricultural consumer durable assets for the rural samples of 1992-93 and 1998-99 are presented in Table 3.10 along with the weights corresponding to these variables for the first principal component. All assets show an increase in ownership over the six-year period. The Principal Components Index

Table 3.10: Results of the first principal component for rural 1992-93 and 1998-99 Uttar Pradesh sample

Variable	*Weights 1992-93*	*Mean 1992-93*	*SD*[a]	*Weight/SD 1992-93*	*Weights 1998-99*	*Mean 1998-99*	*SD*	*Weight/SD 1998-99*
radio	0.354	0.278	0.448	0.789	0.342	0.264	0.441	0.776
TV	0.430	0.077	0.266	1.613	0.419	0.145	0.352	1.189
bicyc	0.221	0.546	0.498	0.443	0.240	0.624	0.484	0.495
sewmach	0.388	0.159	0.365	1.061	0.381	0.156	0.363	1.048
clock	0.363	0.464	0.499	0.728	0.355	0.519	0.500	0.710
fan	0.433	0.107	0.309	1.400	0.413	0.187	0.390	1.059
refrig	0.237	0.009	0.095	2.498	0.254	0.015	0.120	2.124
mobike	0.325	0.034	0.181	1.801	0.341	0.042	0.201	1.697
car	0.101	0.002	0.048	2.099	0.168	0.005	0.068	2.452
index		0.000	1.695			0.000	1.716	

a. SD: Standard deviation

was constructed using the following non-agricultural asset wealth variables: radio, TV, bicycle, sewing machine, clock, fan, refrigerator, motor bike, and car. Ownership of a sofa and VCR were not included here as in the 1998-99 survey these were not covered. Each of these is a dummy variable for whether a household owns the asset or not.[18] The eigenvalues for the first and second principal components in 1992-93 are respectively, 2.87 and 1.29 with the former accounting for 31.92 percent and the latter 14.33 percent of the total variation in the original variables. In 1998-99 the corresponding values were 2.94 and 1.21 for the first and second principal components, respectively, accounting for 33 percent and nearly 14 percent of the total variation. The first Principal Component was used for the analysis and, thus, we are able to account for about 32 percent of the variance of the original data in 1992-93 and 33 percent in 1998-99. The mean value of the asset index is zero as reported in Table 3.10 since we have standardised the data and the standard deviation is 1.69 for 1992-93 and slightly higher at 1.72 for 1998-99, suggesting that there is greater dispersion in wealth over time.

Following Filmer and Pritchett (2001) the weights can be interpreted as follows: a weight of 0.35 for radio indicates that a move from not owning a radio to owning one changes the asset index by 0.78 units (weight/standard deviation presented in Columns 5 and 9 for 1992-93 and 1998-99, respectively) in 1992-93 and by the same amount in 1998-99 when the weight on radio is 0.34. Similarly, a household that owns a TV has an asset index that is 1.61 units higher than a household that does not own a TV in 1992-93.

[18]The standard deviations are reported for the dummy variables because these were used to standardize the variables as mentioned in the previous section. The mean of the asset index is zero because of the standardisation.

In 1998-99 a household that owns a TV relative to one that does not own one has an asset index that is 1.2 units higher. Owning a refrigerator raises the asset index by 2.5 units in 1992-93 but by a lower value of 2.1 in 1998-99. Owning a car raises the index by 2.1 units in 1992-93 and by 2.5 in 1998-99. The ownership of TV sets has doubled over this period from seven percent of households owning a TV in 1992-93 to nearly 15 percent in 1998-99. This registers as a fall in the increase in value of the asset index for a household that owns a TV set relative to one that does not when we compare 1992-93 against 1998-99. The asset index value rises by 1.6 points for a household that owns a TV set relative to one that does not in 1992-93 and rises by a lower amount of 1.2 in 1998-99. Similarly, for the ownership of fans where the percentage of households that own a fan in 1998-99 is higher by nearly nine points relative to 1992-93, the asset index value for 1998-99 for a household that owns a fan relative to one that does not is 24 percent lower. For the ownership of a bicycle the result is the reverse. There is a rise in ownership of bicycles and this registers as an increase in the asset index value for a household that owns a bicycle relative to one that does not over the two periods by nearly 12 percent (a rise from 0.4 to 0.5 in asset index value).

Table 3.11: Summary statistics for quintile groups for the non-agricultural consumer durable asset index for rural Uttar Pradesh, 1992-93 and 1998-99

Variable	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*
	Bottom 20%		*20-40%*		*40-60%*		*60-80%*		*80-100%*	
radio	0.000	0.000	0.000	0.000	0.082	0.054	0.620	0.594	0.689	0.672
TV	0.000	0.000	0.000	0.000	0.000	0.009	0.005	0.059	0.379	0.657
bicyc	0.000	0.000	0.568	0.814	0.591	0.611	0.778	0.840	0.793	0.855
sewmach	0.000	0.000	0.000	0.000	0.049	0.027	0.095	0.162	0.649	0.592
clock	0.000	0.000	0.000	0.035	0.591	0.851	0.795	0.756	0.933	0.950
fan	0.000	0.000	0.000	0.000	0.000	0.059	0.015	0.155	0.520	0.721
refrig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.073
mobike	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.002	0.164	0.208
car	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.023
index	-1.330	-1.531	-1.078	-1.102	-0.521	-0.481	0.222	0.291	2.705	2.820

The sample was divided into quintiles using the value of the asset index with the objective of forming different asset wealth groups. Table 3.11 shows the cut-off points for each quintile and the corresponding mean value of the asset index. As we can see, the bottom 20 percent of the sample had a mean asset index value of -1.33 in 1992-93 and -1.53 in 1998-99. The households belonging to this group did not own any assets or the mean of each asset was zero for both years. In the 20 to 40 percent group the mean asset index value was -1.08 and -1.10 for 1992-93 and 1998-99, respectively. Three-fifths of the households owned a bicycle in 1992-93 and 81 percent in 1998-99. The other variables had zero means in 1992-93 and 3.5 percent also owned a clock in 1998-99. In the third group (40 to 60 percent of the sample) nearly three-fifths of the households owned a bicycle in both time periods and in 1992-93 about 60 percent also owned a clock; in 1998-99 the ownership of clocks rose to about 80 percent. A little less than one-tenth of the households owned a radio and nearly five percent owned a sewing machine in 1992-93. The corresponding figures for 1998-99 are lower at 5.4 percent for a radio and 2.7 percent for a sewing machine. But, at the same time, some households also owned a TV and a fan. The mean asset index value was -0.52 in 1992-93 and slightly higher at -0.48 in 1998-99. In the next segment 84 percent owned a bicycle, about four-fifths a clock and three-fifths a radio. The proportion of households that owned a sewing machine was 0.09. The mean asset index value was 0.22 in 1992-93 and rose to 0.29 in 1998-99 with an increase of six percentage points in the ownership of bicycles, a drop of four percentage points in the ownership of a clock, and a 14 percentage point increase in the number of households that owned a fan. Finally, in the top 20 percent of the asset wealth distribution in 1992-93, a little less than a tenth of the households did not own a clock, nearly 70 percent owned a radio, about two-fifths a bicycle, 65 percent a sewing machine, half the households owned a fan, and nearly 40 percent a TV. Motor bikes were owned by 16 percent of the households and nearly five percent and one percent, respectively, owned a refrigerator and a car. The mean asset index value was 2.71 in 1992-93 and increased to 2.82 in 1998-99. In 1998-99 the sharpest increase for this group of the population was in the ownership of TV sets (by nearly 28 percentage points) followed by fans (20 percentage points). There was also a six and four percentage point increase in the ownership of bicycles and motorbikes, respectively. The difference in the mean asset index value between the top-most wealth group and the lowest wealth-group was 4.04 units in 1992-93 and 4.35 units in 1998-99. The mean asset index value rose over the six-year period for the top three wealth groups and fell for the bottom two.

The summary statistics discussed here indicate how the number of assets owned has increased with increase in the mean asset index value or by wealth groups and shows a clear rise in the asset index value for the top wealth groups over the two time periods. The following subsections discuss the urban samples.

Table 3.12: Results of the first principal component for urban 1992-93 and 1998-99 Uttar Pradesh sample

Variable	*Weights 1992-93*	*Mean 1992-93*	*SD[a]*	*Weight/SD 1992-93*	*Weights 1998-99*	*Mean 1998-99*	*SD*	*Weight/SD 1998-99*
radio	0.355	0.554	0.497	0.714	0.341	0.479	0.500	0.682
TV	0.405	0.548	0.498	0.814	0.394	0.682	0.466	0.846
bicyc	0.233	0.586	0.493	0.472	0.240	0.633	0.482	0.498
sewmach	0.368	0.563	0.496	0.741	0.366	0.550	0.498	0.735
clock	0.355	0.788	0.409	0.867	0.350	0.884	0.320	1.092
fan	0.392	0.700	0.458	0.856	0.382	0.804	0.397	0.960
refrig	0.339	0.241	0.428	0.792	0.362	0.299	0.458	0.791
mobike	0.322	0.222	0.416	0.775	0.334	0.221	0.415	0.803
car	0.149	0.032	0.176	0.846	0.163	0.029	0.168	0.967
index		0.000	1.960			0.000	1.803	

a. SD: Standard deviation

Urban Uttar Pradesh

The Principal Components asset index was constructed using the following non-agricultural asset wealth variables: radio, TV, bicycle, sewing machine, clock, fan, refrigerator, motor bike, and car. Each of these is a dummy variable for whether a household owns that asset or not. The means and standard deviations[19] of these variables are presented in Table 3.12 along with the weights attached to these variables used to arrive at the Principal Components. The weights for the first component are reported. The eigenvalues for the first and second principal components were, respectively, 3.85 and 1.26 in 1992-93 with the former accounting for 43 percent and the latter 14 percent of the total variation. For 1998-99, the eigenvalues were 3.25 (accounting for a variation of 36 percent) and 1.28 (accounting for 14 percent of the variation) for the first and the second principal components, respectively. The mean value of the asset index is zero since we have standardised the data. The standard deviation is 1.96 for 1992-93 and 1.80 for 1998-99.

Table 3.13: Summary statistics for quintile groups for asset index for urban Uttar Pradesh, 1992-93 and 1998-99

Variable	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*	*1992-93*	*1998-99*
	Bottom 20%		*20-40%*		*40-60%*		*60-80%*		*80-100%*	
radio	0.054	0.060	0.315	0.205	0.532	0.449	0.901	0.756	0.966	0.926
TV	0.007	0.080	0.146	0.545	0.626	0.821	0.964	0.964	0.995	1.000
bicyc	0.169	0.259	0.518	0.670	0.721	0.705	0.696	0.667	0.824	0.866
sewmach	0.036	0.060	0.304	0.262	0.570	0.622	0.923	0.851	0.977	0.955
clock	0.149	0.461	0.811	0.964	0.982	0.994	0.998	1.000	1.000	1.000
fan	0.052	0.241	0.486	0.810	0.964	0.973	1.000	0.994	1.000	1.000
refrig	0.000	0.012	0.032	0.030	0.029	0.110	0.169	0.438	0.973	0.905
mobike	0.002	0.012	0.020	0.012	0.054	0.051	0.250	0.202	0.784	0.830
car	0.002	0.000	0.000	0.000	0.005	0.000	0.002	0.018	0.151	0.128
index	-2.880	-2.768	-1.233	-0.813	0.190	0.155	1.285	1.059	2.636	2.367

[19]The standard deviations are reported for dummy variables, as in the rural sample, because these were used to standardize the variables.

The bottom 20 percent of the urban groups (see Table 3.13) showed a slight rise in the mean asset index from -2.880 to -2.768 over the period 1992-93 to 1998-99. In fact, all the groups show an increase in the mean value except for the two top-most groups. In the lowest segment, there is a sizable increase in the percentage of households owning a television from one percent to eight percent. The ownership of clocks and fans tripled. The next group shows a fall in the ownership of radios by nine percentage points, but the ownership of TVs more than tripled from 15 percent to 54 percent over this period. The ownership of clocks also reported an increase of more than 30 percentage points. In the 40 to 60 percent quintile there is a rise in the ownership of refrigerators from three percent to 11 percent. Radios registered a decline by as much as 11 percentage points whereas TV ownership increased by 20 percentage points perhaps suggesting that TVs substitute for radio sets in the higher wealth brackets, as in the previous wealth group. Nearly all households in this group as well as the top two groups owned a clock. Ninety-five percent of the households owned a TV in 1992-93 in this group and in 1998-99 all the households had a TV. The ownership of cars exceeded the two percent mark only for the top group and is as high as 15 percent in 1992-93, but declined to 13 percent is 1998-99. The ownership of refrigerators exhibits a fall of about seven percentage points and sewing machines ownership dropped by two percentage points. In fact, the mean asset index value for this group exhibited a decline for this period. Similarly, all households owned a clock in both time periods. From the above discussion, we may conclude that over the time period considered it is the lowest wealth groups that have shown an improvement in their asset wealth status as measured by the mean asset index value in urban areas.

The Kernel density plots of the asset index for the two time periods for rural and urban Uttar Pradesh are presented in Appendix A. As we can see from Figures A.1 to A.4, none of the distributions appears to follow those generally associated with the distribution of an income measure, such as say a log-normal with a Pareto tail. The distribution in Figure A.4 for the urban asset index for 1998-99 appears to look more like an income distribution, but is not lognormal. The asset indices appear to be weak correlates of the household welfare metric. This might be one of the reasons why we do not obtain consistently well determined results later in the thesis where we use the asset index as a proxy for income in determining the relationship between malaria incidence and the household welfare metric (see p.133 and p.137 for the rural specifications and p.150 and p.152 for the urban specifications).

3.4 District level data

The data on the different district variables were available at the village level and were aggregated to the district level for the districts covered in NFHS used in the earlier household analysis. The variables available are population density, mean distance to a drinking water facility, and the presence of a number of infrastructure facilities such as primary, middle, or high school, adult literacy centre, health centre, and whether the village has a certain type of irrigation facility like a canal or a tube well. The choice of these variables is motivated by reference to the literature discussed in Chapter 2 (see Table 3.3, p.42).

Mean district population density is calculated as the sum of population density in all districts divided by the total number of districts where population density of each district is computed by dividing the total population of all villages in a district by the total area of all villages in the district. The mean distance to drinking water facility is calculated as the sum of the distance to drinking water source (summed over all districts) divided by the number of districts. The distance for each district is calculated as the sum of the mean distance to drinking water source for all villages in a district divided by number of villages in that district.

The mean percentage of villages with a particular facility in a district is given as:

$$\left[\frac{Number\ of\ villages\ with\ facility\ in\ district\ `d'}{Total\ number\ of\ villages\ in\ district\ `d'}\right] \times 100 \qquad (3.10)$$

where *d=1 to 48* and 'facility' includes type of school, health centre, commuting facility or other such variables. Table 3.14) presents the summary statistics for 1991.

The mean percentage of villages with a primary school is 60. The corresponding figures for middle schools and high schools are very small at a little over one-tenth for the former and even lower for high schools. The proportion of villages that have an adult literacy centre is also less than 0.1. On average, in each district, 10 percent of the villages have a primary health sub-centre. The mean distance to collect drinking water was a little over a kilometre. Twenty percent of the villages had some methods of commuting like buses, taxis, or railway. Thirty one percent of the villages had electricity for domestic use and

30 percent for agricultural use.[20] Less than one-tenth of the district area was covered by forest. The average (rural) district population density was a little over four persons per square km. Most of the district area was irrigated by non-electric tube wells (little over a fifth) followed by electric tube wells and government canal at a little over one-tenth. Private canals and non-electric and electric wells share a very small percentage of the total area at 0.21 percent. The average annual rainfall was 805 mm. We now turn to a discussion of the 1991 and 2001 census variables as well as the average for the variables in the two census years.

3.4.1 Census statistics

The 1991 census was used with the 1992-93 NFHS data. The average of the 1991 and 2001 census variables were used with the 1998-99 NFHS data. The lag in the data was not considered a problem since the realizations of such variables as the presence of a school move relatively slowly over time. Table 3.15 presents the summary statistics of the 1991 and 2001 censuses along with the average for each variable for these two periods. Since six districts aside from the 48 selected earlier for 1992-93 were not covered in 1998-99 as mentioned earlier, these were also dropped from the census data. Variables capturing the percentage of villages with domestic and agricultural power supply were not included as the codes for these variables are not clear from the 2001 census. The variable power of all types did not have a well determined effect so it was not included either.

In 2000 the state of Uttaranchal was formed out of Uttar Pradesh so the data for the districts that now fell in Uttarancal, but were covered in the 1992-93 as well as the 1998-99 NFHS survey for Uttar Pradesh were obtained from the Uttranchal census records. These included seven districts: Tehri-Garhwal, Dehradun, Chamoli, Nainital, Garhwal, Pithoragarh, and Almora. In 2001 a number of new districts were also formed. The data for these were merged with the districts these originally fell under. In Table 3.16 the first column gives the name of the new district and the second one the old district from which this was formed. It was decided which district the new one was formed from on the basis of matching villages over the two census periods, the details of which are available in the

[20]Power may be supplied to a village to be used for residential purposes, for agricultural purposes mainly to pump water for irrigation, and for commercial purposes for shops and factories. It is possible for a village to have power supply for just one category and not for the others. Commercial power supply was not included in this analysis as a very small percentage of villages at 1.6 percent had this type of supply and the coefficient on this variable was not significant either when a regression was run including this along with other types of power supplies.

Table 3.14: Summary statistics of district level variables in rural Uttar Pradesh (Census 1991)

Variable	*Description*	*Obs*[a]	*Mean*	*SD*[b]	*Min*	*Max*
PVpsch	Percentage of villages with primary school	48	59.724	13.378	28.042	84.600
PVmsch	Percentage of villages with middle school	48	13.050	5.247	3.175	22.382
PVhsch	Percentage of villages with high school	48	2.993	1.818	0.674	7.353
PVadlit	Percentage of villages with an adult literacy centre	48	6.535	9.831	0.000	58.369
PVphc	Percentage of villages with primary health centre	48	2.215	1.347	0.522	8.663
PVphs	Percentage of villages with primary health sub-centre	48	9.348	7.372	0.237	29.167
AVdistwa	Mean distance to drinking water facility (in kms)	48	1.203	0.889	0.121	4.408
PVcomm	Percentage of villages with facility to commute	48	20.340	9.604	3.491	46.491
PVpowerd	Percentage of villages with domestic power supply	48	31.335	18.571	0.000	95.323
PVpowera	Percentage of villages with power for agriculture	48	29.902	21.262	0.050	87.686
PDforest	Percentage of district area under forest cover	48	7.165	20.844	0.000	129.897[c]
Ddens	District population density	48	4.417	1.526	1.179	7.687
PDirrW	Percentage of district area irrigated by non-electric well	48	0.679	0.884	0.000	3.788
PDirrWE	Percentage of district area irrigated by electric well	48	0.184	0.345	0.000	1.707
PDirrTW	Percentage of district area irrigated by non-electric tube well	48	22.047	24.386	0.001	160.238
PDirrTWE	Percentage of district area irrigated by electric tube well	48	13.377	12.046	0.000	42.847
PDirrGC	Percentage of district area irrigated by government canal	48	13.379	8.974	0.022	31.529
PDirrPC	Percentage of district area irrigated by private canal	48	0.214	0.672	0.000	3.323
rain	Average annual rainfall (mm)	48	804.67	241.050	410.000	1705.500

a. Obs: Number of observations; b. SD: Standard deviation; c. The maximum value is more than 100 per cent because some of the 'villages' included in Tehri Garhwal district are forest areas that are not reported in the village and, hence, district area and appear only in area under forest. Since these forest areas are located close to other populated villages in the district they have been included in this analysis.

Table 3.15: Summary statistics for district level variables from Census 1991, Census 2001, and average of the two censuses for comparable samples

Variable	*Mean (1991)*	*SD*[a]	*Mean (2001)*	*SD*	*Average (1991 and 2001)*
PVpsch	60.90	11.62	82.53	12.04	71.72
PVmsch	13.45	4.97	28.97	9.94	21.21
PVhsch	2.91	1.75	6.98	4.06	4.94
PVadlit	7.10	10.35	7.75	4.70	7.42
PVphc	2.35	1.39	3.67	1.79	3.01
AVdistwa	1.18	0.93	0.02	0.08	0.60
PVcomm	20.81	9.85	19.26	10.39	20.04
PDforest	7.00	22.10	4.57	6.35	5.79
Ddens	4.45	1.47	6.79	2.56	5.62
PDirrW	0.75	0.92	2.45	3.79	1.60
PDirrWE	0.20	0.36	5.59	16.77	2.89
PDirrTW	22.43	25.08	20.40	14.39	21.42
PDirrTWE	13.81	11.93	15.14	14.73	14.47
PDirrGC	13.72	8.36	25.42	50.17	19.57
PDirrPC	0.20	0.70	1.51	1.31	0.86
rain	789.98	240.97	986.49	304.42	888.24
Observations	42		42		42

data for these censuses. The rainfall data for 1998-99 are from the same source as that for 1992-93, the Statistical Abstract of Uttar Pradesh (GoI 1996). The data are available for 1998 and 1999 and an average of the two years was obtained for each district. Rainfall measured for the last available year was used where the data were missing.

As the summary statistics show, on average in each district in Uttar Pradesh the percentage of villages with a primary school as well as a high school increased by 20 percentage points. The simple mean for the two years stood at 71 percent for primary, 21 percent for middle and nearly five percent for high schools. The percentage of adult literacy centres did not show much change over this period. The average distance to a drinking water source declined from 1.18 km in 1991 to 0.02 km in 2001 and the average for the two census years was just a little over half a kilometre. This sharp fall can be attributed to the implementation of the *Swajaldhara* (literally, 'Clean water flow') Programme operated by the State Water and Sanitation Mission of Uttar Pradesh, initially supported by the World Bank. The programme was piloted in seven districts in Uttar Pradesh in 1999 and involved the participation of the communities in planning, construction and operation of the water delivery systems.[21] It specifically involved women in choosing their preferred technology for water supply since they collect water for their households all over India,

[21]The communities also contributed to the total capital cost. For example, where households preferred a private piped water connection, they were expected to contribute Rs 1000 towards the cost.

Table 3.16: New Uttar Pradesh districts reported in 2001

New district (2001)	*Old district (1991)*
Jyotiba Phule Nagar	Rampur
Baghpat	Meerut
Gautam Budh Nagar	Ghaziabad
Auraiya	Etawah
Mahoba	Hamirpur
Chitrakoot	Banda
Kaushambi	Allahabad
Sant Ravidas Nagar	Varanasi
Chandauli	Varanasi
Kushinagar	Gorakhpur
Sant Kabir Nagar	Basti
Balrampur	Gonda
Shrasvati	Gonda
Ambedkar Nagar	Faizabad
Kannauj	Farrukhabad
Hathras	Aligarh

sometimes travelling as much as 3 km two to three times a day. The introduction of the scheme in pilot villages was able to bring water sources closer to their homes cutting the distance to zero in some cases. The success of the programme provided the motivation to extend it to other parts of India.

The percentage of villages with a commuting facility remained nearly the same at about 20 percent. The percentage of district area under forest showed a decline by a little over two percentage points over the decade. The average for the two years indicates that on average for each district nearly six percent of the district area was covered with forests. Population density increased from about four persons per square kilometer to nearly seven and the average for the two census periods was 5.6 persons per sq. km. The percentage of district area under different types of irrigation facilities remained the same or showed a slight increase. The largest increase was seen for the percentage of area under a government canal, which rose by a little over ten percentage points (from 14 to 25) over the decade with the average for the two years at 20 percent. The area under private canal rose from less than 0.2 percent to a little over one percent. The average of these two figures is 0.86. The mean average annual rainfall was 888 mm.

So far we covered the census and NFHS data relevant for the household and district level analysis. The next section discusses the state level panel data used in the aggregate level analysis.

3.5 State level data

As we may recall from the first section of this chapter, the dependent variable used in the inter-state analysis is the annual parasite incidence per thousand population (**API**). These data were calculated from data on annual malaria cases collected by the Ministry of Health and Family Welfare under the National Malaria Eradication Programme and are available from various issues of the Health Information of India, known as Health Statistics of India before 1985 (GoI 1979-1985, GoI 1986-2001). A continuous series of the data that also matched the series available for the explanatory variables used (discussed below) was available from 1978 until 2000. From the year 2000 the states of Bihar, Uttar Pradesh, and Madhya Pradesh were split into Jharkhand, Uttaranchal, and Chattisgarh, respectively. The data on these were available separately. So as to match with the previous years these data were added to the total number of cases for the original states. Number of malaria cases were converted to the more commonly used measure of malaria incidence, **API**, by dividing the number of cases in the state by the population of the state times 1000. As mentioned earlier in this thesis, the method of collecting the data on **API** is controversial and the estimates of malaria incidence arrived at using the data on **API** are different from those using the data from the NFHS.

The explanatory variables employed in the district level analysis were: average annual rainfall, population density, percentage of district area under forest cover, percentage of district area irrigated by different sources (canals, tube wells, and wells), percentage of villages with a primary school, high school, and adult literacy centres, percentage of villages with a commuting facility, and percentage of villages with power supply. At the state level, a match of the above variables was obtained for population density, average annual rainfall and percentage area irrigated. It was not possible to use data on forest area as these are updated only at three-year intervals and the available data are not considered comparable over time because of changes in methodology for measuring forest cover (GoI 2000). The data on number of schools are also not updated every year and the data on adult literacy centres were not available at the state level. However, data on income, mainly net domestic product by state are available as well as data on health and education expenditure by state and rural and urban poverty head count ratios. The analysis for the household and district data was conducted separately for rural and urban

Table 3.17: Explanatory variables used in the state level analysis compared with district variables

Variables used in district level analysis	*Variables used in state level analysis*	*Variable abbreviations for state level analysis*
Variables with exact match at the two levels		
Population density by district (persons per square km)	Population density by state (persons per square km)	**popden**
Average annual rainfall by district (mm)	Average annual rainfall by state (mm)	**rain**
Percentage district area irrigated by different sources	Percentage state area irrigated (all sources)	**irrig**
Variables that did not have an exact match at the two levels but used alternative measure		
Percentage villages with adult literacy centres		**eduex**
Percentage villages with high schools	Real per capita education expenditure by state	
Average distance from drinking water source	Real per capita health expenditure by state (includes expenditure on sanitation and water supply)	**healthex**
Variables that could not be matched but additional state variables could capture some of these effects		
Percentage villages with commuting facility	Real net state domestic product per capita	**stinc**
Percentage villages in district with power supply	Urbanisation	**urban**
	Poverty head count ratio (rural)	**povertyr**
	Poverty head count ratio (urban)	**povertyu**

All data above obtained from Ozler et al. (1996) updated by Besley and Burgess (2004) and Burgess and Pande (2005) with the exception of data on average annual rainfall and percentage irrigated area, which were obtained from various issues of the Statistical Abstract of India (GoI 1979-2001).

areas. The state analysis does not segregate the data into rural and urban but, instead, uses a measure of urbanization (percentage of population in a state living in urban areas) to account for this difference in incidence across the two settlement types.

Data on poverty, state domestic product, state area, price indices, expenditure variables (health and education), and population were compiled by Ozler et al. (1996) for the period 1958 to 1992 (though the current study focuses on the period 1978 to 2000 to match with the availability of the malaria data). These have been updated by Besley and Burgess (2002), Besley and Burgess (2004) and Burgess and Pande (2005) and the current update is to the year 2000. Data on irrigated area and rainfall were collected from various issues of the Statistical Abstract of India (GoI 1979-2001). Table 3.17 summarises the variables used in the state analysis and indicates which variables were available in the district level analysis. The details of how these variables were constructed and the original sources of the data are now discussed.

The population estimates available in Ozler et al. (1996) have been constructed from the census data for 1971, 1981 and 1991 (covering the period 1978 to 2000 used in this study) and are available separately for rural and urban settlement types. Ozler et al. (1996) obtained the data for the intermediate years by interpolation assuming a constant compound rate of growth of population obtained from the census population totals in two successive census years. For this analysis, using the data on population and state area, the variable, population density (**popden**), was obtained by dividing the state population by the state area. The variable urbanisation rate (**urban**) was obtained by dividing the urban population by the total population of the state and expressed as a percentage.

Data on real state income (**stinc**) were originally compiled by Ozler et al. (1996) from published data on net state domestic product (NSDP) available from various reports of the Central Statistical Organisation, Government of India. NSDP measures the volume of goods and services produced within a state during the year accounted for without duplication and obtained net of depreciation from the gross domestic product (GDP). The GDP estimates are compiled separately for the organised and unorganised sectors, which includes the informal sector. It covers the three broad primary, secondary and tertiary sectors. The primary sector includes agriculture, forestry, fishing, mining and quarrying. The secondary sector constitutes registered and unregistered manufacturing, construction, electricity, gas, and water supply. The services sector includes trade, hotels and restaurants; transport, storage, and communication; banking and insurance; real estate, ownership and dwellings, legal and business services; and public administration and other services.

The variables real education expenditure per capita (**eduex**) and real health expenditure per capita (**healthex**) were available from Ozler et al. (1996) updated by Besley and Burgess (2004) and Burgess and Pande (2005) up to 2000 and were originally sourced from various issues of Public Finance Statistics (Ministry of Finance) and Report on Currency and Finance (Reserve Bank of India). The variable **eduex** includes expenditure on education, art and culture, scientific services, and research expenditure and **healthex** includes medical, family planning, public health and sanitation, and water supply expenditure.

The income and expenditure data were available in current prices. These were adjusted to constant prices using the consumer price index for agricultural workers (CPIAL) and the consumer price index for industrial workers (CPIIW) representing the rural and urban cost of living indices, respectively and expressed in terms of a common base of October 1973-June 1974 all-India (rural/urban) prices, which coincides with the year for which the official real poverty line has been set (discussed below). The variables were then presented in per capita terms using the population data. The consumer price index for agricultural workers was corrected for firewood prices. The Labour Bureau has held the price of firewood constant in the construction of the CPIAL since 1960-61. Ozler et al. (1996) corrected the CPIAL to incorporate the changing price of firewood. For the period 1978 to 2000 (used in this study) this involved replacing the constant firewood price sub-series in the official CPIAL with the one based on the actual all-India rural firewood prices. The deflator obtained using the two indices is given as (Burgess and Pande 2005):

$$(popU/(popR + popU)) * CPIIW + (popR/(popR + popU)) * CPIAL \tag{3.11}$$

where popU and popR refer to the urban and rural population respectively and CPIIW and CPIAL are the urban and rural cost of living indices, respectively. To obtain the real measures for the income and expenditure variables, the monetary values were divided by the deflator and multiplied by 100. Further, the real variables so obtained for each year were then divided by the state population for that year to express these in per capita terms.

Average annual rainfall (**rain**) and percentage area irrigated (**irrig**) data are available by state from various issues of the Statistical Abstract of India (GoI 1979-2001). Data are available by state for different categories of irrigated area: government canal, private canal, tank, well, and others. These were added to get the total net area irrigated (as opposed to gross irrigated area where an area irrigated twice in different sowing seasons is entered twice). The unit of measurement in which the data are reported is thousand of hectares. This was converted to square kilometres (one hectare is 10,000 square metres) and then obtained as percentage of total state area available in square kilometres.

The expenditure variables had some missing entries. These missing values were obtained by linear interpolation that is, running a regression of the available data (without the missing years) of these variables on a constant and a time trend variable and then obtaining the fitted value of the variable for the missing years. For health and education expenditure the data were missing for the years 1995, 1998 and 2000 and for net state domestic product for 1998, 1999 and 2000. Thus, the (log) expenditure variable, y,[22] was regressed on a constant and a time trend, t (1 to 23 for the years 1978 to 2000 with a missing entry for the years for which the data are not available) for each state: $y = \alpha + \beta t$ where β is the coefficient on the time trend and α is the constant. To obtain the missing expenditure value, the missing year value was substituted in the above equation for each state.[23] Thus, for example, for 1998 the value of t would be 21.

The poverty head count indices (**povertyr** for rural head count index and **povertyu** for urban head count index) available from the dataset are based on the poverty lines recommended by the Planning Commission (1993) with the rural poverty line at per capita monthly expenditure of Rs. 49 at October 1973-June 1974 all-India rural prices and the urban poverty line at per capita monthly expenditure of Rs. 57 at October 1973-June 1974 all-India urban prices. The measures are based on various rounds of the National Sample Survey (NSS), which are not conducted every year. For the period under consideration in this analysis the surveys were conducted in 1978, 1983, 1987, 1988, 1990, 1991, and 1992. For the intervening years, Ozler et al. (1996) used interpolation. The estimates of head count measures are based on the methodology of Datt and Ravallion (1992), using grouped distributions of per capita expenditure published by the NSS and parameterized Lorenz curves. Parametric specifications of the Lorenz curve have been proposed so as to fit sample Lorenz Curves from actual data. Two such parameterizations used by Datt and Ravallion are by Kakwani (1980) and Villasenor and Arnold (1989)(see Appendix B for details).

3.5.1 Summary statistics

The summary statistics for all variables included in the empirical analysis are presented in Table 3.18 for the 345 observations (15 states times 23 years from 1978 to 2000). The mean

[22] A log-levels model was used in the analysis (see p.172).

[23] This type of interpolation may have implications for the degrees of freedom, but the number of missing data points that required interpolation is only six for each state (three for health and three for education expenditure), the number of the degrees of freedom is still fairly large. The total number of observations when using the one period lag of health expenditure is 330. We lose six degrees of freedom for each state for which we have to interpolate the missing health and education expenditure values losing six times 15 or 90 degrees of freedom and resulting in 240 independent observations.

Table 3.18: Summary statistics of variables used in the state level panel data analysis

Variable[a]	Obs.[b]	Mean	SD[c]	Min	Max
API (per 1000 population)	345	3.84	5.45	0.02	59.44
popden (sq km)	345	345.76	189.81	91.40	886.73
rain (mm)	345	194.22	113.87	30.27	526.76
irrig (percentage)	345	21.22	19.14	2.17	79.50
stinc (Rs per capita per year)	345	1468.25	593.94	584.35	3834.06
healthex (Rs per capita per year)	345	16.22	3.93	5.69	29.54
eduex (Rs per capita per year)	345	53.47	19.95	18.61	118.58
urban (percentage)	345	24.51	8.17	9.53	42.69
povertyr (percentage)	345	42.01	13.43	3.16	78.78
povertyu (percentage)	345	34.53	11.80	6.55	59.75

a. See variable descriptions in Table 7.1; b. Obs. indicates number of observations; c. SD indicates standard deviation.

API for all the 15 states in the period under consideration was 3.84 per 1000 population. The mean population density was 346 persons per square kilometre with a wide variation as can be seen from the standard deviation of 189.81. The average annual rainfall was 194 mm with a standard deviation of 113.87. On average, only a fifth of the state area was irrigated. This conceals considerable variation across states with the largest share of irrigated area being 80 percent for Punjab in 1999 and the lowest for Assam in 2000 at two percent. The average value of the (real) state income was approximately Rs 1,469 (standard deviation of Rs 593.94) per person per year. The mean **healthex** was Rs 16 and the corresponding education expenditure was more than three times higher at Rs 53.47. The respective standard deviations are Rs 3.93 and Rs 19.95 indicating a wider variation in the education measure. Most of the population lived in rural areas and only a little over a fifth lived in urban areas. On average, little less than half the population (42 percent) in rural areas lived under the poverty line and in urban areas the head count ratio was slightly lower at 34 percent.

3.6 Summary

This chapter discussed the three sources of data for the three levels of analysis to be undertaken in this thesis – the NFHS for the household level for 1992-93 and 1998-99, the district level censuses for 1991 and 2001, and the panel based on 15 states for the inter-state analysis. We find that these data provide a sufficiently rich source of information needed to model malaria incidence.

The household level data have been used to address the first two research questions, mainly to investigate the effect of wealth and other socio-economic factors on malaria incidence over time and across settlement-types. Specifically, the explanatory variables identified from the dataset and based on the literature reviewed in the second chapter are social characteristics of the household head mainly caste, gender, age, occupation, and the education level; living conditions including type of house, whether the house has a separate kitchen, whether it has access to sanitation, the type of fuel used for cooking, type of drinking water source used; ownership of agricultural land – irrigated and non-irrigated – and such agricultural assets as a water pump, and finally the ownership of consumer durable assets like a radio or TV. The district data will be used to examine the effect of infrastructure and other variables on household malaria incidence. The explanatory variables identified are the extent of district area irrigated by canals, tube wells, and other sources; percentage of villages with an adult literacy centre and a primary, middle or high school; average distance to a drinking water source in the district; percentage of villages with a commuting facility; percentage of district area under forests; district population density; percentage of villages with power; and extent of rainfall. The inter-state analysis addresses the question of whether it is aggregate income or public health expenditure that has a greater impact on reducing malaria incidence. Aside from these two variables, it has also included education expenditure, poverty head count ratios, rainfall, and population density as explanatory variables.

One limitation of the household data is that these do not have data on any prophylactic measures such as the use of mosquito repellents or insecticide treated nets. District dummies have been used instead to capture some of these unobservables in the household analysis. A second drawback of the NFHS data is the lack of a current income or expenditure measure. However, this limitation has been overcome by the use of a consumer durable asset wealth index using Principal Components Analysis. The use of these assets also helps in countering the potential problem of endogeneity inherent in a study of the health-income relationship. However, as noted, in the inter-state analysis the use of current income is not a problem as we use the data on current state domestic product. A limitation of the state level data, though, is that it does not provide us with separate expenditure data on malaria control measures—the health expenditure category includes expenses on family planning, public health and sanitation, as well as water supply. While the use of this combined category is useful for the purpose of the analysis undertaken here, it would also be informative to know what the effect of expenditure on malaria programmes *per se* is and how it differs from that on more general health expenditure.

The data on the dependent variable, malaria incidence, are drawn from two different sources – the NFHS for the household analysis and the NMEP for the inter-state analysis. There are inherent differences in the method of collecting these data. While the NFHS data are based on the method of recall where the head of household is asked if he/she had malaria in the last three months, the NMEP reporting is based on actual examination of blood smears. There might be an inclination to believe that the latter, based on a blood examination, are a more reliable source. Perhaps the reason why the NFHS has stopped collecting data on malaria incidence from its third phase is apprehension regarding its own estimates. However, as discussed in this chapter, if we were to compare malaria-like fever cases, rather than actual malaria incidence cases, and take into account the imperfections in the NMEP data collection, the discrepancy between the two different estimates narrows down considerably. The NFHS, in fact, provides an excellent source for exploring the relationship of malaria incidence with a host of factors, seldom available in such detail in most datasets. Not only is this demographic survey available for twenty five states in India, but it has been conducted in many African and Asian countries also. We now discuss the methodology used to address the research questions in the next chapter.

Chapter 4

Methodology

This chapter first discusses the probit methodology used in the household analysis and the method used to deal with district data with specific reference to the rural sample. The next section then explains the methodology used for decomposing the malaria difference between the urban and the rural settlement types. The final section discusses the panel econometric methodology used for the state level analysis.

4.1 Household analysis

An analysis of the incidence of malaria at the household level using the NFHS data, where the occurrence of malaria is a dichotomous variable, lends itself suitably to the use of the probit model.[1] We use the probit model rather than the alternative logit model (a) for its historical relevance in the field of entomology, closely tied to the study of malaria caused through the bite of the mosquito and (b) because the probit can be related to an underlying normal distribution of tolerance levels (Aitchison and Brown, 1957 in Cramer 2003). In the current study, this is the normal distribution of the tolerance of Uttar Pradesh household heads towards contracting malaria. A household head's level of tolerance for the disease would depend on such (internal)[2] factors as blood type as discussed earlier in Chapter 2.

The probit (short for probability unit) was introduced by the entomologist, Charles Bliss (Bliss 1934a in Cramer 2003, Bliss 1934b in Cramer 2003) to examine the effect of a poison dosage on a population of insects mainly as a convenient scale for measuring normal deviates. In Bliss's model the dependent variable was whether an insect is unable

[1]This section on methodology is drawn from Maddala (1983), Greene (2003), Pindyck and Rubinfeld (1998), and Kennedy (1998) unless indicated otherwise.

[2]Extraneous factors affecting incidence include such factors as socio-economic status and climate.

to tolerate a given dosage of insecticide and dies (success) or whether it is able to tolerate the dosage and survives (failure). For our model the probability of success is quite the contrary. In a manner of speaking, it captures not the demise of the insect (mosquito) but that of its success in transmitting malaria to humans conditional on a number of factors. Similarly, failure could be seen as the inability to transmit malaria. An important qualification in this study is the underlying assumption that everyone in the sample was actually bitten by a mosquito. In Bliss's model all insects received a poison dosage whereas in the malaria model not all humans may receive the dose of a mosquito bite. There may be some individuals who were not bitten at all, some who were bitten but not affected, and others who were bitten and contracted malaria. However, these distinctions cannot be determined with certainty and neither is it possible to collect data that informs on these issues.

Malaria incidence in the current study is modelled as a function of a number of socio-economic and other factors contained in the vector x_i. It is a model of a binary variable where the dependent variable – whether a head of household i suffers from malaria or not in the last three months – denoted by y_i takes on the values of either zero or one, where zero indicates failure to contract malaria and one indicates 'success' in contracting malaria for head of household i in the last three months. This is a function of a number of socio-economic and other variables. Since the variable y_i can take on only two values, the binary dependent variable model is cast in terms of an underlying latent or unobserved continuous variable y^*, which can then be translated into an observed variable y_i. The probability that the event y occurs is then given by the probability that y*>0 and similarly the probability that it does not occur is the probability that $y^* \leq 0$. Thus:

$$Prob(y_i = 1) = Prob(y_i^* > 0) \tag{4.1}$$

The latent variable y_i^* is in turn given by:

$$\mathrm{y_i^*} = \beta' \mathbf{x_i} + \mathrm{u_i} \tag{4.2}$$

where x_i is the vector of characteristics, β is the parameter vector, u_i is the disturbance term and $E(u_i)=0$ and the variance is σ^2. Further,

$$E(y_i^*/x_i) = \beta' x_i \tag{4.3}$$

In the right hand side of the expression in 4.1, subtracting by the mean of y_i^* (Equation 4.3) we get:

$$Prob(y_i^* - \beta' x_i > -\beta' x_i) \tag{4.4}$$

Dividing by σ throughout we have:

$$Prob(\frac{y_i^* - \beta' x_i}{\sigma} > \frac{-\beta' x_i}{\sigma}) \tag{4.5}$$

From Equation 4.2, $u_i = y_i^* - \beta' x_i$. Substituting in the above equation we get:

$$Prob(\frac{u_i}{\sigma} > \frac{-\beta' x_i}{\sigma}) = Prob(\frac{u_i}{\sigma} \leq \frac{\beta' x_i}{\sigma}) = \Phi(\beta' x_i/\sigma) \tag{4.6}$$

given the symmetric nature of the normal distribution and where $\Phi(.)$ denotes the CDF operator for the standard normal.

Since u_i is distributed normally with mean zero and variance σ^2, the above is a normal CDF with u_i/σ serving as the standardised variable, say z. Thus:

$$\Phi(\beta' x_i) = \int_{-\infty}^{\beta' x_i/\sigma} \frac{1}{2\pi} exp(\frac{-z^2}{2}) dz \tag{4.7}$$

Since this is one equation with two unknowns, β and σ, we can only estimate β/σ but not each separately. Thus, σ is set to one in order to resolve this identification problem. Imposing this condition does not change the observed data because the observed y values (zero or one) are scale invariant. Thus, if u_i is scaled by any positive constant the observed y values are unchanged as these only depend on the sign of $y*$. Thus, there is no information on σ in the data and it cannot be estimated anyway and setting it equal to one enables the estimation of the parameter vector β.

The observed values of y (0 or 1) can be thought of as a binomial process where the probability of an event occurring or not occurring varies across trials according to x_i. The estimation of this probability model is then carried out using the method of maximum likelihood and the likelihood function is given as:

$$\mathcal{L} = \prod_{i=1}^{N_R} \Phi(\beta' x_i)^{y_i}[1 - \Phi(\beta' x_i)^{(1-y_i)}] \quad (4.8)$$

where N_R is the total number of rural households and $\Phi(\beta' x_i)$ is the probability of contracting malaria and 1-$\Phi(\beta' x_i)$ is the probability of failure to contract malaria. $\Phi(.)$ is the CDF of the standard normal distribution. Taking natural logs on both sides:

$$L = \sum_{i=1}^{N_R} y_i ln\Phi(\beta' x_i) + \sum_{i=1}^{N_R} (1 - y_i) ln[1{-}\Phi(\beta' x_i)] \quad (4.9)$$

The first derivatives of the above with respect to β, denoted by *S(β)*, yield the first order conditions that may be used to solve for the parameter vector, β . The first order condition for each β, known as the score function, is:

$$S(\beta_i) = \frac{\partial L}{\partial \beta_i} = 0 \quad (4.10)$$

The first order conditions, as we can see, are non linear in the parameters and are solved by using iterative methods such as the Newton-Raphson method, which utilises the information matrix, *I(β)*, containing the negatives of the expected values of the second order derivatives.

The information matrix is:

$$I(\beta) = E(\frac{-\partial^2 L}{\partial \beta \beta'}) \quad (4.11)$$

The probit coefficients are interpreted in terms of the standardised probit index (standardised using the normal density function as mentioned above). Thus, a probit coefficient of value *+b* implies that a small change in the value of the corresponding x variable raises the standardised probit index by b units of a standard deviation. More informative insights can be obtained by calculating marginal effects for continuous variables and impact effects for dummy variables. The marginal effects are given as:

$$\frac{\partial Prob(y_i = 1)}{\partial x_k} = \frac{\partial \Phi}{\partial x_k} = \phi(\beta' \bar{x})\beta_k \quad (4.12)$$

where x_k indicates the x variable for which the marginal effect is being calculated. The bar on x indicates that the marginal effects are evaluated at the sample average value (averaged over the n households) for each of the x variables. The impact effects, on the other hand, are calculated as a discrete change given by:

$$\Delta = \Phi(\beta'\bar{x} + \delta) - \Phi(\beta'\bar{x}) \tag{4.13}$$

where δ is the coefficient on the dummy variable for which the impact effect is being calculated. Thus, it is computed as the difference between the value of the probit index when the dummy variable takes on the value one and when the dummy takes on the value zero. Like the marginal effects, the impact effects are evaluated at the average value for the continuous variables denoted by $\bar{x}$.

The standard errors are computed using the linear approximation approach or the delta method and conventionally the asymptotic variance-covariance matrix of $\hat{\beta}$ is estimated as $\mathrm{I}(\hat{\beta})^{-1}$. In the current study the Huber correction is made as follows:

$$I(\beta)^{-1}(\frac{\partial lnL}{\partial\beta})^{T}(\frac{\partial lnL}{\partial\beta})I(\beta)^{-1} \tag{4.14}$$

where T is the transpose of the matrix. This correction is generally deemed as appropriate in the face of violation of both normality and homoscedasticity.

There is no universally accepted goodness of fit measure for binary dependent variable models although the pseudo R-squared is commonly used. The measure suggested by McFadden (1974) has been used in this analysis. Let L_{UR} be the maximum of the log-likelihood function when it is maximised with respect to all parameters including the constant. Let L_o be the maximum of the log-likelihood when maximising with respect to the constant only and, hence, restricting the slope parameters to zero. Then McFadden's pseudo R-squared is given as:

$$1 - \frac{L_{UR}}{L_o} \tag{4.15}$$

Thus, this measure compares the likelihood for the intercept only model against the likelihood for the model with the predictors included.

The specification also includes forty-eight district dummies to capture other factors not controlled for in the variables available at the household level. These replace the constant term in the coefficient vector β in the above equation. In order to express this more clearly let α_j be a $K{\times}1$ vector of district dummies then the probit equation may be written as:

$$Prob[y_i = 1] = \Phi(\alpha_j + \beta' x_i) \quad j = 1 \ldots K \tag{4.16}$$

The probit regression does not use sample weights as we are interested in examining the relationship between malaria incidence and various socio-economic and other factors. Deaton (1997) argues that there is no advantage of a weighted regression over an unweighted one with reference to survey data consisting of a number of sectors (in case of the rural analysis, these would be districts and in the urban analysis, regions).

It may be noted that in running the probit model five districts were dropped due to perfect classification of 'failure' implying that in these districts there were no cases of malaria—whenever a household head belonged to any of these districts, the event, failure to contract malaria, always occurred. Such variables are termed perfect classifiers and explain either all or none of the variation in the dependent variable, malaria incidence. The underlying reason for this problem lies in the fact that the probit constrains the fitted probabilities to lie between zero and one. In the case of perfect classifiers the predicted probabilities for these observations are actually equal to one or zero (as in our case: household head does not have malaria or 'failure' to have malaria). In these cases, the unconstrained maximum likelihood for β_k equals ∞ and $|\beta' x_i| = \infty$ and hence, if $\alpha_j + \beta' x_i = \infty$ then $\Phi(\alpha_j + \beta' x_i) = 1$. Similarly, if $\alpha_j + \beta' X_i = -\infty$ then $\Phi(\alpha_j + \beta' x_i) = 0$ (our case). Thus, the maximum likelihood estimates of the probit coefficients are not defined (Ruud 2000). In some computer packages this results in very high values of the coefficient estimates for the perfect classifiers and in the case of Stata, the package used in this study, the observations corresponding to the perfect classifiers are dropped. Since the information contained in these district dummy variables is not useful and cannot yield meaningful estimates, these were dropped and the model was re-estimated with the reduced sample, which contained forty-eight districts and 7,287 households.

In order to allow for ease in interpretation of the coefficients on the district dummies these were converted into deviations from a weighted mean following a methodology adopted by Krueger and Summers (1988). The weights are the proportion of households sampled in each district. Thus, the weighted mean is given as:

$$\sum_{k=1}^{K} \hat{\alpha}_k (n_k / N) \tag{4.17}$$

where K is the total number of districts, equal to 48, $\hat{\alpha}_k$ is the estimated effect for district k and n_k/N is the proportion of households falling in the district k and $N = \sum n_k$. The estimated deviations, $\hat{d}_k$, of the estimated effects of each district from the above weighted average were obtained as:

$$\hat{d}_k = \hat{\alpha}_k - \sum_{k=1}^{K} \hat{\alpha}_k (n_k/N) \tag{4.18}$$

The estimated deviations from the mean indicate whether a particular district lies above or below the district average for malaria incidence. Having transformed the coefficients, the standard errors on these coefficients ought to undergo a transformation as well. Zanchi (1998) provides a way of doing this briefly discussed in the ensuing paragraphs.

Let s be a $K \times 1$ vector with the $K = 48$ rows containing the proportion of households n_k/N sampled in each district. To denote the estimated $\hat{\alpha}$ vector of district dummies in matrix notation let z be a $K \times K$ identity matrix and let e be a $K \times 1$ row vector of ones. Then the deviation matrix can be written as:

$$\begin{aligned} \hat{d} &= \hat{\alpha} - es'\hat{\alpha} \\ &= (z - es')\hat{\alpha} \end{aligned} \tag{4.19}$$

The variance-covariance matrix of $\hat{d}$ is then given as:

$$\Sigma_{\hat{d}} = (z - es')\Sigma_{\hat{\alpha}}(z' - se') \tag{4.20}$$

The diagonal elements of the above variance-covariance matrix give the (transformed) standard errors for the estimated district deviations from the weighted means. The transformed standard errors were used to test whether the deviation from the weighted mean for a particular district is statistically different from the weighted average.

In order to understand what factors are significant in influencing malaria incidence at the district level, an auxiliary regression of the coefficients on the district dummies from the probit regression was run on district level variables. This analysis consists of a two-stage regression procedure where the first stage is the probit regression discussed above and the second stage is a weighted linear regression (explained below) of the district coefficients on district level secondary data.

In the second-stage weighted least squares regression, the inverse of the standard errors on the district coefficients were used as weights in order to provide larger weight in estimation to the more precisely estimated effects from the original probit model. This also accounts, indirectly, for the number of households sampled in the different districts given the inverse relation between standard errors and sample size. The auxiliary regression is, therefore, a weighted least squares regression with the inverse of the standard errors on the district coefficients constituting the weights. Substituting $\hat{d_k}$ with D_k, the estimated model may be expressed as:

$$\hat{D}_k = \hat{\gamma}' Z_k + \varepsilon_i \tag{4.21}$$

where γ is the parameter vector and Z_k is the vector of characteristics for district k.

There is an argument for including the district level variables in the first stage probit regression itself and hence interpret the coefficients on these as the estimated effects of district level characteristics on the malaria incidence at the household level. In other words, use a single equation with both the household and the district level variables. However, this approach has two weaknesses: (1) it does not allow for the separate identification of district level effects; (2) it artificially inflates the sample sizes of the district level variables thus potentially inducing spurious precision; and (3) the district level dummies capture other omitted factors that may be relevant to malaria incidence and all of these are not captured with the selected district level variables available from the census data. At the same time, it is not possible to improve this specification by including district dummies into this equation as it would then not be possible to distinguish between these district level dummy effects from the continuous variables (which are repeated for all households within a particular district).

From a policy perspective, however, using the single equation model provides more interpretable results unlike the two-stage model where the effects of the variables are interpretable as whether the propensity to contract malaria, as measured in terms of the standardised probit index, is above or below the overall district level average expressed in standard deviation units. It may be argued that the single equation model is also econometrically more efficient since the estimation procedure does not involve two steps and the justification for incorporating the district variables into the primary model (in the single equation model) is that of achieving econometric efficiency rather than improving explanatory power. However, the use of the weighted least squares estimation in the second stage model attenuates this problem to some extent.

Given that either model could be used depending on researcher preference, it is worthwhile to check how much of the variation in malaria is explained by each of the two models. For this the goodness of fit measures (McFadden's pseudo R-squared and the squared correlation coefficient between observed and predicted malaria incidence) were compared and the Vuong (1989) test was conducted to see which model, if any, may be the preferred choice. The Vuong test is now briefly explained.

If we treat the probit model with the district level dummy effects as Model 1 and the model with the district level secondary data (the single equation model) as Model 2 we may write these as:

$$Model\ 1 : prob(y_i = 1) = \Phi(x_i'\beta) \tag{4.22}$$

$$Model\ 2 : prob(y_i = 1) = \Phi(z_i'\gamma) \tag{4.23}$$

The competing hypotheses to test which model may be the preferred choice (if any) can be formulated as:

$$H_f : prob(y_i = 1) = \Phi(x_i'\beta) \text{ is true} \tag{4.24}$$

$$H_g : prob(y_i = 1) = \Phi(z_i'\gamma) \text{ is true} \tag{4.25}$$

In the first instance we need to calculate the likelihood ratio after first computing the probit predictions for all n observations for each of the two models. This is given as:

$$LR = 1/n \sum_{i=1}^{n} [y_i \times ln[\frac{\Phi(x_i'\beta)}{\Phi(z_i'\gamma)}] + (1 - y_i) \times ln[\frac{1 - \Phi(x_i'\beta)}{1 - \Phi(z_i'\gamma)}] - [[\frac{p}{2}]ln(n) - [\frac{q}{2}]ln(n)] \tag{4.26}$$

where $\Phi(x_i'\beta)$ is the empirical prediction for observation i using Model 1 and $\Phi(z_i'\gamma)$ is the empirical prediction for observation i using Model 2. y_i is the dependent variable and n is the sample size, p is the number of parameters from Model 1, and q is the number of parameters from Model 2. The variance for the above expression is defined as:

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}[y_i \times ln[\frac{\Phi(x_i'\beta)}{\Phi(z_i'\gamma)}] + (1-y_i) \times ln[\frac{1-\Phi(x_i'\beta)}{1-\Phi(z_i'\gamma)}]]^2 - M^2 \tag{4.27}$$

$$\text{where } \ M = \frac{1}{n}\sum_{i=1}^{n}[y_i \times ln[\frac{\Phi(x_i'\beta)}{\Phi(z_i'\gamma)}] + (1-y_i) \times ln[\frac{1-\Phi(x_i'\beta)}{1-\Phi(z_i'\gamma)}]] \tag{4.28}$$

and the Vuong Test is given as:

$$\frac{LR}{(\sqrt{n}) \times \omega} \tag{4.29}$$

According to Vuong,

$$Under\ H_o : \frac{LR}{(\sqrt{n}) \times \omega} \rightarrow N(0,1) \tag{4.30}$$

$$Under\ H_f : \frac{LR}{(\sqrt{n}) \times \omega} \rightarrow +\infty \tag{4.31}$$

$$Under\ H_g : \frac{LR}{(\sqrt{n}) \times \omega} \rightarrow -\infty \tag{4.32}$$

The test statistic is a z-score such that a value in excess of 1.96 would imply that the data favour the model with the district level fixed effects; a value smaller than -1.96 implies that the data favour the model with the district level variables; and a value in between these two boundaries indicates that the data cannot discriminate between either of the two models leaving the choice to the researcher.

Another considerably important issue requires mention. This is the problem of misclassification of the left hand side variables in the probit regression. Since malaria can mimic the symptoms of a number of diseases it is likely that the respondents reported having malaria when, in fact, it was not the case and vice-versa. In linear regression models such mismeasurement in the left-hand side or dependent variables leads to the estimation of inefficient coefficients. But, in binary dependent variable models – the probit in our case – the estimators are biased and inefficient. It has been proposed that the problem of misclassification can be dealt with by accounting for the probability of error in misclassification in the model both for the case where success is classified as a failure and the case where failure is classified as a success (Hausman 2001). However, in the present model of malaria incidence it is not possible to arrive at these probabilities as the only way to check for the misclassification is through a blood test. Even when that might be possible (and it is not the case in the NFHS dataset being used here) there is likely to be scope for misclassification as those who do not have fever can sometimes test positive for malaria. Further, there are a number of problems in collecting blood samples and in reporting accurately the malaria cases to arrive at malaria incidence as discussed in Chapter 3. Thus, the problem of misclassification is recognised and merits caution in interpreting the coefficient estimates, but nothing can be done to address this empirically.

4.2 Decomposition

Before beginning the explanation for the methodology used for decomposing the malaria difference between the rural and urban sectors, the probit methodology is repeated briefly here with specific reference to the urban sample.

Urban malaria is modelled using a probit regression model with the binary dependent variable, y_i, capturing the incidence of malaria. The explanatory variables constitute socio-economic and other factors and are contained in Z_i where i ranges from 1 to 2,220 households. The model may be written as:

$$Prob[y_i = 1] = \Phi(Z_i\gamma) \tag{4.33}$$

where y_i is an $N_u{\times}1$ vector of the dependent variable, malaria incidence, Z_i is an $N_U{\times}$j matrix of independent variables and γ is a $J{\times}\mathit{1}$ vector of unknown parameters. N_U indicates the number of households in the urban sample and J is the number of explanatory variables used in this specification. The likelihood function is given as:

$$\mathcal{L} = \prod_{i=1}^{N_U} \Phi(\gamma' Z_i)^{y_i}[1 - \Phi(\gamma' Z_i)]^{(1-y_i)} \tag{4.34}$$

where $\Phi(\gamma' Z_i)$ is the probability of contracting malaria and 1-$\Phi(\gamma' Z_i)$ is the probability of failure to contract malaria. Taking natural logs on both sides:

$$L = \sum_{i=1}^{N_U} y_i ln\Phi(\gamma' Z_i) + \sum_{i=1}^{N_U} (1 - y_i) ln[1-\Phi(\gamma' Z_i)] \tag{4.35}$$

The specification also included region dummies to capture other factors not controlled for in the variables available at the household level. District dummies as in the case of the rural analysis were not used because several of these did not have any malaria cases and seven were dropped due to perfect classification. The decomposition was then conducted for comparable rural and urban specifications using region dummies rather than district dummies for the rural as well as the urban sector. The same approach has been applied to the temporal decomposition pursued in Chapter 5. In the ensuing discussion, the subscripts R and U can be interchanged with the time periods, 1992-93 and 1998-99.

The decomposition analysis aims to capture whether the difference in rural and urban malaria incidence is attributable to differences in treatment or endowments. This technique has its origins in the analysis of racial and gender wage differences, using linear regression models. But, in the current study non-linear models have been used to undertake the decompositions. The decomposition entails an estimation of the overall or aggregate effect of both the characteristic and the coefficient differences as well as of the differences at the individual variable level (or groups of variables where these are categorical). The method to obtain these differences is explained below. A Wald's χ^2 will be conducted to test for separation of the data where the null hypothesis being tested is that the effect on malaria incidence does not differ by settlement type.

We begin with the malaria equation where malaria incidence, the dependent variable, is a function of a number of explanatory variables. It may be noted that while the original formulation of the decomposition analysis was for continuous variables, its application is not restricted to these alone and can be applied to binary dependent variable models as in Al-Samarrai and Reilly (2000). The ensuing paragraphs draw on this reference, Yun (2004), and Yun (2005).

$$Prob[y_S = 1] = \Phi(X_S \beta_S) \tag{4.36}$$

where S=R,U represents the rural or urban sector, respectively. X is an $N \times K$ matrix of explanatory variables and β is a $K \times 1$ vector of coefficients.

The average difference between rural and urban malaria incidence can then be expressed as:

$$\Delta = \overline{prob[y_{\mathrm{R}}=1]} - \overline{prob[y_{\mathrm{U}}=1]} = \overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_U)} \tag{4.37}$$

where the overbar indicates that sample average values have been obtained. Thus, for each sector the malaria incidence is obtained as the average over all the N households calculated using the estimated vector β_S for that sector.

The difference in malaria incidence between the rural and urban areas can be decomposed into differences between coefficients and characteristics. Adding and subtracting the term $\Phi(X_U\beta_R)$ and re-arranging we obtain:

$$\Delta = [\overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_R)}] + [\overline{\Phi(X_U\beta_R)} - \overline{\Phi(X_U\beta_U)}] \tag{4.38}$$

In the right hand side of the above equation, the term in the first square bracket represents the difference in the characteristics between rural and urban areas (with estimated coefficients fixed at their rural values) and the term in the second square bracket is the difference in malaria attributable to the difference in the coefficients (using urban characteristics), which indicates a shift in the relationship between the rural and urban sectors. This equation uses estimated rural coefficients to obtain the characteristic differences and urban characteristics for arriving at the coefficient differences. Alternatively, estimated urban coefficients and rural characteristics could also be used as weights to get the respective characteristic and coefficient differences. This method is known as the index number approach and the final estimates are likely to differ depending on which set of weights are used to capture the component parts. It is also possible to use a weighted average of the two sets of coefficients as suggested by Cotton (1988) and Neumark (1988). Thus, for example, Cotton uses a linear function of the two least squares regression coefficients as weights in the decomposition (in his case, representing the wage structure of blacks and whites) weighted by the respective proportions of white and black males employed in the labout force. The weighted coefficient, B*, would then be equal to fwBw + fbBb where Bw and Bb are the wage rates of whites and blacks, respectively, and fw and fb are the proportions of whites and blacks, respectively, employed in the labour force. This method provides extra components to the decomposition, but resolves the conventional index number problem generally encountered in this type of work. It is also worth noting that some of the aggregate estimates may be sensitive to the base group (see, for example, Oaxaca and Ransom (1999)). These two issues have not been pursued here.

Aside from the decomposition at the aggregate level, the contribution of each variable to the difference in malaria can also be worked out by assigning weights to each variable. A linearisation is proposed using Taylor's theorem and the derivation of malaria difference using this is outlined in Appendix C. The characteristic and coefficient weights in the equations below are obtained using this derived equation.

For each variable, k, the characteristic weight is obtained as:

$$W_{\Delta C}^{k} = \frac{(\overline{X}_{R}^{k} - \overline{X}_{U}^{k})\hat{\beta}_{R}^{k}}{(\overline{X}_{R} - \overline{X}_{U})\hat{\beta}_{R}} \tag{4.39}$$

where $\sum_{k=1}^{K} W_{\Delta C}^{k} = 1$

Similarly, for the difference in coefficients, the weight is:

$$W_{\Delta D}^{k} = \frac{(\hat{\beta}_{R}^{k} - \hat{\beta}_{U}^{k})\overline{X}_{U}^{k}}{(\hat{\beta}_{R} - \hat{\beta}_{U})\overline{X}_{U}} \tag{4.40}$$

where $\sum_{k=1}^{K} W_{\Delta D}^{k} = 1$

The detailed decomposition may then be expressed as:

$$\Delta = \sum_{k=1}^{K} W_{\Delta C}^{k}[\overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_R)}] + \sum_{k=1}^{K} W_{\Delta D}^{k}[\overline{\Phi(X_U\beta_R)} - \overline{\Phi(X_U\beta_U)}] \tag{4.41}$$

where W stands for the respective weights for each variable and the subscript, ΔC, indicates the difference in characteristics for each variable, *k=1* to K, and the subscript, ΔD, indicates the difference in coefficients. $W_{\Delta C}$ and $W_{\Delta D}$ are calculated using the mean values of the characteristics and coefficients (see expressions C.10 and C.11).

The next step is to obtain the sampling variances associated both with the aggregate effects as well as the component effects. These are obtained using the Delta method. Using the first part in the right-hand side of Equation 4.38, the aggregate difference in characteristics is:

$$C = \overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_R)} \tag{4.42}$$

The first derivative of the above equation with respect to β_R is:

$$\frac{\delta C}{\delta \beta_R} = G_C = \phi(\overline{X}_R\beta_R)X_R \tag{4.43}$$

The variance is then given as:

$$\hat{\sigma}_C^2 = G_C\hat{\Sigma}_{\hat{\beta}}G'_C \tag{4.44}$$

To obtain the variance for the coefficient effect we need to first obtain its derivative, G_D, with respect to $\hat{\beta}$, where for convenience we have suppressed the settlement type subscript. The coefficient effect from Equation 4.38 is:

$$D = \overline{\Phi(X_U\beta_R)} - \overline{\Phi(X_U\beta_U)} \tag{4.45}$$

The first derivative with respect to $\hat{\beta}_R$ is:

$$\frac{\delta D}{\delta \hat{\beta}_R} = \phi(\overline{X}_U\hat{\beta}_R)\overline{X}_U \tag{4.46}$$

Similarly, the first derivative with respect to $\hat{\beta}_U$ is:

$$\frac{\delta D}{\delta \hat{\beta}_U} = -\phi(\overline{X}_U\hat{\beta}_U)\overline{X}_U \tag{4.47}$$

The total derivative G_D is the sum of the above two partial derivatives:

$$G_D = \frac{\delta D}{\delta \hat{\beta}_R} + \frac{\delta D}{\delta \hat{\beta}_U} \tag{4.48}$$

The variance of the coefficient effect is expressed as:

$$\hat{\sigma}_D^2 = G_D \begin{bmatrix} \hat{\Sigma}_{\hat{\beta}_R} & 0 \\ 0 & \hat{\Sigma}_{\hat{\beta}_U} \end{bmatrix} G'_D = \frac{\delta D}{\delta \beta'_R}\hat{\Sigma}_{\hat{\beta}_R}\frac{\delta D'}{\delta \beta_R} + \frac{\delta D}{\delta \beta'_U}\hat{\Sigma}_{\hat{\beta}_U}\frac{\delta D'}{\delta \beta_U} \tag{4.49}$$

where

$$G_C = \frac{\delta C}{\delta \acute{\beta}} = [\frac{\delta C}{\delta \beta_R^1}\frac{\delta C}{\delta \beta_R^2}\cdots\frac{\delta C}{\delta \beta_R^K}] \tag{4.50}$$

and

$$G_D = \frac{\delta D}{\delta \acute{\beta}} = [\frac{\delta D}{\delta \beta_R^1} \frac{\delta D}{\delta \beta_R^2} \cdots \frac{\delta D}{\delta \beta_R^K} \vdots \frac{\delta D}{\delta \beta_U^1} \frac{\delta D}{\delta \beta_U^2} \cdots \frac{\delta D}{\delta \beta_U^K}] \tag{4.51}$$

Finally, the null hypothesis that there is no difference between the urban and rural characteristics and coefficients can be tested using a t-statistic.

The individual variable characteristic effects, C_k, and coefficient effects, D_k, for each variable/coefficient, k, are given by the expressions:

$$C_k = \sum_{k=1}^{K} W_c^k [\overline{\Phi(X_R \beta_R)} - \overline{\Phi(X_U \beta_R)}] \tag{4.52}$$

$$D_k = \sum_{k=1}^{K} W_d^k [\overline{\Phi(X_U \beta_R)} - \overline{\Phi(X_U \beta_U)}] \tag{4.53}$$

The corresponding variances are then given by:

$$\hat{\sigma}^2_{C_k} = G_{C_k} \Sigma_{\hat{\beta}_k} G'_{C_k} \tag{4.54}$$

where $\Sigma_{\hat{\beta}_k}$ represents the variance-covariance matrix associated with $\hat{\beta}_k$. Correspondingly, G_{C_k} is the partitioned derivative vector associated with the variable k.

$$\hat{\sigma}^2_{D_k} = G_{D_k} \begin{bmatrix} \hat{\Sigma}_{\hat{\beta}_R} & 0 \\ 0 & \hat{\Sigma}_{\hat{\beta}_U} \end{bmatrix} G'_{D_k} = \frac{\delta D_k}{\delta \beta'_R} \hat{\Sigma}_{\hat{\beta}_R} \frac{\delta D'_k}{\delta \beta_R} + \frac{\delta D_k}{\delta \beta'_U} \hat{\Sigma}_{\hat{\beta}_U} \frac{\delta D'_k}{\delta \beta_U} \tag{4.55}$$

and

$$G_{C_k} = \frac{\delta C_k}{\delta \beta'_R} = [\frac{\delta C_k}{\delta \beta_R^1} \frac{\delta C_k}{\delta \beta_R^2} \cdots \frac{\delta C_k}{\delta \beta_R^K}] \tag{4.56}$$

$$G_{D_k} = [\frac{\delta D_k}{\delta \beta'_R} \vdots \frac{\delta D_k}{\delta \beta'_U}] = [\frac{\delta D_k}{\delta \beta_R^1} \frac{\delta D_k}{\delta \beta_R^2} \cdots \frac{\delta D_k}{\delta \beta_R^K} \vdots \frac{\delta D_k}{\delta \beta_U^1} \frac{\delta D_k}{\delta \beta_U^2} \cdots \frac{\delta D_k}{\delta \beta_U^K}] \tag{4.57}$$

A t-test can be used here to test statistical effects.

The decomposition analysis is subject to some caveats since the regression models used are assumed to be well specified. One of the caveats is due to the restrictions inherent in any dataset (in this case, the NFHS) that may not cover all the variables that may explain malaria incidence. For example, information on such variables as the use of bed nets and other measures employed to prevent mosquito bites at the household level was not collected in the NFHS survey. The second reason is attributable to the fact that a number of variables could not be included in the current analysis as explained in Chapter 3 (Table 3.6, p.54). Finally, a third factor relates to the modification of the rural specification to a more parsimonious one. These omissions may result in a larger unexplained or treatment component than justified if such variables were available. Despite our caveats, however, we believe the specification chosen provides a good approximation of malaria incidence in both urban and rural areas.

We now discuss the method used for the state level panel study.

4.3 State level analysis

The dataset used in this study is a panel based on 15 major states covering a period of 23 years from 1978 to 2000, yielding 345 observations (15 states times 23 years). One of the advantages of panel data is that it can take into account unobserved effects that vary across states such as geography and climate but are assumed constant through time.[3] The model can be written as:

$$y_{it} = x_{it}\beta + \alpha_i + \sum_{t=1}^{T-1} D_t\gamma + \epsilon_{it} \quad t = 1\ldots 23 \text{ and } i = 1\ldots 15 \tag{4.58}$$

where y_{it} is the incidence of malaria (**API**) for state i in year t; x_{it} is a $1{\times}K$ vector of characteristics affecting malaria incidence. This vector, x, consists of the following variables: **popden**, **urban**, **rain**, **irrig**, **stinc**, **eduex**, **healthex** with a one period lag or **healthlag**, **povertyr**, and **povertyu**.[4] A one period lag of real health per capita expenditure is used rather than the contemporaneous measure based on the argument that the effect of health expenditure is likely to be realized one period later in terms of health outcomes. This means we lose 15 observations since we lose one observation for each cross-section on account of using the one period lag of the health measure rather than the contemporaneous measure, thus resulting in a total of 330 rather than 345 observations;[5] α_i represents unobserved heterogeneity or the unobserved effect varying across states but not over time; D_i are the set of time dummies and γ are the corresponding parameters for the time dummies for 21 years (since we lose one year in using the **healthlag** measure and

[3] The main reference used for this section is Wooldridge (2002a).

[4] See Table 3.17, p.79 for full forms of these abbreviations.

[5] The regression was also run with the contemporaneous health measure but the sign on health was positive though insignificant.

one to avoid the dummy variable trap) using 1979 as the base; and u_{it} are idiosyncratic disturbances. The time period dummies control for aggregate time effects like technological changes (such as the introduction of new kinds of mosquito repellents) that may have the same influence on y_{it} for all i.[6]

Three methodological issues have been addressed here: one, of choosing the appropriate functional form; two, of choosing the appropriate estimator (fixed effects, pooled OLS or random effects), and three, testing whether the variables **stinc**, **eduex**, and **healthlag** are endogenous or exogenous. To address the first of these issues, two alternative models were considered: Model 1 in which the dependent variable was logged and the independent variables were entered as levels (a log-levels model) and Model 2 where all the variables, including the dependent variable, were entered in log form (a log-log model). The dependent variable was logged in order to compress its scale, reduce the potential for heteroscedasticity and to allow ease in interpretation since logging the variable gives us estimated effects that are proportional changes. The use of the logarithm on the dependent variable also follows Anand and Kanbur (1995) where the dependent variable was used in log form.

The second methodological issue was that of choosing a pooled OLS, fixed effects or a random effects estimator. In a pooled OLS model, the unobserved effect α is treated as a constant across states and over time. If this (restrictive) assumption is violated, the estimator would be biased and inconsistent. The random effects model assumes that the unobserved effect is a random variable, thus implying zero correlation between the observed explanatory variables and the unobserved effect. Again, if this condition is violated, the random effects estimator would be inconsistent.

In the fixed effects model the unobserved effect can be eliminated through time-demeaning the data as described below. This model also has the advantage of not assuming independence of the explanatory variables from the fixed effects. The T=22 equations from Equation 4.58 (excluding the time dummies for simplicity) can be written as:

$$\mathbf{y_i} = e\alpha_i + \mathbf{x_i}\beta + \mathbf{u_i} \tag{4.59}$$

where y_i is a 22×1 vector, e is a 22×1 vector of ones, u_i is a 22×1 vector of disturbances which is $IID(0,\sigma^2)$, x_i is a $22{\times}K$ matrix of K explanatory variables and β

[6]The joint significance of the time effects can be tested using an F test under the null hypothesis that the time effects are equal to zero.

is a $K \times 1$ vector of unknown parameters. This vector can be estimated by eliminating the unobserved effects using the 'within' or 'fixed effects' transformation. To allow this, Equation 4.59 is averaged over the 22 time periods so that for each state we get:

$$\bar{y}_i = \bar{x}_i\beta + \alpha_i + \bar{u}_i \tag{4.60}$$

$$\text{where } \; \bar{y}_i = 22^{-1}\sum_{t=1}^{22} y_{it}, \;\; \bar{x}_i = 22^{-1}\sum_{t=1}^{22} x_{it}, \;\; \bar{u}_i = 22^{-1}\sum_{t=1}^{22} u_{it} \tag{4.61}$$

Subtracting Equation 4.60 from Equation 4.58 for each t (without the vector for time dummies) we get:

$$(y_{it} - \bar{y}_i) = (\bar{x}_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i) \tag{4.62}$$

The above time demeaning eliminates α_i. The fixed effects estimator $\hat{\beta}$ is then obtained by OLS estimation from the above time-demeaned equation for $N = 15$ cross sections. In matrix notation we can write Equation 4.62 by pre-multiplying Equation 4.59 by a $T \times T$ symmetric idempotent matrix $Q_T = I_T - \frac{1}{T}ee'$:

$$Qy_i = Qe\alpha_i + Qx_i\beta + Qu_i \tag{4.63}$$

where,

$$Qy_i = y_{it} - \bar{y}_i \;\; Qx_i = x_{it} - \bar{x}_i \;\; Qu_i = u_{it} - \bar{u}_i \text{ and } Qe = 0 \tag{4.64}$$

Therefore, Equation 4.63 simplifies to:

$$Qy_i = Qx_i\beta + Qu_i \tag{4.65}$$

The fixed effects estimator $\hat{\beta}$ is:

$$\hat{\beta_{FE}} = [\sum_{i=1}^{N}(x_i^{'}Qx_i)]^{-1}[\sum_{i=1}^{N}(x_i^{'}Qy_i)] \tag{4.66}$$

or,

$$[\sum_{i=1}^{N}\sum_{i=1}^{T}(x_{it}-\bar{x_i})(x_{it}-\bar{x_i})^{'}]^{-1}[(x_{it}-\bar{x_i})(y_{it}-\bar{y_i})] \tag{4.67}$$

Since the estimator uses time variation within each cross section it is also called the within estimator. The expression for the variance of the estimator is:

$$var(\hat{\beta}_{FE}) = \sigma^2[\sum_{i=1}^{N}(x_i^{'}Qx_i)]^{-1} \tag{4.68}$$

or,

$$\sigma^2[\sum_{i=1}^{N}\sum_{i=1}^{T}(x_{it}-\bar{x_i})(x_{it}-\bar{x_i})^{'}]^{-1} \tag{4.69}$$

where

$$\sigma^2 = \frac{1}{NT-N-K}\sum_{i}^{N}\sum_{i=1}^{T}[(y_{it}-\bar{y}_i-(x_{it}-\bar{x}_i))\hat{\beta}_{FE}]^2 \tag{4.70}$$

All the estimates were obtained using the robust option in Stata, which corrects for heteroscedasticity. Under the null hypothesis of homoscedasticity, the error, u_{it}, is uncorrelated with any function, h_{it}, of x_{it}. If the R-squared from a regression of $\hat{u}_{it}^2$ on 1 and the $1\times Q$ vector h_{it} is R_c^2 then the Breusch-Pagan Lagrange multiplier statistic is given as NTR_c^2 which is asymptotically distributed as χ^2 with P degrees of freedom. In the context of a panel since the deviation from homoscedastic errors is most likely to come from the error variances specific to the cross-sectional units, the hypothesis tested is:

$$\sigma_i^2 = \sigma \,\forall\, i = 1\ldots N \tag{4.71}$$

where N is the number of cross-section units or states. The test statistic is distributed as χ^2 with N degrees of freedom (Greene 2003, Baum and Schaffer 2002). The robust variance covariance matrix adjusted using the Huber procedure to account for heteroscedasticity of unknown form is given as:

$$(\sum_{i=1}^{n}x_i'x_i)^{-1}(\sum_{i=1}^{n}x_i'\hat{u}_i\hat{u}_i'x_i)(\sum_{i=1}^{n}x_i'x_i)^{-1} \tag{4.72}$$

The Hausman test can be conducted to decide whether the appropriate model is random or fixed effects. The null hypothesis under this test assumes the random effects model.

The test statistic is asymptotically distributed as χ^2 with K degrees of freedom:

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})' [v\hat{a}r(\hat{\beta}_{FE}) - v\hat{a}r(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \overset{a}{\approx} \chi^2_K \tag{4.73}$$

An alternative approach to eliminating the unobserved effects is by using first differences of the variables. Lagging the model in Equation 4.58 by one period and subtracting gives:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it} \qquad t = 2 \ldots 23 \text{ and } i = 1 \ldots 15 \tag{4.74}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta x_{it} = x_{it} - x_{it-1}$, and $\Delta u_{it} = u_{it} - u_{i,t-1}$. First-differencing eliminates the unobserved effect. The first time period for each state is lost. Since the total number of observations in the fixed effects model was 330, in the first difference model the number of observations will be 15 less than this number or 315. The above model is then run using pooled OLS and the resulting estimator $\hat{\beta}_{FD}$ is known as the first-difference estimator. The first difference model allows us to to observe if there is a relationship between changes in malaria incidence and changes in any of the explanatory variables – for example, if there is a change in the amount of public health expenditure between two years, how did it affect the growth in malaria incidence?

If any one or more variables in the regression are suspected to be potentially endogenous this may lead to biased and inconsistent estimates. The model could then be estimated using instrumental variables. The set of regressors, X, in our model may be thought of as comprising two subsets X_1 and X_2 where the former consists of potentially endogenous variables (**stinc**, **eduex**, and **healthlag**) and the latter of the exogenous variables. The endogenous variables can be instrumented by a new set Z_1. The entire set, Z, of exogenous variables could then be considered to consist of Z_1 and X_2. The instruments chosen must meet two conditions (a) they must be relevant, which means they must be highly correlated with the potentially endogenous variables and (b) they must be orthogonal to the error process. Thus, (a) and (b) together yield valid instruments.

The first condition is tested by specifying reduced form regressions of the endogenous variables on all the exogenous variables, which is the set Z. The relevant test statistic is partial R-squared suggested by Bound et al. 1995 in Baum and Schaffer (2002) for a single endogenous variable or the Shea R-squared for multiple endogenous variables. The former statistic can take the alternative form of an F-test of the joint significance of the instruments, Z_1, in the first stage regression. A rule of thumb to indicate whether the instruments are valid is that the F-statistic should not be less than 10 (Staiger and Stock 1997 in Baum and Schaffer 2002). The Bound et al. statistic, however, is not suitable when there are multiple (potential) endogenous variables (as is the case in this study). Instead, Shea's (1997 in Baum and Schaffer 2002) partial R^2 is preferred as it takes into account the intercorrelations among the instruments. This is a goodness of fit measure and not a formal test statistic. If the Shea partial R^2 measure is small compared to the Bound et al partial R^2 then the instruments may not be relevant to explain the endogenous regressors. Additionally, a value of at least 0.1 for the Shea measure may be considered indicative of instrument relevance.[7]

The second condition that the instruments must meet is that of orthogonality – the instruments must be independent from the error process. This requirement can be tested using the Hansen J or the Sargan statistic. The former corrects for possible heteroscedasticity and the latter assumes a homoscedastic error process. It is possible to conduct these tests if the number of instruments excluded from the model, which is the subset Z_1, exceeds the number of potentially endogenous variables. In that case, the number of elements, L, in the set Z (which includes the exogenous regressors, X_2 and the instruments Z_1) is greater than the total number of explanatory variables, K, that we started out with (the set X). These overidentifying restrictions are tested using the Hansen J or Sargan statistics. The J statistic is obtained by using the generalised method of moments and is distributed as χ^2 with L-K degrees of freedom. The Sargan statistic is a special case of the Hansen J statistic assuming a homoscedastic error process and is also distributed as χ^2 with L-K degrees of freedom. The null hypothesis is that all instruments are orthogonal to the error process in the primary equation. Therefore, a rejection of the null would be a violation of the second condition noted above.

After testing the above two conditions the next step is to test whether the regressors are, in fact, endogenous or not. This can be done by comparing the OLS coefficient vectors against those obtained using the instrumental variables estimation (Hausman 1978 in Wooldridge 2002b). The original test statistic takes the form of a quadratic in the differences between the two coefficient vectors. This study used an alternative regression-

[7] An example of this appears in Bussea and Groizardb (2006).

based form of the test proposed by Hausman, which is asymptotically equivalent to the original test (Hausman 1983 in Wooldridge 2002a). It entails the following steps:

1. Estimating the reduced form regression for each potentially endogenous variable on all the exogenous variables (the set Z) and obtain the residuals.

2. Adding these residuals to the regression of the dependent variable (annual parasite incidence per 1000 population in this study) on all the potentially endogenous variables, X_1 and the exogenous variables, which together constituted the original set X of K explanatory variables, but not the instruments Z_2. For a single potentially endogenous variable if the coefficient on the residual is insignificant (a t-test) it would indicate that the variable is not endogenous. For multiple endogenous variables an F-test of joint significance is conducted. A significant test statistic would be evidence that at least one of the potentially endogenous variables is, in fact, endogenous.

Thus, the above essentially tests if the reduced form error from the first step above is correlated with the disturbance term. If not, then exogeneity is assumed for the regressor or regressors of interest. These tests can be conducted within a pooled OLS or fixed effects framework, but not for a random effects model, which requires a generalised least squares framework.

4.4 Summary

In this chapter we reviewed the details of the methodology followed for the analysis of factors affecting malaria incidence in rural and urban Uttar Pradesh at the household and district level, the method of decomposition of malaria incidence between the rural and urban sectors and over time and finally the method followed to examine state level malaria incidence using a panel of state level indicators. The next three chapters discuss the empirical results – Chapter 5 for the rural probit and district analysis for both time periods, Chapter 6 for the urban analysis for these time periods and the decomposition of malaria difference across the rural and urban sectors and Chapter 7 for the state level analysis.

Chapter 5

Analysis of rural malaria incidence in Uttar Pradesh

The research questions addressed in this chapter are:

1. What are the socio-economic factors that affect the incidence of malaria at the household and district level? Are the effects of these factors robust over time?

2. Does household wealth and other socio-economic status variables have a negative impact on malaria incidence?

The above questions have been addressed here with respect to rural Uttar Pradesh. The analysis of urban malaria incidence and differences in factors affecting malaria across settlement types is dealt with in the next chapter. We may recall that there are intrinsic differences in the incidence of malaria between rural and urban areas primarily because of differences in the habitat of the main vectors found here – *An.culicifacies* in rural and *An.stephensi* in urban areas. The breeding of the former vector is confined to stagnant, fresh water collections such as rivers, ponds, lakes, and agricultural fields, mostly found in rural settings. The urban vector, on the other hand, can breed in these locales, which are not predominant in urban areas. But, it has also adapted to breeding in all kinds of containers to collect water found in dry areas and particularly in towns and cities (see Chapter 2). It is this adaptive behaviour that has led to a rise in malaria cases in urban settlement types. The importance of focusing considerable attention on rural malaria incidence comes also from the fact that 80 percent of the population of Uttar Pradesh resides in rural areas. The incidence of rural malaria is also higher at 8.5 percent for

1992-93 as compared to 2.4 percent for urban areas (see Table 3.7, p.55) and 5.1 percent and 1.4 percent, respectively for rural and urban Uttar Pradesh in 1998-99 (Tables 3.5 and 3.8). We now present a quick review of the methodology followed in this chapter. For details, see Chapter 4.

The analysis of malaria incidence has been undertaken using the probit model since we are dealing with a binary dependent variable. The explanatory variables were chosen on the basis of the literature explaining factors that influence malaria incidence (see Table 3.3). These variables include the social characteristics of the household head, the living conditions, and the ownership of agricultural land, agricultural assets and a consumer durable asset wealth index. These reflect socio-economic status and the asset wealth index has been used as a proxy for permanent income as explained in Chapter 3.

For an analysis of district level factors influencing malaria incidence, the coefficients on the district dummies from the household regression provide the dependent variable. Thus, this part of the analysis constituted a two-stage regression procedure where the first stage is the probit regression and the second stage the district level regression. The explanatory variables included in this second stage regression are such variables as percentage of district area irrigated by different types of irrigation systems, percentage of villages in the district with commuting facilities, percentage of villages with power supply, average distance to drinking water source in the district, as well as controls for population density and rainfall. These explanatory variables were also included in the first stage probit regression and the results of this exercise were found to be very similar to the two-stage model.

A comparison of the factors determining malaria incidence in 1992-93 with those in 1998-99 is also undertaken here both for the household and the district level data. This part of the analysis addresses the research question of whether the factors affecting malaria incidence are robust over time.

We now begin with the household analysis for 1992-93. The subsequent section deals with the district level regression. The third section compares the factors affecting malaria incidence at the household and district level in 1992-93 with those in 1998-99 for comparable specifications. The last section presents a summary of our findings.

5.1 Household level analysis: 1992-93

The selected probit specification consists of forty-eight district dummies to capture climate, ecology, and other unobservables for variables that are not available in the NFHS. The other explanatory variables are: (a) household characteristics and social status: dummy for caste, dummy for gender, five dummies for age of head, three dummies for level of education, and three dummies for occupation (OC) status; (b) living conditions: two dummies for type of house, dummy for whether house has a separate kitchen, dummy for whether house has electricity, dummy for sanitation, two dummies for type of fuel used, and three dummies for type of drinking water source; (c) land: two dummies each for size of irrigated and non-irrigated area and four interaction dummies for combinations of irrigated and non-irrigated land; (d) agricultural assets: dummy for whether household owns a water pump; and (e) non-agricultural consumer durable asset index.

Table 5.1 reports the results from the probit regression and Appendix D reports the corresponding marginal and impact effects. At the outset, we can identify the factors that determine malaria incidence in 1992-93. These are caste, education, electricity, fuel-type, and ownership of a water pump and irrigated area. The coefficient on the asset index is not significant. However, further analysis using quintiles for this variable was pursued. We first discuss the results of this exercise in the next subsection. This addresses the research question of whether household wealth has a negative impact on malaria incidence. The subsequent subsection discusses the impact of the different socio-economic status variables included in the specification on malaria incidence.

5.1.1 Impact of wealth on malaria incidence

We find that the coefficient on the asset index variable is not significant indicating that wealth does not influence malaria incidence in rural Uttar Pradesh in 1992-93. In order to assess whether the relationship between asset wealth as measured here and malaria incidence is linear or non-linear, a quadratic was introduced into the specification. However, the quadratic term did not yield a well determined effect. Therefore, linear splines for the consumer durable index were used instead. These test for a linear relationship for different levels of the index and malaria incidence. Each spline is assigned a threshold value by which to differentiate one from the other and these are referred to as knots. The knots in

Table 5.1: Probit of malaria incidence: 1992-93, rural

Variable	*Coefficient*	*SE*[a]
Asset index	-0.003	0.018
Caste	-0.112**	0.055
Gender	0.190*	0.107
Age dummies (Reference category: 18-27 years)		
28 to 37 years	-0.007	0.086
38 to 47 years	0.093	0.084
48 to 57 years	0.048	0.089
58 to 67 years	0.041	0.093
Above 67 years	-0.017	0.110
Education (Reference category: Illiterate)		
High school	-0.220***	0.082
Middle school	-0.124	0.078
Primary school	0.027	0.056
Occupation (Reference category: production and transport)		
Agriculture	0.096	0.065
Wage	-0.093	0.095
Other	0.196*	0.104
House type (Reference category: semi-*pucca*)		
Pucca house	-0.010	0.093
Kachcha house	-0.031	0.053
Separate kitchen	0.065	0.049
Electricity	-0.178**	0.070
Sanitation	-0.051	0.107
Fuel type (Reference category: dung)		
Miscellaneous	-0.013	0.168
Wood	-0.186***	0.061
Drinking water (Reference category: protected public)		
Open private	0.011	0.114
Protected private	-0.010	0.058
Open public	0.208*	0.114
Owns a water pump	0.201***	0.076
Irrigated area (Reference category: no irrigated area)		
Less than 1 to 2 acres	-0.051	0.066
Above 2 acres	-0.153*	0.085
Non-irrigated area (Reference category: no non-irrigated area)		
Less than 1 to less than 2 acres	0.157	0.101
2 acres and above	-0.119	0.130
Irrigated and non-irrigated area interactions (Reference category: none of either)		
Irrigated till 2, non-irrigated less than 2	-0.033	0.119
Irrigated till 2, non-irrigated 2 and above	-0.206	0.203
Irrigated above 2, non-irrigated less than 2	-0.011	0.219
Irrigated above 2, non-irrigated 2 and above	0.213	0.197
Number of observations	7287	
Log likelihood	-1963.1525	
Pseudo R-squared	0.0944	

The above regression includes 48 district effects as discussed in the methodology; Pseudo R-squared reported is McFadden's pseudo R-squared; a. SE: Robust standard errors adjusted using Huber's procedure, reported in separate column for all explanatory variables and in brackets for district dummies; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis of the coefficient being equal to zero is tested using an asymptotic t-test.

our case were equally spaced at different percentiles of the data based on the consumer durable index.

When breaking up the sample by quintiles using splines, we find that the poorest group has a positive coefficient, indicating that at this level of income an increase in wealth fails to reduce malaria incidence. The 60–80 percent group also shows a positive and significant coefficient. It is only for the top-most wealth group that the effect is negative and significant implying that in the richest 20 percent category, a rise in wealth would result in reducing malaria incidence (see Table 5.2). The results here support the view that wealth has a negative impact on malaria incidence, confirming evidence from the wider literature on the subject of income and health outcomes as discussed in Chapter 1. In particular, the finding here points to the fact that the health and wealth relationship is non-monotonic, as noted by Deaton (2002), and the negative impact of wealth on malaria incidence takes effect only in the top most wealth category in rural Uttar Pradesh.

5.1.2 Impact of socio-economic status variables on malaria incidence

As noted above, all the variables included reflect socio-economic status. We discuss these in turn below beginning with the characteristics of caste, gender, and age. Then we discuss the effects of occupation and education followed by living conditions and, finally, the ownership of land and agricultural assets with reference to Table 5.1. It may be noted, though, that the signs on the coefficients on the different variables are unchanged and the same variables have well determined effects in Table 5.2 with splines except the variable middle-school educated, which yields a significant negative effect in this specification being not significant in the specification reported in Table 5.1.

It may be noted that the estimated effects for the house-type categories are not significant in the current specification. These results are in tandem with Banguero's (1984) study for Colombia as well as Bhati et al.'s (1996) study in Kheda district in Gujarat. Subramanian et al. (1991), on the other hand, in their study of a rural community in Koraput district in Orissa found that malaria incidence was highest in thatched houses without a temporary ceiling, relatively lower in thatched houses with tin or asbestos roofs and the lowest in tiled houses with proper ceilings since the latter do not have eaves for entry. Thus, varied results have been obtained on this variable and these depend on the precise nature of the dwellings.[1]

The livestock variable was not used in the specification although it is viewed as a potentially important factor in influencing malaria incidence as discussed in the literature in Chapter 2. It was dropped because the coefficient on this variable was not well determined in the preliminary regressions. A plausible reason for this could be that it was not

[1]The variable, whether animals are kept inside the house at night was also included in preliminary regressions but was not found to have a well determined effect and is not reported here.

Table 5.2: Probit of malaria incidence with quintiles instead of an aggregated asset index: 1992-93, rural

Variable	*ME/IE*[a]	*SE*[b]
Asset index quintiles		
Bottom 20%	0.140***	0.038
20–40%	-0.012	0.024
40–60%	-0.007	0.016
60–80%	0.034**	0.014
Top 20%	-0.006*	0.003
Social characteristics		
Caste	-0.014**	0.007
Gender	0.023*	0.011
Age dummies (Reference category: 18-27 years)		
28 to 37 years	0.000	0.011
38 to 47 years	0.013	0.012
48 to 57 years	0.007	0.012
58 to 67 years	0.005	0.013
Above 67 years	-0.002	0.014
Education (Reference category: Illiterate)		
High school	-0.026***	0.008
Middle school	-0.016*	0.009
Primary school	0.003	0.008
Occupation (Reference category: production and transport)		
Agriculture	0.013	0.008
Wage	-0.011	0.011
Other	0.030**	0.017
House type (Reference category: semi-*pucca*)		
Pucca house	0.000	0.012
Kachcha house	-0.004	0.007
Separate kitchen	0.008	0.006
Electricity	-0.021**	0.008
Sanitation	-0.005	0.013
Fuel type (Reference category: dung)		
Miscellaneous	0.004	0.024
Wood	-0.027***	0.009
Drinking water (Reference category: protected public)		
Open private	0.001	0.015
Protected private	-0.002	0.008
Open public	0.031*	0.019
Owns a water pump	0.028**	0.013
Irrigated area (Reference category: no irrigated area)		
Less than 1 to 2 acres	-0.006	0.009
Above 2 acres	-0.018*	0.010
Non-irrigated area (Reference category: no non-irrigated area)		
Less than 1 to less than 2 acres 0.023	0.015	
2 acres and above	-0.014	0.015
Irrigated and non-irrigated area interactions (Reference category: none of either)		
Irrigated till 2, non-irrigated less than 2	-0.005	0.015
Irrigated till 2, non-irrigated 2 and above	-0.023	0.019
Irrigated above 2, non-irrigated less than 2	-0.003	0.027
Irrigated above 2, non-irrigated 2 and above	0.033	0.034
No. households	7287	
Log likelihood	-1959.501	
Pseudo R-squared	0.0961	

The above regression includes 48 district effects as discussed in the methodology; Pseudo R-squared reported is the McFadden's pseudo R-squared; a. ME/IR indicates marginal/impact effects; b. SE: robust standard errors adjusted using Huber's procedure, reported in separate column for all explanatory variables; * significant at 10%; ** significant at 5%; *** significant at 1%.

included as a continuous measure.[2] As we can see from Table 3.4 (p.49), 83 percent of the households own livestock. A continuous variable would have more variation. Moreover, through a continuous measure it would also be possible to include the cattle-human ratio in the model, a more precise indicator of the effect of the ownership of livestock on human malaria incidence in view of Macdonald's equation (see Chapter 2, p.29).

Among the social characteristics, gender and caste (whether a household head belongs to a scheduled caste or not) have well determined effects. The coefficient on caste has a negative sign. This indicates that if a head belongs to the lower socio-economic categories of scheduled caste (SC) or scheduled tribe (ST), they are less likely to suffer from malaria as compared to other castes. Our result is in conflict with Sharma's (1998) observation that more than a third of the total malaria cases in India are from the tribal areas. This section of the population also contributes 60 percent of all *falciparum* cases in the country. The conflicting results can be explained by the fact that Sharma's results are primarily based on studies in Orissa and the north-eastern states where a large proportion of scheduled tribes are found and are susceptible to the forest species of *Anopheles*, *An. dirus*, which is not common in Uttar Pradesh (Srivastava et al. 2001). Futher, even though we are controlling for land ownership and specifically the ownership of different categories of irrigated and non-irrigated areas, it is likely that the lower likelihood of malaria in the SC population is driven by the fact that a small proportion of the SCs own irrigated fields, which are one of the main resting and breeding grounds for mosquitoes—the Pearson product moment correlation coefficient indicates a negative relation between these variables and the value of the Phi correlation coefficient[3] between caste (0: does not belong to SC/ST; 1: belongs to SC/ST) and ownership of irrigated land is (0: does not own irrigated land; 1: owns irrigated land) 0.15. As Table 5.3 shows, 45 percent of the SC/ST households do not own any irrigated land whereas the corresponding figure for other castes is 28 percent. Similarly, a relatively higher percentage of other castes (those that are not SC/ST) have larger irrigated tracts as compared to the SC/STs. Even among those SC/STs who own land, as Parker and Kozel (2004) argue, the land may lie fallow due to a lack of access to irrigation. Thus, the lower incidence of malaria among the SC/ST population is not an indication that they are better off but that they are less likely to contract malaria because of the smaller proportion working in agriculture or owning agricultural fields, relative to other castes.

Another reason, as discussed in Chapter 2, for the relatively lower incidence of malaria in the SCs and STs could be genetic. A study in Kheda district of Gujarat found that 5.9

[2]The NFHS (IIPS, 1995) reports whether a household owns livestock or not rather than the total number owned.

[3]See Appendix E for an explanation of the Phi correlation coefficient.

Table 5.3: Ownership of different irrigated area categories by caste

Irrigated area in acres	*Head belongs to SC/ST*		*Total*
	0: No	1: Yes	
None	1,580 (27.91)	733 (45.11)	2,313 (31.74)
Less than 1	1,310 (23.14)	431 (26.52)	1,741 (23.89)
1 and 2	1,373 (24.25)	318 (19.57)	1,691 (23.21)
3 to 5	750 (13.25)	87 (5.35)	837 (11.49)
5 to 10	412 (7.28)	41 (2.52)	453 (6.22)
Above 10	237 (4.19)	15 (0.92)	252 (3.46)
Total	5,662 (100)	1,625 (100)	7,287 (100)

Percentages reported in parentheses.

percent of the SCs among 429 male blood samples showed G6PD deficiency associated with lower malaria incidence. In the upper castes for the same area they found the percentage to be lower at 3.8 percent. From 769 blood samples covering all castes and both gender groups a larger proportion of STs (0.15) had the sickle cell trait, known to be negatively associated with malaria incidence, as compared to upper castes. In fact, among the upper castes three percent of the Rajputs had the sickle cell trait whereas none of the other castes showed this trait. Only one malaria case (4.3 percent of the sample) occurred in those blood samples that were G6PD deficient and in others the percentage of malaria cases was nearly 13 (Pant et al. 1992, Pant et al. 1993). The results on the caste variable reveal the anomaly that the occurrence of malaria is lower in SC/STs because of such problems as a sickle-cell trait likely to be associated with those of a lower socio-economic status. Thus, the sign on the caste coefficient does not support the view that better socio-economic status is associated with better health outcomes. However, when exploring the possible reasons for the unexpected sign we find that it is precisely because of their lower socio-economic status that the SC/STs are less likely to contract malaria, highlighting the complex nature of malaria incidence.

The coefficient on gender is positive and significant, though not as well determined as the caste effect, at the 10 percent level of significance. The positive sign indicates that male-headed households (relative to female headed ones) are more likely to suffer from malaria. This result may be in conflict with one's expectations since one would expect male-headed households to enjoy better socio-economic status and, hence, better health outcomes. However, women in north India have been observed to spend the evening hours cooking while men usually sit outdoors, thus reducing the exposure to bites among women (Rueben 1992 in Vlassoff and Bonilla 1994, Rueben and Panicker 1979 in Vlassoff and Bonilla 1994, Jotkar et al. 1997). Bhati et al. (1996) also support this finding from their study on Kheda district of Gujarat giving the extensive clothing of women and limited outdoor activities as a reason for lower exposure to bites and lower incidence of malaria. Given that the behaviour of *An. culicifacies* has shown exophilic and exophagic traits, as noted earlier, this provides further reason for possible exposure to bites for men who spend their time outdoors in the evenings at the peak biting time of this species (Saxena et al. 1992).

Attention now turns to the effects of occupation and education. The results for both these variables support the negative socio-economic status and health relationship. The reference for the education variables is the illiterate category. The coefficient on having attained high school or higher education level is well determined with a negative coefficient, relative to the base—a household head educated at least up to the high school level, on average and *ceteris paribus*, is 2.5 percentage points less likely to contract malaria. In Table 5.2 the coefficient on middle-school educated is also negative and significant at the 10 percent level. It is likely that a person who is relatively more educated also has a higher level of awareness on malaria incidence and takes better and more preventive measures. Macintyre et al. (2002) in a recent study conducted in Kenya find the level of education to be an important predictor of whether a household uses multiple methods to protect against mosquito bites.

Among the occupation categories, the coefficients on the category, household head in agriculture, as well as the 'other' category (conflated from the occupation categories unemployed, disabled, students, household duties, retired, or other) are positive and well determined, relative to the higher socio-economic status category of whether a household head is in 'production and transport' or not. [4] Bhati et al. (1996) also found a higher incidence of malaria among dependents in their study on households in Kheda district in Gujarat.[5] The results on this variable, however, need to be treated with caution because

[4]The well determined coefficient on the 'other' category seems to be driven by two groups – household heads who are students and those who are disabled. When a probit regression was run using the disaggregated categories the coefficients on disabled and students were significant with a positive sign. The other coefficients were not well determined. Both students and the disabled heads of households might have lower immunity – the disabled because of a medical condition and students because they belong to a younger age group and the maturity of the immune system depends on the age of an individual (Tongren et al. 2006, Sharma et al. 2004).

[5]Bhati et al. (1996) do not specify which categories they include in dependents although this term

of the heterogeneous nature of the category.

Among living conditions, the variables with well determined effects are type of fuel used, type of drinking water source, and whether the household has access to electricity. The results for all these variables support the negative socio-economic status and health relationship.

The negative and significant sign on the electricity coefficient indicates that a household head with an electricity connection in the house is less likely to have suffered from malaria in the last three months as compared to a head whose house does not have electricity, on average and *ceteris paribus*. The impact effect (see Appendix D) suggests that a household head with an electricity connection would have a lower probability of contracting malaria (relative to the base category of not having electricity) by two percentage points, *ceteris paribus*. Having electricity supply allows households to use mosquito repellents that run on electricity and to run fans as a way of avoiding mosquito bites—the Phi correlation coefficient between fan and electricity is + 0.54. It is also common to use non-electric mosquito coils and/or create smoke by burning weeds to avoid nuisance biting by mosquitoes, as observed in villages in Mexico (Rodriguez et al. 2003) and Kenya (Macintyre et al. 2002). Having electricity (thus enabling the use of fans and electric coils) adds to the set of possible preventive measures that can be used by the household and is likely to bring down the number of infective bites, and hence the incidence of malaria, on average and *ceteris paribus*.

Among the fuel categories, relative to the base category of dung, the negative and significant coefficient on firewood indicates that a household that uses wood as fuel is less likely to have a malaria episode, on average and *ceteris paribus*. It could be conjectured that this is because of the higher socio-economic status associated with households that use firewood compared to those that use dung given that such measures as occupation and education (which are controlled for here) may not completely capture the socio-economic status of the households. Another explanation could lie in the fact that firewood may be used to prevent mosquito bites among those households that can afford to use it. Dung is not a preferred energy source and is used once the household begins to face scarcity of firewood.[6] In that case a household that uses dung is already faced with the problem of

is normally used for those who do not have their own income and are thus dependent on others. Their classification of respondents by profession is into: farmer, labour services, business, and dependents.

[6]According to Leach and Gowen (1987) once the supply of firewood starts declining households adapt the following sequence of response mechanisms: economizing on fuel, using crop residues and dung, lowering

economising the use of fuel and is not likely to use it for purposes other than cooking.[7] Thus, the negative association of malaria with firewood is indicative possibly of both better socio-economic status as well as the use of this fuel for preventing mosquito bites, relative to the base category, dung (Leach and Gowen 1987, Lefevre et al. 1997).

The reference category for drinking water source is a protected public water source. The coefficient on open public source of drinking water is significant and positive relative to the base. *Ceteris paribus* and on average, a head of household who uses an open public water source for drinking needs is three percentage points more likely to suffer from malaria relative to one who uses a protected source (see Appendix D). This result is as expected since such open sources as rivers and streams are known breeding grounds for *An. culicifacies*, the primary vector for transmitting malaria in north India, making those who collect water from these sources more susceptible to bites (Konradsen et al. 2003, Piyaratne et al. 2005, Ansari and Razdan 2004).

Ownership of agricultural land is an important socio-economic status variable in rural areas. The base category for irrigated/non-irrigated land is no irrigated/non-irrigated area. On average and *ceteris paribus*, having more than two acres of irrigated land[8] as compared to not having any irrigated area reduces malaria incidence by two percentage points. In case of non-irrigated area, none of the estimated effects are well determined. Several interaction terms were also employed in the regression model to control for ownership of various combinations of irrigated and non-irrigated land (with household owning neither irrigated nor non-irrigated land as the base). The coefficients for these categories were not well determined either.

Thus, we find that heads who own larger areas of irrigated tracts are less conducive to contracting malaria, indicating that better socio-economic status does have a negative impact on malaria incidence. The reason for the negative effect here could lie in the fact that farmers with larger tacts of irrigated land can afford to run off water not allowing it to stagnate in the fields thus preventing the breeding of mosquitoes. Landowners who do not have any irrigation facilities, on the other hand, do not possess the luxury to drain and re-water their fields. Watson (1937) alludes to this way of 'irrigating' fields in those places

living standards and applying restrictions on their diet when faced with acute scarcity, and finally resorting to substituting firewood that is collected free of cost from forest and common areas to purchasing biomass fuels or other fuel substitutes.

[7] This is possible despite the fact that dung can be used for repelling mosquitoes since poor households are likely to prioritise the use of fuel for cooking purposes rather than to prevent mosquito bites.

[8] This category is also referred to later as 'three or more acres'.

of Africa where rainfall is uneven and suggests that, in such settings, draining fields would be wasteful and unwise. Sogoba and Doumbiaa (2007), in a recent study in Niger, also find evidence to support the view that the type of water management and drainage system can affect malaria transmission in areas that cultivate rice. They compared rice plots with controlled and uncontrolled water depth for the presence of *Anopheles* mosquitoes and found that larval density was generally higher in the uncontrolled water depth areas as compared to the controlled plots.

A possible explanation for the, *ceteris paribus*, positive significant impact of owning a water pump relative to not owning one on malaria incidence could be the presence of stagnant water in areas around water pumps. Such fresh water sites are ideal breeding areas for *An. culicifacies*, which has been observed to scatter eggs over transiently available water pools to attain success in breeding (Piyaratne et al. 2005). This result is interesting in that although relatively wealthier houses are more likely to own a water pump, this symbol of wealth in itself does not imply a lower likelihood of malaria incidence. On the contrary, such sites may create pools of water that breed mosquitoes. According to the impact effects calculated (see Appendix D), owning a water pump makes a household head three percentage points more likely to contract malaria, relative to a head who does not own a water pump, *ceteris paribus* and on average.

Thus, we may conclude that better socio-economic status does not always guarantee lower malaria incidence. The living conditions variables that do support the negative relationship are having an electricity connection, using firewood rather than dung as fuel, and using a protected public water source instead of open water sources for drinking needs. Being high-school educated rather than illiterate and having more than two acres of irrigated land rather than none also offer better protection against malaria. However, the effects for caste, gender, and ownership of water pump do not support the negative socio-economic status and health relationship. The next section deals with the district level analysis.

5.2 District level analysis: 1992-93

The coefficients on all the district dummies in the probit regression reported above are statistically significant indicating that the district level dummy effects capture a number of unobservable factors that affect malaria incidence (see Appendix D). Table 5.4 displays the raw data for the incidence of malaria across the different districts in the Uttar Pradesh malaria model as well as the percentage of malaria cases after controlling for a number of other factors using the probit model. The district, Etah, has the highest incidence of malaria at 24 percent when we look at the raw data and also ranks at the top even after controlling for other factors. For other districts the ranking is not always the same. For example, Aligarh has higher incidence than Rae Bareilly when looking at the raw data, but when controlling for other factors the percentage of cases is nearly the same.

Table 5.4: Percentage of malaria cases (controlling for other factors) across districts ranked by descending order of incidence with corresponding incidence using raw data (without controls)

R[a]	District	Percentage malaria cases[b]		R	District	Percentage malaria cases		R	District	Percentage malaria cases	
		With controls	Raw data			With controls	Raw data			With controls	Raw data
1	Etah	23.45	24.07	17	Rae Bareili	9.33	9.49	33	Dehradun	3.48	3.13
2	Shahjahanpur	19.6	20.78	18	Banda	9.08	10.34	34	Pratapgarh	3.23	4.17
3	Mainpuri	18.92	21.33	19	Bijnor	8.18	8.59	35	Mau	2.98	2.78
4	Budaun	18.67	22.22	20	Unnao	7.77	9.43	36	Bahraich	2.64	3.03
5	Kanpur Nagar	18.41	19.49	21	Faizabad	7.59	7.64	37	Basti	2.61	2.27
6	Bareilly	17.41	20.00	22	Agra	7.55	8.22	38	Saharanpur	2.41	2.68
7	Deoria	15.22	14.49	23	Hardoi	7.31	7.09	39	Siddhartha Nagar	2.40	2.45
8	Firozabad	15.17	17.02	24	Meerut	7.06	8.47	40	Almora	2.37	4.12
9	Muzaffarnagar	13.94	15.47	25	Pilibhit	7.02	6.37	41	Fatehpur	2.21	2.58
10	Jalaun	13.09	13.46	26	Etawah	6.73	7.20	42	Hamirpur	1.88	1.89
11	Tehri Garhwal	11.01	14.29	27	Kanpur Dehat	5.64	6.32	43	Ghaziabad	1.82	1.52
12	Jhansi	10.88	13.89	28	Nainital	4.61	4.36	44	Garhwal	1.67	2.17
13	Barabanki	10.28	10.46	29	Maharajganj	4.26	4.73	45	Mathura	1.43	1.47
14	Bulandshahar	9.68	12.77	30	Moradabad	4.17	4.65	46	Gorakhpur	1.36	1.35
15	Aligarh	9.59	13.04	31	Allahabad	3.69	3.77	47	Gonda	1.04	1.01
16	Sitapur	9.58	9.30	32	Sultanpur	3.62	3.70	48	Rampur	0.63	0.61

a. R: Ranking of districts in descending order of malaria incidence; b. Percentage malaria without controls is from the raw data and obtained as number of malaria cases/number of households times 100; percentage malaria cases with controls is obtained as the standard normal cumulative density function of the district coefficient times 100 where the coefficients on the district dummies are from the probit regression of malaria incidence in household heads on socio-economic and other factors and district dummy effects.

An additional exploratory exercise was undertaken where pseudo R-squareds were calculated for the probit with district dummies and without district effects to observe the degradation in the pseudo R-squared that occurs when we do not include district effects. For the former, the value for the McFadden's pseudo R-squared is 0.0944 and for the latter case (without dummies or district level variables) the value is 0.0351, a degradation of 65.6 per cent (see Table 5.5). This indicates the importance of including district effects which incorporate the district level development indicators and other omitted factors. These results indicate a clear role for district level factors captured using the district dummies and provide another reason for examining the influence on malaria incidence of the variables embodied in the district dummies in greater detail. As pointed out in Chapter 4, we could examine the role of district variables either by running a second-stage regression of the coefficients on the district effects from the (first-stage) probit regression or by including district level data in the first-stage equation itself using this single equation model without district dummies. The pseudo R-squared is slightly lower for this latter specification at 0.0717 as compared to that which uses district dummies at 0.0944. As we can see from Table 5.5, the district effects model yields, not surprisingly, the highest pseudo R-squared and correlation coefficient (between predicted and observed malaria incidence). This is likely because the district dummies add more information than using just the available district level variables from the census. Both models, of course, are superior to one that does not use either the district level variables or the dummy effects.

Table 5.5: Goodness of fit measures

Model	*McFadden's pseudo R-squared*	*Squared correlation coefficient (between predicted and observed malaria)*
No district-level effects or district-level variables	0.0351	0.021
District-level effects	0.0944	0.047
Single equation (district-level variables)	0.0717	0.041

In addition to the goodness of fit measures, the Vuong test explained in Chapter 4 (p.94) may be used to assess which model is preferable. The Vuong test statistic has a value of -0.6279, which is larger than -1.96 and indicates that we cannot discriminate between the district dummy effects model and the model with district level data. In other words, the choice is left to researcher preference. Table 5.6 reports the estimates for the explanatory variables based on a single equation that includes the district level variables in lieu of district effects and the corresponding estimates for these variables based on the use of the two-step regression approach.[9] As we can see, the results are similar for the two models leaving us to choose the model of our choice. The two-stage regression is

[9]Note that the results in Column 2 of Table 5.6 for the two-stage model differ from those presented in Table 5.8 as the former does not make any adjustment for standard errors such as by using the Zanchi transformation. It uses the coefficients from the district effects equation directly without such a transformation in order to compare similar models to conduct the Vuong test.

the model chosen for the current study given the higher R-squareds as discussed in the previous paragraph, though the single-equation results are presented in Appendix F.

Table 5.6: Coefficients from single equation model and the two-stage district regression: 1992-93, rural

Variable	*Coefficients from single equation model*	*District effects model*[a]
PV with a high-school	-0.095***	-0.091***
	[0.022]	[0.035]
PV with an adult-literacy centre	-0.014***	-0.016***
	[0.003]	[0.005]
Average distance to drinking water	-0.173***	-0.153***
	[0.043]	[0.055]
PV with a commuting facility	0.008**	0.007
	[0.003]	[0.004]
PV with domestic power	-0.005**	-0.006**
	[0.002]	[0.003]
PV with agricultural power	0.006***	0.006**
	[0.002]	[0.003]
PD irrigated by a government canal	0.015***	0.016*
	[0.004]	[0.007]
PD irrigated by a private canal	0.677***	0.610**
	[0.137]	[0.270]
PD irrigated by a well	0.145***	0.140***
	[0.034]	[0.049]
PD irrigated by a tube well	0.008***	0.007**
	[0.001]	[0.002]
PD irrigated by an electric tube well	0.008**	0.007
	[0.003]	[0.005]
PD under forest	-0.007**	-0.005
	[0.003]	[0.007]
District population density	-0.100***	-0.114**
	[0.027]	[0.036]
Log rain	-0.404***	-0.418**
	[0.135]	[0.195]
Constant	1.495	1.628
	[0.929]	[1.326]
Number of observations	7287	48
Log likelihood	-2012.53	
R-squared	0.7595	

PV stands for percentage of villages (in the district) and PD for percentage of district area; a. This model uses the coefficients from the second stage district dummy effects equation without any adjustment for the standard errors (unlike Table 5.8); standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis of the coefficient being equal to zero is tested using an asymptotic t-test.

5.2.1 Two-stage model

In order to allow for ease in interpretation, the coefficients on district dummies were transformed into deviations from the mean following a method introduced by Krueger and Summers (1988), discussed in the methodology chapter. The deviations are presented in Table 5.7 along with the transformed standard errors as calculated using the procedure proposed by Zanchi (1998). Using these standard errors a t-test was conducted to check whether the malaria incidence in a particular district is statistically different from the average incidence across all the districts in the state (described in the methodology section). For twenty-three districts the deviations are significantly different from zero. Nine of these fall below and the other fourteen are above the Uttar Pradesh average malaria incidence.

Table 5.7: Deviations of district malaria effects from mean weighted effect, 1992-93 rural

District	*Deviations from mean*[a]	*District*	*Deviations from mean*	*District*	*Deviations from mean*
Tehri-Garhwal	0.26***	Firozabad	0.485***	Hamirpur	-0.566
	[-0.1000]		[-0.130]		[-0.422]
Dehradun	-0.286	Etah	0.808***	Banda	0.183
	[-0.437]		[-0.138]		[-0.319]
Garhwal	-0.625	Mainpuri	0.646***	Fatehpur	-0.499**
	[-0.448]		[-0.169]		[-0.209]
Almora	-0.475*	Budaon	0.651***	Pratapgarh	-0.321
	[-0.253]		[0.139]		[-0.264]
Nainital	-0.181	Bareilly	0.578***	Allahabad	-0.283
	[-0.121]		[-0.134]		[-0.227]
Bijnor	0.115	Pilibhit	0.034	Bahraich	-0.436
	[-0.167]		[-0.158]		[-0.431]
Moradabad	-0.209	Shahjahanpur	0.664***	Gonda	-0.799***
	[-0.332]		[-0.118]		[-0.261]
Rampur	-0.973*	Sitapur	0.185	Barabanki	0.249*
	[-0.344]		[-0.192]		[-0.137]
Saharanpur	-0.465	Hardoi	0.041	Faizabad	0.077
	[-0.219]		[-0.165]		[-0.093]
Muzaffarnagar	0.428***	Unnao	0.084	Sultanpur	-0.280**
	[-0.119]		[-0.244]		[-0.145]
Meerut	0.053	Rae Bareili	0.207*	Siddhartha Nagar	-0.47***
	[-0.137]		[-0.117]		[-0.187]
Ghaziabad	-0.578	Etawah	0.018	Maharajganj	-0.204
	[-0.415]		[-0.076]		[-0.182]
Bullandshahar	0.231	Kanpur Dehat	-0.059	Basti	-0.415
	[-0.139]		[-0.142]		[-0.419]
Aligarh	0.206***	Kanpur Nagar	0.62***	Gorakhpur	-0.694**
	[-0.044]		[-0.092]		[-0.269]
Mathura	-0.688*	Jalaun	0.39*	Deoria	0.479***
	[-0.393]		[-0.219]		[-0.135]
Agra	0.074	Jhansi	0.299	Mau	-0.374
	[-0.153]		[-0.199]		[-0.246]

a. Deviations from the mean = Incidence in district less weighted average incidence over all districts, weighted by proportion of households sampled in each district = -1.5127 (see methodology chapter for details); Coefficient on district dummy is from the probit of malaria incidence on socio-economic, other factors, and district dummies. *** Deviation significantly different from mean at 1% level of significance; ** Deviation significantly different from mean at 5% level of significance; * Deviation significantly different from mean at 10% level of significance in the t-test for the null hypothesis that there is no difference between the malaria incidence of a particular district and the average malaria incidence; standard errors (in brackets) are the transformed standard errors on district dummies (as per Zanchi (1998)).

In order to assess which district level factors influence the estimated district effects, a weighted least squares regression of the deviations was run on secondary-level district census data, as discussed earlier. It may be recalled that these deviations are in terms of the standardised probit index. The weights used were the inverse of the (transformed) standard errors on the district coefficients in order to account for differences in the precision of these estimates from the original probit model. Thus, by using the inverse of the standard error as a weight, the highest weight was given to the district coefficients that are the most precisely estimated. This may also account, indirectly, for the number of households sampled in the different districts given the inverse relation between standard errors and sample size. Table 5.8 presents the results of the final specification for this second-stage regression after experimenting with a number of different specifications. The estimated coefficients on primary school, middle school, primary health centre/subcentre, and electric well were not well determined and these were not included in the final specification presented in Table 5.8. The estimated effects that were found to be robust to different specifications are discussed below. The dependent variable reflects the underlying propensity to contract malaria at the district level measured in terms of standardised probit index units.

This part of the analysis focuses on the effects of a number of infrastructure variables. The percentage of high schools and adult literacy centres yield significant negative effects. The average distance travelled by households to access drinking water also has a negative and significant coefficient. Whereas the effect of the percentage of villages with domestic power supply is negative and well determined, that of agricultural power supply is positive and significant. The effects of the extent of district area irrigated by different irrigation systems are positive and well determined. Controls for forest area, rainfall, and population density were also included and these, too, have well determined effects.

The negative signs on the well determined coefficients on high school and adult literacy centres indicate that as the presence of these educational facilities rises, the incidence of malaria at the district level, on average and *ceteris paribus*, declines. The presence of adult literacy centres and high schools could reduce the propensity to contract malaria as these institutions are attended by the relatively older age groups who are likely to have a better grasp of the nature of the disease and using preventive measures. In a study carried out at the higher secondary school, secondary school, and primary school level in Nadiad *taluka* in Kheda district of Gujarat with 79 people falling in a control group and 489 included in a target group, subject to a malaria education programme, showed that students at the higher secondary level showed better knowledge and understanding perhaps because of their maturity as compared to the other groups (Bhati et al. 1995).

Table 5.8: Results of the two-stage regression for 1992-93 using Zanchi's transformation[a]

Variable	*Coefficient*[a]
PV with a high-school	-0.102***
	[0.024]
PV with an adult literacy centre	-0.014***
	[0.004]
Average distance to drinking water	-0.147***
	[0.043]
PV with a commuting facility	0.008***
	[0.003]
PV with domestic power	-0.005***
	[0.002]
PV with agricultural power	0.006***
	[0.002]
PD irrigated by a government canal	0.015***
	[0.005]
PD irrigated by private canal	0.827***
	[0.221]
PD irrigated by a well	0.133***
	[0.036]
PD irrigated by a tube well	0.008***
	[0.001]
PD irrigated by an electric tube well	0.007*
	[0.004]
PD under forest	-0.010**
	[0.005]
District population density	-0.106***
	[0.026]
Log rain	-0.369***
	[0.133]
Constant	2.727***
	[0.898]
R-squared	0.6678
Observations	48

PV stands for percentage of villages (in the district) and PD for percentage of district area; a. Coefficients are from variance-weighted least squares ('vwls' command in Stata) regression of coefficients on district dummies (from the probit regression of malaria incidence on socio-economic and other factors and district fixed effects) on district-level variables (column 1 above) and variances used as weights are from the squares of the transformed standard errors on the coefficients for the district dummies as per Zanchi (1998); the correlation coefficient for this model is 0.6776; standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis of the coefficient being equal to zero is tested using an aymptotic t-test.

Adult education classes through the use of mass media, mainly radio and TV to promote awareness on the national polio campaign and such aspects as controlling and preventing diarrhoea in children, often presented in the form of short films or advertisements on TV, may provide a positive externality for malaria prevention by making people aware of the different measures that may be taken to prevent mosquito bites and to prevent the breeding of mosquitoes (Parhar 1998). Short films for the purpose of creating malaria awareness have also been used by Ahmedabad TV in Gujarat. It telecast ten short films at various times for this purpose as a part of a malaria eradication campaign undertaken in the late 1980s. These films are reported to have produced positive responses from the communities. The campaign also included organising health camps in the target area, Nadiad *taluka*, along with arranging participatory exercises and interpersonal discussions to generate awareness. As a result, several villagers participated in vector control work such as filling out pits and puddles (Sharma and Sharma 1989). The presence of such institutions is also likely to create positive externalities in the form of an increase in the knowledge base even for those who are not participating in these programmes. Bhati et al. (1996) also support imparting health education to create awareness to encourage prophylaxis and in Amazonia, given the exophilic nature of the *Anopheles* species found there, Tadei et al. (1998) emphasize the need to educate people on the habitat of the species so that they can avoid these places thereby reducing the transmission rate of malaria.

The results from the current study support the above arguments. The coefficient on 'PV adult literacy' (Table 5.8) indicates that a one percentage point increase in villages with an adult literacy centre in a particular district would result in reducing the propensity to contract malaria by 0.014 of a standard deviation below the overall district level average, *ceteris paribus.* The effect of a one-percentage point increase in villages with a high school in a district is stronger and leads to a decline in the propensity to contract malaria by 0.102 of a standard deviation below the district level average, *ceteris paribus.*[10] Thus, education is a significant influence in reducing malaria incidence at the household level, as discussed earlier, as well as at the district level, through the presence of educational institutions.

Households that do not have piped water supply in their houses use such drinking water facilities as rivers, streams or public tube wells and taps. If the mean distance to a

[10]On average for each district, the percentage of villages that have an adult literacy centre is 6.5 and that for high schools is lower at three (Table 3.14, p.75).

drinking water facility for a district increases by a kilometre, the propensity to contract malaria at the district level is likely to decline by 0.147 of a standard deviation below the average. This negative relationship reflects the possibility that locations where drinking water is collected are a breeding ground for mosquitoes, consistent with the literature on the subject since mosquitoes utilise both open sources like rivers and streams to lay eggs as well as fresh water 'puddles' that can be formed from residual water collection at public sites. Surface water areas or areas where surplus water may collect, including temporarily available water pools, are commonly exploited as breeding areas by many *Anopheles* species including *An. culicifacies*, the main vector found in Uttar Pradesh (Tyagi 2004, Piyaratne et al. 2005). In fact, a study conducted in some villages in Sri Lanka found that houses that were less than 750 metres from a stream had more *An. culicifacies* as compared to those that were farther away (Konradsen et al. 2003). The result on this variable clearly suggests a role for improving the infrastructure around public water collection sites so as to prevent mosquito breeding.

The positive sign on the well determined coefficient on the percentage of villages with a commuting facility indicates that the greater the number of villages in a particular district with a commuting facility, the more likely it is to have malaria incidence, on average and *ceteris paribus*. While any district would welcome an increase in commuting facilities, this seems to be accompanied with increased incidence of malaria. This positive relationship between a rise in commuting facilities and malaria incidence seems to point to the possible transmission of malaria through transport networks. It is likely that mosquitoes use dark corners of buses or other means of transport as 'hiding' places, consistent with the literature that highlights the fact that mosquitoes generally rest in dark corners inside houses, cattle sheds and other such places and the possibility of mosquitoes travelling in cargo holds in planes. The presence of commuting facilities may also enable people to travel regularly for work to other areas that may have an efficient malaria vector, thus exposing them to the disease, contracting malaria, and acting as active transmitters (Robert et al. 2003). There are not many studies looking at the possible transmission of malaria through ground transportation. One such entomological and parasitological study was conducted on the Djibouti-Ethiopian railway as a possible mediator for transmitting malaria. While the study did not find any infected mosquitoes on the passenger or the goods train, some passengers tested positive for *P. falciparum* indicating that infected passengers could act as passive importers of malaria triggering the spread of the disease in the new area (Fox et al. 1991). While it may not be possible to control transmission by restricting passengers who carry the parasite, insecticide spraying in commuting facilities, as a part of the government's malaria eradication programme, is feasible and could help in controlling malaria transmission.

The negative and significant coefficient on domestic power supply,[11] *ceteris paribus*, reinforces the result obtained in the probit regressions at the household level that an electricity connection in the house increases possibilities of preventing mosquito bites through use of fans or electric coils. While the coefficient on domestic power supply is negative, that on agricultural power supply is positive.[12] Usually half the geographic area within the jurisdiction of a State Electricity Board gets power supply during the night while the other half receives it during the day. Further, the rates charged at night are lower providing an incentive to use irrigation systems at night, thus exposing farmers to mosquito bites (Kalra et al. 2004).

The coefficients on percentage of district area irrigated by government canal, private canal, wells, non-electric wells, and tube wells are significant and positive. The irrigated area variables have been introduced at the district level (even though irrigated area was used in the household level analysis) in order to account for the influence of the different types of irrigation systems. The high likelihood of irrigation systems contributing to the spread of malaria has been confirmed through various studies in different regions such as Herrel et al. (2004) and Mukhtar et al. (2003) for Punjab in Pakistan; Tyagi (2004) for the Thar desert in Rajasthan; Bhatt et al. (1990) for Kheda district in Gujarat; and Premasiri et al. (2005) for southern Sri Lanka. The main argument in all these studies is the ability of the *Anopheles* vector to exploit temporary pools of waste water from irrigation as well as irrigation systems like canal linings for breeding. It is in response to these observed facts that Konradsen et al. (1997) emphasize the importance of efficient water run-off or drainage facilities in irrigated areas to prevent the proliferation of malaria. A study conducted by Shukla et al. (1995) in Nainital district in Uttar Pradesh found *An. fluviatilis* in irrigation drains and streams. Herrel et al. (2004) found that *An. culicifacies* occurred

[11]This variable was introduced at the district level as well as the household level as in the former it proxies for development and in the latter it proxies for the use of such devices as electric fans and electric coils used to prevent mosquito bites.

[12]As explained earlier power supply in a village can be for domestic or agricultural purposes alone often at different tariff rates. In fact, some villages may have one kind of supply but not the other.

at the head of the irrigated canal where the waterlogged and irrigated fields were present in south Punjab. The relation between malaria and irrigation is well known and not new. In Spain's Ebro delta malaria epidemics in 1862, 1912, and 1919-20 were believed to be a result of lack of drainage in irrigation canals. As Russell (1952) in Najera (2001) points out, it is not uncommon to have a large amount of water flowing into an area because of irrigation but no way of taking it away often resulting in the spread of malaria.

Results from the present study show that an increase of one percentage point in the district area irrigated by a government canal increases the propensity to contract malaria by 0.015 of a standard deviation; increasing private irrigated area at the district level by the same amount results in a 0.827 of a standard deviation rise in malaria, a larger effect than that for a government canal possibly because private canals usually fall under the category of small-scale irrigation works and their construction is quite rudimentary. Government canals, on the other hand, cover a much larger area often spreading over a number of districts and involve more complex engineering and drainage facilities. The statistics in Table 3.14 (p.75) show that on average only 0.21 percent of the total district area is irrigated by a private canal as opposed to 13 percent by a government canal, which also points to the smaller scale of the former relative to the latter.

The coefficient on forest area is well determined and negative, which indicates that the greater the forest cover in the district the lower the likelihood of contracting malaria. Forest cover could prevent the possibility of a greater interaction of forest *Anopheles* species with humans. On the other hand, deforestation can create more breeding sites as Tadei et al. (1998) found in Amazonia—breeding of *Anopheles* perpetuated in stagnant water left after deforestation. As Patz et al. (2000) observe, forest floors in primary growth do not support mosquito breeding like cleared areas. In case of the former, a combination of shade with organic matter that absorbs water makes the site too acidic for mosquito breeding. On the other hand, water puddles formed on cleared lands with exposure to sunlight develop a less acidic and more neutral pH that favours the development of certain *Anopheles* larvae.

Another factor that might contribute to the negative relationship of malaria with forest cover in Uttar Pradesh might be that the composition of forests in the state may not be conducive to mosquito breeding. The largest proportion of the forest plant species found in Uttar Pradesh is that of *Eucalyptus* at 7.2 percent (FSI 1999), which has natural mosquito repellent properties (Moore et al. 2002). In marshy areas *Eucalyptus* trees are, in fact, planted to absorb excess water, preventing the formation of marshy areas and mosquito breeding as was done in Nadiad *taluka* in Kheda district (Sharma and Sharma 1989).

The coefficient on population density is negative and well determined, *ceteris paribus*. Thus, after controlling for other factors such as presence of irrigated area, mean distance to drinking water facilities and other such factors, the more densely populated a particular district is, the less likely it is for its inhabitants to contract malaria. This can be supported by the argument that the more densely populated an area is, the lower is the number of infective bites per person per year (Robert et al. 2003) as discussed earlier in Chapter 2. An increase in district population density by one person per sq. km reduces the propensity to contract malaria by 0.106 of a standard deviation below the district average.

The coefficient on the logarithm of rainfall is negative and significant implying that there is a decline in the propensity to contract malaria with increased rainfall. This result is consistent with the premise that more rainfall can flush out mosquito larvae, thus preventing the spread of malaria.

The above discussion points to the possibility of improving infrastructure around water collection points such as public taps and wells and improving canal linings to prevent the breeding of mosquitoes. It also supports provision of domestic electricity supply and increasing the presence of educational institutions as factors that can reduce malaria incidence. The educational institutions, particularly adult literacy classes, could be used to inform people of preventive measures. The possibility of greater exposure to bites at night when using agricultural power to irrigate fields could be highlighted in such programmes so that farmers can take prophylactic measures. While the government's malaria eradication programme does engage in insecticide spraying, there is no mention in the literature that this is also targeted to commuting facilities, which may harbour the malaria vector. This is another area where there may be a role for government intervention.

5.3 Comparing rural malaria incidence in 1992-93 and 1998-99

The research question addressed in this section is to determine if the factors identified as affecting malaria incidence are the same for the two time periods both for the household and the district level. We first investigate the robustness of household level factors affecting malaria incidence over time including the effect of wealth. The next subsection compares the district results.

5.3.1 Household level analysis

The incidence of malaria nearly halved over the six-year period from 1992-93 to 1998-99 (see Table 3.5, p.52) from 9.6 percent to 5.1 percent. It is possible that this decline occurred as a result of an increase in efforts of the malaria eradication programme over this time period, data for which are not available in the datasets we are using. There was particular focus on some malarious areas of Uttar Pradesh including the districts Hardwar, Nainital,[13] Allahabad, and Shahjahanpur and areas surrounding these. A significant fall in malaria was registered after the launch of the programme by the end of the 1990s. For example, in the Shankargarh mining area, located in Allahabad district, the API was found to be very high and an intensive malaria eradication programme (involving environmental management and insecticide spraying as well as health education) was launched in 1987 in six villages and was gradually extended to cover 107 villages. The slide positivity rate recorded at the malaria clinic located in Shankargarh was maximum in 1990 at 57.5 percent and by 1999 it fell to just over 30 percent (NIMR 2007).[14]

The 1992-93 rural household specification was modified to match that of 1998-99. The comparable specification did not include detailed irrigated and non-irrigated categories but a dummy for whether the household owns agricultural land since the former variables are not reported correctly in the second phase of the NFHS (see p.45). Additionally, ownership of a bullock cart and a tractor were included in the comparable specifications since these variables were found to have significant effects in 1998-99. It may be recalled that six districts were not covered in 1998-99 and these had to be dropped from the comparable 1992-93 specification as well, resulting in 6,071 observations rather than 7,287. At the same time, new districts included in 1998-99 had to be dropped to make this dataset comparable with 1992-93 (pp.50). The results for the comparable specifications are discussed now.

[13] Both Hardwar and Nainital are a part of the state of Uttaranchal since 2000.

[14] The report does not provide further details on falls registered in the other districts but indicates that there was a decline in malaria prevalence after the launch of the programme.

Table 5.9: Probit of malaria incidence for 1992-93 and 1998-99: comparable rural specifications

Variable	*1992–93 Coefficient*	*1998–99 Coefficient*	*t–ratio*
Asset index	-0.007	-0.054*	0.86
	[0.046]	[0.029]	
Caste	-0.105*	0.047	-1.54
	[0.059]	[0.079]	
Gender	0.174	-0.504***	3.99
	[0.113]	[0.127]	
Age dummies (Reference category: 18-27 years)			
28 to 37 years	-0.003	0.129	-0.86
	[0.091]	[0.123]	
38 to 47 years	0.084	-0.036	0.77
	[0.089]	[0.127]	
48 to 57 years	0.039	0.178	-0.84
	[0.095]	[0.135]	
58 to 67 years	0.038	0.186	-0.86
	[0.098]	[0.141]	
Above 67 years	-0.056	0.274*	-1.71
	[0.117]	[0.154]	
Education (Reference category: Illiterate)			
High school	-0.224**	0.206**	-3.12
	[0.088]	[0.106]	
Middle school	-0.094	0.112	-1.48
	[0.082]	[0.113]	
Primary school	0.047	0.040	0.07
	[0.061]	[0.086]	
Occupation (Reference category: production and transport)			
Agriculture	0.077	0.065	0.10
	[0.069]	[0.096]	
Wage	-0.123	-0.093	-0.18
	[0.100]	[0.132]	
Other	0.155	-0.217	2.12
	[0.109]	[0.137]	
House type (Reference category: semi-*pucca*)			
Pucca house	-0.027	-0.060	0.19
	[0.096]	[0.146]	
Kachcha house	-0.047	0.090	-1.49
	[0.056]	[0.073]	
Separate kitchen	0.003	-0.087	0.92
	[0.053]	[0.082]	
Electricity	-0.170**	-0.076	-0.73
	[0.076]	[0.104]	
Sanitation	-0.110	0.017	-0.74
	[0.116]	[0.127]	
Fuel type (Reference category: dung)			
Miscellaneous	0.034	-0.269*	1.26
	[0.173]	[0.166]	
Wood	-0.181***	0.169*	-3.14
	[0.066]	[0.090]	
Drinking water (Reference category: protected public)			
Open	0.312**	0.187	0.23
	[0.127]	[0.527]	
Protected private	-0.003	-0.095	
	[0.059]	[0.078]	
Agricultural land	-0.037	-0.122	0.75
	[0.068]	[0.090]	
Owns a water pump	0.138*	0.203*	-0.48
	[0.081]	[0.110]	
Owns a bullock cart	-0.014	0.218**	-1.62
	[0.093]	[0.109]	
Owns a tractor	0.030	0.312*	-1.23
	[0.153]	[0.170]	

Continued on next page...

Table 5.9 continued

Variable	*1992–93 Coefficient*	*1998–99 Coefficient*	*t–ratio*
No. of households	6071	4588	
Log likelihood	-1742.89	-836.24	
Pseudo R-squared	0.0891	0.1009	

District effects were included in both specifications above; robust standard errors reported in brackets; Pseudo R-squareds reported are the McFadden's psuedo R-squareds; * significant at 10%; ** significant at 5%; *** significant at 1%, where the null hypothesis of the coefficient being equal to zero is tested using an asymptotic t–test; t–ratios presented in the last column test for differences in estimates across the two years; t-statistic at the 5% level of significance is ±1.96.

Evidence of robustness at the household level

We find little evidence of robustness over the two time periods. With the exception of the variable, ownership of a water pump, other variables either did not have well determined effects in both the time periods or the signs on the coefficients were different in case these were significant.

The estimated effects for ownership of a water pump are robust—the coefficient is positive and significant for both the time periods. However, the other agricultural assets included in these comparable specifications have significant positive coefficients in 1998-99 but not in 1992-93. The explanation for the positive association between ownership of a water pump and malaria incidence could lie in the collection of stagnant water in areas around water pumps, an ideal source for mosquito breeding, as explained earlier. The reason for the positive association with tractor and bullock cart ownership may be similar and could be associated with there being potential resting or hiding places for mosquitoes in such places as bullock cart wheels or even in hoof marks. For the other variables, discussed now, we do not find any evidence of robustness.

The coefficient on gender is well determined in 1998-99 but not in 1992-93 for the comparable specifications. In fact, the probit results for the 1992-93 rural sample are the same in terms of the signs on the coefficients and the same variables have well determined effects as before (for the specification with a larger sample size) except for gender where the coefficient is no longer significant (see Table 5.9). However, this variable does not seem to be robust to the two different specifications of 1992-93 – in the specification with splines, the coefficient on gender is positive at the ten percent level of significance. In 1998-99, the coefficient on gender switches signs and is negative and well determined implying that male heads of households are now less likely to contract malaria as compared to female heads. This could be because of the small percentage of female heads of households, perhaps being a more heterogeneous sub-set over the two years and, thus, resulting in the switch in coefficient sign observed—in 1992-93 there were seven percent female heads and in 1998-99 there were eight percent. It is also likely that male heads of households have better economic status in 1998-99 and this could alternatively explain the change in sign on the gender coefficient. This is not implausible given that economic status as measured by the non-agricultural consumer durable asset wealth index has risen over this time period, when India also underwent rapid economic reforms in the post-liberalisation era since 1991. However, it is impossible to provide a definitive explanation for the change in sign given the data available to us.

The estimated effect for caste is also well determined (negative) in 1992-93 but not so in 1998-99. None of the age categories had well determined effects in 1992-93, but in 1998-99 the category, above 67 years, relative to 18 to 27 years, has a positive and significant coefficient, *ceteris paribus* and on average, possibly because of lower immunity in this oldest age bracket.

There is also a reversal in signs on the coefficient on high-school educated heads as well as on heads using wood as fuel. The result for the education variable in 1992-93 indicates that *ceteris paribus* and on average a high-school educated household head is less likely to contract malaria compared to one who is illiterate by 2.7 percentage points (see Appendix G). In 1998-99 this result is reversed: a high-school educated head is more likely to get malaria than an illiterate head, *ceteris paribus* and on average, by 1.9 percentage points. It is not clear why this change has occurred especially since the proportion of heads of households in each category has not changed much over the two time periods. The percentage of heads in the high school category exhibits an increase of a mere one percentage point in 1998-99 as compared to 1992-93 (see Table 3.5, p.52). Also, illiterate heads are a much larger percentage in both years at 54 percent in 1992-93 and 50 percent in 1998-99 — the χ^2 test fails to reject the null of no difference over the two years.[15]

[15]This result is in conflict for the two years and it is hard to explain why. It may be related to the ownership of water pumps, tractors and bullock carts (even though we are controlling for these variables), all of which have a positive significant impact in 1998-99. The Phi correlation coefficient between tractor and high-school educated is 0.104, between high-school educated and water pump ownership is 0.108, and between high-school educated and bullock cart ownership is 0.025. Corresponding correlation coefficients between illiterate heads and the above agricultural assets ownership are: tractor: -0.098, water pump: -0.125 and bullock cart: -0.048. If all three of these variables are dropped from the specification, the coefficient on high-school is positive significant not at ten percent but at the five percent level. The

Among the fuel categories, for 1992-93 the coefficient on firewood is negative and significant, which means that relative to the base category, dung, heads of households using firewood are less likely to contract malaria, *ceteris paribus* and on average. The reasons for this were discussed earlier in this chapter. In 1998-99 the coefficient on wood is positive and significant contrary to expectations indicating that heads of households using this type of fuel are more likely to contract malaria relative to those using dung, *ceteris paribus* and on average, by 1.3 percentage points. However, there is a seven percentage point increase in households using the miscellaneous category of fuels (which includes superior fuels like liquefied petroleum gas) in 1998-99 and this category has a negative significant effect. There is a corresponding drop of seven percentage points in using wood as fuel. The proportion of households using dung is similar at 0.20 for both years. It is possible that the reporting of the wood category also includes households that use a mix of dung and wood. Better-off households who used only wood are likely to have switched to the miscellaneous category in the second time period. The finding here supports the negative malaria incidence and socio-economic status association,

In the drinking water source categories, using an open public water source has a positive and significant coefficient in the 1992-93 specification (reasons for which were explained earlier) whereas none of the estimated effects are well determined in 1998-99. This is the case for electricity as well even though the proportion of households having electricity is nearly the same for both time periods at approximately 20 percent. We fail to reject the null hypothesis of no difference in the means – see Table 3.5 (p.52). Finally, the coefficient on asset index is not well determined in 1992-93, but is negative and significant in 1998-99. The next subsection discusses the estimated effects of this wealth variable in greater detail.

Impact of wealth on malaria incidence

The non-agricultural consumer durable asset wealth index constructed using Principal Components Analysis is significant with a negative sign in 1998-99, indicating a negative association with malaria as the value of the index rises, *ceteris paribus* and on average, which may be interpreted as wealthier households being less susceptible to malaria. The marginal effect indicates that a one standard deviation increase in the asset index would

agricultural assets ownership is being considered here as these have well determined effects in 1998-99 but did not in 1992-93 (except water pump). Thus, if a household owns these assets, they are more likely to get malaria and these heads also happen to be high-school educated rather than illiterate. But, the interaction terms of high-school with each of these agricultural assets were not significant.

Table 5.10: Probit of malaria incidence for 1992-93 and 1998-99 with asset index quintiles: comparable rural specifications

Variable	*1992–93* *ME/IE*[a]	*1998–99* *ME/IE*
Asset index quintiles		
Bottom 20%	0.179***	0.063***
	[0.048]	[0.021]
20–40%	-0.018	-0.015
	[0.022]	[0.013]
40–60%	-0.012	-0.017
	[0.016]	[0.019]
60–80%	0.039***	0.012
	[0.015]	[0.008]
Top 20%	-0.011**	-0.009*
	[0.005]	[0.005]
Social characteristics		
Caste	-0.014*	0.004
	[0.008]	[0.006]
Gender	0.023*	-0.055***
	[0.012]	[0.019]
Age dummies (Reference category: 18-27 years)		
28 to 37 years	0.001	0.011
	[0.013]	[0.011]
38 to 47 years	0.013	-0.002
	[0.013]	[0.010]
48 to 57 years	0.007	0.017
	[0.014]	[0.013]
58 to 67 years	0.006	0.017
	[0.014]	[0.014]
Above 67 years	-0.008	0.026
	[0.015]	[0.018]
Education (Reference category: Illiterate)		
High school	-0.028***	0.018*
	[0.010]	[0.011]
Middle school	-0.013	0.011
	[0.010]	[0.011]
Primary school	0.006	0.004
	[0.009]	[0.007]
Occupation (Reference category: production and transport)		
Agriculture	0.011	0.005
	[0.009]	[0.008]
Wage	-0.015	-0.007
	[0.012]	[0.009]
Other	0.025	-0.015*
	[0.018]	[0.008]
House type (Reference category: semi-*pucca*)		
Pucca house	-0.003	-0.004
	[0.013]	[0.011]
Kachcha house	-0.006	0.007
	[0.008]	[0.006]
Fuel type (Reference category: dung)		
Miscellaneous	0.012	-0.018**

Continued on next page...

Table 5.10 continued

Variable	*1992–93 ME/IE*	*1998–99 ME/IE*
	[0.028]	[0.008]
Wood	-0.028***	0.013**
	[0.011]	[0.006]
Drinking water (Reference category: protected public)		
Open	0.053**	0.017
	[0.026]	[0.056]
Protected private	-0.001	-0.007
	[0.008]	[0.006]
Other characteristics		
Separate kitchen	0.001	-0.006
	[0.007]	[0.006]
Electricity	-0.021**	-0.007
	[0.009]	[0.008]
Sanitation	-0.012	0.001
	[0.015]	[0.010]
Agricultural land and assets		
Agricultural land	-0.005	-0.010
	[0.010]	[0.008]
Owns a water pump	0.020	0.018
	[0.013]	[0.011]
Owns a bullock cart	-0.002	0.021
	[0.013]	[0.012]
Owns a tractor	0.013	0.032
	[0.025]	[0.022]
Number of observations	6071	4588.000
Log Likelihood		
Pseudo R-squared	0.0915	0.1036

a. ME/IE indicates marginal or impact effects; district-effects were included in both specifications above; robust standard errors reported in brackets; Pseudo R-squareds reported are the McFadden's psuedo R-squareds; * significant at 10%; ** significant at 5%; *** significant at 1%

reduce malaria incidence by 0.4 percentage of a point (see Appendix G).

The sample was divided into quintiles on the basis of their value for the index, thus placing knots at the 20th, 40th, 60th, and 80th percentiles of the data.[16] Table 5.10 shows that the coefficient on the asset index for the bottom 20 percent of the sample remains positive and significant at the one percent level indicating that a small increase in asset wealth at this level would fail to bring down malaria incidence. For the top most group the effect is negative and significant – a rise in wealth would reduce malaria incidence. These results are consonant with the quintile results obtained for the 1992-93 specification used earlier in this chapter as well as for the comparable specification and supports the view that household wealth and health outcomes are negatively associated.

[16]Introducing a quadratic term to see if the effect of the asset index was linear or non-linear did not yield well determined effects for either year.

There is considerable discussion in the literature on the negative association between health and poverty. The findings for the measure of wealth, the consumer durable asset index, for the year 1998-99 in our study support this view. However, we need to exercise caution in making generalisations in this context. One apparent support for this lies in the behaviour of the agricultural asset variables. It is the more wealthy households that are likely to own a tractor, a bullock cart, or a water pump. Yet, these are the households that are also more likely to contract malaria. It was, in fact, to identify the specific impact of different agricultural variables that these were not collapsed into an index in our specification.[17]

Having compared the probit results for 1992-93 and 1998-99 rural Uttar Pradesh separately we now turn to the decomposition of the malaria differences over time.

Decomposition of malaria difference over time

In order to identify the factors that contribute to differences in malaria incidence over time, a decomposition analysis can be undertaken, as described in Chapter 4 (p.98).[18] An aggregate decomposition (Table 5.11) of the malaria difference over these two periods showed that structural differences dominated and the explained component was very small as we can see from the table below. A detailed decomposition of the malaria differential into component effects is presented in Table 5.12 and is dicussed now.

Table 5.11: Aggregate decomposition of malaria difference over time (1992-93 to 1998-99) into differences in characteristics and coefficients

	1998-99 coefficients and 1992-93 characteristics	*1992-93 coefficients and 1998-99 characteristics*
Characteristics difference	-0.00058	0.0074
Coefficients difference	0.0447	0.0367
Aggregate difference	0.0441	0.0441

The aggregate decomposition effects (Table 5.12, last two rows) show that the coefficient effects dominate the 1992-93 and 1998-99 malaria incidence difference and widen the gap in malaria incidence between these two time periods. It is worth noting that the constant is the largest contributor, possibly capturing some of the effects not included in the specification including policy variables, for which the data were not available.

[17] Additionally, the effect for the index of agricultural assets was not well determined for either period.

[18] The decomposition methodology is presented in Chapter 4 with reference to settlement types, but it can be equally applied to temporal data.

Table 5.12: Detailed decomposition of 1992-93 and 1998-99 malaria difference into differences in characteristics and coefficients for rural Uttar Pradesh

Variable	*C92*	*D98*	*C98*	*D92*
Caste	0.0003***	-0.0025***	-0.00004	-0.0023***
	[0.00003]	[0.0006]	[0.00006]	[0.0004]
Gender	0.0001	0.0562	-0.0004	0.0594
	[0.00009]	[0.0963]	[0.0004]	[0.0859]
Age (aggregate)	0.0002***	-0.0027**	-0.0001	-0.0024**
	[0.00003]	[0.0012]	[0.0009]	[0.0011]
Education (aggregate)	0.00008***	-0.0058***	-0.0004	-0.0053***
	[0.00002]	[0.0014]	[0.0003]	[0.0010]
Occupation (aggregate)	-0.0002	0.0059***	0.0012	0.0044***
	[0.0007]	[0.0021]	[0.0028]	[0.0014]
House type	-0.0003***	-0.0072	0.0003	-0.0082*
	[0.00005]	[0.0046]	[0.0007]	[0.0047]
Separate kitchen	0.0008***	0.0054***	-0.0007	0.0073***
	[0.0003]	[0.0016]	[0.0034]	[0.0024]
Sanitation	0.0019***	-0.0057***	-0.0004	-0.0039***
	[0.0004]	[0.0012]	[0.0013]	[0.0004]
Electricity	0.0003**	-0.0050***	0.00003*	-0.0050***
	[0.00002]	[0.0014]	[0.00002]	[0.0011]
Fuel (aggregate)	-0.0021**	-0.0329	0.0032	-0.0404
	[0.0010]	[0.0289]	[0.0129]	[0.0310]
Drinking water source (aggregate)	0.0041*	-0.0023*	0.0008	-0.00008
	[0.0023]	[0.0013]	[0.0034]	[0.0009]
Agricultural land	0.00003***	0.0100	-0.00034	0.0109
	[0.00005]	[0.0092]	[0.0007]	[0.0088]
Water pump	0.00012**	-0.0015***	0.0002*	-0.0017***
	[0.00007]	[0.0002]	[0.0001]	[0.0003]
Bullock cart	0.0002*	-0.0040***	-0.0002	-0.0038***
	[0.00001]	[0.0006]	[0.0001]	[0.0005]
Tractor	0.00002	-0.0012	-0.0002	-0.0009**
	[0.000017]	[0.0001]	[0.0003]	[0.00006]
Asset index	-0.00012***	0.0169	0.0012	0.0161**
	[0.00003]	[0.0123]	[0.0040]	[0.0088]
Region (aggregate)	0.0020	-0.0249**	-0.0034	-0.0198***
	[0.0022]	[0.0100]	[0.0213]	[0.0062]
Constant		0.0362		0.0378
		[0.1117]		[0.0974]
Aggregate	0.0075	0.0351*	0.0005	0.0420***
	[0.0072]	[0.0186]	[0.0118]	[0.0205]

Standard errors for the characteristic and coefficient effects are reported in brackets; C92: Characteristic effect using the 1992-93 coefficient vector; D98: Coefficient effect using 1998-99 characteristics; C98: Characteristic effect using the 1998-99 coefficient vector; D92: Coefficient effect using the 1992-93 characteristic vector; Characteristic (or coefficient) effect of each variable is calculated as the product of the aggregate characteristic (or coefficient effect) and the weight assigned to each variable; Total effect of each variable (or groups of variables) = Sum of characteristic and coefficient effect of that variable; Share of characteristic effect of each variable= (Characteristic effect of that variable/Total effect of the variable)*100; Share of coefficient effect of each variable = (Coefficient effect of that variable/Total effect)*100; Sum of the individual (or group) characteristics (or coefficients) effects = Total characteristic or coefficients effect using this set of weights (urban characteristics and rural coefficients); Aggregate characteristic and coefficient effects reported are calculated using mean values of the characteristic vector and differ from those reported in Table 5.11.

Among the individual or group effects, a number of characteristic estimates are signifcant at a conventional level when using the 1992-93 coefficient estimates but not when using the 1998-99 coefficients. Differences in caste, age, education, saeparate kitchen, electricity, sanitation, drinking water source, and the ownership of agricultural land, water pump, and bullock cart, and asset index widen the inter-temporal malaria incidence gap whereas house type and fuel type narrow it. Among the coefficient effects when using the 1992-93 characterisitcs caste, age, education, electricity, sanitation, drinking water source, water pump, bullock cart and region narrow the malaria incidence gap. Occupation and separate kitchen widen it. Using the alternative set of weights, that is, the 1998-99 coefficient vector and 1992-93 characteristics, none of the characteristic effects with the exception of owning a water pump attain significance. But, among the corresponding coefficient effects, caste, age, education, house type, electricity, sanitation, water pump, bullock cart, and region narrow the gap. On the other hand, occupation, separate kitchen, water pump, and asset index widen the gap. The results for the different sets of weights are not in conflict even if most of the characteristic effects are not well determined for the latter set of weights using 1992-93 coefficients.

Using 1992-93 characteristics, the average 1992-93 and 1998-99 difference in households belonging to a scheduled caste or tribe (SC/ST) widens the malaria incidence gap by 0.03 of a percentage point. A higher percentage of SC/ST households in 1998-99 (three percentage points more than 1992-93, see p.52, Table 3.5) registers as a widening of the malaria gap. However, the larger negative coefficient effect or unexplained component narrows the malaria incidence gap between the two time periods, seeming to indicate that SC/STs were better off in 1992-93 as compared to 1998-99.

The positive characteristic effect for age indicates a widening of the inter-temporal malaria incidence gap using the 1992-93 characteristics, once again counteracted by the negative unexplained component, as in the case of caste. The statistically significant change in age composition seems to work favourably in 1998-99 in terms of reducing malaria incidence, but the unexplained effect counteracts this positive effect. Using the alternative set of weights, too, the coefficient effect narrows the malaria incidence gap.

For education, too, the characteristic effect for 1992-93 widens the malaria incidence gap, attributed likely to the higher proportion of educated household heads and a concomitant decline in those who are illiterate, but structural differences dominate this effect, i.e., the coefficient effect is negative, thus narrowing the gap. In the case of occupation, the characteristic effects are not significant with either set of weights and the coefficient effects are negative implying a narrowing of the malaria incidence gap.

The characteristic effect for house type using the 1992-93 coefficients is negative and significant, indicating a narrowing of the malaria incidence gap between the two time periods. The corresponding coefficient effects are not significant. As we can see from Table 3.5, p.52, there is a decline in the percentage of kachcha houses by five percentage points over this six year period. There is, however, a corresponding rise in the ownership of semi-pucca houses while the pucca house percentage is nearly the same. Using the alternative set of weights, the coefficient effect also indicates a narrowing of the malaria incidence gap. Having a semi-pucca house, although somewhat superior in construction as compared to a kachcha house does not seem to bear any advatange over a kachcha house for preventing malaria.

Having a separate kitchen widens the malaria incidence gap as indicated by the characterisic and coefficient effect possibly because of a decline in houses with a separate kitchen in 1998-99. As dicussed, not having a separate kitchen might help in deterring mosquitoes because of the smoky environment in rural households that mostly use smoky fuels.

The characteristic effect for electricity using either set of weights widens the difference in inter-temporal malaria difference. The corresponding coefficient effects, which are larger, contribute to narrowing the malaria difference indicating that the behavioural response of households to having electricity connections in 1998-99 is not working in favour of bringing malaria incidence down even though there is a small (though insignificant: see Table 3.5) rise in the number of households that have access to electricity. As noted, a larger percentage of households with electricity in 1998-99 as compared to 1992-93, registers as a widening of the malaria incidence gap but seems to be counteracted by the negative coefficient effect. For sanitation, too, the characteristic effect widens the malaria incidence gap and the unexplained differential narrows it.

The negative characteristic effect for type of fuel using 1998-99 coefficients works towards narrowing the malaria incidence differential between the two time periods likely because of a decline in the use of wood as fuel in households by seven percentage points from 1992-93 to 1998-99. This decline seems to be accompanied with a rise in the use of the miscellaneous fuel category, primarily Liquified Petroleum Gas or LPG. Since wood is a smoky fuel and would, therefore, act as a deterrent for mosquitoes, the decline in use of wood as fuel registers as a narrowing of the malaria incidence gap between the two time periods.

For drinking water source the characteristic effect serves to widen the malaria incidence gap, as one would expect, since there are fewer open water sources in 1998-99 as compared to 1992-93 (Table 3.5) and a large percentage of households use a private source for drinking water, less likely to be breeding grounds for mosquitoes as compared to open or public sources. The positive characteristic effect is also larger than the negative coefficient effect when using 1998-99 coefficients.

The decline in ownership of agricultural land by 3.5 percentage points registers as a positive characteristic effect, serving to widen the inter-temporal malaria incidence gap, as expected. For bullock cart and water pump ownership, the positive characteristic effect that widens the malaria incidence gap is dominated by the coefficient effect or the unexplained component serving to narrow the malaria incidence gap.

Finally, the average difference in asset index characteristics narrows the malaria incidence gap between the two time periods. But, the unexplained component reveals a widening of the gap, which is a larger effect than the former, indicating that structural or behavioural variables seem to have a much greater effect on malaria incidence than a mere increase in the ownership of assets.

The overall conclusion from the above analysis is that coefficient effects or the unexplained component of the malaria differential between the two time periods dominates, and in most cases results in a narrowing of the differential. On the whole, however, there is a decline in malaria incidence over the time period and as the aggregate effects reveal these are attributable to the unexplained differential, serving to widen the malaria incidence gap (rather than narrowing it). It is possible that this is due to a change in policy over the time period, a step-up in insecticide spraying and other such measures, data on which are not available. We now turn to the district level analysis.

5.3.2 District level analysis

Evidence of robustness at the district level

As discussed in Chapter 3, the 1998-99 household data were combined with the census data for 1991 and 2001. The 1992-93 data were used with the 1991 census data. The specifications for the two time periods were matched. Since six districts were not covered in 1998-99 as mentioned in the previous section, these were dropped from the district-analysis as well.

Table 5.13: Results of the second stage district regression for 1992-93 and 1998-99: comparable samples

Variable	*1992-93*	*1998-99*	*t-ratios*[a]
PV with a high-school	-0.072***	0.019	-2.57
	[0.026]	[0.024]	
PV with an adult literacy centre	-0.019***	0.021***	-4.68
	[0.003]	[0.008]	
Average distance to drinking water	-0.154***	-0.022	-1.49
	[0.042]	[0.078]	
PV with a commuting facility	0.005*	-0.005	1.05
	[0.003]	[0.009]	
PD irrigated by a government canal	0.009*	-0.001	1.71
	[0.005]	[0.003]	
PD irrigated by a private canal	0.614**	-0.073	2.60
	[0.239]	[0.112]	
PD irrigated by a well	0.144***	0.040*	2.37
	[0.038]	[0.022]	
PD irrigated by a tube well	0.007***	-0.0002	2.28
	[0.001]	[0.003]	
PD irrigated by an electric tube well	0.012***	0.013***	-0.20
	[0.003]	[0.004]	
PD under forest	-0.01	-0.005	-0.54
	[0.006]	[0.007]	
District population density	-0.106***	-0.04	-1.55
	[0.027]	[0.033]	
Log rain	-0.471***	-0.055	-2.44
	[0.131]	[0.109]	
Constant	3.502***	0.483	
	[0.890]	[0.810]	
Number of observations	42	42	
R-squared	0.6487	0.4954	

PV stands for percentage of villages (in the district) and PD for percentage of district area; the t-statistic to test for differences in estimates across the two time-periods at the 5% level of significance is ±1.96.

In 1992-93, the same variables have well determined effects and with the same signs as before, mainly percentage of villages with a high school, an adult literacy centre, and a commuting facility, average distance to a drinking water source, percentage of district area irrigated by different irrigation types, rainfall, and population density (see Table 5.13). When comparing the results with 1998-99, once again we find little evidence of robustness. In 1998-99, only three variables have well determined effects. These are percentage of villages with an adult literacy centre, percentage of district area irrigated by a well and percentage area irrigated by an electric tube well. The effect of the latter variable on malaria incidence is robust over the time period considered since the t-test fails to reject the null of no difference in estimates even though the coefficient estimates are nearly the same at 0.012 for 1992-93 and 0.013 in 1998-99. The sign on the coefficient on adult literacy centre is positive and the t-test rejects the null of no difference in estimates in this variable over the two time periods. The result for adult literacy is in conflict with that obtained for 1992-93. The reason for this reversal is not clear and cannot be determined from the data available. However, it might be a result of targeting health education classes to malaria-prone districts to raise awareness as a part of the malaria eradication programme. There was, in fact, an increase in such classes as a part of malaria-control efforts of the NMEP over this period as noted at the beginning of this section.

The variables, average distance to a drinking water source, percentage of villages with a commuting facility, percentage of district area irrigated by a government canal, and population density had significant coefficients in 1992-93 but are not well determined in 1998-99. For each of these, the t-test does not reject the null hypothesis of no difference in estimates meaning that there is no statistical difference in the estimates, which provides some evidence of robustness in the relationship over time. Finally, for percentage of villages with a high school, log rain, and percentage of district area irrigated by a tube well, the null of no difference is decisively rejected, largely because one or other of the estimates is not different from zero.

There is more noise in the determination process for the second time period – the R-squared is lower at 0.4954 in 1998-99 as compared to 0.6487 in 1992-93. Moreover, there might have been changes in unobservables over time, not captured in the specification presented here. The analysis of the kind followed here can indicate how variables may behave in a particular period but it is hard to determine what the behaviour will be like for other periods. It may depend on conditions prevailing in that particular period and these may not be captured by the data available for analysis. Thus, generalisations are much more difficult to make.

5.4 Summary

The research questions we have addressed in this chapter were (1) what socio-economic factors affect malaria incidence at the household and district level and assess if these are robust over time for rural Uttar Pradesh; (2) is the impact of wealth on malaria incidence negative and is the relationship between other socio-economic factors and malaria incidence negative?

With reference to the first research question about whether the factors identified as affecting malaria incidence are robust over time or not, we can conclude that there is little evidence of robustness at the household or district level, indicating that factors determining malaria change over time.

In answering the second research question, the findings from this chapter caution us in generalising the negative link between indicators of socio-economic status and disease incidence. This generalisation does not always hold true for malaria incidence in Uttar Pradesh. However, we do find that wealth has a negative impact on malaria incidence both for 1992-93 and 1998-99. The bottom group when using quintiles in the probit regression had a positive significant coefficient indicating that this group is more likely to contract malaria whereas the coefficient on the asset wealth index for the top-most group was negative and well determined implying that this group is less likely to contract malaria, *ceteris paribus* and on average.

At the household level the conventionally held belief of a negative link between socio-economic status and malaria incidence is questioned by the result for the caste and ownership of water pump variables. Household heads from a scheduled caste/tribe (of a lower socio-economic status) are less likely to contract malaria as compared to other castes in Uttar Pradesh, *ceteris paribus*, possibly because of certain genetic traits. On the other hand, owning a water pump is indicative of higher wealth status and yet a head of household head who owns this asset is more likely to contract malaria, *ceteris paribus*, possibly because of the presence of stagnant water pools around pumps providing potential areas where the mosquito vector can thrive.

Findings from both the household level probit and the district level regression also seem to point to the positive role played by some of the socio-economic development indicators (used in this study) in countering malaria incidence. At the household level, having an electricity connection and having access to protected public water supply (relative to an open one) exerts a negative impact on malaria incidence.

Education and, implicitly, awareness is also identified as a significant factor associated with a lower likelihood of malaria incidence both at the household level as well as at the district level. *Ceteris paribus*, a household head educated to the high school level relative to an illiterate household head has a lower likelihood of contracting malaria according to the 1992-93 specification but not with reference to the 1998-99 results. At the district level for the same year, the greater the percentage of villages with high schools and adult literacy centres, the lower is the propensity to contract malaria, *ceteris paribus*.

The impact of socio-economic development indicators or infrastructure variables on malaria incidence may, however, not always be negative. At the district level, area irrigated by canals shows a positive association with malaria incidence (for 1992-93). However, the lower (though positive and significant) coefficient for area irrigated by government canals as compared to private canals may be indicative of the (relatively) lower positive impact of government infrastructure development (in the form of canals) on malaria incidence, likely because private canals are mostly rudimentary in nature and may, therefore, have inferior drainage systems as compared to government canals.

Another socio-economic development indicator operating at the district level that yields a positive association with malaria (in 1992-93) is the percentage of villages with a commuting facility. Although this issue needs to be explored further it may be conjectured that this is because transport vehicles harbour mosquitoes or commuters could act as carriers of *Plasmodium* from one area to another.

The results for the district level variables point towards a governmental intervention in such areas as improving infrastructure around public water collection sites and improving canal linings to prevent mosquito breeding, introducing insecticide spraying in public commuting facilities, and imparting education on prophylactic measures.

Having summarised the main findings from this chapter, we now turn to an analysis of urban malaria in the next chapter and conduct the decomposition of the rural and urban malaria difference.

Chapter 6

Analysis of differences in rural and urban malaria incidence in Uttar Pradesh

The previous chapter dealt with the analysis of malaria in rural Uttar Pradesh for 1992-93 and 1998-99. This chapter compares rural and urban malaria and presents the decomposition of the malaria difference between these settlement types for the two NFHS years. The research questions are: What are the socio-economic factors affecting malaria incidence in urban Uttar Pradesh? Are these robust over time? How do the factors differ by settlement types? Does wealth and other socio-economic factors have a negative impact on malaria incidence? We only focus on households and not districts here since a large number of districts were dropped from the urban samples due to perfect classification of failure as there was no malaria in these districts.

As explained in the beginning of the previous chapter and in Chapter 2, the factors that affect rural and urban incidence of malaria differ because of differences in the nature of the malaria-causing *Anopheles* vectors found in these locales. Rural settings are more conducive to the spread of malaria providing fresh water sources for breeding of *An. culicifacies* in such breeding areas as agricultural fields, ponds, rivers, and canal linings. The urban vector, *Anopheles stephensi*, on the other hand, can exploit these water bodies for breeding, commonly found in rural settings, but has increasingly been able to adapt to breeding even in small water collections and particularly in tanks used to store water in urban areas.

The first section of this chapter discusses the probit results for urban Uttar Pradesh for 1992-93 and 1998-99 and compares these with rural Uttar Pradesh and the second section presents the decomposition of the malaria difference between the two settlement types. The Wald's χ^2 test was carried out to assess whether the effects of the different variables on malaria incidence differed by the rural and urban settlement types. The null hypothesis being tested was that the effect on malaria does not differ by settlement type. The χ^2 statistic (with 25 degrees of freedom) is 83.55 so that we reject the null of no settlement type differences since the critical value at 25 degrees of freedom for the five percent level of significance is much lower at 37.652.

6.1 Comparing urban and rural results

The rural results presented in the previous chapter were obtained from a different specification. Therefore, the sample size and results differ from those presented earlier. The changes made to the original rural specification in order to match it with the urban one were discussed in Chapter 3 and presented in Table 3.6. Briefly, the changes made are: (1) combining age groups 18 to 27 years with 28 to 37 years since the former predicted failure perfectly in the urban sample; (2) combining private water source with public source as individually these were dropped due to perfect prediction of failure in the urban sample; (3) using ownership of agricultural land rather than split up irrigated and non-irrigated area categories since in urban areas the land owned was mostly irrigated; (4) not using agricultural assets in the specification since in urban areas less than one percent of the households owned these assets; and (5) using regions instead of districts since 38 districts did not have any malaria in the urban sample and of those that had malaria, seven were perfectly classified. This is likely because of the low incidence of malaria in urban areas – 2.30 cases per 100 population in 1992-93 and 1.43 per 100 population in 1998-99.

The results for the changed rural sample are not in conflict even if some coefficients do not retain significance (e.g., gender and caste). The variables yielding significant coefficients are discussed here. Table 6.1 presents the urban probit results for 1992-93 and 1998-99. Table 6.2 presents these for comparable rural and urban samples for the two time-periods.

Table 6.1: Probit of malaria incidence for 1992-93 and 1998-99 using quintiles for asset index: comparable urban specifications

Variable	*1992–93 Coefficient*	*1998–99 Coefficient*	*t–ratio*
Asset index quintiles			
Bottom 20%	-0.291	0.385**	-2.66
	[0.199]	[0.158]	
20–40%	0.156	0.095	0.34
	[0.152]	[0.231]	
40–60%	-0.213	-0.813**	1.54
	[0.266]	[0.381]	
60–80%	-0.526	0.891**	-2.45
	[0.432]	[0.385]	
Top 20%	0.160	[a]	
	[0.454]		
Caste	-0.334	-0.704**	0.90
	[0.236]	[0.336]	
Gender	-0.255	0.372	-1.48
	[0.250]	[0.342]	
Age dummies (Ref: above 57 years)			
18 to 37 years	-0.092	0.059	-0.46
	[0.210]	[0.248]	
38 to 47 years	0.249	-0.33	1.76
	[0.189]	[0.270]	
48 to 57 years	0.195	-0.027	0.67
	[0.206]	[0.258]	
Education (Ref: illiterate)			
High school	-0.101	-0.168	0.20
	[0.237]	[0.226]	
Middle school	0.014	-0.497	1.17
	[0.237]	[0.367]	
Primary school	0.428**	-0.172	2.08
	[0.177]	[0.227]	
Occupation (Ref: production and transport)			
Agriculture	0.061	0.489	-0.95
	[0.304]	[0.330]	
Wage	0.143	-0.320*	1.94
	[0.156]	[0.181]	
Other	0.284	-0.267	1.60
	[0.229]	[0.256]	
House type (Ref: semi–*pucca*)			
Pucca house	0.300	0.039	0.88
	[0.201]	[0.220]	
Kachcha house	0.450**	0.181	0.72
	[0.224]	[0.297]	
Separate kitchen	-0.434***	-0.415*	-0.07
	[0.149]	[0.223]	
Electricity	-0.115	-0.388	0.89
	[0.191]	[0.241]	
Sanitation	0.402**	0.287	0.33
	[0.186]	[0.294]	
Fuel type (Ref: dung)			
Miscellaneous	0.347	-0.217	1.23
	[0.313]	[0.337]	
Wood	0.461*	-0.020	1.10
	[0.264]	[0.347]	
Drinking water (Ref: protected public)			
Open	-0.244	0.139	-0.70
	[0.296]	[0.459]	
Protected private	-0.120	-0.039	-0.28

Continued on next page. . .

Table 6.1 continued

Variable	*1992–93 Coefficient*	*1998–99 Coefficient*	*t–ratio*
	[0.172]	[0.235]	
Agricultural land	0.184	0.116	0.24
	[0.179]	[0.227]	
Regions			
Hill	-0.565**	-0.429	-0.33
	[0.260]	[0.313]	
East	-0.578*	-0.279	-0.61
	[0.318]	[0.373]	
West	-0.137	-0.461**	1.20
	[0.184]	[0.199]	
Bundelkhand	0.137	-0.716	1.75
	[0.213]	[0.439]	
Constant	-3.125***	-0.754	-2.10
	[0.820]	[0.775]	
Number of observations	2220	1399	
Log likelihood	-207.86	-102.71	
Pseudo R–squared	0.1696	0.1835	

a. The top 20% quintile drops out because of perfect prediction of failure; Robust standard errors reported in brackets; Pseudo R-squareds reported are the McFadden's psuedo R-squareds; * significant at 10%; ** significant at 5%; *** significant at 1%, where the null hypothesis of the coefficient being equal to zero is tested using an asymptotic t–test; t–ratios presented in the last column test for differences in estimates across the two years; t-statistic at the 5% level of significance is ± 1.96.

The socio-economic factors affecting urban malaria can be identified at the outset as: caste, education, occupation, separate kitchen, sanitation, and fuel-type. The only factor found to be robust over time is separate kitchen. We now discuss these findings in detail.

6.1.1 Impact of socio-economic status on malaria incidence

The urban results are as expected for most variables, given the literature on the subject. The coefficient on caste is negative in both the time periods for the urban sample but is significant only in 1998-99 (though the t-test does not reject the null hypothesis of no difference in the estimates) and indicates that SC or ST household heads are less likely to contract malaria, as compared to those who are not. This refutes the negative socio-economic status and health relationship, as we found in the previous chapter. The reasons for the negative impact of caste have been discussed earlier in the thesis and probably lie in the presence of the sickle cell or other genetic abnormalities. Caste does not have a significant effect for the corresponding rural samples.

In the education categories, the coefficient on household heads with primary level of schooling is significant and positive for the 1992-93 urban sample implying that they are more likely to contract malaria as compared to the base group of illiterate heads, *ceteris*

Table 6.2: Probit of malaria incidence for rural and urban Uttar Pradesh: comparable specifications, 1992-93 and 1998-99

Variable	*Rural Coefficient 1992-93*	*Urban Coefficient 1992-93*	*t-ratio*	*Rural Coefficient 1998-99*	*Urban Coefficient 1998-99*	*t-ratio*
Asset index	0.0003	-0.238**	2.14	-0.042	0.023	-1.05
	[0.043]	[0.103]		[0.030]	[0.054]	
Caste	-0.039	-0.327	1.17	0.031	-0.686**	2.01
	[0.056]	[0.240]		[0.080]	[0.348]	
Gender	0.083	-0.252	1.25	-0.446***	0.403	-2.35
	[0.104]	[0.247]		[0.127]	[0.338]	
Age dummies (Ref: above 57 years)						
18 to 37 years	0.001	-0.082	0.38	-0.096	0.112	-0.81
	[0.066]	[0.209]		[0.099]	[0.237]	
38 to 47 years	0.097	0.239	-0.71	-0.218**	-0.222	1.57
	[0.067]	[0.187]		[0.105]	[0.259]	
48 to 57 years	0.031	0.199	-0.77	-0.101	0.027	-0.48
	[0.072]	[0.205]		[0.110]	[0.242]	
Education (Ref: illiterate)						
High school	-0.119	-0.128	0.03	0.170	-0.164	1.29
	[0.087]	[0.247]		[0.112]	[0.233]	
Middle school	-0.112	0.002	-0.45	0.061	-0.440	1.26
	[0.083]	[0.237]		[0.120]	[0.380]	
Primary school	0.082	0.405**	-1.73	0.018	-0.080	0.40
	[0.058]	[0.177]		[0.090]	[0.230]	
Occupation (Ref: production and transport)						
Agriculture	0.070	0.055	0.05	0.077	0.539*	-1.37
	[0.065]	[0.303]		[0.093]	[0.325]	
Wage	-0.072	0.127	-1.08	-0.042	-0.297*	1.20
	[0.094]	[0.158]		[0.132]	[0.166]	
Other	0.165	0.285	-0.48	-0.227	-0.239	0.04
	[0.104]	[0.227]		[0.141]	[0.251]	
House type (Ref: semi-*pucca*)						
Pucca house	0.085	0.309	-0.98	-0.063	0.036	-0.37
	[0.098]	[0.207]		[0.151]	[0.222]	
Kachcha house	-0.030	0.459**	-2.10	0.055	0.074	-0.07
	[0.053]	[0.227]		[0.076]	[0.276]	
Separate kitchen	0.098**	-0.441***	3.40	-0.04	-0.410*	1.57
	[0.049]	[0.151]		[0.083]	[0.220]	
Electricity	-0.252***	-0.066	-0.94	-0.025	-0.268	0.93
	[0.074]	[0.184]		[0.101]	[0.242]	
Sanitation	-0.159	0.455**	-2.84	0.156	0.319	-0.53
	[0.107]	[0.188]		[0.113]	[0.287]	
Fuel type (Ref: dung)						
Miscellaneous	-0.15	0.355	-1.41	-0.456**	-0.215	-0.63
	[0.165]	[0.319]		[0.186]	[0.335]	
Wood	-0.324***	0.482*	-2.92	0.146*	-0.021	0.48
	[0.058]	[0.270]		[0.088]	[0.338]	
Drinking water (Ref: protected public)						
Open	0.101	-0.294	0.77	-0.076	0.043	-0.25
	[0.077]	[0.504]		[0.076]	[0.470]	
Protected private	-0.119**	-0.103	-0.09	-0.814**	-0.056	-1.67
	[0.054]	[0.168]		[0.387]	[0.236]	
Agricultural land	-0.018	0.220	-1.26	-0.078	0.061	-0.59
	[0.061]	[0.179]		[0.085]	[0.218]	
Regions						
Hill	0.002	-0.541**	1.99	0.019	-0.368	1.02

Continued on next page...

Table 6.2 continued

Variable	*Rural Coefficient 1992-93*	*Urban Coefficient 1992-93*	*t-ratio*	*Rural Coefficient 1998-99*	*Urban Coefficient 1998-99*	*t-ratio*
	[0.085]	[0.260]		[0.240]	[0.292]	
East	-0.322***	-0.538*	0.67	0.045	-0.278	0.86
	[0.074]	[0.315]		[0.119]	[0.355]	
West	-0.014	-0.137	0.64	0.383***	-0.421**	3.49
	[0.064]	[0.180]		[0.112]	[0.201]	
Bundelkhand	-0.136	0.119	-1.01	0.299**	-0.629	2.08
	[0.145]	[0.208]		[0.136]	[0.425]	
Constant	-1.127***	-2.304***	2.12	-1.430***	-1.749***	0.46
	[0.159]	[0.533]		[0.226]	[0.660]	
Number of observations	6582	2220		4103	1680	
Log pseudo likelihood	-1840.75	-210.03		-763.9	-109.88	
Pseudo R-squared	0.0342	0.1609		0.0453	0.1261	

Robust standard errors in brackets; Pseudo R-squareds reported are the McFadden's psuedo R-squareds; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis is a t-test of the coefficient being equal to zero.

paribus and on average, once again conflicting the negative socio-economic status and health link. Heads of households in the illiterate category in urban areas are 10 percentage points lower than the primary school category (see Table 3.8, p.57). The relatively larger number of potential blood meals in the latter category may be responsible for driving the positive effect of primary school level on malaria relative to illiterate heads.

The results for occupation are consistent with the negative socio-economic status and health relationship. In 1998-99 (though not in 1992-93), working in the agricultural sector in urban areas had a significant positive effect on contracting malaria in urban areas (two percentage points more likely compared to the base group of production and transport – see Appendix H). The effect of those heads of households who are wage earners is significant and negative relative to the base for 1998-99 but is not significant for 1992-93. The heads of households who belong to this category are 0.6 of a percentage point less likely to get malaria as compared to those who work in production and transportation, *ceteris paribus* and on average. Wage earners are likely to be better-off and may proxy for higher income groups. At the same time, they may also be better-off than those in production and transportation because the latter are likely to have greater exposure to malaria by travelling to areas that may have a high incidence of the disease and, thus, act as both active and passive carriers of malaria. These effects are not found to be significant for the rural samples.

A household head who lived in a *kachcha* house in 1992-93, relative to one who resided in a semi-*pucca* house in urban areas, was more likely to contract malaria, as one would expect. This result can be explained by the fact that a *kachcha* house may have eaves and other open spaces to permit entry of mosquitoes as compared to a semi-*pucca* house. However, this effect is not significant in 1998-99 or for either of the rural samples.

Having a separate kitchen in the house has a well determined negative effect on malaria incidence in both time periods in the urban settlement type. The impact effect indicates that a household head owning a separate kitchen in urban areas is 1.5 percentage points less likely to contract malaria in 1992-93 and 0.9 percentage points in 1998-99, on average and *ceteris paribus*, relative to one who does not have a house with a separate kitchen (see Appendix H). This result is possibly because of the potential collection of water in the kitchen coupled by the fact that since most urban households do not use firewood or dung, there is no deterrent for mosquitoes caused by smoke produced by these fuels as is the case in rural areas. In the rural probit model the effect is positive and significant in 1992-93, a plausible explanation for which may be the use of smoky fuels in rural areas that may deter mosquitoes as compared to those households that have a separate kitchen. This result is robust over time.

The coefficient on 'using wood as main fuel source', is positive and significant for 1992-93 but not 1998-99. The effect for 1992-93 is perhaps attributable to the higher percentage of households that use wood as fuel (at 37 percent) relative to dung (at six percent) in urban areas (see Table 3.8, p.57). On the other hand, in rural areas the coefficient on wood is negative in 1992-93 and positive in 1998-99. At the same time, the estimate for the miscellaneous category of fuel in 1998-99 yields a significant negative effect. These results indicate that a household head using a superior fuel (wood or miscellaneous compared to dung) is less likely to contract malaria. This result is the same as that obtained for the earlier rural specification in the previous chapter.

The coefficient on sanitation is significant with a positive sign in 1992-93 (but not in 1998-99) in urban areas—those household heads who have access to sanitation (a toilet) are more likely to contract malaria, *ceteris paribus* and on average. Even though this is in conflict with a negative socio-economic status and health relationship, it is plausible given that the urban vector hides in dark humid places and has been found to hide in public lavatories as discussed in Chapter 2. A household head with access to sanitation is one percentage point more likely to contract malaria relative to one that does not have access to sanitation, *ceteris paribus* and on average.

Drinking water source, unlike rural areas, does not register well determined effects in either of the two time periods for the urban samples. Among regions, heads of households living in eastern Uttar Pradesh and the Hill area are less likely to have suffered from malaria relative to the Central region, *ceteris paribus* and on average, in 1992-93 and in the Hill and West regions in 1998-99. This result has to be treated with caution because a large number of districts were dropped from the final model specifications.

In conclusion, we find that the socio-economic factors differ over the two time periods and there is very little evidence of robustness. The negative socio-economic status and health relationship holds true for a separate kitchen in both time periods. It is also supported for fuel and type of house though only in one time period.

Table 6.3: Proportion of rural and urban households owning different non-agricultural consumer durable assets in 1992-93

Variable	*Rural*	*Urban*	*t-ratio*
HH[a] owns a radio	0.254	0.554	-26.00
HH owns a TV	0.055	0.548	-52.33
HH owns a bicycle	0.505	0.586	-6.61
HH owns a sewing machine	0.130	0.563	-41.36
HH owns a clock or a watch	0.424	0.788	-29.68
HH owns a fan	0.081	0.700	-59.30
HH owns a sofa set	0.024	0.289	-37.58
HH owns a refrigerator	0.005	0.241	-39.14
HH owns a motorcycle	0.022	0.222	-31.45
HH owns a car	0.001	0.032	-13.39
HH owns a VCR or a VCP	0.009	0.072	-16.48

a. HH: Household

6.1.2 Impact of wealth on malaria incidence

The effect of the non-agricultural consumer durable asset index is negative and significant for the urban regression in 1992-93, unlike rural areas. Assets seem to reflect wealth status much better in urban areas as compared to the rural settlement type perhaps because of a more diverse set and higher value assets owned in urban areas including such assets as TVs and cars as we can see from Table 6.3. The t-test decisively rejects the null of no difference between rural and urban means for the different assets at the five percent level of significance. It could be that in rural areas the ownership of land substitutes for the ownership of non-agricultural consumer durable wealth assets. However, in the rural malaria specification the estimated effects of land variables may or may not be well determined—in the 1992-93 specification (Table 5.1, p.112) that uses a disaggregated land variable, only the category owning three acres or more irrigated area had a negative impact on malaria and the estimated effects for the other categories were not well determined. The marginal effect indicates that a one standard deviation increase in the index leads to a 0.7 percentage point reduction in malaria incidence. However, the estimates for the quintiles are not significant. In 1998-99, the coefficient on asset index is not well determined but the estimates for the different quintile categories show that the poorest and the 60-80 percent wealth groups are more likely to contract malaria. The 40 to 60 percent quintile, on the other hand, yields a negative and significant coefficient implying that this group is less likely to contract malaria. The result for the 60-80 percent group is in conflict with expectations. However, as noted in Chapter 2, even the wealthiest are not spared from the possible breeding of mosquitoes in places like air coolers and air conditioning equipment or in pots or ponds found in urban gardens. We can conclude that malaria spreads across all groups but this result must be treated with caution given that in specifying our model we lost a number of observations. However, this may not be an issue if the observations lost are random, which appears to be the case here when we observe the summary statistics for the two distributions.[1]

Yet another observation from the results here relates to the non-monotonicity of the health-wealth (income) gradient, as observed in the rural analysis. In Table 6.1 for 1998-99, while the coefficient on the asset index is positive and significant, it is negative for the 40-60 percent category and positive again for the top most wealth group. One might expect a concave relationship between health and income as noted by Deaton (2002, 2003), so that income has a larger effect on health among the poor as compared to the rich forming the basis for the argument that income redistribution would then improve health outcomes of the poor. However, our finding here indicates that inequaltiy per se may not play a signifcant role in determining health outcomes in the context of malaria given the negative coefficients on the asset index for the rich as well as the poor wealth segments. Malaria does not seem to discriminate between the rich and the poor in urban Uttar Pradesh. This is in line with Deaton's (2003) finding. While arguing that inequality (of race or other social inequalites as well as income) is inherently undesirable and may indirectly affect heath outcomes, Deaton finds that there is no direct link between income inequality and

[1] With the larger sample size, the percentage of households in each quintile (starting with the bottom 20 percent) is 71, 39, 20, 19 and 5. With the final specification used here the corresponding percentages are 72, 37, 19, 20, and 5.

ill health. To quote, "all else equal, individuals are no more likely to be sick or to die if they live in places or in periods where income inequality is higher." This conclusion persuades us to look at other socioeconomic factors that might affect malaria incidence directly and where policies to counter malaria may be directed.

We now turn to an analysis of the rural and urban decomposition to assess what contributes to the difference in malaria incidence across the two settlement types.

6.2 Decomposition analysis

The decomposition analysis aims to capture whether the difference in rural and urban malaria incidence is attributable to differences in treatment or endowments, also referred to as the unexplained and explained differentials, respectively. The malaria incidence difference between the rural and urban settlement types can be decomposed into differences between coefficients (treatment effects) and characteristics (endowments):

$$\Delta = [\overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_R)}] + [\overline{\Phi(X_U\beta_R)} - \overline{\Phi(X_U\beta_U)}] \tag{6.1}$$

In the right hand side of the above equation, the term in the first square bracket represents the difference in endowments between rural and urban areas (with estimated coefficients fixed at their rural values) and the term in the second square bracket is the difference in malaria attributable to the differences in treatment effects (using urban characteristics), which indicates a shift in the relationship between the rural and urban settlement types. Estimated rural coefficients have been used above to obtain the characteristic differences and the average urban characteristics are used as the weights for arriving at the coefficient differences. Alternatively, estimated urban coefficients and rural characteristics could also be used as weights to get the respective characteristic and coefficient differences. The final estimates in this index number approach are likely to differ depending on which set of weights are used to capture the component parts. Aside from the above decomposition at the aggregate level, the contribution of each variable to the difference in malaria can also be worked out by assigning weights to each variable (see Chapter 4, p.99).

The final specification used for the decomposition analysis, compatible across the two settlement types, included the variables caste and gender; age, occupation, and education categories; type of house; whether the household has a separate kitchen or not; different fuel types and drinking water sources; whether the household has access to sanitation and electricity; whether it owns agricultural land; non-agricultural consumer durable asset wealth index, and dummies for regions in Uttar Pradesh. In 1992-93, for the rural specification, the McFadden R-squared is 0.0342. For the earlier rural specification used in the previous chapter the measure was higher at 0.0944. For the 1998-99 rural specification employed in this chapter, the measure is 0.0442, lower than that for the specification used in the previous chapter, which was 0.1009. Thus, the comparable rural specifications appear to have lower goodness of fit. The comparable urban specifications for 1992-93 and 1998-99 have McFadden R-squareds of 0.1609 and 0.1265, respectively.

As mentioned in Chapter 3, while for the rural analysis the usable sample consisted of 7,287 households, the sample used to compare rural to urban comprised of 6,582 households after making the changes necessary to match variables in the two samples (as discussed in the previous section). The urban sample, on the other hand, comprised of 2,220 households in 1992-93. The changes covered only five of the 14 groups of variables. In 1998-99, the initial rural sample consisted of 4,588 households and the sample used in the decomposition analysis, after matching it with the urban settlement type, had 4,103 households. The urban sample consisted of 1,680 households

Table 6.4: Aggregate decomposition of rural and urban malaria difference into differences in characteristics and coefficients

	Rural coefficients and urban characteristics		Urban coefficients and rural characteristics	
	1992-93	1998-99	1992-93	1998-99
Characteristics difference	0.0311	0.0406	0.0048	0.0350
	(51.32%)	(118%)	(7.89%)	(119%)
Coefficients difference	0.0295	-0.0062	0.0558	-0.00057
	(48.68%)	(-18%)	(92.11%)	(-19%)
Aggregate difference	0.0606	0.0344	0.0606	0.0344
	(100%)	(100%)	(100%)	(100%)

Table 6.4 presents the contributions of the (aggregate) characteristic and coefficient effects to the differential in malaria incidence between rural and urban areas using rural coefficients and urban characteristics as well as urban coefficients and rural characteristics. The observed and predicted malaria difference is 6.06 percentage points in 1992-93[2]

[2]The observed rural malaria incidence is 0.08447 and the corresponding urban malaria incidence is 0.02387. Using predicted probabilities, the rural malaria incidence is 0.08446 and the urban incidence is 0.02389 in 1992-93.

and 3.4 percentage points in 1998-99. Using rural coefficients the characteristic effect is 3.1 percentage points in 1992-93, contributing 51 percent to the rural-urban malaria incidence gap. The remaining 49 percent is contributed by the coefficient effect of 2.9 percentage points. In 1998-99, the coefficient effect contributes to all the difference in malaria incidence (see Table 6.4). When urban coefficients are used, the contribution of the characteristic effect in 1992-93 is only eight percent and in 1998-99 the coefficient effect continues to contribute to all the difference. Using mean characteristic vectors (used in the decomposition as discussed in the methodology) the difference in malaria is slightly higher at 6.62 percentage points in 1992-93 (see last rows of Tables 6.6 and 6.7). Neither the characteristic nor the coefficient effect is significant for either the rural or the urban coefficient or characteristic weights though a number of individual variables yield significant effects and these are discussed in the subsequent subsection.

The method used to arrive at the asset index for this analysis is based on using the covariance matrix rather than standardising the original variables to a mean of zero and variance of one (see pp.65-66 of the data chapter), which also allows an easier interpretation of the results for the asset index in standard deviation units. However, it is not possible to use this method when dealing with the decomposition since the difference between the averages would be zero not making it possible to compute the characteristic and the coefficient effect for the asset index (see pages 100-101 of the methodology chapter). If we were to drop the asset index from the specification, other coefficients would be biased since the coefficient on the asset index is significant in 1992-93 for the urban sample. The means are presented in Table 6.5.

Table 6.5: Mean asset index values for rural and urban Uttar Pradesh samples using the covariance option

Year	*Rural*	*Urban*
1992-93	0.69	1.49
	(0.63)	(0.95)
1998-99	0.80	1.53
	(0.67)	(0.80)

Standard deviations reported in parentheses.

Table 6.6: Detailed decomposition of rural and urban malaria difference into differences in characteristics and coefficients for 1992-93

Variable	*Cr*	*Du*	*Cu*	*Dr*
Caste	-0.0006***	0.0016***	-0.0014	0.005***
	[0.00006]	[0.0004]	[0.0042]	[0.002]
Gender	-0.0002***	0.019	0.0001***	0.026
	[0.000004]	[0.049]	[0.0001]	[0.041]
Age (aggregate)	-0.0006***	-0.0028**	-0.0004	-0.003***
	[0.00003]	[0.002]	[0.0004]	[0.001]
Education (aggregate)	0.0055***	-0.0036***	0.0021	-0.007***
	[0.002]	[0.0009]	[0.0105]	[0.002]
Occupation (aggregate)	0.007***	-0.0067	-0.0008	-0.002
	[0.002]	[0.005]	[0.018]	[0.001]
House type (aggregate)	-0.0079	-0.0144	0.0009	-0.028
	[0.005]	[0.016]	[0.125]	[0.025]
Separate kitchen	-0.0031***	0.020	0.0038	0.016***
	[0.0006]	[0.020]	[0.014]	[0.006]
Sanitation	0.0145	-0.0299	-0.0114	-0.003***
	[0.016]	[0.056]	[0.147]	[0.0003]
Electricity	0.0196	-0.0092	0.0014	-0.003***
	[0.013]	[0.015]	[0.016]	[0.0007]
Fuel (aggregate)	-0.0068	-0.0355	0.0004	-0.055***
	[0.016]	[0.052]	[0.172]	[0.076]
Drinking water (aggregate)	0.0078***	-0.0004	0.0007	0.003***
	[0.002]	[0.001]	[0.013]	[0.001]
Agricultural land	-0.0014	-0.0024	0.0046	-0.016
	[0.0007]	[0.0008]	[0.047]	[0.015]
Asset index	-0.000003**	0.022	0.0066	0.014**
	[0.000001]	[0.037]	[0.040]	[0.007]
Region (aggregate)	-0.0061***	0.010	-0.0021	0.0144**
	[0.002]	[0.007]	[0.020]	[0.007]
Constant		0.0714		0.099
		[0.410]		[0.355]
Aggregate	0.0276	0.0387	0.0043	0.0620
	[0.048]	[0.059]	[0.068]	[0.059]

Standard errors for the characteristic and coefficient effects are reported in brackets; Cr: Characteristic effect using the rural coefficient vector; Du: Coefficient effect using urban characteristics; Cu: Characteristic effect using the urban coefficient vector; Dr: Coefficient effect using rural characteristics; Characteristic (or coefficient) effect of each variable is calculated as the product of the aggregate characteristic (or coefficient) effect and the weight assigned to each variable; Aggregate characteristic and coefficient effects reported are calculated using mean values of the characteristic vector and differ from those reported in Table 6.4.

Table 6.7: Detailed decomposition of rural and urban malaria difference into differences in characteristics and coefficients for 1998-99

Variable	*Cr*	*Du*	*Cu*	*Dr*
Caste	0.0001***	0.0018	-0.0031	0.0147
	[0.00001]	[0.016]	[0.002]	[0.056]
Gender	-0.0004***	-0.0128	0.0004***	-0.0618
	[0.00001]	[0.663]	[0.00004]	[0.853]
Age (aggregate)	0.0004***	-0.0015	0.0006***	-0.0074
	[0.00003]	[0.016]	[0.00008]	[0.023]
Education (aggregate)	-0.0023**	0.0038	0.0018	0.0102
	[0.0009]	[0.052]	[0.003]	[0.013]
Occupation (aggregate)	0.0027***	0.0017	0.0148	-0.0151
	[0.0006]	[0.030]	[0.022]	[0.116]
House type (aggregate)	0.0029	-0.0013	0.0005	-0.0016
	[0.0018]	[0.039]	[0.004]	[0.006]
Separate kitchen	0.0005***	0.0034	0.0045	0.0081
	[0.0001]	[0.068]	[0.004]	[0.022]
Sanitation	-0.0047	-0.0023	-0.0094	-0.0015
	[0.0042]	[0.092]	[0.031]	[0.002]
Electricity	0.0007	0.0036	0.0074	0.0040
	[0.0005]	[0.129]	[0.019]	[0.009]
Fuel (aggregate)	0.0133	-0.0018	0.0046	0.0076
	[0.0120]	[0.095]	[0.016]	[0.095]
Drinking water (aggregate)	0.0246	-0.0100	0.0025	-0.0062
	[0.0825]	[0.562]	[0.005]	[0.039]
Agricultural land	-0.0019	-0.0004	0.0014	-0.0081
	[0.0009]	[0.003]	[0.003]	[0.057]
Asset index	0.0035**	-0.0032	-0.0005	-0.008
	[0.002]	[0.079]	[0.0009]	[0.031]
Region (aggregate)	0.0003***	0.0079	-0.0015	0.0405
	[0.0001]	[0.078]	[0.003]	[0.108]
Constant		0.0054		0.0254
		[0.603]		[0.739]
Aggregate	0.0364	-0.0025	0.0244	0.0096
	[0.126]	[0.139]	[0.125]	[0.128]

Standard errors for the characteristic and coefficient effects are reported in brackets; Cr: Characteristic effect using the rural coefficient vector; Du: Coefficient effect using urban characteristics; Du: Coefficient effect using urban characteristics; Cu: Characteristic effect using the urban coefficient vector; Dr: Coefficient effect using rural characteristics Characteristic (or coefficient) effect of each variable is calculated as the product of the aggregate characteristic (or coefficient) effect and the weight assigned to each variable; Aggregate characteristic and coefficient effects reported are calculated using mean values of the characteristic vector and differ from those reported in Table 6.4.

6.2.1 Decomposition into component effects

The break-up into component effects is presented in Table 6.6 for 1992-93 and in Table 6.7 for 1998-99. We first discuss the results for 1992-93 followed by those for 1998-99. A positive effect indicates a widening of the malaria incidence gap between rural and urban areas. The largest contributor is the constant term, which could be an indication of inherent structural differences in rural and urban areas (bearing in mind the fact that this term may also capture a number of other factors not explicitly included in our specification). This may be a reflection of the fact that rural areas are considered to be more prone to malaria incidence as discussed earlier in Chapter 2. One major structural factor contributing to the difference would be the fact that rural areas are more conducive to mosquito breeding. Even though such variables as ownership of agricultural land may capture some of these structural differences, other factors like presence of possible pits in agricultural fields, around houses, and around water collection points, which serve as ideal breeding grounds for mosquitoes, are not likely to be captured in our specification.

1992-93

A number of characteristic estimates in 1992-93 attain significance at a conventional level when using rural coefficients. Differences in caste, gender, age, separate kitchen, and the asset index narrow the rural-urban malaria differential whereas characteristic differences in education, occupation, and water source contribute to widening the malaria incidence gap. Among the corresponding coefficient effects using urban characteristics, caste widens the malaria incidence gap whereas age and education narrow the gap. None of the characteristic effects are significant with the exception of gender when using urban coefficients though a much larger set of coefficient effects is significant when using rural characteristics—caste, separate kitchen, asset index, and region widen the gap whereas age, education, sanitation, and electricity narrow the malaria incidence gap. This reflects the standard index number problem where the results are sensitive to the weights used.

Using rural coefficients, we find that the average rural-urban difference in households belonging to an SC or ST narrows the malaria incidence gap by 0.06 of a percentage point. Due to the large difference in the proportion of households from lower castes in rural areas (22 percent) the malaria incidence is lower in the rural settlement type as compared to the urban areas (nine percent: see Table 3.7, p.55) but this difference in characteristics (that registers as a negative impact on malaria incidence in the rural probit) is countered-balanced by the coefficient effect or unexplained component of 1.6 percentage points, indicating that SC/STs are better-off in urban areas. Using the alternative set of weights – urban coefficients and rural characteristics – only the coefficient effect is significant and positive. We may conclude, therefore, that it is the unexplained component of caste that is driving the malaria difference and contributing to widening the malaria incidence gap between urban and rural areas.

Both the explained and the unexplained components for age contribute to narrowing the rural-urban malaria incidence gap using rural coefficients and urban characteristics. The characteristic effect indicates that the age composition of urban areas is such as to be more susceptible to contracting malaria. In rural areas there is a lower percentage in the younger age group of 38 to 47 years, a more mobile population group, which is more likely to transmit malaria. The treatment or coefficient effect of age group, as already mentioned, also contributes to narrowing the gap. Using urban coefficients and rural characteristics, only the coefficient effect is significant but negative.

A large proportion of urban households have a separate kitchen in the house (61 percent in urban and 35 percent in rural areas). The probit reveals a negative impact of having a separate kitchen on malaria incidence in urban areas and positive in rural areas, *ceteris paribus*. This registers as a narrowing of the malaria incidence gap as we can see from the negative and significant characteristic effect when using rural coefficients. The corresponding coefficient effect is not statistically significant. When using the alternate set of weights, that is, urban coefficients and rural characteristics, the differences in characteristics do not explain the malaria incidence gap whereas the coefficient effect is significant though positive.

Using rural coefficients the characteristic effect of the asset index is negative and significant, narrowing the malaria incidence gap although by a very small magnitude of 0.0003 of a percentage point. The Principal Components index value is much larger for urban as compared to rural areas and one would have expected a widening of the gap due to this difference in characteristics. This result is, therefore, in conflict with expectations. However, looking at the results using urban coefficients and rural characteristics, we find that the characteristic effect is not significant whereas the coefficient effect is positive and significant, thus contributing to widening the rural-urban malaria gap, as one would expect. This can be explained by the probit results where the effect of the asset index on malaria incidence is negative and significant for urban areas and insignificant for rural areas. We now discuss the results for education, occupation, and drinking water source.

The coefficient (or treatment) effect for education is negative and significant whether we use urban or rural characteristics, thus narrowing the malaria incidence gap between the two settlement types. This indicates that the response of household heads to education in rural areas (as compared to urban areas) is more favourable to reducing malaria incidence. However, when using rural coefficients the differences in education characteristics also make a contribution although in widening the malaria incidence gap. This effect of 0.5 of a percentage point outweighs the corresponding negative coefficient effect of nearly -0.4 of a percentage point. This is not surprising as the proportion of high school and middle school educated heads of households is lower in rural as compared to urban areas (see Table 3.7, p.55). The probit analysis shows that although not well determined these register a negative impact on malaria incidence and, thus, the urban negative effect seems to outweigh the rural effect contributing to a widening of the gap between rural and urban malaria incidence.

For occupation, the positive sign on the characteristic effect using rural coefficients indicates a net widening of the rural and urban malaria incidence gap. This result is not surprising and is a reflection of the distribution of occupations among heads of households in the two settlement types—rural areas have a concentration of heads working in the agricultural sector and in the urban settlement type a large percentage of heads of households work as wage earners (58 percent and 48 percent, respectively); the latter have a lower likelihood of contracting malaria, as discussed earlier.

The characteristic effect for the drinking water source category using rural coefficients is positive and significant. Thus, differences in rural and urban characteristics in drinking water source contribute to widening the rural-urban malaria incidence gap, as one would expect given that rural households source drinking water from public taps, rivers, ponds, and lakes, which are likely to be breeding grounds for mosquitoes. Using urban coefficients, the characteristic effect is not significant, but the corresponding treatment effect is positive and significant, also widening the malaria gap. Attention now turns to the 1998-99 decomposition results.

1998-99

As noted earlier from Table 6.4, the difference in malaria incidence between rural and urban areas in 1998-99 is contributed by differences in characteristics and not in treatment effects. This is clear when we observe the results for the component effects as well. Caste, age, occupation, asset index, separate kitchen, and region have statistically significant characteristic effects in 1998-99 and contribute to widening the malaria incidence gap between rural and urban areas when using rural coefficients. The characteristic effects for gender and education narrow the malaria incidence gap. None of the coefficient effects (using either rural or urban characteristics) are significant. Using urban coefficients, only gender and age have significant (positive) characteristic effects. The results for gender and occupation match with those found for 1992-93. However, for all other variables, the point estimates show a reversal in signs. Thus, while in 1998-99 differences in characteristics for caste, age, asset index, separate kitchen, and region contribute to widening the malaria incidence gap, in 1992-93 these narrowed the gap. Similarly, while differences in education in 1998-99 narrow the malaria incidence gap, in 1992-93, these widened the gap. In other words, there is little evidence of robustness over the two time periods.

There is a change in the proportion of SC/ST households in 1998-99 as compared to 1992-93 (see Table 3.7, p.55 and Table 3.9, p.59). While in 1992-93 only nine percent of the urban households belonged to this category, in 1998-99 it increased to 15 percent. The corresponding rural statistics are 22 percent for 1992-93 and 25 for 1998-99. At the same time, as we can see from the probit results, the coefficient on caste for the urban households is negative and significant in 1998-99 and the difference in rural and urban estimates is statistically significant. These factors seem to contribute to the positive though small characteristic effect of 0.01 of a percentage point in 1998-99, thus widening the rural-urban malaria incidence gap.

The result for the age characteristic effect has also switched signs as compared to 1992-93 even though the distribution by age groups does not seem to have changed much over the two time periods. For education, the differences in characteristics narrow the malaria incidence gap, which might be because even though the percentage of high school educated heads continues to be higher in urban areas, there is a smaller proportion of heads from the middle school educated category in urban areas and an even smaller fraction in the primary school educated category (see p.57). In 1992-93 in rural areas six percent of the heads had completed primary education, in urban Uttar Pradesh 15 percent fell in this group. In 1998-99 the percentage was nearly equal at 16 percent and 18 percent for rural and urban Uttar Pradesh, respectively.

Differences in characteristics in separate kitchen (using rural coefficients) contribute to widening the malaria incidence gap in 1998-99, in conflict with the 1992-93 result. This seems to be driven by the fact that while the coefficient on a separate kitchen is negative and significant in urban areas it is negative though not well determined in the rural settlement type in 1998-99.

Finally, the characteristic effect for asset index in 1998-99 widens the malaria gap by as much as 3.5 percentage points whereas in 1992-93 it narrowed the gap though by a very small amount of 0.03 of a percentage point. The widening of the gap agrees with our intuition since the mean asset index value for urban areas is higher than that for rural areas (1.53 versus 0.80: see Table 6.5).

6.3 Summary

The main aim of this chapter was to understand what factors contribute to differences in rural and urban malaria incidence by undertaking a decomposition analysis. We also compared urban malaria incidence over time, where we once again examined the impact of socio-economic factors and wealth on malaria incidence.

There is scant evidence of robustness over time, with the exception of one variable, separate kitchen. Aside from the fact that factors determining malaria may change over time, a plausible explanation for this lack of robustness could be that a large number of factors that drive urban malaria are not included in the specification. In Chapter 2, we made reference to the possible breeding of mosquitoes around construction sites and the association of malaria with other developmental activities. Data on these variables are not available.

The hypothesis suggesting that poor socio-economic status has a negative impact on malaria incidence is supported by the negative and significant caste coefficient (though only in 1998-99), the positive effect of a household head working in the agricultural sector in urban areas and the negative coefficient on wage earners, relative to the production and transport category in 1998-99. A household head living in a *kachcha* house is also more likely to contract malaria in urban areas. However, the result for sanitation, where the sign on the coefficient is positive, implying that presence of sanitation is favourable to contracting malaria, conflicts the negative socio-economic status and health relationship generally found in the literature.

In the analysis of the relationship between wealth and malaria incidence for urban Uttar Pradesh, we find that in 1992-93 the effect of the asset index is negative and significant though individually the quintiles do not have well determined effects. In 1998-99, even though the asset index does not bear a significant coefficient, when using splines, we find that the 60-80 percent and the poorest wealth groups are more likely to contract malaria. This result points to the spread of malaria across different wealth groups, not just being confined to the poorest, once again contradicting conventional wisdom.

The decomposition analysis warrants a cautious interpretation of the results since the explained differentials or characteristic effects are found to be sensitive to the coefficients used and the unexplained differentials or coefficient effects are sensitive to the characteristics used, which is a standard index number problem. Moreover, the regression specifications used have poor goodness of fit and may not produce informative findings. As we observed, the McFadden's R-squared for the rural specification adjusted to match the urban one is very low (0.0342 for 1992-93 and 0.0442 for 1998-99, much lower than the measures for corresponding specifications used in the earlier rural analysis chapter at 0.0944 and 0.1009, respectively). Further, in Chapter 2 we discussed that there are basic differences in the epidemiology of malaria in rural and urban locales so that using a common set of variables may not be appropriate.

The decomposition results do not show evidence of robustness over time. This lack of robustness makes it hard to obtain a clear answer to the research question of which factors contribute to the rural-urban malaria difference in Uttar Pradesh. For example, caste, age, separate kitchen, and asset index characteristic differences narrow the malaria incidence gap in 1992-93 but widen it in 1998-99. A clear conclusion that does emerge, however, is that in 1998-99, the difference in the malaria incidence gap is explained by differences in characteristics including the component effects, whereas in 1992-93 it is structural differences or differences in treatment effects that appear to drive the malaria difference. One important result with some policy implication emerges for drinking water source for 1992-93 (though in 1998-99 the coefficient and characteristic effects are not well determined). Differences in drinking water characteristics serve to widen the malaria differential using rural coefficients. Using rural characteristics, the treatment effects also contribute to widening the gap. These results are as anticipated since the drinking water infrastructure in urban areas is superior. Urban households depend mainly on private piped water supply unlike rural areas where they source their water needs from public taps and open sources, which can be ideal breeding grounds for mosquitoes because of the collection of stagnant water. An improvement in infrastructure by way of preventing the collection of stagnant water or in the long run moving to private piped water supply would help in narrowing the rural-urban malaria incidence gap.

Chapter 7

State level analysis of malaria incidence

In this chapter we address the research question: Does aggregate income have a negative impact on malaria incidence? How does it compare with the effect of public health expenditure? The basis of addressing the first part of this research question comes from the belief that poorer countries or regions are more likely to suffer from poor health. McCarthy et al. (2000), for example, found a negative association between high malaria morbidity and GDP growth per capita in their investigation of 187 malarial countries and Gwatkin and Guillot (2000) found three-fifths of global malaria deaths to occur amongst the poorest fifth of the world's population. The second part of the above research question draws on studies by Anand and Ravallion (1993), Anand and Kanbur (1995), and Haddad et al. (2003). These studies, discussed in Chapter 1, suggest that improvements in health outcomes are likely to be realised faster by investing directly in health rather than wait for the benefits of income growth to accrue.

The dependent variable used in the current analysis is malaria incidence, measured as the annual parasite incidence per 1000 population (**API**) at the state-level. As noted earlier in Chapter 3, while **API** estimates are based on examining actual blood smears for the malaria parasite, the malaria estimates used for the household and district analysis are based on reporting whether a head of household suffered from malaria in the last three months. The explanatory variables include real state domestic product per capita (**stinc**), real health expenditure per capita (**healthex**), real education expenditure per capita (**eduex**), rural and urban poverty headcount ratios (**povertyr** and **povertyu**), rate of urbanisation (**urban**), population density (**popden**), average annual rainfall (**rain**), and percentage of state area irrigated (**irrig**). These explanatory variables were drawn from the dataset prepared originally by Ozler et al. (1996) and is a panel of 15 states over the period 1958 to 1992, further updated by Besley and Burgess (2002), Besley and Burgess (2004) and Burgess and Pande (2005). The data on malaria incidence were not available in this compilation. These were collected for this study from various issues of the Health Information of India (GoI 1979-1985, GoI 1986-2001). These data, added to the updated Ozler data, were available from 1978, whereas the Ozler data are available till 2000. The resulting dataset used in this analysis is, therefore, a panel of 15 states spanning a period of 23 years from 1978 to 2000.

While the focus of the analysis here is on the relationship between (a) aggregate income with malaria incidence and (b) health expenditure with malaria incidence, as far as possible the specification was chosen to also enable a comparison with the district level analysis. Table 7.1 (repeated here from Chapter 3) presents the variables used in the state level analysis and indicates if these were also used in the district level analysis. The income relationship with malaria incidence could not be examined explicitly in the earlier chapters due to a lack of data. This chapter fills this gap by examining the relationship between a measure of aggregate income and malaria incidence. The summary statistics for the variables included in the regression are presented in Table 7.2, also repeated from Chapter 3.

The methodology used here was discussed in Chapter 4. To recapitulate briefly, the model can be written as:

$$y_{it} = x_{it}\beta + \alpha_i + \sum_{t=1}^{T-1} D_t\gamma + \epsilon_{it} \quad t = 1 \dots 23 \text{ and } i = 1 \dots 15 \tag{7.1}$$

where y_{it} is the incidence of malaria (**API**) for state i in year t; x_{it} is a $1\times K$ vector of characteristics affecting malaria incidence. This vector, x, consists of the following variables: **popden**, **urban**, **rain**, **irrig**, **stinc**, **eduex**, **healthex** with a one period lag or **healthlag**, **povertyr**, and **povertyu**. α_i represents unobserved heterogeneity or the unobserved effect varying across states but not over time, D_i are the set of time dummies and γ are the corresponding parameters for the time dummies. It may be argued that both state income and **povertyr** and **povertyu** should not be included in the specification

Table 7.1: Explanatory variables used in the state level analysis compared with district variables

Variables used in district level analysis	*Variables used in state level analysis*	*Variable abbreviations for state level analysis*
Variables with exact match at the two levels		
Population density by district (persons per square km)	Population density by state (persons per square km)	**popden**
Average annual rainfall by district (mm)	Average annual rainfall by state (mm)	**rain**
Percentage district area irrigated by different sources	Percentage state area irrigated (all sources)	**irrig**
Variables that did not have an exact match at the two levels but used alternative measure		
Percentage villages with adult literacy centres		**eduex**
Percentage villages with high schools	Real per capita education expenditure by state	
Average distance from drinking water source	Real per capita health expenditure by state (includes expenditure on sanitation and water supply)	**healthex**
Variables that could not be matched but additional state variables could capture some of these effects		
Percentage villages with commuting facility	Real net state domestic product per capita	**stinc**
Percentage villages in district with power supply	Urbanisation	**urban**
	Poverty head count ratio (rural)	**povertyr**
	Poverty head count ratio (urban)	**povertyu**

All data above obtained from Ozler et al. (1996) updated by Besley and Burgess (2004) and Burgess and Pande (2005) with the exception of data on average annual rainfall and percentage irrigated area, which were obtained from various issues of the Statistical Abstract of India (GoI 1979-2001).

since they are highly correlated—the correlation coefficient between **povertyu** and **stinc** is -0.65 and between **povertyr** and **stinc** is -0.64. However, both income and poverty measures are included because the former is a broad measure and does not indicate how malaria incidence may be affected by poverty *per se.* Moreover, in the current analysis, we found that state income was robust to the inclusion of the poverty measures (see Columns 1 to 4 in Table 7.4). Poverty gap index and the squared poverty gap measures were also used instead of and along with the poverty headcount ratios in preliminary regressions but these did not have significant effects and were dropped from the final specification. A one period lag of real health per capita expenditure is used rather than the contemporaneous measure based on the argument that the effect of health expenditure on particular health outcomes is likely to be realized one period later.[1]

[1] The regression was also run with the contemporaneous health measure. The sign on health expenditure

Table 7.2: Summary statistics of variables used in the state level panel data analysis

Variable[a]	Obs.[b]	Mean	SD[c]	Min	Max
API (per 1000 population)	345	3.84	5.45	0.02	59.44
popden (sq km)	345	345.76	189.81	91.40	886.73
rain (mm)	345	194.22	113.87	30.27	526.76
irrig (percentage)	345	21.22	19.14	2.17	79.50
stinc (Rs per capita per year)	345	1468.25	593.94	584.35	3834.06
healthex (Rs per capita per year)	345	16.22	3.93	5.69	29.54
eduex (Rs per capita per year)	345	53.47	19.95	18.61	118.58
urban (percentage)	345	24.51	8.17	9.53	42.69
povertyr (percentage)	345	42.01	13.43	3.16	78.78
povertyu (percentage)	345	34.53	11.80	6.55	59.75

a. See variable descriptions in Table 7.1; b. Obs. indicates number of observations; c. SD indicates standard deviation.

7.1 Results

This section consists of two parts – the first one deals with model specification issues and the second one with the empirical findings.

7.1.1 Model specification issues

Three model specification issues have been addressed here:[2] (1) choosing the appropriate functional form (whether to use a log-levels model or a log-log model); (2) choosing the appropriate estimator – fixed effects, pooled OLS or random effects; (3) testing whether the variables **stinc**, **eduex**, and **healthlag** are endogenous or not (since endogeneity can cause biased and inconsistent estimates). The Hausman test has been used to decide whether to use the fixed effects or the random effects estimator. The variables **stinc**, **healthex**, and **eduex** are likely to impact on malaria incidence as well as be influenced by changes in malaria incidence rates—the decline in malaria incidence would be expected to bring with it lower expenditure on public health and possibly that part of public education expenditure aimed at raising awareness regarding malaria spread and prophylaxis. Similarly, decline in malaria incidence can be expected to raise income levels by improving productivity.

The log-levels functional form was chosen for the analysis, rather than the log-log model, since it had the highest R-squared.[3] The fixed effects model was preferred over

was positive though insignificant.

[2]See Chapter 4, pp.102, for details.

[3]Under the pooled OLS model, the log-log model (Model 2) had an R-squared of 0.54 and the log-levels model (Model 1) had an R-squared of 0.48. Under the fixed effects model the within R-squared for the

the pooled OLS model since the latter involves the restrictive assumption of treating unobserved effects as homogeneous across all states and over time. The results of the pooled OLS are, however, reported for comparison. Further, the fixed effects model was chosen over the random effects model as on the basis of the Hausman test to determine whether to use a fixed or a random effects estimator, the difference between the variance covariance matrix was negative definite rendering the test invalid. The most circumspect approach would, thus, be to use the fixed effects rather than the random effects model, as confidence in using the latter cannot be obtained.

Table 7.3: Validity, orthogonality and exogeneity tests for instrumental variables for per capita real health expenditure, per capita real net state domestic product and per capita real education expenditure

	First stage test for validity of instruments for potentially endogenous variables[a]: Shea Partial R^2 (Partial R^2)[b]		
	stinc	**eduex**	**healthlag**
OLS	0.5041(0.5405)	0.3627(0.5375)	0.5106(0.3425)
FE	0.2717(0.2962)	0.1386(0.1916)	0.2703(0.4047)
	Overidentification test for all instruments: Hansen J statistic		
	χ^2 value	Degrees of freedom	P value
OLS	3.672	2	0.1594
FE	3.432	2	0.1798
	Overidentification test for all instruments: Sargan statistic		
	χ^2 value	Degrees of freedom	P value
OLS	3.321	2	0.1901
FE	3.837	2	0.1468
	Exogeneity test[c]		
OLS	χ^2_2 = 0.31 (P> χ^2=0.8556)=> cannot reject null of exogenenity		
FE	F(2,293)=1.16(P>F=0.3145)=> cannot reject null of exogeneity		

a. Instruments used are one and two period lags of **stinc**, one-period lag of **eduex** and two and three period lags of **healthex**. b. Bound et al. partial R^2 measure is reported in parentheses. c. The values reported are from the test of joint significance of the residuals from a regression of each potentially endogenous variable on all other regressors included in a regression of malaria incidence on the endogenous and exogenous regressors and the residuals. Stata reports the χ^2 value for the OLS regression and the F value for the fixed effects model.

With reference to the issue of endogeneity, Table 7.3 reports the validity, orthogonality, and exogeneity tests for the three potentially endogenous variables, **stinc**, **eduex** and **healthlag**. Instruments used are one and two period lags of **stinc**, one-period lag of **eduex** and two and three period lags of **healthex**. The difference between the Bound et al. partial R^2 measure reported in parentheses and the Shea Partial R^2 measure is not large indicating that the instruments pass the relevance test especially for the fixed effects

log-log model was 0.25 and for the log-levels model it was 0.42. The Davidson-MacKinnon test as well as the Mizon and Richard tests were run for these competing functional forms but these did not help discriminate the models.

model where the difference between the two measures is even smaller. The orthogonality condition is also satisfied since the test values are less than the critical value for both the Hansen J as well as the Sargan test statistic (see p.107). Finally, the exogeneity test also indicates that we cannot reject the null of exogeneity.

7.1.2 Empirical findings

The main findings from this chapter relate to the effect of aggregate income and health expenditure on malaria incidence. The results pertaining to the other variables, the time effects, and the fixed effects are also discussed. We begin with a discussion of the latter and then focus on the income and health effects, followed by the other variables. The results are presented in Table 7.4, which reports the fixed effects (first four columns) and pooled OLS regressions (fifth column).

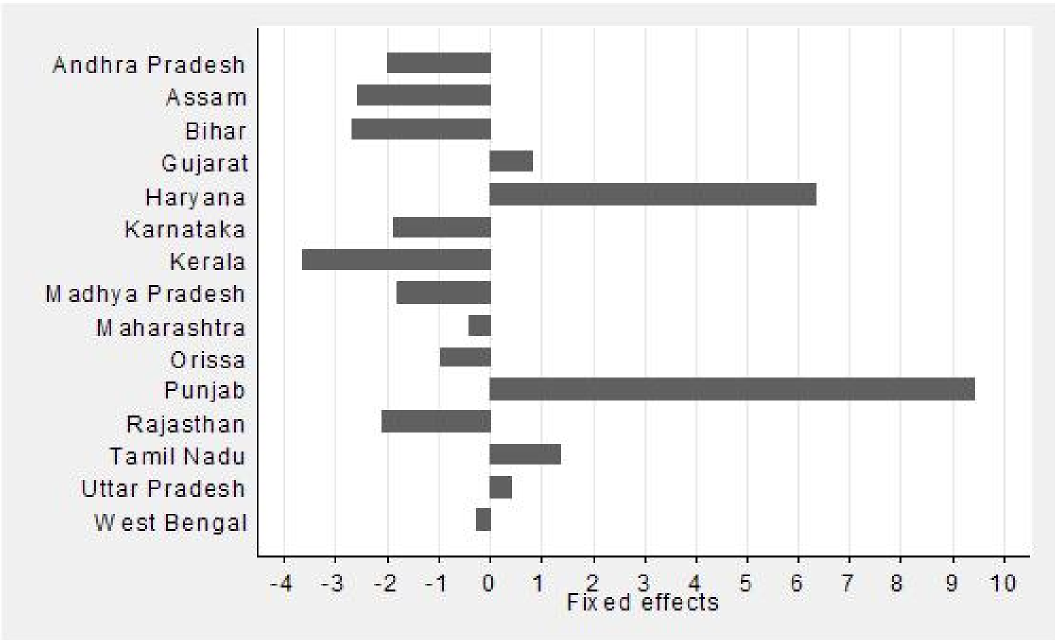


Figure 7.1: State fixed effects

The within R-squared for the fixed effects model (Column 4) is 0.42 indicating that 42 percent of the variation in the demeaned data is explained by the included regressors. The correlation between the fixed effects and the explanatory variables is quite large at -0.95. This could mean that the omitted heterogeneity effects are highly correlated with some of the variables (such as population density and urbanisation which are negative and significant under the pooled OLS but not under the fixed effects model). Fixed effects may absorb some of the variation that was attributable to the correlated variables.

Figure 7.1 plots the fixed effects for each state. The fixed effect is the impact on malaria incidence of factors specific to that state, given other explanatory variables we have included in the specification. The state fixed effects are the largest for Punjab and Haryana. The mean **API** in Punjab for the period 1979 to 2000 was 5.49 per 1000 population with very high (above average) incidence of 20.31 in the year 1979 and above nine over the years 1980 to 1986. From 1988 until 2000 the **API** was much lower and ranged between 0.02 and 1.74. The excessively high malaria incidence rates in the early years strongly impact the magnitude of the estimated fixed effect, which is nothing more than an average for the entire period. For Haryana the incidence rates are similar. The high values attained only for a few years would also explain the high standard deviations for these two states (6.05 for Punjab and 9.26 for Haryana).

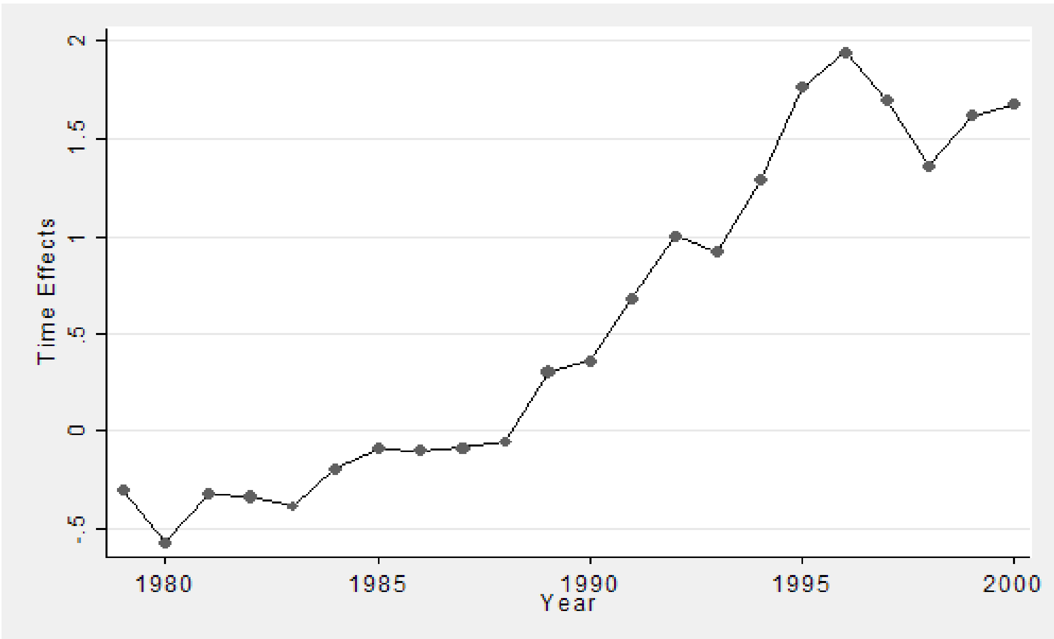


Figure 7.2: Time effects

Figure 7.2 plots the coefficients on the time dummies (from 1980 to 2000 with 1979 as the base year). There is an upward trend in malaria incidence. The time effects potentially capture the effect of factors that change over time and have not been captured in the other variables included in the regression model. The time effects are positive and significant from 1990 till 2000 indicating that there is an increase in malaria incidence (relative to the base 1979) over this period, after controlling for other factors. The effects are insignificant (though positive except for 1980 and 1983) for earlier years – individually as well as jointly.[4] This is consistent with the literature reviewed on the subject that malaria cases were on the rise after the 1970s, reached a peak in 1996, declined till 1998, and once again showed a rise in 2000. The all-India upward trend till 1998 is likely to be driven by high malaria incidence in such states as Madhya Pradesh and Orissa where the **API** is above 6 and 12, respectively, after 1995 (see Appendix K). We now turn to discussing the research question set out for this chapter relating to the effect of income and public health expenditure on malaria incidence.

7.1.3 Impact of aggregate income and health expenditure on malaria incidence

The effect of income on malaria is robust to the inclusion of public expenditure (mainly health and education expenditure) and poverty head count ratios.[5] These variables were introduced one at a time in the regressions starting with the base regression in the first column of Table 7.4, which includes the variables **popden**, **urban**, **rain**, **irrigation**, and **stinc**. The regression in the second column includes **healthlag**. In the third column it also includes **eduex** and finally the fourth column reports the regression with all the above variables and **povertyr** and **povertyu**. The estimated effect of **stinc** is the same for the different variants of this model with an estimated coefficient between -0.0012 and -0.0013 for the four models. In all cases it is significant at the one percent level. The signs and significance on the coefficients of other variables also remain unchanged.

The coefficient on **stinc** (Column 4, Table 7.4) indicates that an increase of a rupee per person in **stinc** would reduce **API** by 0.1 percent.[6] On the other hand, the coefficient of -0.0399 on **healthlag** translates as a four percent fall in **API** for an increase of a rupee on real health expenditure. Clearly, the effect of real income on malaria incidence, though significant, is very small especially as compared to the effect of an equivalent increase in

[4] For the 10-year period F(10,285)=0.76 and for 21 years and F(21,285)=3.50 (significant at the one percent level) for the overall significance of time effects.

[5] As noted earlier in Chapters 5 and 6, the relationship between malaria incidence and income (wealth) is non-monotonic, but in the state-level analysis undertaken here we have assumed a monotonic relationship: an increase in income by one rupee at the bottom end would be the same as a rupee increase at the top end of the distribution. Admittedly, this is a constraint in the panel analysis.

[6] A log-log model was also estimated to match the log-linear specification reported in Column 4 of Table 7.4 and is reported in Table L.1 in Appendix L. The coefficient on the log of health lag was found to be insignificant. The income elasticity is -1.052 with a standard error of 0.56 and the health elasticity is -0.049 with a standard error of 0.35, statistically insignifcant from zero. It may be noted that the log-linear model was chosen for detailed analysis, as mentioned earlier in this chapter, because the within R-squared for the log-linear model is 0.42, higher than the one for the log-log model, which is 0.25. Is is possible to compare these R-squareds because the dependent variable is the same for both the models.

Table 7.4: Effect of income and other variables on malaria incidence (log(**API**) = dependent variable)

Variable	(1) FE[a]	(2) FE	(3) FE	(4) FE	(5) OLS[a]
popden	-0.0008	-0.0017	-0.0018	-0.0027	-0.0045***
	[0.0025]	[0.0026]	[0.0027]	[0.0026]	[0.0004]
urban	-0.0553	-0.0479	-0.0473	-0.0656	-0.0259**
	[0.0468]	[0.0495]	[0.0498]	[0.0496]	[0.0123]
rain	0.0003***	0.0003***	0.0003***	0.0003***	0.00012
	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0004]
irrig	-0.1503***	-0.1525***	-0.1537***	-0.1383***	0.0041
	[0.0261]	[0.0277]	[0.0298]	[0.0295]	[0.0051]
stinc	-0.0013***	-0.0013***	-0.0013***	-0.0014***	-0.0001
	[0.0004]	[0.0004]	[0.0004]	[0.0004]	[0.0003]
healthlag		-0.0392*	-0.0392	-0.0399*	0.0251
		[0.0237]	[0.0239]	[0.0230]	[0.0275]
eduex			-0.0021	-0.006	0.0080
			[0.0106]	[0.0100]	[0.0092]
povertyr				-0.0252***	-0.0141*
				[0.0087]	[0.0075]
povertyu				0.0051	0.0070
				[0.0094]	[0.0069]
Constant	5.9322***	6.6058***	6.6835***	8.3206***	2.8200***
	[1.1015]	[1.2904]	[1.4138]	[1.6922]	[0.8986]
Observations	345	330	330	330	330
Number of states	15	15	15	15	
R-squared[b]	0.42	0.40	0.40	0.42	0.47
Correlation[c]	-0.95	-0.96	-0.96	-0.95	
Heteroscedasticity tests[d]	631.05	530.52	552.78	276.67	8.48

All the above regressions include time dummies as per Equation 7.1 from 1979 to 2000 using 1979 as the base for Models 2 to 5. Model 5 uses time dummies from 1978 to 2000 with 1978 as base; a. FE indicates fixed effects and OLS indicates pooled OLS; b. Within R-squareds reported for the FE regressions; c. Correlation indicates the correlation between the fixed effects and the explanatory variables; d. The null hypothesis of a homoscedastic error process is rejected in all the models. The first four columns for the fixed effects models report the modified Wald statistic for groupwise heteroscedasticity in residuals distributed as χ^2_{15} and the last column for the pooled OLS model reports the Breusch-Pagan test for heteroscedasticity distributed as χ^2_1; Robust standard errors reported in brackets; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis is a t-test of the coefficient being equal to zero.

real health expenditure. The coefficient on **healthlag** is larger than that on **stinc** by a factor of 28. Thus, if the government decides not to invest the one rupee per capita on (real) health expenditure, to bring about the same fall in malaria incidence through income growth would need an equivalent increase of Rs 28 in real income per capita.[7] This result is very similar to the Rs 33 trade-off obtained by Anand and Kanbur (1995) for their study on Sri Lanka. As they point out, large and sustained increases of income over a long period are likely to have a positive impact on health outcomes, but in the short-run income growth is not likely to achieve the desired effect on health outcomes and direct public health expenditure would be more effective in meeting these goals. The view that the benefits of income growth accrue slowly is supported by Haddad et al. (2003), who find that income growth is not sufficient to bring down malnutrition rates by half (the target set out in the Millennium Development Goals) in developing countries over a twenty five year period, even if the growth rate is as high as 2.5 percent barring the case of three (of 12) developing countries in their study.

Thus, our findings relating to the impact of income on malaria support the above views that income growth may not be sufficient to bring about the desired improvement in health outcomes, in our case a decline in malaria incidence. Direct expenditure on health succeeds in bringing down malaria incidence far more than an increase in income. We now turn attention to the effects of other variables employed in the regression models.

7.1.4 Impact of other variables on malaria incidence

Based on the district level evidence and on the literature on the subject, **popden** and **rain** may be expected to be negatively associated with malaria incidence, whereas **irrig** may be expected to be positively associated. The variable **urban** would be expected to be negatively associated with malaria incidence and so would **eduex**. In the first instance, it may seem that poverty measures should also be associated with higher malaria incidence because of the literature on the broader subject of income and health (and not specifically on malaria) that talks of poor health being positively associated with poverty. But, based on household level evidence and the specific nature of malaria poverty may be expected to have a negative association with malaria. This follows from the observation made

[7] Although per capita income is not an explicit policy instrument, its evolution can be affected by policy changes in regard to, for example, trade openness and privatization. Thus, any policies that impact on economic growth through these channels are likely to indirectly impact health.

by Gallup and Sachs (2001) that being constantly subjected to malaria, many ethnic groups tend to develop a resistance to the disease over time in the form of the sickle cell abnormality. In India this trait is generally found in lower caste households and among the scheduled tribes, who are also likely to have lower living standards (see, for example, Kijima, 2006). Further, we found at the household level that in rural areas a household head who owns a water pump is more likely to contract malaria and it is mostly the relatively wealthy households that own water pumps. Given these factors, there may be a negative association between poverty and malaria incidence.

As we can see from Column 5 in Table 7.4, **popden** has a negative and significant effect on malaria incidence under the pooled OLS model (an increase in population density by 10 persons per square kilometer reduces malaria incidence by 4.5 percent), whereas in the fixed effects model (Column 4) the coefficient on population density is not significant. At the district level, under the single equation model for rural Uttar Pradesh, an increase in population density by one person per square kilometer reduces the likelihood of a head of household contracting malaria by 0.1 of a percentage point (see Table 5.6, p.123). The urbanization measure has a negative and significant effect in the pooled OLS. The coefficient of -0.023 on **urban** in the OLS model translates as a fall in **API** of 2.3 percent for a one percentage point increase in urbanization. The negative association obtained here is consonant with both the literature reviewed as well as the results of the analysis comparing rural and urban malaria incidence where we found that the prevalence of malaria is much lower in urban areas. However, under the FE model, while the coefficient is negative, it is not statistically significant at any conventional level of significance. It is possible that such factors as the level of infrastructure development that was not captured at the district level gets absorbed within the fixed effects. In some states poor infrastructure development (such as bad roads or poor canal systems may increase the formation of stagnant water pools and hence the breeding of mosquitoes) could attenuate the negative effect of population density and urbanisation. As mentioned above the correlation between the fixed effects and the explanatory variables is very high (-0.95). This indicates the presence of such factors that may be correlated with these variables. Using a fixed effects model possibly absorbs some of the variation that might be more legitimately attributable to these variables so that the coefficient is poorly determined.

The coefficient on **rain** is not significant under the pooled OLS model and is positive and significant under the FE model (a 100 mm increase in average annual rainfall can increase **API** by as much as three percent, *ceteris paribus* and on average). At the district level, the coefficient was negative and significant, which could be on account of rainfall patterns prevalent across districts in Uttar Pradesh in those years. As noted in the literature on the subject, more rainfall can flush mosquito eggs and less rainfall can encourage breeding. At the same time, heavy rains that are followed by formation of stagnant water pools can encourage breeding of mosquitoes. The state level regressions also incorporate the time effects for different states and the sign on the coefficient indicates that *ceteris paribus* and on average over time across states for the period under consideration more rainfall is associated with higher malaria incidence.

The result for **irrig** is in conflict with the result at the district level under the fixed effects model and is not significant when we use pooled OLS. At the district level the effect of percentage irrigated area on the likelihood of contracting malaria is positive and significant, whereas at the state level, the fixed effects model shows a negative and significant effect: an increase of one percentage point in area irrigated can reduce **API** by 14 percent. Possible effects of technology (such as improvements in the construction of canals or other irrigation systems that prevent breeding of mosquitoes) that are not included at the district level are incorporated at the state level in the time dummies and may account for the negative association.

The variable **eduex** does not have a significant effect, contrary to the negative and significant effects obtained specifically for percentage of villages with adult literacy centres and high schools at the district level in the 1992-93 specification. The difference in the estimates could be attributed to the difference in the measure of education—expenditure at the state level and percentage of villages with adult education centres and high schools at the district level.

The coefficient on **povertyr** is negative and significant regardless of the model being used. In the FE model the coefficient on **povertyr** is -0.025 translating into a decline in **API** of 2.5 percent for a one percentage point increase in **povertyr**. This result is not unexpected given some of the results for the household analysis as well as the literature reviewed on the subject. The result might be an indication of the fact that there are factors associated with rural poverty (as opposed to relatively well-off sections in the rural society) that are less inclined to support breeding grounds for mosquitoes and these unobservables cannot be controlled for. In the household level analysis we found that heads of households who owned water pumps (an asset owned by relatively well-off households) were more likely to contract malaria. But, the ownership of water pumps is likely to change every year and across states. There is likely to be an increase in the percentage of water pumps owned and we do not have data on this aspect. Another explanation for this negative link lies in the possibility of rural poverty being associated with the presence of genetic abnormalities like the sickle cell trait. As Gallup and Sachs (2001) argue, the presence of this trait in certain ethnic groups is actually a result of their being poor and is acquired because of the subjection of these groups to repeated attacks of malaria. The body develops the defense against malaria in the form of this genetic defect. As discussed in Chapter 2 (p.19), having the G6PD hereditary enzyme defect also partly protects against malaria. The annual parasite incidence data are for the entire population of adults and children and over a twenty three year period there will be 'new additions' to the population that may or may not carry the sickle cell trait or the G6PD deficiency resulting in a change in the distribution of the people with this trait. These unobservables that are likely to vary over time and across states cannot be controlled for in our model because of a lack of data. It has been found that both the sickle cell trait and the G6PD deficiency in India is more common among scheduled castes (SCs), scheduled tribes (STs), and other 'backward' castes (OBCs). Since data on state-wise (or even all-India) percentage of OBCs over time are not available, it is not possible to control for these traits.[8] Data on percentage of SCs and STs by state are available for the 1971, 1981, and 1991 censuses and by interpolation using all-India decadal growth rates of these population groups (since state-wise growth rates are not available), the data for the intervening years could be filled. Even this procedure may not control for the sickle cell trait because of heterogeneity within these groups. As Balgir (2005) found in his study of scheduled tribes in Sundargarh district of Orissa that while the sickle cell trait was present in some tribes, blood tests did not show any evidence of the trait in other tribes. Further, even within a tribe, while one sub-tribe had the trait, it was absent in another. In another study conducted by Kate and Lingokwar (2002) in Maharashtra found that while the prevalence of the sickle cell abnormality was higher among the SCs, STs, and OBCs, it was not absent among other

[8]While yearly data on caste distribution across states is not available, data on rural and urban distribution of SCs, STs, OBCs, and other castes and groups are available for 2003 (Indiastat.com 2008a, Indiastat.com 2008b). These data show that the lower castes are concentrated in rural areas. The combined SC and ST percentage in rural India was 16.8 and that in urban India was 30.9.

communities. Thus, the utility of the caste measure is questionable given the available data on it and has not been used in the current study.

Table 7.5: Effect of income and other variables on malaria incidence: First difference model (D1log(**API**) = dependent variable)

Variable	*FD*
D1popden[a]	-0.0002
	[0.0076]
D1urban	-0.2138**
	[0.0990]
D1rain	0.00002
	[0.00004]
D1irrig	-0.0247
	[0.0250]
D1stinc	0.0001
	[0.0003]
D2health	-0.0080
	[0.0092]
D1eduex	-0.0067
	[0.0053]
D1povertyr	-0.0018
	[0.0031]
D1povertyu	0.0031
	[0.0047]
Constant	0.0175
	[0.0791]
Observations	315
R-squared	0.02

a. FD: First-difference b. D1 indicates first-difference and D2 indicates first-difference for the one period lag of health; standard errors reported in brackets; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis is a t-test of the coefficient being equal to zero.

In the model in first differences, presented in Table 7.5, only **urban** has a well determined coefficient with a negative sign indicating that a change of one percentage point in **urban** results in a 21 percent fall in **API**. All other variables (including the growth in health expenditure and income) do not yield well determined coefficients. Thus, it seems that the changes in the different explanatory variables (except **urban**) do not affect the change in **API** over the period under consideration. This might be attributed to lower variation in the differenced variables, as compared to the variables in levels, resulting in large standard errors. The means and standard deviations of the variables in levels and first differences are presented in Table J.1 in Appendix J and as we can see, the standard deviations are much higher in case of the variables in levels.

7.2 Summary

The analysis in this chapter has addressed the research question: Does aggregate income have a negative impact on malaria? How does it compare with the effect of public health expenditure? We find that aggregate income does have a negative impact on malaria incidence. The association with income is robust even when including controls for education and health expenditure, poverty, irrigation, population density, as well as some other variables like climate variation over time through time dummies and across states through fixed effects, and by inclusion of average annual rainfall by state. However, even though the impact of income is negative it is of a small magnitude. Comparatively, the effect of direct health expenditure is much larger. The coefficient on **healthlag** indicates that an increase in real health expenditure per capita by as little as one rupee annually can reduce annual parasite incidence per 1000 population by as much as four percent. In contrast, a one rupee increase in state income reduces malaria incidence by a mere 0.1 percent. The coefficient on health is larger than that on income by a factor of 28 in absolute terms. Thus, the effect of a one rupee increase in expenditure on health has the equivalent effect of 28 rupees spent on aggregate income, similar to the result obtained by Anand and Kanbur (1995) for infant mortality in Sri Lanka – the coefficient on health was 32 times larger than that on income. While income growth would certainly have positive effects on health outcomes, these are likely to be realised over the longer term. For developing countries, waiting for the benefits of income growth to accrue would certainly not be a prudent decision and direct expenditure on health would bring about quicker returns. This finding underlines a clear role for government spending in bringing down malaria incidence. As argued in Chapter 1 (p.6), an increase in income in the hands of individuals may not bring about the desired fall in malaria incidence since individuals may not choose to invest in prophylactic measures.

The negative sign on the coefficient for rural head count index is indicative of the fact that the rural poor[9] may have developed a resistance to malaria over time by acquiring the sickle cell abnormality. The estimates for irrigation, population density and rainfall are different from the district analysis. These results are not strictly comparable as the district analysis is for Uttar Pradesh alone whereas the state level analysis uses data from

[9]As noted in this chapter, the SC and ST often have lower living standards than other groups and are also concentrated in rural areas.

15 major states (which could cause aggregation bias) and includes time dummies and state fixed effects. Moreover, the district and state level results are not strictly comparable as the variables are measured differently at the two levels.

Chapter 8

Conclusion

Malaria eradication is a complex and challenging task. Despite more than a century of knowledge on the mechanisms that cause malaria, it has not been possible to eradicate the disease. A major reason for this is the ability of the parasite, *Plasmodium*, and the malaria-casing *Anopheles* vector to develop immunity to drugs and insecticides, respectively. In addition, the *Anopheles* mosquito can adapt to breed in varied and new habitats. Coupled with these factors, the recent trend towards warm and humid climate creates conditions conducive to mosquito breeding and suitable for the survival of *Plasmodium* and particularly its more deadly strain, *falciparum*.

There is evidence to show that poor countries bear the burden of malaria, leading to the obvious deduction that income growth would bring down malaria incidence. However, it is argued that instead of waiting for the benefits of income growth to accrue, in the short run it is more beneficial to invest in public health expenditure. Studies on the impact of aggregate income on infant and child mortality (Anand and Ravallion 1993, Anand and Kanbur 1995) and reducing malnutrition (Haddad et al. 2003) find that investing in public health expenditure brings quicker improvements in health outcomes as compared to income. Furthermore, increases in income at the household level may not necessarily bring down malaria incidence. Individuals or households may choose not to invest in prophylactic measures like using insecticide-treated nets or in improving their private infrastructure such as repairing leaking pipes, which form stagnant water pools that support mosquito breeding. These interventions also have positive externalities for other households and given their 'public' nature, it requires government intervention, which could be in the form of insecticide spraying, raising awareness, and both indoor and outdoor surveillance to check for breeding areas. These are some of the areas where public spending could be directed. In order to know where public spending may be effective, it is important to study the factors that determine malaria incidence at the household and district level, as has been done in this thesis. This, however, is not the focus of the Roll Back Malaria initiative which emphasizes artemisinin-based combination therapy, use of insecticide-treated nets, and development of a vaccine as ways of attacking malaria. This thesis argues that while these are essential for malaria control and, eventually, eradication, a holistic approach that also involves identifying specific factors that influence malaria at the micro-level would go a long way in identifying areas of intervention. Based on this background the research questions we set out were:

1. What are the socio-economic factors that affect the incidence of malaria at the household and district level? How do these differ by settlement types and are the effects of these factors robust over time?

2. Does household wealth and other socio-economic status variables have a negative impact on malaria incidence?

3. Does aggregate income have a negative impact on malaria incidence? How does it compare with the effect of public health expenditure?

The first two research questions, relating to the relationship of malaria incidence with household and district level socio-economic factors, were addressed using the National Family Health Survey (NFHS) for 1992-93 and 1998-99 and the census data for 1991 and 2001. The third research question employed a panel of 15 major states in India to study the relationship of malaria incidence with aggregate income and public health expenditure at the state level.

The focus of the household and district analysis was on Uttar Pradesh since this had the highest number of malaria cases in 1992-93 according to NFHS estimates, aside from having a number of characteristics that are found to be associated with malaria incidence such as a large population, high levels of poverty, low levels of literacy, and a large proportion of land under irrigation. Instead of focusing on other high incidence states like Madhya Pradesh, Gujarat and Orissa, where the analysis of malaria is effectively that of tribal and forest malaria, we preferred to focus on Uttar Pradesh in order to investigate the effect of other factors on malaria incidence. Our choice of Uttar Pradesh was further supported by the results we obtained in our panel study. We found that after controlling for other factors, Uttar Pradesh had the fifth largest fixed effect, indicating that there are a large number of other factors that affect malaria incidence vindicating the micro-level analysis undertaken for this particular state in the thesis.

8.1 Addressing the research questions

In this section we summarise the findings relating to the different research questions in turn, starting with the last one.

8.1.1 Effect of aggregate income and public health expenditure on malaria incidence

We found that an increase of a rupee per person in aggregate income per person reduces malaria incidence by 0.1 percent, whereas an equivalent increase in real health expenditure per capita brings about a decline of four percent. The coefficient on real health expenditure per capita is larger than that on aggregate income per capita by a factor of 28, implying that to bring about the same fall in malaria incidence, real health expenditure per capita would need to be increased by one rupee as opposed to Rs 28 in real income per capita. This result is very similar to the Rs 33 trade-off obtained by Anand and Kanbur (1995) for their study on Sri Lanka. Thus, our study finds that direct expenditure on health would be more reliable in bringing about a decline in malaria incidence, a view also supported by Haddad et al. (2003). They find that income growth alone cannot bring down malnutrition rates in developing countries to half the target set by the Millennium Development Goals over a 25-year period even with a growth rate of 2.5 percent. Our findings also support the argument made earlier that income increase in the hands of individuals may not bring about the desired fall in malaria incidence. Government intervention is more likely to bring the desired results, at least in the short-run.

8.1.2 Factors determining malaria incidence at household and district levels

Our analysis identifies a number of key factors that determine malaria incidence at the household and district level. These include caste, education, drinking water source, fuel-type, house-type, owning a water pump, sanitation, asset wealth, occupation, irrigation, population density, and rainfall. Since these vary across different specifications, Table 8.1 provides a summary of these factors for the household rural and urban specifications for the two time periods and for the district level regressions. As we can see, there is little evidence of robustness over time. When we compare the 1992-93 and 1998-99 household rural specifications, the only variable that has the same sign and significance is whether a household owns a water pump. In the comparable urban specifications for the two years, the only variable that has a robust effect is separate kitchen. The reason for such limited evidence of robustness is likely to be the fact that the factors determining malaria change depending on particular conditions prevailing that year. Even though we used district dummies that may capture several unobservables, factors relating to government policy, use of bed nets, and other prophylactic measures are absent from the data. These factors could help explain why certain variables ceased to have significant effects in the latter period.

The district level results also show a lack of robustness over time. Whereas the effects for a number of variables are found to be well determined in 1992-93, in 1998-99 most effects were not well determined. However, the coefficient on percentage of villages with an adult literacy centre was significant with a positive sign in conflict with the 1992-93 specification. We made the argument earlier that this was probably because of directing such classes to malaria-prone areas. The lack of robustness, however, does not imply a dismissal of the results we have obtained for the different years. Instead it indicates that factors affecting malaria incidence can vary over time.

8.1.3 Effect of household wealth and socio-economic factors on malaria incidence

To interrogate the effect of household income on malaria incidence, we used a proxy measure of permanent income because the NFHS does not include data on a current income or expenditure measure. The measure used was a non-agricultural consumer durable asset

Table 8.1: Factors determining malaria at the household and district level in Uttar Pradesh

Year/settlement type	*Factors determining malaria incidence*
Household analysis	
1992-93 rural (a)	Caste (-), gender (+), high-school educated (-), electricity (-), wood as fuel (-), open public water source (+), own water pump (+), above two acres irrigated area (-)
1992-93 rural (b)	Caste (-), gender (-), high-school educated (-), electricity (-), wood as fuel (-), open source drinking water (+), own water pump (+)
1998-99 rural	High-school educated (+), wood as fuel (+), miscellaneous fuel category (-), own water pump (+), own bullock cart (+), own tractor (+), asset-index (-), above 67 years (+)
1992-93 urban	Primary-school educated (+), *kachcha* house, separate kitchen (-), sanitation (+), wood as fuel (+), asset-index (-)
1998-99 urban	Caste (-), agriculture occupation (+), wage-earner (-), separate kitchen (-)
District-analysis (rural)	
1992-93 (a)	Percentage of villages (PV) with an adult literacy centre (-), PV with a commuting facility (-), PV domestic power supply (-), average distance to water source (-), percentage of district area under forest (-), district population density (-), log rainfall (-), percentage area irrigated by government canal (+) and private canal (+), by tube well (+) and electric and non-electric well (+), and PV with power for agriculture (+).
1992-93 (b)	All of the above variables had the same signs and significance except for forest area that ceased to have a well determined effect
1998-99	PV adult literacy centre (+), percentage of district area irrigated by an electric tube well (+)

The rural (b) specification is a modified version of (a) to compare with 1998-99. The signs on the coefficients of the variables are indicated in parentheses. The reference categories for the relevant categorical variables in the household specifications are: education: illiterate; fuel-type: dung; drinking water source: protected public; house-type: semi-*pucca*; occupation: producion and transport.

wealth index constructed from data on ownership of consumer durables using Principal Components Analysis. The estimated effect for the asset index was not well determined in the 1992-93 rural specification, but was negative and significant in 1998-99, indicating a negative association with malaria as the value of the index rises, *ceteris paribus* and on average. Thus, wealthier households are likely to be less susceptible to contracting malaria. In urban Uttar Pradesh, the effect of the asset index in 1992-93 was negative and significant and in 1998-99, it did not yield a significant coefficient.

To assess how the relationship differs across different wealth groups we broke up the sample populations into quintiles using splines. In rural Uttar Pradesh in 1992-93 and in urban Uttar Pradesh in 1998-99, the poorest group as well as the 60-80 percent group had a positive coefficient, indicating that at these levels of income an increase in wealth would fail to reduce malaria incidence. The positive coefficient for the 60-80 percent group conflicts our expectations but seems to indicate that malaria can also affect the relatively wealthier population groups. At the same time, the effect for the top-most wealth group was negative and significant for these years implying that among the richest a rise in wealth would result in reducing malaria incidence, supporting the view that household wealth exerts a negative impact on malaria incidence though only for the richest set of households.

We now focus on the relationship between socio-economic status and malaria incidence. Caste yielded a negative and significant coefficient in rural Uttar Pradesh in 1992-93 and in urban Uttar Pradesh in 1998-99, refuting the negative socio-economic status and health link—household heads from the lower socio-economic status groups of scheduled caste/tribe, relative to other castes and groups are less likely to contract malaria, *ceteris paribus*, possibly because of certain genetic traits. However, the reason for this genetic immunity against malaria is possibly a result of being subjected to malaria repeatedly over several generations.

Owning a water pump, indicative of higher wealth status, had a positive effect on malaria incidence and was robust to different rural specifications. This result very clearly refutes the negative socio-economic status and health relationship. Owning a water pump increases the likelihood of contracting malaria because areas around water pumps are likely to have stagnant water pools providing breeding grounds for mosquitoes.

In urban areas, the result for sanitation, where the sign on the coefficient is positive, implying that presence of sanitation is favourable to contracting malaria, also conflicts with the negative socio-economic status and health relationship.

At the district level, area irrigated by canals had a positive association with malaria incidence in 1992-93. Similarly, in the same year, the results reveal that the higher the percentage of villages with a commuting facility in a district, the higher is the propensity to contract malaria at the household level. This could be because transport vehicles harbour mosquitoes or commuters act as passive carriers of *Plasmodium*. Both these results fail to indicate a negative link between socio-economic status and health outcomes.

There are, however, other variables that support the negative socio-economic status and health link. At the household level, having an electricity connection and having access to protected public water supply (relative to an open one) exerts a negative impact on malaria incidence in 1992-93. Education is also found to be a significant factor associated with a lower likelihood of malaria incidence both at the household level as well as at the district level in 1992-93. In urban Uttar Pradesh working in the agricultural sector is found to have a positive effect on malaria incidence whereas wage earners are less likely to contract malaria, relative to the production and transport category in 1998-99. A household head living in a *kachcha* house, relative to a semi-*pucca* house is also more conducive to contracting malaria in urban areas in both 1992-93 and 1998-99.

In conclusion, we need to be wary of making generalisations regarding the negative relationship between malaria incidence and wealth or other socio-economic factors at the household level, given our conflicting set of results.

8.1.4 Factors driving the difference in malaria incidence across the two settlement types

As we observed, there are essential differences between rural and urban malaria incidence that can be attributed, to a large extent, to the differences in the habitats of the malaria-causing vectors in these locales. We conducted a decomposition analysis to understand whether these differences could be attributed to differences in characteristics between rural and urban households (endowments or characteristic effects) or because of the differences in the underlying relationships (coefficient or treatment effects). In 1998-99, the difference in the malaria incidence gap is explained by differences in characteristics even for the component effects, whereas in 1992-93, it is structural differences or differences in treatment effects that seem to drive the malaria difference.

Our results for the component effects do not indicate evidence of robustness over time. At the same time, within each year, the characteristic effects are found to be sensitive to the coefficients used and the coefficient effects are sensitive to the characteristics used, a standard index number problem. These factors together make it hard to infer consistent findings from this analysis.

As we concluded in Chapter 6, the inability to obtain definitive evidence from the decomposition analysis might also be a result of the fact that the regression specifications used have poor goodness of fit. Moreover, there are intrinsic differences in the epidemiology of malaria in rural and urban locales, so that using a common set of variables may not be appropriate.

Despite our caveats, the result for drinking water source for 1992-93 has some policy relevance. The difference in malaria incidence between rural and urban areas is widened because of differences in drinking water characteristics as well as the treatment effects. This is likely because of the superior drinking water infrastructure in urban areas—urban households depend on private piped water supply whereas rural households get drinking water from public taps and open sources, which can be ideal breeding grounds for mosquitoes because of the collection of stagnant water. Thus, an improvement in infrastructure in rural areas so as to prevent collection of stagnant water around water collection points and, in the long run, a move towards piped private supply would narrow the rural-urban malaria incidence gap. We now discuss possible policy interventions based on our findings.

8.2 Policy interventions

On the basis of the analysis undertaken at the micro-level it is possible to identify areas where public spending could be directed for effective malaria prevention.

Using Macdonald's equation, discussed in Chapter 2, while the parameter, p, the mosquito survival probability for a day can be affected by such measures as insecticide spraying or through natural changes such as heavy rainfall that can flush mosquito larvae, there are other parameters in the equation that can be influenced by acting on socio-economic development variables used in the current study. These parameters are b (proportion of infective mosquito bites) and m (anopheline density in relation to human population).

Both these parameters are likely to be influenced, for example, by controlling the breeding areas of mosquitoes, in turn a function of irrigated area, the type of irrigation system, and the kind of drinking water sources used as well as the management of these sources.

Given the negative association of education (and implicitly awareness) with malaria incidence, programmes that provide information on the peak biting times of the *Anopheles* vector and whether it is endophilic or exophilic could help people in preventing bites. Behavioural changes could then be encouraged such as spending less time outdoors at the time of peak biting activity of the vector.

The results obtained for source of drinking water at the household level, the decomposition analysis, and the average distance from drinking water source at the district level, reinforce the fact that such open water sources as rivers and streams are breeding areas for the *Anopheles* vector. Other water collection sites such as public taps may suffer from the tragedy of the commons (Hardin 1968) and have residual water collection resulting in mosquito proliferation. In open water source areas it might be possible to control breeding of anophelines by introducing larvivorous fish as a part of a vector control programme. The introduction of guppy fishes controlled larvae in Nadiad *taluka*, Kheda district, Gujarat: one to seven percent of the containers in target areas in the sub-district showed mosquito breeding and in control areas the range was between seven to 29 percent. The difference in breeding between the two types of areas was significant at the one percent level (Bhati et al. 1995). At water collection points such as public taps and water pumps there may be a role for better provision of drainage facilities.

Variables related to irrigation at the district level in 1992-93, whether power supply for agriculture purposes or area irrigated by canals or tube wells, show a positive association with malaria incidence. This result is in agreement with the literature on the subject and it has been proposed by several studies to devise such methods as lining canals and improving water flow so as not to allow stagnant water collection. At the state level and for the district regression for 1998-99, irrigation does not have a positive effect and might actually be a result of improved irrigation works.

Positive examples of involving communities and undertaking environment-friendly measures in controlling mosquito breeding sites resulting from water collection are provided by the experiment undertaken in Nadiad *taluka*. Leakages in taps, pipelines, and tube wells were repaired by the staff working on the project undertaken in this sub-district, but were maintained by villagers who were trained so that they had the capacity to sustain this once the project ended. Soakage pits were constructed on the streets to stop flowing water and prevent mosquito breeding. The experiment also ran social forestry programmes involving the planting of *Eucalyptus* trees in marshy areas to control mosquito breeding (Bhati et al. 1995).

At the district level we found that the greater the percentage of villages with high schools and adult literacy centres, the lower was the likelihood of contracting malaria, identifying a clear role for education in attacking malaria incidence. Adult literacy centers could, in fact, play a crucial role in raising awareness to control malaria. The Nadiad *taluka* experiment also recognised the role of health education in malaria control. It organised health camps and interpersonal discussions to increase knowledge on the disease. These motivated people to participate in such vector control programmes as filling pot holes found to be breeding areas for mosquitoes (Sharma and Sharma 1989).

8.3 Limitations of current research and future direction

The analysis undertaken in this thesis at the household and district level is unique in using the NFHS, the Demographic Health Survey (DHS) for India, to identify the factors associated with malaria incidence. Even though a number of these factors are not robust over time, individually for each year this exercise contributes to an understanding of the incidence of malaria incidence in Uttar Pradesh and identifying areas where intervention is likely to bring better returns. The analysis undertaken here can be replicated for other Indian states and for other countries given the availability of the DHS for many African and Asian countries.

The DHS, however, requires substantial improvement in the collection of explanatory variables likely to affect malaria incidence such as data on prophylactic measures like using insecticide-treated nets, spraying, and electric and non-electric mosquito coils at the household level. It would also be useful to include district level information including details on specific government intervention programmes. Another drawback of the data is the lack of a current income or expenditure measure. Even though this limitation was overcome by the use of a consumer durable asset wealth index, it would be useful to have data on current income or expenditure.

Another limitation encountered was the lack of a break-up of health expenditure data into various components and specifically on malaria control measures. The health expenditure data used in the inter-state analysis includes expenses on family planning, public health and sanitation, as well as water supply. While the use of this combined category is useful for the purpose of the analysis undertaken here, it would be informative to know what the effect of expenditure on malaria programmes *per se* is and how it differs from that on more general public health expenditure.

Finally, as we noted earlier on in the thesis, the data on the dependent variable, malaria incidence, are drawn from two different sources – the NFHS for the household analysis and the National Malaria Eradication Programme (NMEP) for the inter-state analysis. The NFHS data are collected by asking the head of household if he/she had malaria in the last three months, whereas the NMEP reporting is based on actual examination of blood smears. There might be an inclination to trust the latter source since the estimates are based on a blood examination. This might be one of the reasons why the NFHS ceased collecting data on malaria incidence from its third phase. However, as argued in Chapter 3, comparing malaria-like fever cases, rather than actual malaria incidence cases, and taking into account the imperfections in the NMEP data collection, the discrepancy between the two different estimates narrows down considerably.

In conclusion, we cannot under-emphasize the fact that the DHS provides an excellent source for exploring the relationship of malaria incidence with a number of socio-economic factors for different countries and for different states in India. With the improvements in data collection suggested here, it would go a long way in contributing to an understanding of socio-economic factors affecting malaria and identifying areas for effective intervention. It is this belief that motivated this thesis and, despite data limitations, we have been able to make a new and relevant contribution to the subject. As we found, there are many factors that affect malaria incidence, thus narrowing down the focus on providing insecticide-treated nets or using drug therapy alone will not be sufficient to control let alone eradicate malaria. A holistic approach, which understands the different socio-economic factors that affect malaria, how these change over time and by location, and infer areas for effective intervention, would make a bigger dent on malaria control and, eventually, eradication.

Bibliography

Asterisks below refer to secondary references.

Adak, T., O.P. Singh, M.K. Das, S. Wattal, and N. Nanda. 2005. 'Comparative susceptibility of three important malaria vectors *Anopheles stephensi*, *Anopheles fluviatilis*, and *Anopheles sundaicus* to *Plasmodium vivax*.' *The Journal of Parasitology*, 91:1, pp. 79-82.

*Aitchison, J. and J.A.C. Brown. 1957. *The lognormal distribution*. Cambridge: Cambridge University Press.

Akhtar, R. and A.J. McMichael. 1996. 'Rainfall and malaria outbreaks in western Rajasthan.' *Lancet*, 348:9039, pp. 1457-58.

Al-Samarrai, S. and B. Reilly. 2000. 'Urban and rural differences in primary school attendance: an empirical study for Tanzania.' *Journal of African Economies*, 9:4, pp 430-474.

Ambedkar. Unpublished. 'Essays on untouchables and untouchability: Social.' Available from: http://www.dr-ambedkar.com/home.htm. Centre of Socio-economic Development for SC/ST: Pune.

Anand, S. and M. Ravallion (1993). 'Human development in poor countries: On the role of private income and public services.' *The Journal of Economic Perspectives* 7(1): 133-150.

Anand, S. and R. Kanbur (1995). 'Public policy and basic needs provision: Intervention and achievement in Sri Lanka.' in *The Political Economy of Hunger: Selected Essays*. J. Dreze, A. Sen and A. Hussain eds. Oxford: Oxford Clarendon Press, pp 626.

*Annon (1981). ‘A report on the Indian drug industry: 1980-2000 AD.’ Bombay, Indian Drug Manufacturer’s Association.

Ansari, M.A. and R.K. Razdan. 2004. ‘Impact of residual spraying of *Reldan* against *Anopheles culicifacies* in selected villages of District Ghaziabad (Uttar Pradesh), India.’ *Journal of Vector Borne Diseases*, 41:3-4, pp. 54-60.

Arthreya, L.A. and L.L. Coriell. 1967. ‘Relation of blood groups to infection.’ *American Journal of Epidemiology*, 86:2, pp. 293-304.

Bang, Y.H. 1985. ‘Integrated management of urban mosquito vectors of human diseases.’ *Journal of Communicable Diseases*, 17:1, pp. 1-10.

Banguero, H. ‘Socioeconomic factors associated with malaria in Colombia.’ *Social Science and Medicine*,19:10, pp.1099-1104.

Basher, R.E. and M.A. Cane. 2002. ‘Climate variability, climate change, and malaria,’ in *The Contextual Determinants of Malaria*. E.A. Casman and H. Dowlatabadi eds. Washington, DC: Resources for the Future, pp. 382.

Batra, C.P., T. Adak, V.P. Sharma, and P.K. Mittal. 2001. ‘Impact of urbanization on bionomics of *An.culcifacies* and *An.stephensi* in Delhi.’ *Indian Journal of Malariology*, 38, pp. 61-75.

Baum, C. F. and M. Schaffer (2002). ‘Instrumental variables and GMM: Estimation and testing.’ *Boston College of Economics Working Paper* 545.

Besley, T. and R. Burgess (2002). ‘The political economy of Government responsiveness: Theory and evidence from India.’ *The Quarterly Journal of Economics* 117(4): 1415-1452.

Besley, T. and R. Burgess (2004). ‘Can labour regulation hinder economic performance? Evidence from India.’ *The Quarterly Journal of Economics* 19(1): 91-134.

Bhalotra, S. (2006). Health and growth in India: Panel data estimates. Presented at conference entitled ‘Reducing poverty and inequality: How can Africa be included?’, Centre for the study of African economies (CSAE), St. Catherine’s College, Oxford. Available from: http://www.csae.ox.ac.uk/conferences/2006-EOI-RPI/papers/csae/Bhalotra.pdf.

Bhati, P.G., R. Kant, H.C. Srivastava, V.S. Malaviya, and P.K. Pujara. 1995. 'Role of health education in school children with particular reference to malaria.' *Indian Journal of Malariology*, 32:3, pp. 93-98.

Bhati, P.G., V.S. Malaviya, R. Kant, H.C. Srivastava, S.K. Sharma, and V.P. Sharma. 1996. 'Socio-economic aspects of malaria in Kheda district, Gujarat.' *Indian Journal of Malariology*, 33:4, pp. 200-8.

Bhatt, R.M., R.C. Sharma, and V.K. Kohli. 1990. 'Interspecific associations among anophelines in different breeding habitats of Kheda district, Gujarat. Part I: Canal irrigated area.' *Indian Journal of Malariology*, 27:3, pp. 167-72.

Bhattacharya, S., C. Sharma, R. C. Dhiman, and A. P. Mitra. 2006. 'Climate change and malaria in India.' *Current Science*, 90:3, pp. 369-75.

*Bliss, C. 1934a. 'The method of probits.' *Science*, 79, pp. 38-39.

*Bliss, C. 1934b. 'The method of probits.' *Science*, 79, pp. 409-10.

*Bound, J., D. A. Jaeger, et al.. (1995). 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak.' *Journal of the American Statistical Association* 90(430): 443-450.

Breslau, N., N. Paneth, and V.C. Lucia. 2004. 'The lingering academic deficits of low birth weight children.' *Pediatrics*, 114:4, pp. 1035-40.

Burgess, R. and R. Pandey (2005). 'Do rural banks matter? Evidence from the Indian social banking experiment.' *The American Economic Review* 95(3): 780-795.

Bussea, M. and J. Groizardb (2006). 'Foreign direct investment, regulations, and growth.' *World Bank Policy Research Working Paper* 3882.

Centers for Disease Control and Prevention. 2006. '*Anopheles* mosquitoes.' Available from: http://www.cdc.gov/malaria/biology/mosquito. Department of Health and Human Services, CDC: Atlanta.

CDC. 2004. 'Malaria: glossary.' Available from: http://www.cdc.gov/malaria/glossary.html [accessed 23 January 2007] Centers for Disease Control and Prevention (CDC).

Chakraborty, S., S. Ray, and N. Tandon. 1998. 'Seasonal prevalence of *Anopheles stephensi* larvae and existence of two of forms of the species in an urban garden in Calcutta city.' *Indian Journal of Malariology*, 35, pp. 8-14.

Chauhan, V. S. (2008). 'A step forward in malaria vaccine development.' *Current Science* 94(8): 973.

*Clyde, D. 1931. 'Report on the control of malaria during Sarda canal construction (1920-1929).' *Rec. Malaria survey India*, 2, pp. 49-110.

Conover, W.J. 1999. *Practical Non-parametric Statistics*. New York: John Wiley & Sons, Inc.

Cotton, J. 1988. 'On the decomposition of wage differentials.' *The Review of Economics and Statistics*, 70:2, pp. 236-243.

*Covell, G. 1955. 'Developments in malaria control methods during the past forty years.' *Indian Journal of Malariology*,9, pp. 305-12.

Cramer, J.S. 2003. 'The origins and development of the logit model.' Available from: www.cambridge.org/resources/0521815886. University of Amsterdam and Tinbergen Institute: Amsterdam.

D'Alessandro, U., B.O. Olaleye, W. McGuire, P. Langerock, S. Bennett, M.K. Aikins, M.C. Thomson, M.K. Cham, B.A. Cham, and B.M. Greenwood. 1995. 'Mortality and morbidity from malaria in Gambian children after introduction of an impregnated bednet programme.' *Lancet*, 345:8948, pp. 479-83.

Das, R., Z. Khan, P. Nath, and A. Amir. 2005. 'Epidemiological assessment of the trend of malaria in rural western U.P.' *Indian Journal of Community Medicine*, 29:3, pp. 134-135.

Dasgupta, P. 1997. 'Nutritional status, the capacity for work, and poverty traps.' *Journal of Econometrics*, 77:1, pp. 5-37.

Datt, G. and M. Ravallion (1992). 'Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s.' *Journal of Development Economics* 38(2): 275-295.

Datt, G. 1998. 'Poverty in India and Indian states: an update.' *IFPRI FCND Discussion Paper No. 47.* Food Consumption and Nutrition Division (FCND), International Food Policy Research Institute: Washington, D.C.

Deaton, A. 1997. *The analysis of household surveys: A microeconometric approach to development policy.* Baltimore and London: The John Hopkins University Press for the World Bank, Washington DC.

Deaton, A. 2002. 'Policy implications of the gradient of health and wealth.' *Health Affairs*, 21.

Deaton, A. 2003. 'Health, inequality, and economic development.' *Journal of Economic Literature*, 41:1, pp. 113-158.

Deaton, A. and J. Dreze. 2002. 'Poverty and inequality in India: a re-examination.' *Economic and Political Weekly*, September(Special articles): 3729-3748.

Drakeley, C.J., I. Carneiro, H. Reyburn, R. Malima, J.P.A. Lusingu, J. Cox, T.G. Theander, W.M.M.M. Nkya, M.M. Lemnge, and E.M. Riley. 2005. 'Altitude-dependent and -independent variations in *Plasmodium falciparum* prevalence in northeastern Tanzania.' *The Journal of Infectious Diseases* 191:10, pp. 1589-98.

Ehiri, J. E and E.C. Anyanwu. 2004. 'Mass use of insecticide-treated bednets in malaria endemic poor countries: Public health concerns and remedies.' *Journal of Public Health Policy*, 25:1, pp. 9-22.

Encyclopaedia Britannica. 2006. 'Tarai.' Availabe from: http://www.britannica.com/eb/article-9071265. Britannica: Chicago

Encyclopaedia Britannica. 1997. 'Nobel prizes.' Available from: http://www.britannica.com/nobel/micro. Britannica: Chicago.

Filmer, D. and L.H. Pritchett. 2001. 'Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India.' *Demography*, 8:1, pp. 115-32.

Flessa, S. 1999. 'Decision support for malaria-control programmes: A system dynamics model.' *Health Care Management Science*, 2:3, pp. 181-91.

Forest Survey of India (FSI). 1999. 'State of Forest Report, 1999.' Available from: http://www.envfor.nic.in/fsi/sfr99/sfr.html. FSI, Ministry of Environment and Forests: Dehradun.

Fouarge, D. and R. Muffels. 2000. 'Persistent poverty in the Netherlands, Germany, and the UK: A model-based approach using panel data for the 1990s.' *OSA-Working paper WP2000-4*. The Institute for Labour Studies (OSA), Tilburg University: Tilburg.

Fox, E., J. Bouloumie, J.G. Olson, D. Tible, M. Lluberas, S.O. Shakib, J.P. Parra, and G. Rodier. 1991. '*Plasmodium falciparum* travels by train from Ethiopia to Djibouti.' *Medicine Tropicale: revue du corps de sante colonial*, 51:2, pp. 185-9.

Gallup, J. L. and J. D. Sachs (2001). 'The economic burden of malaria.' *The American Journal of Tropical Medicine and Hygiene* 64(1(Supplement)): 85-96.

Gertler, P. and J. Gruber. 2002. 'Insuring consumption against illness.' The American Economic Review, 92:1, pp. 51-70.

Ghosh, A. 1995. The Calcutta chromosome: a novel of fevers, delirium and discovery. New York: HarperCollins Publishers Inc.

*Government of India (GoI). 1979. 'Report of the task force on projections of minimum needs and effective consumption.' Planning Commission, GoI: New Delhi.

Government of India (GoI). 1979-1985. 'Health Statistics of India.' Directorate General of Health Services, Ministry of Health and Family Welfare: New Delhi.

Government of India (GoI). 1979-2001. 'Statistical Abstract of India.' Central Statistical Organisation, Ministry of Statistics and Programme Implementation: New Delhi.

Government of India (GoI). 1986-2001. 'Health Information of India.' Directorate General of Health Services, Ministry of Health and Family Welfare: New Delhi.

Government of India (GoI). 1996. 'Statistical Abstract, Uttar Pradesh.' Economics and Statistical Division, State Planning Institute, GoI, Uttar Pradesh: Lucknow.

Government of India (GoI). 2000. 'Population of India, Appendix Table 9.1,' in Economic Survey, 1999-2000. New Delhi: Ministry of Finance, GoI.

Government of India (GoI). 2000. 'Annual Report 1999-2000.' New Delhi: Ministry of Environment and Forests, GoI.

Government of India (GoI). 2001. 'Census 2001.' Available from: http://upgov.nic.in/upinfo/census01/cen01-4.htm. New Delhi: National Informatics Centre, GoI.

Government of India (GoI). 2004. 'Agricultural Statistics at a Glance, August 2004.' Available from: http://agricoop.nic.in/statatglance2004/AtGlance.pdf. New Delhi: Agricultural Statistics Division, Directorate of Economics and Statistics, GoI.

Government of India (GoI). 2005. 'States at a Glance.' Available from: http://mha.nic.in/stat.htm. New Delhi: Ministry of Home Affairs, GoI.

Government of India (GoI). 2006a. 'Definitions (nos. 24 and 25), Article 366, Miscellaneous, Part XIX: The Constitution of India updated up to the 93rd Amendment Act, 2003, 2nd edition of the Constitution of India.' Available from: http://indiacode.nic.in/coiweb. New Delhi: Ministry of Law and Justice (Legislative Department), GoI.

Government of India (GoI). 2006b. 'Scheduled Castes and Scheduled Tribes Lists.' Available from: http://socialjustice.nic.in/schedule/welcome.htm. (under FAQ). New Delhi: Ministry of Social Justice and Empowerment, GoI.

Government of India (GoI). 2006. 'Malaria Surveillance.' Available from: namp.gov.in/Doc/Director%20Desk/Malaria%20Surveillance%20-%203.doc. New Delhi: National Vector Borne Disease Control Programme, Directorate General of Health Services, Ministry of Health and Family Welfare.

Government of Uttar Pradesh. 1999. *Swajaldhara* Programme State Water and Sanitation Mission, Uttar Pradesh: Lucknow. Available from: http://www.swajal.org.

Greene, W.H. 2003. *Econometric Analysis*. New Jersey: Prentice Hall.

Grizzle, J.E., C.F. Starmer, and G.G. Koch. 1969. 'Analysis of categorical data by linear models.' *Biometrics*, 25, pp. 489-504.

Grossman, M. 1972. 'On the concept of health capital and the demand for health.' Journal of Political Economy, 80:2, pp. 223-255.

Guillet, P., M.C. Germain, T. Giacomini, F. Chandre, M. Akogbeto, O. Faye, A. Kone, L. Manga, and J. Mouchet. 1998. 'Origin and prevention of airport malaria in France.' *Tropical Medicine and International Health*, 3:9, pp. 700-05.

Gupta, S., R.W. Snow, C.A. Donnelly, K. Marsh, and C. Newbold. 1999. 'Immunity to non-cerebral severe malaria is acquired after one or two infections.' *Nature Medicine*, 5:3, pp. 340-43.

Guthmann, J.P., A.J. Hall, S. Jaffar, A. Palacios, J. Lines, and A. Llanos-Cuentas. 2001. 'Environmental risk factors for clinical malaria: A case-control study in the Grau region of Peru.' *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95:6, pp. 577-83.

Guthmann, J.P., A. Llanos-Cuentas, A. Palacios, and A.J. Hall. 2002. 'Environmental factors as determinants of malaria risk: A descriptive study on the northern coast of Peru.' *Tropical Medicine and International Health*, 7:6, pp. 518-25.

Guyatt, H.L. and R.W. Snow. 2004. 'Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa.' *Clinical Microbiology Reviews*, 17:4, pp. 760-69.

Gwatkin, D.R. and M. Guillot. 2000. *The burden of disease among the global poor: Current situation, future trends, and implications for strategy.* Health, Nutrition, and Population (HNP) Division. Human Development Network, The World Bank: Washington, DC.

Haddad, L., H. Alderman, S. Appleton, L. Song, and Y. Yohannes. (2003). Reducing child malnutrition: how far does income growth take us? *The World Bank Economic Review* 17(1): 107-131.

Hamoudi, A. and J.D. Sachs. 1999. 'The changing global distribution of malaria: A review.' *CID Working Paper No. 2.* Centre for International Development (CID) at Harvard University: Cambridge.

Hardin, G. 1968. 'The Tragedy of the Commons.' *Science*, 162, pp. 1243-48.

Hati, A.K. 1997. 'Urban malaria vector biology.' *Indian J. Med. Res.*, 106, pp. 149-63.

*Hausman, J. A. (1978). 'Specification tests in econometrics.' *Econometrica* 46(6): 1251-1271.

*Hausman, J. (1983). 'Specification tests and estimation of simultaneous equation models' in *Handbook of Econometrics*. Z. Griliches and M. D. Intrilligator eds. Amsterdam, North Holland. 1: 391-448.

Hausman, J. 2001. 'Mismeasured variables in econometric analysis: problems from the right and problems from the left.' *Journal of Economic Perspectives*, 15, pp. 57-67.

Heggenhougen, H.K. and V. Hackethal, and P. Vivek. 2003. The behavioural and social aspects of malaria and its control: An introduction and annotated bibliography. *Report no. TDR/STR/SEB/VOL/03.1: 214*. World Health Organisation (WHO) on behalf of the Special Programme for Research and Training in Tropical Diseases (TDR): Geneva.

Hentschel, J., J.O. Lanjouw, P. Lanjouw, and J. Poggi. 1998. 'Combining census and survey data to study spatial dimensions of poverty.' *Policy Research Working Paper No. 1928*. Development Research Group and Poverty Reduction and Economic Management Network, Poverty Division, The World Bank: Washington, DC.

Herrel, N., F.P. Amerasinghe, J. Ensink, M. Mukhtar, W. van der Hoek, and F. Konradsen. 2004. 'Adult *anopheline* ecology and malaria transmission in irrigated areas of South Punjab, Pakistan.' *Medical and Veterinary Entomology*, 18:2, pp. 141-52.

Holding, P.A. and P.K. Kitsao-Wekulo. 2004. 'Describing the burden of malaria on child development: what should we be measuring and how should we be measuring it?' *American Journal of Tropical Medicine and Hygiene*, 71:2 Supplement, pp. 71-79.

Hueveline, P., M. Guillot, and D. Gwatkin. 2002. 'The uneven tides of the health transition.' *Social Science and Medicine*, 55:2, pp. 313-22.

Hyma, B., A. Ramesh, and K.P. Chakrapani. 1983. 'Urban malaria control situation and environmental issues, Madras city, India.' *Ecology of Disease*, 2:4, pp. 321-35.

International Institute for Population Sciences. 1995. *National Family Health Survey (MCH and Family Planning), India 1992-93*. IIPS: Bombay, pp 333.

International Institute for Population Sciences and ORC Macro. 2000. *National Family Health Survey (NFHS-2), India 1998-99*. IIPS: Mumbai, pp 340.

International Institute for Population Sciences and ORC Macro. 2008. *National Family Health Survey (NFHS-3), India 2005-06*. IIPS: Mumbai, pp 340.

Indiastat.com. 2008. State-wise distribution of households by social group (urban) in India. Available from: www.indiastat.com

Indiastat.com. 2008. State-wise distribution of households by social group (rural) in India. Available from: www.indiastat.com

Johnson, R.A and D.W. Wichern. 1998. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Jotkar, R., P. Ambhurkar, and R. Mugade. 1997. 'Malariometry in District Ratnagiri during 1988-1993.' *Indian Journal of Public Health*, 41:2, pp. 36-42.

Kakwani, N. (1980). 'On a class of poverty measures.' *Econometrica* 48(2): 437-446.

Kalra, P.K., A.K. Saxena, T.C. Kalra, Y.K. Bichpuriya, and V.P. Singh. 2004. 'Technology and regulation in electricity,' in *India Infrastructure Report 2004*. S. Morris ed. New Delhi and New York: 3iNetwork, Oxford University Press, pp. 348.

Karn, S.K. and H. Harada. 2002. 'Field survey on water supply, sanitation and associated health impacts in urban poor communities – a case from Mumbai city, India.' *Water Science and Technology*, 46:11-12, pp. 269-75.

Kennedy, P. 1998. *A Guide to Econometrics*. Oxford: Blackwell Publishers, pp. 468.

Kijma, Y. 2006. Kijma, Y. (2006). 'Caste and tribe inequality: evidence from India, 1983-1999.' *Economic Development and Cultural Change* 54(2): 369-404.

Kimbi, H.K., N.W. Awah, K.J. Ndamukong, and J.V. Mbuh. 2005. 'Malaria infection and its consequences in school children.' *East African Medical Journal*, 82:2, pp. 92-7.

Konradsen, F., P. Amerasinghe, W. van der Hoek, F. Amerasinghe, D. Perera, and M. Piyaratne. 2003. 'Strong association between house characteristics and malaria vectors in Sri lanka.' *American Journal of Tropical Medicine and Hygiene*, 68:2, pp. 177-81.

Konradsen, F., W. van der Hoek, F.P. Amerasinghe, C. Mutero, and E. Boelee. 2002. 'Engineering and malaria control: learning from the past 100 years,' in Malaria and irrigated agriculture, *IWMI Working Paper 47.* E. Boelee, F. Konradsen and W. van der Hoek eds. Colombo: International Water Management Institute (IWMI).

Konradsen, F., W. van der Hoek, F.P. Amerasinghe, C. Mutero, and E. Boelee. 2004. 'Engineering and malaria control: Learning from the past 100 years.' *Acta Tropica*, 89:2, pp. 99-108.

Konradsen, F., W. van der Hoek, P.H. Amerasinghe, and F.P. Amerasinghe. 1997. 'Measuring the economic cost of malaria to households in Sri Lanka.' *American Journal of Tropical Medicine and Hygiene*, 56:6, pp. 656-60.

Krueger, A.B. and L.H. Summers. 1988. 'Efficieny wages and the inter-industry wage structure.' *Econometrica*, 56:2, pp. 259-93.

Lal, S. , B.M. Vashisht, M.S. Punia, et al.. 1996. 'Containment of outbreak of P. falciparum malaria in community development block Lakhanmajra.' *Indian Journal of Community Medicine*, 21:1-4, pp. 28-33.

Launiala, A. and T. Kulmala. 2006. 'The importance of understanding the local context: women's perceptions and knowledge concerning malaria in pregnancy in rural Malawi.' *Acta Tropica*, 98:2, pp. 111-17.

Leach, G. and M. Gowen. 1987. 'Household Energy Handbook: An Interim Guide and Reference Manual.' *World Bank Technical Report No. 67.* World Bank: Washington, DC.

Lefevre, T., J. L. Todoc, and J. R. Timilsina. 1997. 'Micro factors affecting wood energy consumption.' The role of wood energy in Asia. Food and Agriculture Organisation of the United Nations: Rome, pp 107.

Ludvigson, S., C. Steindel, and M. Lettau. 2002. 'Monetary policy transmission through the consumption-wealth channel.' *FRBNY Economic Policy Review.* The Federal Bank of New York: New York.

Macdonald, G. 1957. *The Epidemiology and Control of Malaria.* London, New York, Toronto: Oxford University Press.

Macintyre, K., J. Keating, S. Sosler, L. Kibe, C.M. Mbogo, A.K. Githeko, and J.C. Beier. 2002. 'Examining the determinants of mosquito-avoidance practices in two Kenyan cities.' *Malaria Journal*, 1:14, pp. 10. Available from: http://www.malariajournal.com/content/pdf/1475-2875-1-14.pdf.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, New York: Cambridge University Press.

Malakooti, M.A., K. Biomndo, and G.D. Shanks. 1998. 'Reemergence of epidemic malaria in the highlands of western Kenya.' *Emerging Infectious Diseases*, 4:4, pp. 661-76.

Malaria Vaccine Initiative (MVI) at PATH (2008). 'Accelerating malaria vaccine development.' Available from http://www.malariavaccine.org/, PATH (Programme for Appropriate Technology in Health).

Malaria Research Centre. 2003. 'Vector biology' in *A Profile of Malaria Research Centre*. Delhi: National Institute of Malaria Research (ICMR), pp. 234.

Mariappan, T., N. Arunachalam, et al.. 1992. 'Note on urban malaria vector *Anopheles stephensi* (Liston) in Cochin.' *Indian Journal of Malariology* 29: 247-249.

Mathur, K.K., G. Harpalani, N.L. Kalra, G.G.K. Murthy, and M.V.V.I. Narasimham. 1992. 'Epidemic of malaria in Barmer district (Thar desert) of Rajasthan during 1990.' *Indian Journal of Malariology*, 29:1, pp. 1-10.

McCarthy, D.F., H. Wolf, and Y. Wu. 2000. 'Malaria and growth.' *Policy Research Working Paper No. 2303*. Development Research Group, World Bank: Washington, DC.

*McFadden, D.L. 1974. 'Conditional logit analysis of qualitative choice analysis,' in *Frontiers in Econometrics*. P. Zarembka ed. New York: Academic Press, pp. 105-42.

MedlinePlus. 2003. 'Medical encyclopedia.' Available from: http://www.nlm.nih.gov/medlineplus/ency/article/000528.htm. Bethesda: U.S. National Library of Medicine.

MedlinePlus. 2006. 'Medical encyclopedia.' Available from: http://www.nlm.nih.gov/medlineplus/ency/article/000528.htm. Bethesda: U.S. National Library of Medicine.

Mishra, V. K., R. D. Retherford, et al.. (1999). Cooking with biomass fuels increases the risk of tuberculosis. *NFHS Bulletin No.13.* Hawaii, East-West Center.

Mishra, A.K., N. Singh, and V.P. Sharma. 1995. 'Use of Neem oil as a mosquito repellent in tribal villages of Mandla district, Madhya Pradesh.' *Indian Journal of Malariology*, 32:3, pp. 99-103.

Moore, S.J., A. Lenglet, and N. Hill. 2002. 'Field evaluation of three plant-based insect repellents against malaria vectors in Vaca Diez Province, the Bolivian Amazon.' *Journal of the American Mosquito Control Association*, 18:2, pp. 107-10.

Mukhtar, M., N. Herrel, F.P. Amerasinghe, J. Ensink, W. van der Hoek, and F. Konradsen. 2003. 'Role of wastewater irrigation in mosquito breeding in South Punjab, Pakistan.' *The Southeast Asian Journal of Tropical Medicine and Public Health*, 34:1, pp. 72-80.

Najera, J.A. 2001. 'Malaria control: achievements, problems and strategies.' *Parassitologia*, Vol. 43: 1-89.

Najera, J.A. and J. Hempel. 1996. 'The Burden of Malaria.' Available from: http://www.rbm.who.int/docs/burden.htm. Roll Back Malaria (RBM) Department, World Health Organisation (WHO): Geneva.

Narayan, D. 2001. 'Consultations with the poor.' *Journal of the Society for International Development*, 44:1, pp. 15-21.

Neumark, D. 1988. 'Employer's discrimination behaviour and the estimation of wage discrimination.' *The Journal of Human Resources*, 23:3, pp. 279-295.

National Institutes of Health (NIH), 2008. 'What is sickle-cell anaemia?' Available from: http://www.nhlbi.nih.gov/health/dci/Diseases/Sca. National Institutes of Health, US Department of Health and Human Services: Washington DC.

National Institute of Malaria Research (NIMR). 2007. Integrated Disease Vector Control Project: A Profile. Available from: http://www.mrcindia.org/idvc-profile. National Institute of Malaria Research: Delhi.

Njie, H.A.B. 2001. 'Poverty and ill-health: The Ugandan national response.' *Journal of the Society for International Development*, 44:1, pp. 93-98.

NMEP. 1985. 'In-depth evaluation report of the modified plan of operation under the National Malaria Eradication Programme of India (NMEP).': 52. NMEP: New Delhi.

Oanda. 2006. 'Oanda.com: The Currency Site.' Available from: http://www.oanda.com/convert/classic. Oanda Corporation: New York.

Oaxaca, R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review*, 14:3, pp. 693-709.

Oaxaca, R. and Ransom, M 1999. 'Identification in detailed wage decomposition.' *The Review of Economics and Statistics*, 81:1, pp. 154-157.

Ozler, B., G. Datt, et al. 1996. A data base on poverty and growth in India. mimeo. Washington, DC, The World Bank.

Pampana, E. 1969. *A Textbook of Malaria Eradication.* London: Oxford University Press.

Pannarunothoi, S. and A. Mills. 1997. 'The poor pay more: Health-related inequality in Thailand.' *Social Science and Medicine*, 44:12, pp. 1781-90.

Pant, C.S., D.K. Gupta, R.M. Bhatt, A.S. Gautam, and R.C. Sharma. 1993. 'Three genetic markers and malaria in upper caste Hindus of Kheda District of Gujarat State.' *Indian Journal of Malariology*, 30:4, pp. 229-33.

Pant, C.S., D.K. Gupta, R.C. Sharma, A.S. Gautam, and R.M. Bhatt. 1992. 'Frequency of ABO blood groups, sickle-cell haemoglobin, G-6-PD deficiency and their relation with malaria in scheduled castes and scheduled tribes of Kheda district, Gujarat.' *Indian Journal of Malariology*, 29:4, pp. 235-39.

Parhar, M. 1998. 'Application of educational technology in adult education.' *Conference on 'Innovation and Professionalization in Adult Literacy: A Focus on Diversity'*, New Delhi, India, February 9-13, 1998. Second Asia Regional Literacy Forum.

Parker, B. and V. Kozel. 2004. 'Understanding poverty and vulnerability in India's Uttar Pradesh and Bihar: A mixed method approach.' Q-squared in practice. *Conference on experiences of combining quantitative and qualitative methods in poverty appraisal*, Toronto, Canada, May 15-16, 2004. Munk Centre for International Studies at Trinity College, University of Toronto and Cornell University, Toronto.

Pattanayak, S., V.P. Sharma, N.L. Kalra, V.S. Orlov, and R.S. Sharma. 1994. 'Malaria paradigms in India and control strategies.' *Indian Journal of Malariology*, 31:4, pp. 141-99.

Patz, J.A., T.K. Graczyk, N. Geller, and A.Y. Vittor. 2000. 'Effects of environmental change on emerging parasitic diseases.' *International Journal for Parasitology*, 30:12-13, pp. 1395-405.

Patz, J.A., K. Strzepek, S. Lele, M. Hedden, S. Greene, B. Noden, S.I. Hay, L. Kalkstein, and J.C. Beier. 1998. 'Predicting key malaria transmission factors, biting and entomological inoculation rates, using modeled soil moisture in Kenya.' *Tropical Medicine and International Health*, 3:10, pp. 818-27.

Pindyck, R.S. and D.L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts.* New York: McGraw-Hill.

Pinikahana J, Dixon RA. 1993. Trends in malaria morbidity and mortality in Sri Lanka. *Indian Journal of Malariology*,30:2, pp. 51-55.

Piyaratne, M.K., F.P. Amerasinghe, P.H. Amerasinghe, and F. Konradsen. 2005. 'Physico-chemical characteristics of Anopheles culicifacies and Anopheles varuna breeding water in a dry zone stream in Sri Lanka.' *Journal of Vector Borne Diseases*, 42:2, pp. 61-7.

Planning Commission (1993). 'Report of the Expert Group on Estimation of Proportion and Number of Poor.' Government of India. New Delhi.

Poolsuwan, S. 1995. 'Malaria in prehistoric southeastern Asia.' *Southeast Asian Journal of Tropical Medicine and Public Health*, 26:1, pp. 3-22.

Prakash, A., P.K. Mohapatra, D.R. Bhattacharyya, B.K. Goswami, and J. Mahanta. 2003. 'Plasmodium ovale: First case report from Assam, India.' *Current Science*, 84:9, pp. 1187-8.

Premasiri, D.A., A.R. Wickremasinghe, D.S. Premasiri, and N. Karunaweera. 2005. 'Malarial vectors in an irrigated rice cultivation area in southern Sri Lanka.' *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99:2, pp. 106-14.

Pritchett, L. and L. Summers. 1996. 'Wealthier is healthier.' *The Journal of Human Resources* 31(4): 841-868.

*Rao, B.A. and M. Nassiruddin. 1945. 'Malaria in Irwin canal area, Mysore state. Part II.' *Journal of Malaria Institute India*, 6, pp. 109-28.

Rich, S. M., M. C. Licht, R. R. Hudson, and F. J. Ayala. 1998. 'Malaria's eve: Evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum.' *Proceedings of the National Academy of Science, USA*, 95:8, pp. 4425-30.

Rickman, L.S., T.R. Jones, G.W. Long, S. Paparello, I. Schneider, C.F. Paul, R. Beaudoin, and S.L. Hoffman. 1990. 'Plasmodium falciparum-infected Anopheles stephensi inconsistently transmit malaria to humans.' *American Journal of Tropical Medicine and Hygiene*, 43:5, pp. 441-45.

Robert, V., K. Macintyre, J. Keating, J. Trape, J. Duchemin, M. Warren, and J.C. Beier. 2003. 'Malaria transmission in urban sub-Saharan Africa.' *American Journal of Tropical Medicine and Hygiene*, 68:2, pp. 169-76.

Roberts, D.R., L.L. Laughlin, P. Hshieh, and L.J. Legters. 1997. 'DDT, global strategies, and a malaria control crisis in South America.' *Emerging Infectious Diseases*, 3:3, pp. 295-302.

Roberts, D.R., S. Manguin, and J. Mouchet. 2000. 'DDT house spraying and re-emerging malaria.' *Lancet*, 356:9226, pp. 330-32.

Rodriguez, A.D., R.P. Penilla, M. Henry-Rodriguez, J. Hemingway, A.F. Betanzos, and J.E. Hernandez-Avila. 2003. 'Knowledge and beliefs about malaria transmission and practices for vector control in southern Mexico.' *Salud publicade de Mexico*, 45:2, pp. 110-16.

*Rueben, R. 1992. 'Women and malaria,' in *Women and Tropical Diseases.* P. Wijeyaratne, E.M. Rathgeber and E. St-Onge eds. Ottawa: International Development Research Centre.

*Rueben, R. and K.N. Panicker. 1979. 'A study of human behaviour influencing man-mosquito contact, and of biting activity on children in a south Indian village community.' *Indian Journal of Medical Research*, 70:Nov, pp. 723-32.

*Russell, P.F. 1952. *Malaria: Basic principles briefly stated.* Oxford: Blackwell.

*Russell, P.F. and F.W. Knipe. 1942. 'A demonstration project in the control of rural malaria by anti-larval measures.' *Journal of Malaria Institute India*, 4, pp. 615-31.

Russell, S. 2005. 'Illuminating cases: understanding the economic burden of illness through case study household research.' *Health Policy and Planning*, 20:5, pp. 277-89.

Ruud, P.A. 2000. *An Introduction to Classical Economic Theory.* New York: Oxford University Press.

Sachs, J. 2002. 'Perspective on improving access to anti-malarial treatment: verbatim of video conference delivered on 2 October 2002.' Available from: www.emro.who.int/RBM/publications/antimalariatreatment.pdf. WHO: Geneva

Sachs, J. and P. Malaney. 2002. 'The economic and social burden of malaria.' *Nature Medicine*, 415:6872, pp. 679-85.

Sahn, D. and D. Stifel. 2000. 'Poverty comparisons over time and across countries in Africa.' *World Development*, 28:12, pp. 2123-55.

Saxena, V.K., M.V. Narasimham, and N.L. Kalra. 1992. 'Critical appraisal of entomological data of Madhya Pradesh for 1991 and its relevance to the National Malaria Eradication Programme.' *The Journal of Communicable Diseases*, 24:2, pp. 97-108.

Sen, A.K. 1987. 'The standard of living: Lecture II, lives and capabilities,' in *The Standard of Living.* G. Hawthorn ed. Cambridge: Cambridge University Press.

Sethi, N.K., Y. Choudhiri, and C.S. Chuttani. 1990. 'Role of migratory population in keeping up endemicity of malaria in metropolitan cities of India.' *Journal of Communicable Diseases*, 22:2, pp. 86-91.

Sharma, S. 1996. *Applied multivariate techniques.* New York, J. Wiley.

Sharma, S.K., R. Chattopadhyay, K. Chakrabarti, S.S. Pati, V.K. Srivastava, P.K. Tyagi, S. Mahanty, S.K. Misra, T. Adak, B.S. Das, and C.E. Chitnis. 2004. 'Epidemiology of malaria transmission and development of natural immunity in a malaria-endemic village, San Dulakudar, in Orissa state, India.' *American Journal of Tropical Medicine and Hygiene*, 71:4, pp. 457-65.

Sharma, S.K., P. Pradhan, and D.M. Padhi. 2001. 'Socio-economic factors associated with malaria in a tribal area of Orissa, India.' *Indian Journal of Public Health*, 45:3, pp. 93-8.

Sharma, S.N., S.K. Subbarao, D.S. Choudhury, and K.C. Pandey. 1993. 'Role of *An. culcifacies* and *An.stephensi* in malaria transmission in urban Delhi.' *Indian Journal of Malariology*, 30, pp. 155-68.

*Sharma, V.P. 1996. 'Malaria: cost to India and future trends.' *South East Asian Journal of Tropical Medicine and Public Health*, 27:1, pp. 4-14.

Sharma, V.P. 1998. 'Fighting malaria in India.' *Current Science*, 75:11, pp. 1127-40.

Sharma. 1999. 'Current scenario of malaria in India.' *Parassitologia*, 41:1-3, pp. 349-53.

Sharma, V.P. 2002. 'Determinants of malaria in South Asia,' in *The contextual determinants of malaria.* E.A. Casman and H. Dowlatabadi eds. Washington, DC: Resources for the future, pp. 382.

Sharma, V.P. and R.C. Sharma. 1989. 'Community based bioenvironmental control of malaria in Kheda district, Gujarat, India.' *Journal of the American Mosquito Control Association*, 5:4, pp. 514-21.

*Shea, J. (1997). 'Instrument relevance in multivariate linear models: A simple measure.' *Review of Economics and Statistics* 79(2): 348-352.

Shukla, R.P., A.C. Pandey, V.K. Kohli, V.P. Ojha, and V.P. Sharma. 1995. 'Bionomics of vector anophelines in District Naini Tal, Uttar Pradesh.' *Indian Journal of Malariology*, 32:4, pp. 153-63.

Singh, N., O. Kataria, and M.P. Singh. 2004a. 'The changing dynamics of Plasmodium vivax and P. falciparum in central India: Trends over a 27-year period (1975-2002).' *Vector-Borne and Zoonotic Diseases*, 4:3, pp. 239-48.

Singh, N., A.C. Nagpal, A. Saxena, and M.P. Singh. 2004b. 'Changing scenario of malaria in central India, the replacement of Plamodium vivax by Plasmodium falciparum (1986-2000).' *Tropical Medicine and International Health*, 9:3, pp. 364-71.

Singh, N. and V.P. Sharma. 2002. 'Patterns of rainfall and malaria in Madhya Pradesh, central India.' *Annals of Tropical Medicine and Parasitology*, 96:4, pp. 349-59.

Snowden, F.M. *The Conquest of Malaria: Italy, 1900–1962* London: Yale University Press

Sogoba, N., S. Doumbiaa, et al.. (2007). 'Malaria transmission dynamics in Niono, Mali: The effect of the irrigation systems.' *Acta Tropica* 101(3): 232-240.

Srivastava, A., B.N. Nagpal, R. Saxena, and S.K. Subbarao. 2001. 'Predictive habitat modeling for forest malaria vector species *An. Dirus* in India: A GIS-based approach.' *Current Science*, 80:9, pp. 1129-34.

Srivastava, A., B.N. Nagpal, R. Saxena, A. Eapen, K.J. Ravindran, S.K. Subbarao, C. Rajamanikam, M. Palanisamy, N.L. Kalra, and N.C. Appavoo. 2003. 'GIS based malaria information management system for urban malaria scheme in India.' *Computer Methods and Programs in Biomedicine*, 71:1, pp. 63-75.

*Staiger, D. and J. H. Stock (1997). 'Instrumental variables regression with weak instruments.' *Econometrica* 65(3): 557-586.

Strauss, J. and D. Thomas. 1998. 'Health, nutrition, and economic development.' *Journal of Economic Literature*, 36:2, pp. 766-817.

Subramanian, S., A. Manoharan, S. Sahu, P. Jambulingam, P. Govardhini, S.S. Mohapatra, and P.K. Das. 1991. 'Living conditions and occurrence of malaria in a rural community.' *Indian Journal of Malariology*, 28:1, pp. 29-37.

Sumodan, P.K., A. Kumar, and R.S. Yadav. 2004. 'Resting behaviour and malaria vector incrimination of *Anopheles stephensi* in Goa, India.' *Journal of the American Mosquito Control Association*, 20:3, pp. 317-18.

Swaminathan, M. 1989. 'Household wealth mobility and its implications for policy: An illustration from a south Indian village.' *Applied Economics Discussion Paper No. 66.* Institute of Economics and Statistics: Oxford.

Tadei, W.P., B.D. Thatcher, J.M.M. Santos, V.M. Scarpassa, I.B. Rodrigues, and M.S. Rafael. 1998. 'Ecologic observations on anopheline vectors of malaria in the Brazilian Amazon.' *American Journal of Tropical Medicine and Hygiene*, 59:2, pp. 325-35.

Thankappan, K.R., C.A.K. Yesudian, P. Sankara Sarma, O. Philio, P. Ramachandran, and A. Joseph. 2006. 'Independent review of national vector borne disease control programme.' Sree Chitra Tirunal Institute for Medical Sciences and Technology.

The Hindu. 2008. 'Malaria death toll reaches 16 in Kanpur.' August 20, 2008.

he World Bank. 2007. 'India - Rural Water Supply.' New Delhi: The World Bank. Available from: www.worldbank.org.in/WBSITE/EXTERNAL/COUNTRIES/SO

Tongren, J.E., C.J. Drakeley, McDonald S.L., H.G. Reyburn, A. Manjurano, W.M. Nkya, M.M. Lemnge, C.D. Gowda, J.E. Todd, P.H. Corran, and E.M. Riley. 2006. 'Target antigen, age, and duration of antigen exposure independently regulate immunoglobulin G subclass switching in malaria.' *Infection and Immunity*, 74:1, pp. 257-64.

Tyagi, B.K. 2004. 'A review of the emergence of Plasmodium falciparum-dominated malaria in irrigated areas of the Thar Desert, India.' *Acta Tropica*, 89:2, pp. 227-39.

Van den Ende, J., L. Lynen, P. Elsen, R. Colebunders, H. Demey, K. Depraetere, K. De Schrijver, W.E. Peetermans, P. Pereira de Almeida, and Vogelaers. D. 1998. 'A cluster of airport malaria in Belgium in 1995.' *Acta Clinica Belgica*, 53:4, pp. 259-63.

Villasenor, J. A. and B. C. Arnold (1989). 'Elliptical Lorenz curves.' *Journal of Econometrics* 40(2): 327-338.

Vlassoff, C. and E. Bonilla. 1994. 'Gender-related differences in the impact of tropical diseases on women: What do we know?' *Journal of Biosocial Science*, 26:1, pp. 37-53.

Vuong, Q. H. 1989. 'Likelihood ratio tests for model selection and non-nested hypotheses.' *Econometrica* 57(2): 307-333.

Wagstaff, A. 2002. 'Poverty and health sector inequalities.' Bulletin of the World Health Organization, 80:2, pp. 97-105.

Watson, M. 1937. 'Malaria and nutrition in Africa.' *Journal of the Royal African Society*, 36:145, pp. 405-20.

Wikipedia. 2006. 'Terai.' Available from: http://en.wikipedia.org/wiki/Terai. Wikimedia: St. Petersburg.

Wilkes, A., Y. Hao, G. Bloom, and G. Xingyuan. 1997. 'Coping with the costs of severe illness in rural China.' *IDS Policy Working Paper 58*. Institute of Development Studies: Brighton.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA, MIT Press.

Wooldridge, J. M. (2002). *Introductory Econometrics: A Modern Approach.* Mason, Ohio, South-Western.

*World Health Organisation (WHO). 1993. 'A global strategy for malaria control.' WHO: Geneva.

World Health Organisation. 2003. 'Global Atlas of Infectious Diseases.' Available from: http://www.who.int/GlobalAtlas/DataQuery/home.asp. Communicable Disease Surveillance and Response (CSR), WHO: Geneva.

World Health Organisation. 2004. 'Frequenty asked questions about malaria.' Available from: http://malaria.who.int. WHO: Geneva

World Health Organisation. 2005a. 'World Malaria Report, 2005.' Available from: http://rbm.who.int/wmr2005/html/toc.htm. Roll Back Malaria, WHO: Geneva.

World Health Organisation. 2005b. 'India: Country Profile' in *The World Malaria Report*. Available from: http://rbm.who.int/wmr2005/html/toc.htm. Roll Back Malaria, WHO: Geneva.

World Health Organisation. 2005c. 'What is malaria? WHO Fact Sheet, No. 94.' Available from: http://mosquito.who.int. Roll Back Malaria (RBM), WHO: Geneva.

World Health Organisation, India. 2005d. *Roll Back Malaria: National anti-malaria programme*. Available from: http://www.whoindia.org/CDS/CD/RBM. Communicable diseases and disease surveillance, Roll Back Malaria (RBM), WHO Representative to India: New Delhi.

World Health Organisation, AFRO. 2006. *Africa Malaria Report 2006*, Available from: http://www.afro.who.int/malaria/publications. WHO Regional Office for Africa: Brazzaville, Congo.

World Health Organisation. 2007. 'The use of anti-malarial drugs: artemisinin and its derivatives'. Available from: http://www.rbm.who.int/cmc upload/0/000/014/923/am2 1-8. Roll Back Malaria (RBM), WHO: Geneva.

World Health Organisation. 2008. 'What is the Roll Back Malaria (RBM) partnership?' Available from: http://www.rbm.who.int/aboutus.html, RBM, WHO: Geneva.

Yadav, R.S., R.M. Bhatt, V.K. Kohli, and V.P. Sharma. 2003. 'The burden of malaria in Ahmedabad city, India: a retrospective analysis of reported cases and deaths.' Annals of Tropical Medicine and Parasitology, 97:8, pp. 793-802.

Yun, M. 2004. 'Decomposing differences in the first moment.' *Economics Letters*, 82, pp. 275-80.

Yun, M. 2005. 'Hypothesis tests when decomposing differences in the first moment.' *Journal of Economic and Social Measurement*, 30, pp. 295-304.

Zanchi, L. 1998. 'Interindustry wage differentials in dummy variable models.' *Economics Letters*, 60:3, pp. 297-301.

Appendix A

Kernel density plots for the asset index

Figure A.1: Kernel density plot for rural asset index, 1992-93

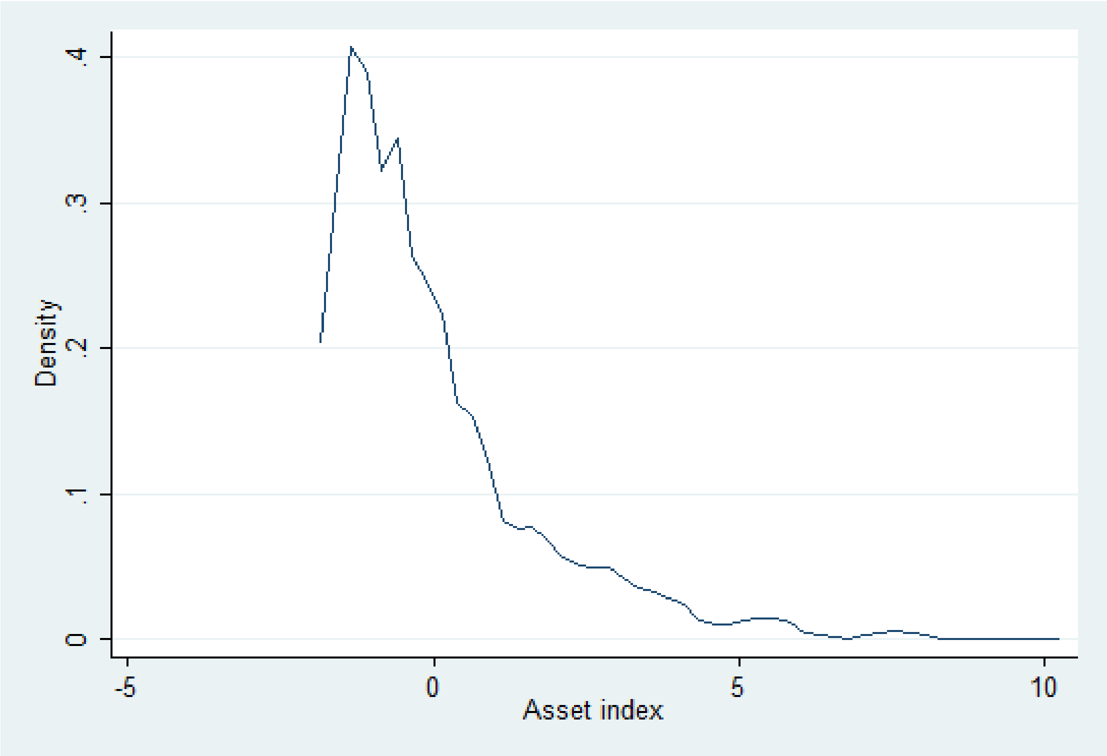


Figure A.2: Kernel density plot for rural asset index, 1998-99

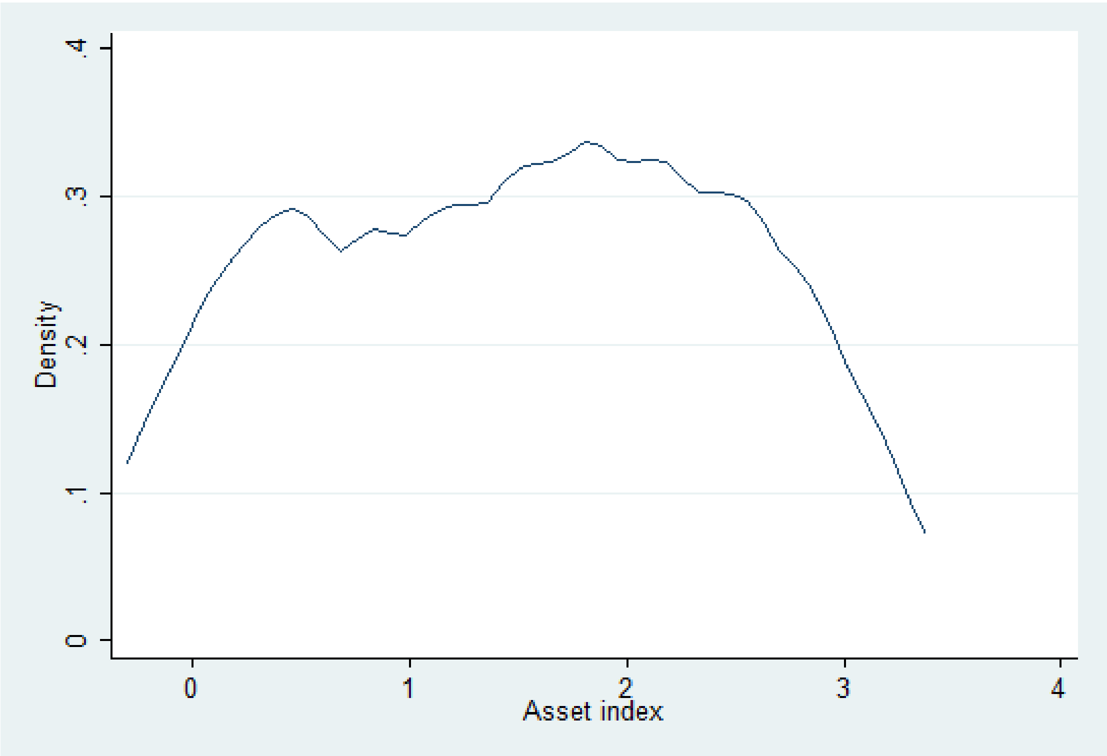


Figure A.3: Kernel density plot for urban asset index, 1992-93

Figure A.4: Kernel density plot for urban asset index, 1998-99

Appendix B

Lorenz curve functional forms used by Ozler et al

The Lorenz curve provides a graphical representation of the data on income distribution by plotting a parametric relation between cumulative income proportions on the y-axis and cumulative population proportions on the x-axis, both arranged in increasing order. The graph is presented in a square box and the line joining (0,0) and (1,1) is the line of perfect equality or the egalitarian line. The LC falls below this line (Vellasenor and Arnold 1989). The LC defined by Gastwirth (1971 in Vellasonor and Arnold 1989) can be given as:

$$L(x) = \int_0^x F^{-1}(y)dy/\mu \quad x \in [0, 1] \tag{B.1}$$

Then using the above equation, $L'(x) = F^{-1}(y)\mu$ where μ is the mean of F.

The head count index, H, which gives the proportion of the population that lives below the poverty line, z, can be obtained using the above relation where:

$$\mu L'(H) = z \tag{B.2}$$

Parametric specifications of the Lorenz curve (LC) have been proposed so as to fit sample LCs from actual data. Two such parameterizations used by Datt and Ravallion are by Kakwani (1980) and Vellasenor and Arnold (1989). Kakwani suggests the following form (evaluated at the value p on the x-axis):

$$L(p) = p - \theta p^{\gamma}(1-p)^{\delta} \tag{B.3}$$

where θ, γ and β are parameters and each is greater than zero. If $p = 0$ and $p = 1$ then $L(p) = 0$ and for L(p)to be convex to the x axis $0 < \gamma \leq < 1$ and $0 < \delta \leq 1$. Using the relation $\mu L'(H) = z$, H can then be obtained as:

$$\theta H^{\gamma}(1-H)^{\delta}[\frac{\gamma}{H} - \frac{\delta}{1-H}] = 1 - z/\mu \tag{B.4}$$

The Vellasenor and Arnold representation derived from the solution to a general quadratic form:

$$ax^2 + bxy + cy^2 + dx + ey + 1 = 0 \tag{B.5}$$

The above equation must satisfy the conditions L(0)=0 and L(1)=1 and the curve must be convex to the x-axis to represent a Lorenz curve. Solving for y in the above equation evaluated at x=p we get:

$$L(p) = -(bp + e) - (mp^2 + np + e^2)^{1/2}/2 \tag{B.6}$$

where m and n are the roots of the quadratic equation and b and e are parameters. Vellasenor and Arnold argue, based on observations of several income datasets, that the elliptical form of the LC is the one that fits the actual data best. If c=0 the above quadratic would present a hyperbola. Since we are dealing with an elliptical curve a standardization of c=1 is used. Further conditions to ensure an elliptical Lorenz curve are: $m = b^2 - 4a < 0$, $a + b + d + 1 > 0$, $d \geq 0$, and $a + d - 1 \leq 0$.

Applying Equation B.2 to Equation B.7, H can be obtained as:

$$H = -[n + r(b + 2z/\mu)(b + 2z/\mu)^2 - m^{1/2}]/2m \tag{B.7}$$

where $r = (n^2 - 4me^2)^{1/2}$

Ozler et al used the LC that gave the best fit to the data. In fact, Datt and Ravallion (1992) find that both the functional forms provide good fits ranging from R-squareds of 0.955 to 1.000.

Appendix C

Taylor series expansion to obtain malaria difference between rural and urban areas

According to the Taylor thorem, given a function $g(.)$ if the value of this function at $x = x_0$ and the value of the derivatives at x_o are known, then the function, g, can be expanded around the point x_o. Thus, for any positive integer n:

$$g(x) = \frac{g(x_o)}{0!} + \frac{g'(x_o)}{1!}(x - x_o) + \frac{g''(x_o)}{2!}(x - x_o)^2 + \ldots... + \frac{g^n(x_o)}{n!}(x - x_o)^n \qquad (C.1)$$

If the second and higher order terms are neglected and we restrict the expansion to n=1, the above equation will be left with the first two terms on the right hand side. This approximation is used to obtain a detailed decomposition of the difference in malaria and the mean value of the characteristics is used in the expansion. The Taylor Series expansion for $n = 1$ is:

$$g(x) = \frac{g(x_o)}{0!} + \frac{g'(x_o)}{1!}(x - x_o) \qquad (C.2)$$

Applying this to the malaria difference equation i.e., substitute the malaria difference from the right hand side of Equation 6.1 as $g(x)$ and let this be linearised around mean characteristic values. Then the part corresponding to $g(x_o)/0!$ in Equation C.2 and evaluated at the mean values of the characteristics vectors would be:

$$[\Phi(\bar{X}_R\beta_R) - \Phi(\bar{X}_U\beta_R) + [\Phi(\bar{X}_U\beta_R) + \Phi(\bar{X}_U\beta_U)] \tag{C.3}$$

The part corresponding to $g'(x_o)(x - x_o)$ is worked out below. Starting with $g'(x_o)$ i.e., obtaining the first derivative, we have :

$$[\phi(\bar{X}_R\beta_R) - \phi(\bar{X}_U\beta_R)] + [\phi(\bar{X}_U\beta_R) - \phi(\bar{X}_U\beta_U)] \tag{C.4}$$

The term, $(x - x_o)$, for the malaria difference function would be:

$$[X_R\beta_R - X_U\beta_R + X_U\beta_R - X_U\beta_U] - [\bar{X}_R\beta_R - \bar{X}_U\beta_R + \bar{X}_U\beta_R - \bar{X}_U\beta_U] \tag{C.5}$$

which, on collecting terms together gives:

$$-(\bar{X}_R - \bar{X}_U)\beta_R\phi(\bar{X}_R\beta_R) - (\beta_R - \beta_U)\bar{X}_U\phi(\bar{X}_U\beta_U) + Rem \tag{C.6}$$

where 'Rem' are the remaining terms.

Following Yun (2004) the malaria difference (let this be represented by $g(x)$) may be expressed as:

$$g(x) = g(x) + g(x_0) - g(x_o) + g'(x_o)(x - x_o) - g'(x_o)(x - x_o) \tag{C.7}$$

Let $R_M = g(x) - g(x_o)$ and $R_T = g(x_o) + g'(x_o)(x - x_o)$ then we have:

$$g(x) = -g'(x_0)(x - x_o) + R_T + R_M \tag{C.8}$$

R_T is the residual from the Taylor Series approximation and R_M from evaluating the function at mean values. Applying this to the malaria difference equation using Equations C.4 and C.6, we have:

$$\bar{y}_R - \bar{y}_U \cong (\bar{X}_R - \bar{X}_U)\beta_R\phi(\bar{X}_R\beta_R) + (\beta_R - \beta_U)\bar{X}_U\phi(\bar{X}_U\beta_U) \tag{C.9}$$

Using the above equation, for each variable, k, the characteristic weight may be obtained as:

$$W^k_{\Delta C} = \frac{(\overline{X}^k_R - \overline{X}^k_U)\hat{\beta}^k_R}{(\overline{X}_R - \overline{X}_U)\hat{\beta}_R} \tag{C.10}$$

where $\sum_{k=1}^{K} W^k_{\Delta C} = 1$

Similarly, for the difference in coefficients, the weight is:

$$W^k_{\Delta D} = \frac{(\hat{\beta}^k_R - \hat{\beta}^k_U)\overline{X}^k_U}{(\hat{\beta}_R - \hat{\beta}_U)\overline{X}_U} \tag{C.11}$$

where $\sum_{k=1}^{K} W^k_{\Delta D} = 1$

The detailed decomposition may then be expressed as:

$$y_R - y_U = \sum_{k=1}^{K} W^k_{\Delta C}[\overline{\Phi(X_R\beta_R)} - \overline{\Phi(X_U\beta_R)}] + \sum_{k=1}^{K} W^k_{\Delta D}[\overline{\Phi(X_U\beta_R)} - \overline{\Phi(X_U\beta_U)}] \tag{C.12}$$

where W stands for the respective weights for each variable, ΔC indicates the difference in characteristics for each variable, k (k=1 to K), and ΔD indicates the difference in coefficients. $W_{\Delta C}$ and $W_{\Delta D}$ are calculated using the mean values of the characteristics and coefficients.

Appendix D

Marginal and impact effects for rural probit regression model for 1992-93

Table D.1: Marginal and impact effects of the probit of malaria incidence with district dummies, 1992-93

Variable	*ME/IE*[a]	*SE*[b]
Asset-index	-0.0003	0.002
Caste	-0.0139**	0.007
Gender	0.0219*	0.011
Age dummies (Reference category: 18-27 years)		
28 to 37 years	-0.0009	0.011
38 to 47 years	0.0125	0.012
48 to 57 years	0.0063	0.012
58 to 67 years	0.0054	0.013
Above 67 years	-0.0021	0.014
Education (Reference category: Illiterate)		
High school	-0.0254***	0.008
Middle school	-0.0150	0.009
Primary school	0.0036	0.007
Occupation (Reference category: production and transport)		
Agriculture	0.0123	0.008
Wage	-0.0115	0.011
Other	0.0287*	0.017
House type (Reference category: semi-*pucca*)		
Pucca house	-0.0012	0.012
Kachcha house	-0.0041	0.007
Separate kitchen	0.0085	0.006
Electricity	-0.0213**	0.008
Sanitation	-0.0064	0.013
Fuel type (Reference category: dung)		
Miscellaneous	-0.0017	0.021
Wood	-0.0263***	0.009

Continued on next page...

Table D.1 continued

District	*ME/IE*	SE
Drinking water (Reference category: protected public)		
Open private	0.0014	0.015
Protected private	-0.0012	0.008
Open public	0.0310*	0.019
Owns a water pump	0.0294***	0.012
Irrigated area (Reference category: no irrigated area)		
Less than 1 to 2 acres	-0.0066	0.009
Above 2 acres	-0.0186*	0.010
Non-irrigated area (Reference category: no non-irrigated area)		
Less than 1 to less than 2 acres	0.0222	0.015
2 acres and above	-0.0143	0.015
Irrigated and non-irrigated area interactions (Reference category: none of either)		
Irrigated till 2, non-irrigated less than 2	-0.0042	0.015
Irrigated till 2, non-irrigated 2 and above	-0.0231	0.019
Irrigated above 2, non-irrigated less than 2	-0.0014	0.028
Irrigated above 2, non-irrigated 2 and above	0.0322	0.034
District fixed effects		
Tehri Garhwal	-0.0744***	0.005
Dehradun	-0.0677***	0.003
Garhwal	-0.0688***	0.003
Almora	-0.0704***	0.003
Nainital	-0.0780***	0.004
Bijnor	-0.0683***	0.004
Moradabad	-0.0679***	0.003
Rampur	-0.0747***	0.004
Saharanpur	-0.0723***	0.004
Muzaffarnagar	-0.0654***	0.004
Meerut	-0.0706***	0.004
Ghaziabad	-0.0695***	0.003
Bulandshahar	-0.0677***	0.004
Aligarh	-0.0656***	0.004
Mathura	-0.0697***	0.003
Agra	-0.0692***	0.004
Firozabad	-0.0637***	0.005
Etah	-0.0551***	0.007
Mainpuri	-0.0594***	0.006
Budaun	-0.0601***	0.006
Bareilly	-0.0614***	0.005
Pilibhit	-0.0698***	0.004
Shahjahanpur	-0.0598***	0.006
Sitapur	-0.0666***	0.004
Hardoi	-0.0692***	0.004
Unnao	-0.0667***	0.004
Rae Bareili	-0.0711***	0.004
Etawah	-0.0843***	0.005
Kanpur Dehat	-0.0716***	0.004
Kanpur Nagar	-0.0626***	0.006
Jalaun	-0.0637***	0.005
Jhansi	-0.0655***	0.004
Hamirpur	-0.0690***	0.003
Banda	-0.0655***	0.004
Fatehpur	-0.0727***	0.004

Continued on next page...

Table D.1 continued

District	*ME/IE*	SE
Pratapgarh	-0.0691***	0.003
Allahabad	-0.0701***	0.004
Bahraich	-0.0680***	0.003
Gonda	-0.0757***	0.004
Barabanki	-0.0676***	0.004
Faizabad	-0.0764***	0.004
Sultanpur	-0.0756***	0.004
Siddhartha Nagar	-0.0744***	0.004
Maharajganj	-0.0711***	0.004
Basti	-0.0684***	0.003
Gorakhpur	-0.0731***	0.004
Deoria	-0.0636***	0.005
Mau	-0.0705***	0.004
Number of observations	7287	
Log Likelihood	-1963.1525	

a. ME/IE: Marginal effects (ME) for continuous variables and impact effects (IE) for dummy variables; b. SE: Robust standard errors adjusted using Huber's procedure; McFadden's adjusted pseudo R-squared = 0.0575; * significant at 10%; ** significant at 5%; *** significant at 1% where the null hypothesis is a test of the coefficient being equal to zero.

Appendix E

Phi Correlation coefficient

In various instances in order to explain the results reference has been made to the correlation coefficient between two variables. Pearson's product moment correlation coefficient is one way of measuring the correlation between two variables X and Y and is given by the following formula:

$$r = \frac{\Sigma_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{[\Sigma_{i=1}^{n}(X_i - \bar{X})^2 \Sigma_{i=1}^{n}(Y_i - \bar{Y})^2]^{1/2}} \tag{E.1}$$

This measure is more suitable for continuous variables but it may also be used to study the relationship between two binary variables (such as in most variables in the current study where the variables take on the value zero or one) provided each of these is not classified into more than two categories (a 2 x 2 contingency table). A more appropriate measure (even though it yields the same result as the Pearson measure for a 2 x 2 table) for categorical variables, which may be employed for reporting the correlation is the Phi coefficient and has been used to report the correlations in the current study given that most variables are dichotomous. For a pair of values (Xi, Yi) depending on which cell each observation belongs to this is effectively the Pearson's product moment correlation coefficient. The method is explained below (Conover, 1999).

Consider a 2-way contingency table between household ownership of a fan and household having access to electricity. Fan takes on the value zero when a household does not own a fan and one when a household owns a fan. Similarly, having electricity takes the value one and not having electricity in the house takes on the value zero. This is shown below where a, b, c, and d refer to the number of households that fall in each category, n is the total number of households, r1 and r2 are the row totals and c1 and c2 are the column totals.

Then the Phi coefficient, Phi r, is given as:

$$Phi\ r = (ad - bc)/(r_1 r_2 c_1 c_2)^{1/2} \tag{E.2}$$

Thus, the Phi r is reported as a positive number. To comment on the sign of the correlation coefficient we need to refer to the Pearson's product moment correlation coefficient.

Appendix F

Marginal effects from the single equation model, 1992-93 rural

Table F.1: Marginal effects from the single equation model, 1992-93 rural

	Single equation (Marginal effects)
PV with a high-school	-0.013***
	[0.003]
PV with an adult literacy centre	-0.002***
	[0.000]
Average distance to drinking water	-0.024***
	[0.006]
PV with a commuting facility	0.001**
	[0.000]
PV with domestic power	-0.001***
	[0.000]
PV with agricultural power	0.001***
	[0.000]
PD irrigated by a government canal	0.002***
	[0.001]
PD irrigated by private canal	0.094***
	[0.019]
PD irrigated by a well	0.020***
	[0.005]
PD irrigated by a tube well	0.001***
	[0.000]
PD irrigated by an electric tube well	0.001**
	[0.000]
PD under forest	-0.001**
	[0.000]
District population density	-0.014***
	[0.004]
Log rain	-0.056***
	[0.019]
Number of observations	7287

PV stands for percentage of villages (in the district) and PD for percentage of district area; other variables from the first-stage probit regression were also included in the above regression; * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix G

Marginal and impact effects for comparable household rural specifications of 1992-93 and 1998-99

Table G.1: Marginal effects for 1992-93 and 1998-99: comparable rural specifications

Variable	*ME/IE:1992–93*	*ME/IE: 1998–99*
Caste	-0.014*	0.004
	[0.008]	[0.007]
Gender	0.022	-0.058***
	[0.013]	[0.020]
Age dummies (Reference category: 18-27 years)		
28 to 37 years	-0.001	0.011
	[0.013]	[0.011]
38 to 47 years	0.012	-0.003
	[0.013]	[0.010]
48 to 57 years	0.006	0.016
	[0.014]	[0.013]
58 to 67 years	0.006	0.017
	[0.014]	[0.014]
Above 67 years	-0.007	0.027*
	[0.015]	[0.018]
Education (Reference category: Illiterate)		
High school	-0.027**	0.019*
	[0.010]	[0.011]
Middle school	-0.012	0.010
	[0.010]	[0.011]
Primary school	0.007	0.003
	[0.009]	[0.007]

Continued on next page...

Table G.1 continued

Variable	*ME/IE: 1992–93*	*ME/IE: 1998–99*
Occupation (Reference category: production and transport)		
Agriculture	0.011	0.005
	[0.010]	[0.008]
Wage	-0.016	-0.007
	[0.012]	[0.009]
Other	0.024	-0.015
	[0.018]	[0.008]
House type (Reference category: semi-*pucca*)		
Pucca house	-0.003	-0.005
	[0.013]	[0.011]
Kachcha house	-0.007	0.007
	[0.008]	[0.006]
Separate kitchen	0.001	-0.007
	[0.007]	[0.006]
Electricity	-0.021**	-0.006
	[0.010]	[0.008]
Sanitation	-0.013	0.001
	[0.015]	[0.010]
Fuel type (Reference category: dung)		
Miscellaneous	0.007	-0.018
	[0.026]	[0.009]
Wood	-0.028***	0.013*
	[0.011]	[0.006]
Drinking water (Reference category: protected public)		
Open	0.053**	0.018
	[0.026]	[0.057]
Protected private	.00002	-0.008
	[0.008]	[0.006]
Agricultural land	-0.005	-0.010
	[0.010]	[0.008]
Owns a water pump	0.022*	0.019*
	[0.013]	[0.012]
Owns a bullock cart	-0.002	0.020**
	[0.013]	[0.012]
Owns a tractor	0.007	0.032*
	[0.023]	[0.022]
Asset index	-0.002	-0.004*
	[0.003]	[0.002]
District fixed effects		
Tehri Garhwal	-0.082***	-0.039***
	[0.006]	[0.003]
Dehradun	-0.075***	-0.036***
	[0.004]	[0.003]
Nainital	-0.088***	-0.038***
	[0.005]	[0.003]
Bijnor	-0.076***	-0.038***
	[0.004]	[0.003]
Moradabad	-0.075***	-0.037***
	[0.004]	[0.004]
Rampur	-0.084***	-0.015
	[0.004]	[0.014]
Saharanpur	-0.081***	-0.037***

Continued on next page. . .

Table G.1 continued

Variable	*ME/IE: 1992–93*	*ME/IE: 1998–99*
	[0.004]	[0.003]
Muzaffarngar	-0.072***	-0.040***
	[0.005]	[0.003]***
Meerut	-0.078***	-0.034
	[0.004]	[0.004]***
Ghaziabad	-0.077***	-0.038
	[0.004]	[0.003]***
Bullandshahar	-0.075***	-0.039
	[0.005]	[0.003]***
Aligarh	-0.072***	-0.034
	[0.005]	[0.004]***
Agra	-0.077***	-0.04
	[0.004]	[0.003]***
Firozabad	-0.070***	-0.036***
	[0.006]	[0.003]
Etah	-0.059***	-0.038***
	[0.009]	[0.004]
Mainpuri	-0.064***	-0.038***
	[0.008]	[0.003]
Budaon	-0.065***	-0.034***
	[0.007]	[0.004]
Bareilly	-0.067***	-0.039***
	[0.007]	[0.003]
Pilibhit	-0.078***	-0.035***
	[0.004]	[0.004]
Shahjahanpur	-0.064***	-0.042***
	[0.008]	[0.004]
Sitapur	-0.073***	-0.041***
	[0.005]	[0.003]
Hardoi	-0.077***	-0.040***
	[0.004]	[0.003]
Unnao	-0.074***	-0.037***
	[0.004]	[0.003]
Rae Bareili	-0.080***	-0.038***
	[0.005]	[0.003]
Kanpur Dehat	-0.080***	-0.036***
	[0.004]	[0.003]
Kanpur Nagar	-0.068***	-0.039***
	[0.008]	[0.003]
Jalaun	-0.070***	-0.033***
	[0.006]	[0.004]
Jhansi	-0.072***	-0.038***
	[0.005]	[0.004]
Hamirpur	-0.077***	-0.041***
	[0.004]	[0.003]
Banda	-0.072***	-0.040***
	[0.005]	[0.003]
Fatehpur	-0.081***	-0.037***
	[0.004]	[0.003]
Pratapgarh	-0.077***	-0.041***
	[0.004]	[0.003]
Allahabad	-0.078***	-0.041***

Continued on next page...

Table G.1 continued

Variable	*ME/IE: 1992–93*	*ME/IE: 1998–99*
	[0.004]	[0.004]
Bahraich	-0.075***	-0.039***
	[0.004]	[0.004]
Gonda	-0.085***	-0.042***
	[0.004]	[0.004]
Barabanki	-0.075***	-0.040***
	[0.005]	[0.003]
Faizabad	-0.086***	-0.038***
	[0.005]	[0.003]
Sultanpur	-0.085***	-0.039***
	[0.004]	[0.004]
Maharajganj	-0.079***	-0.038***
	[0.004]	[0.003]
Basti	-0.076***	-0.044***
	[0.004]	[0.004]
Deoria	-0.070***	-0.044***
	[0.006]	[0.004]
Mau	-0.079***	-0.034***
	[0.004]	[0.004]
	6071	4588
Number of observations	6071	4588
Log likelihood	-1742.7	-836.24

ME/IE: Marginal effects for continuous and impact effects for dummy variables; Robust standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix H

Marginal and impact effects for comparable rural and urban probit models for 1992-93 and 1998-99

Table H.1: Marginal and impact effects for comparable rural and urban probit models for 1992-93 and 1998-99

	Rural ME/IE 1992-93	*Urban ME/IE 1992-93*	*Rural ME/IE 1998-99*	*Urban ME/IE 1998-99*
Caste	-0.006	-0.007	0.003	-0.008***
	[0.008]	[0.004]	[0.007]	[0.003]
Gender	0.011	-0.010	-0.054***	0.006
	[0.014]	[0.012]	[0.020]	[0.003]
Age dummies (Ref: above 57 years)				
18 to 37 years	0.000	-0.002	-0.008	0.002
	[0.010]	[0.006]	[0.008]	[0.005]
38 to 47 years	0.015	0.008	-0.018**	-0.004
	[0.010]	[0.007]	[0.008]	[0.004]
48 to 57 years	0.005	0.007	-0.008	0.001
	[0.011]	[0.008]	[0.009]	[0.005]
Education (Ref: illiterate)				
High school	-0.016	-0.004	0.017	-0.003
	[0.011]	[0.007]	[0.012]	[0.005]
Middle school	-0.015	0.000	0.006	-0.006*
	[0.011]	[0.007]	[0.011]	[0.003]
Primary school	0.012	0.017**	0.002	-0.002
	[0.009]	[0.009]	[0.008]	[0.004]
Occupation (Ref: production and transport)				
Agriculture	0.010	0.002	0.007	0.020*
	[0.009]	[0.010]	[0.008]	[0.020]
Wage	-0.010	0.004	-0.004	-0.006*
	[0.013]	[0.005]	[0.011]	[0.004]
Other	0.026	0.011	-0.017	-0.004

Continued on next page...

Table H.1 continued

	Rural ME/IE 1992-93	*Urban ME/IE 1992-93*	*Rural ME/IE 1998-99*	*Urban ME/IE 1998-99*
	[0.018]	[0.011]	[0.009]	[0.004]
House type (Ref: semi-*pucca*)				
Pucca house	0.013	0.008*	-0.005	0.001
	[0.016]	[0.005]	[0.012]	[0.004]
Kachcha house	-0.004	0.019*	0.005	0.002
	[0.008]	[0.013]	[0.007]	[0.006]
Separate kitchen	0.015**	-0.015**	-0.004	-0.009*
	[0.007]	[0.006]	[0.007]	[0.005]
Electricity	-0.033***	-0.002	-0.002	-0.007
	[0.008]	[0.006]	[0.009]	[0.008]
Sanitation	-0.021	0.010**	0.015	0.005
	[0.013]	[0.003]	[0.012]	[0.004]
Fuel type (Ref: dung)				
Miscellaneous	-0.020	0.010	-0.029***	-0.005
	[0.019]	[0.009]	[0.008]	[0.008]
Wood	-0.054***	0.017*	0.012*	-.0004
	[0.011]	[0.012]	[0.007]	[0.007]
Drinking water (Ref: protected public)				
Open	0.016	-0.006	-0.007	0.001
	[0.013]	[0.008]	[0.007]	[0.010]
Protected private	-0.017**	-0.003	-0.038**	-0.001
	[0.007]	[0.006]	[0.007]	[0.005]
Agricultural land	-0.003	0.008	-0.007	0.001
	[0.009]	[0.007]	[0.008]	[0.005]
Asset index	0.0004	-0.007**	-0.004	0.0004
	[0.006]	[0.003]	[0.003]	[0.001]
Regions Hill	0.000	-0.012**	0.002	-0.005*
	[0.012]	[0.005]	[0.022]	[0.003]
East	-0.042***	-0.010*	0.004	-0.004
	[0.009]	[0.003]	[0.011]	[0.004]
West	-0.002	-0.004	0.037***	-0.008**
	[0.009]	[0.005]	[0.012]	[0.004]
Bundelkhand	-0.018	0.004	0.032**	-0.007
	[0.017]	[0.008]	[0.018]	[0.003]
Number of observations	6582	2220	4103	1680
Log pseudo likelihood	-1840.75	-210.10	-763.9	-109.88
Pseudo R-squared	0.0342	0.1607	0.0105	-0.0882

Note: Robust standard errors in parantheses; ME/IE indicates marginal effects for continuous and impact effects for dummy variables; Pseudo R-squareds reported are the McFadden's adjusted pseudo R-squareds; * significant at 10%; ** significant at 5%; *** significant at 1%

Appendix I

Marginal and impact effects of urban probit using quintiles instead of an aggregated asset index: 1992-93 and 1998-99

Table I.1: Marginal and impact effects of urban probit using quintiles instead of an aggregated asset-index: 1992-93 and 1998-99

Variable	*ME/IE[a] 1992-93*	*ME/IE 1998-99*
Asset-index quintiles		
Bottom 20%	-0.008	0.007**
	[0.005]	[0.004]
20–40%	0.004	0.002
	[0.004]	[0.005]
40–60%	-0.006	-0.016**
	[0.007]	[0.007]
60–80%	-0.014	0.017**
	[0.011]	[0.008]
Top 20%	0.004	
	[0.012]	
Caste	-0.007	-0.008**
	[0.004]	[0.003]
Gender	-0.009	0.005
	[0.011]	[0.003]
Age dummies (Ref: above 57 years)		
18 to 37 years	-0.002	0.001
	[0.005]	[0.005]
38 to 47 years	0.008	-0.005

Continued on next page. . .

Table I.1 continued

Variable	*ME/IE*[a] *1992-93*	*ME/IE 1998-99*
	[0.007]	[0.004]
48 to 57 years	0.006	-0.0005
	[0.007]	[0.005]
Education (Ref: illiterate)		
High school	-0.003	-0.003
	[0.006]	[0.004]
Middle school	0.0004	-0.006
	[0.006]	[0.003]
Primary school	0.016**	-0.003
	[0.009]	[0.004]
Occupation (Ref: production and transport)		
Agriculture	0.002	0.017
	[0.009]	[0.017]
Wage	0.004	-0.005*
	[0.004]	[0.003]
Other	0.010	-0.004
	[0.010]	[0.003]
House type (Ref: semi-*pucca*)		
Pucca house	0.007	0.001
	[0.004]	[0.004]
Kachcha house	0.017**	0.004
	[0.011]	[0.008]
Separate kitchen	-0.013***	-0.008*
	[0.006]	[0.004]
Electricity	-0.003	-0.011
	[0.006]	[0.009]
Sanitation	0.008**	0.005
	[0.003]	[0.003]
Fuel type (Ref: dung)		
Miscellaneous	0.009	-0.004
	[0.008]	[0.007]
Wood	0.015*	0.0004
	[0.010]	[0.006]
Drinking water (Ref: protected public)		
Open	-0.005	0.003
	[0.005]	[0.012]
Protected private	-0.003	-0.0008
	[0.006]	[0.005]
Agricultural land	0.006	0.002
	[0.006]	[0.005]
Regions		
Hill	-0.011**	-0.005
	[0.004]	[0.003]
East	-0.009*	-0.004

Continued on next page...

Table I.1 continued

Variable	*ME/IE[a] 1992-93*	*ME/IE 1998-99*
	[0.003]	[0.004]
West	-0.004	-0.008**
	[0.005]	[0.004]
Bundelkhand	0.004	-0.007
	[0.007]	[0.003]
Number of observations	2220	1680
Log likelihood	-207.86	-102.71
Pseudo R-squared	0.1696	0.1835

Note: a. ME/IE indicates impact/marginal effects; robust standard errors in brackets, adjusted using Huber's procedure; * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix J

Means and standard deviations of levels and first-differenced variables

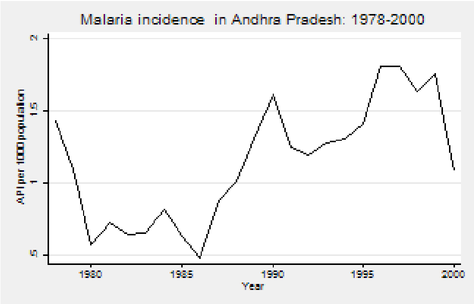
Table J.1: Means and standard deviations of levels and first differenced variables used in the regression analysis

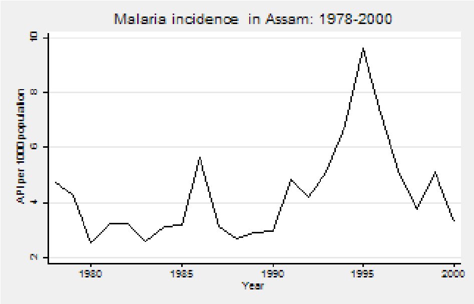
First differenced variables	Mean	SD[a]	Variables in levels	Mean	SD
D1lAPI	-0.07	0.42	**lAPI**	0.65	1.25
D1popden	6.53	3.39	**popden**	345.76	189.81
D1urban	0.26	0.20	**urban**	24.51	8.17
D1rain	-13.75	560.60	**rain**	2330.67	1366.40
D1irrig	0.28	1.23	**irrig**	21.22	19.14
D1stinc	59.16	107.12	**stinc**	1468.25	593.94
D2health	0.23	2.32	**healthlag**	16.07	3.83
D1eduex	2.80	5.15	**eduex**	53.47	19.95
D1povertyr	-0.93	5.61	**povertyr**	42.01	13.43
D1povertyu	-1.08	3.81	**povertyu**	34.53	11.80

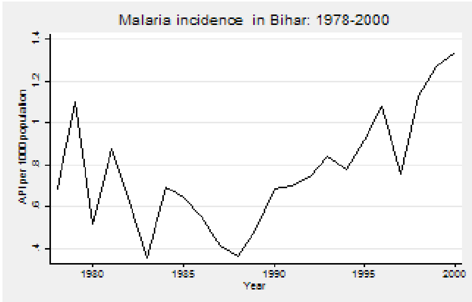
a. SD: Standard deviation

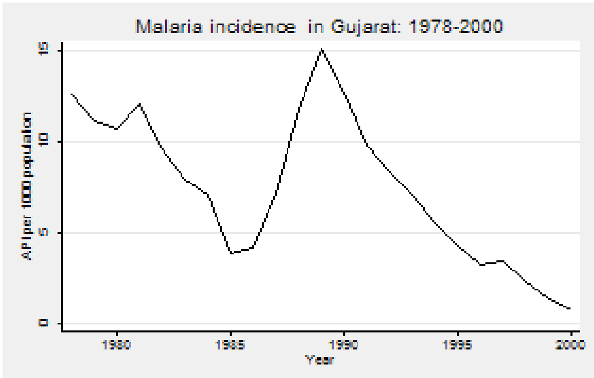
Appendix K

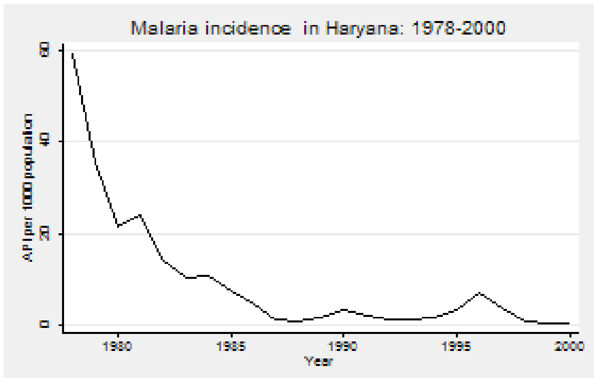
Malaria incidence measured as annual parasite incidence (API) per 1000 population for 15 major states in India: 1978-2000

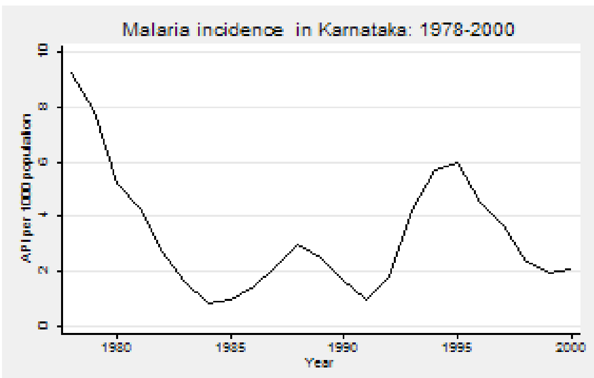


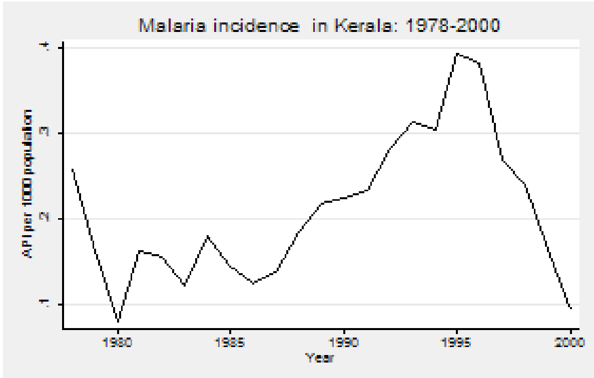


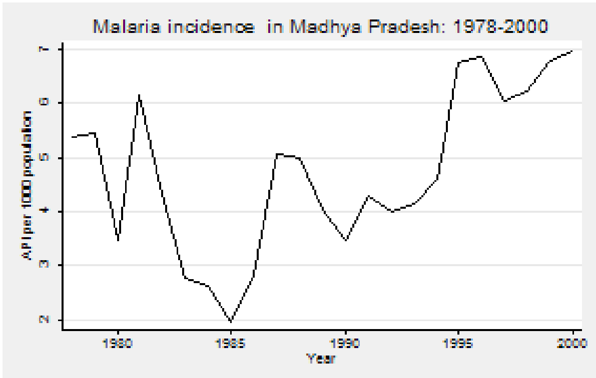

Malaria incidence in Bihar: 1978-2000
API per 1000 population
Year

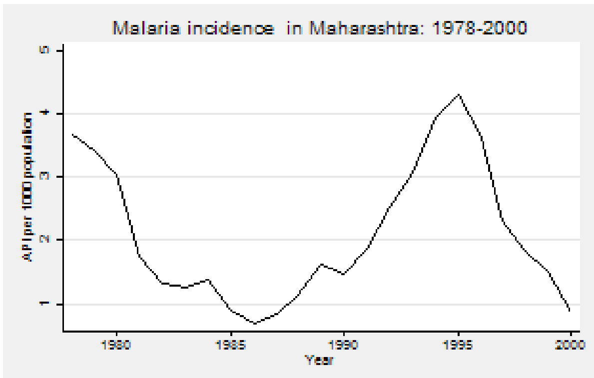

Malaria incidence in Gujarat: 1978-2000
API per 1000 population
Year

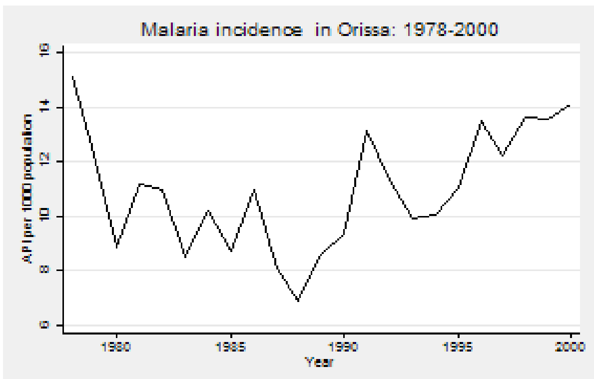

Malaria incidence in Haryana: 1978-2000
API per 1000 population
Year

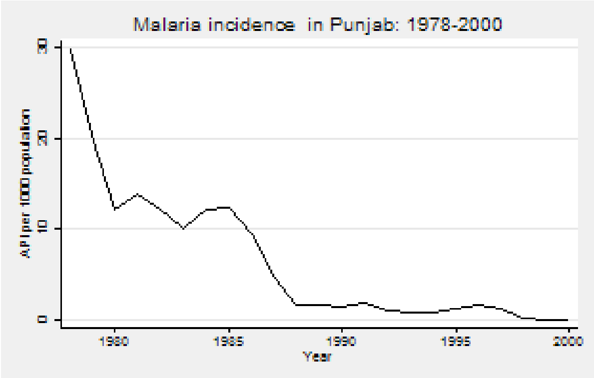

Malaria incidence in Karnataka: 1978-2000
API per 1000 population
Year

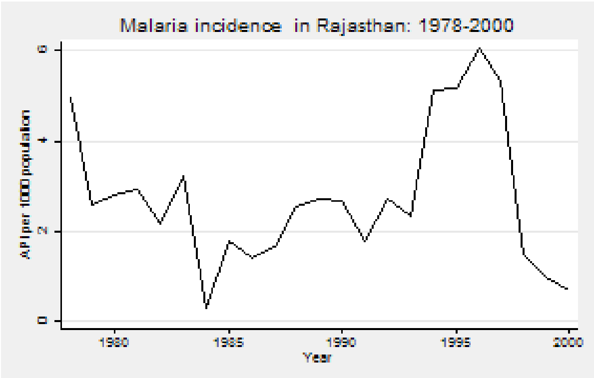

Malaria incidence in Kerala: 1978-2000
API per 1000 population
.1
.2
.3
.4
1980
1985
1990
1995
2000
Year

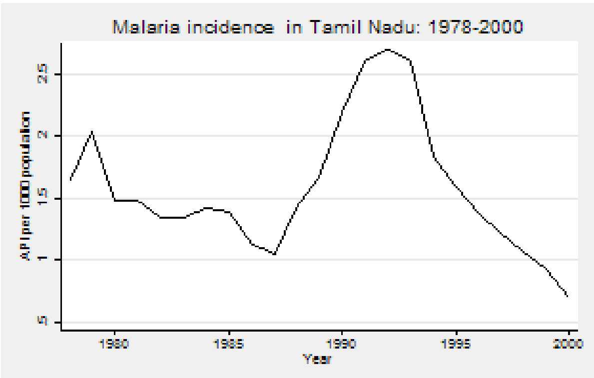

Malaria incidence in Madhya Pradesh: 1978-2000
API per 1000 population
2
3
4
5
6
7
1980
1985
1990
1995
2000
Year

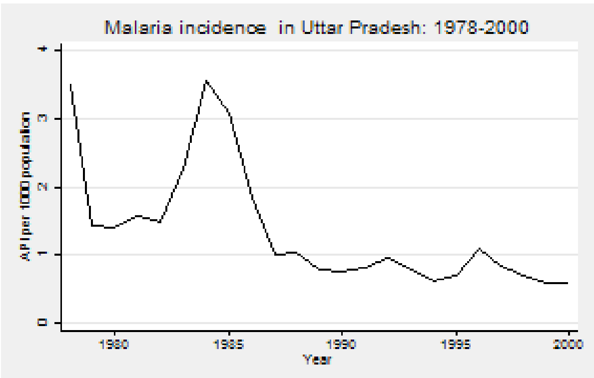

Malaria incidence in Maharashtra: 1978-2000
API per 1000 population
1
2
3
4
5
1980
1985
1990
1995
2000
Year

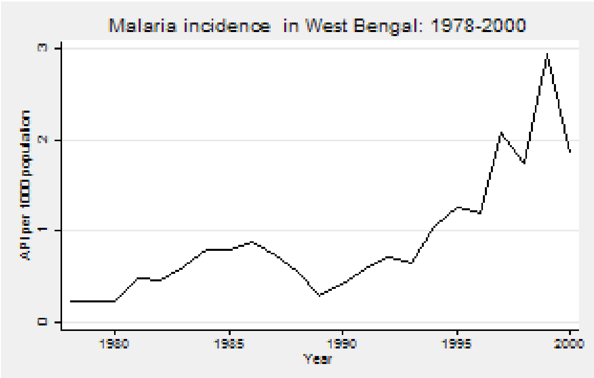

Malaria incidence in Orissa: 1978-2000
API per 1000 population
6
8
10
12
14
16
1980
1985
1990
1995
2000
Year


Malaria incidence in Punjab: 1978-2000
API per 1000 population
0
10
20
30
1980
1985
1990
1995
2000
Year


Malaria incidence in Rajasthan: 1978-2000
API per 1000 population
0
2
4
6
1980
1985
1990
1995
2000
Year


Malaria incidence in Tamil Nadu: 1978-2000
API per 1000 population
.5
1
1.5
2
2.5
1980
1985
1990
1995
2000
Year


Malaria incidence in Uttar Pradesh: 1978-2000
API per 1000 population
0
1
2
3
4
1980
1985
1990
1995
2000
Year


Malaria incidence in West Bengal: 1978-2000
API per 1000 population
Year
0
1
2
3
1980
1985
1990
1995
2000

Appendix L

Effect of income and other variables on malaria incidence

Table L.1: Effect of income and other variables on malaria incidence using a log-log model where log(API)=dependent variable and all explanatory variables are also logged.

Variable	*Coefficient*	*Robust standard error*
lpopden	-8.865***	2.75
lpercU	-4.307***	1.94
lrain	0.197	0.26
lirrPerc	0.471	0.43
lstinc	-1.052*	0.56
lhealthlag	-0.049	0.35
leduex	1.084**	0.54
lpovertyr	-0.106	0.16
lpovertyu	0.915***	0.34
Constant	60.809***	20.85
Observations	330	
Number of states	15	
Within R-squared	0.25	

The above regression includes time dummies as per Equation 7.1 from 1979 to 2000 using 1979 as the base. For full forms of the variables refer to Table 7.1, p.171.