EFFECT OF HOUSEHOLD DYNAMICS ON RISK OF DISEASE ASSOCIATED WITH HOUSEHOLD CONTACT

Tobias Freeman Chirwa

Department of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine

Thesis submitted for the degree of Doctor of Philosophy
Faculty of Medicine, University of London
October 2001

ABSTRACT

Household contact is known to be a risk factor for transmission of many infections, and the magnitude of such contact-associated risk is a classic measure of transmissibility. The risk associated with household contact may be underestimated due to misclassification of contact status of some individuals, particularly in cohort studies of diseases with long incubation period. Such studies generally begin with contact defined at a single point in time, thus a "snap-shot" of dynamic households. However, individuals change households, form new households, die, or migrate over time. Thus, some individuals who experienced household contact may be misclassified as non-contacts.

Published analyses of Karonga Prevention Study (KPS) data have indicated that household contact with active paucibacillary or multibacillary leprosy as assessed during a survey carried out 1980-84 (LEP-1) imparted two- or fivefold increased risk of leprosy, respectively, compared to individuals not living in such households. This was as assessed in a second survey carried out 1986-89 (LEP-2). The current project began as an investigation of the implications of household dynamics on these measures of household contact associated risk, and evolved into a broad consideration of household dynamics, touching upon a variety of demographic and epidemiological issues.

The approach included detailed analysis of KPS data and development of a simulation model of household dynamics tracing contact status over a period of time, to quantify contact status misclassification and estimate the "true" underlying rate ratios adjusting for this misclassification. Not even such a model captures all the selective household changes of a rural society, and there will be many unrecorded and unrecordable contacts.

A total of 112886 individuals were interviewed in LEP-1, of whom about 85,000 were examined in LEP-2. 46% of this population was under 15 years of age.

Procedures for smoothing the age distribution initially characterized by age heaping (a direct result of birth estimates) and for correcting for the under-ascertainment of infants, especially common in studies when reporting of birth dates is poor, are explained.

The crude birth and death rate were estimated to be 49 births per 1000 persons and 10 deaths per 1000 persons respectively. Under-5 mortality was estimated to be about 250 deaths per 1000 live births. Estimates of mortality adjusted for age, sex and socio-economic factors show interesting patterns. Mortality was higher in north rather than south Karonga (rate ratio of 1.29, 95% CI: 1.19, 1.38); and in those with estimated rather than precise years of birth (rate ratio of 1.14, 95% CI: 1.03, 1.25). No significant differences in mortality were found between leprosy cases and non-cases, 1.14 (95% CI: 0.84, 1.54). The finding of significantly lower mortality among those with compared to those without a BCG scar, rate ratio of 0.70 (95% CI: 0.64, 0.76) was surprising though we suspect that it reflects residual socio-economic confounding rather than a biological effect of the vaccine.

The mean and median household size (6.42 and 5 respectively) were similar for the LEP-1 and LEP-2 surveys. 85% of heads of households were male with their mean age, 47 (s.d. 14.6) lower than that for females, 55 (s.d. 14.2). There was a very high rate of household change among 15-29 year olds with a higher rate for females in the lower part of this age band (approximately 63% over 5 years).

A household dynamics model was constructed in order to simulate births, deaths, in- and out-migrations, marriages and movement between households on an annual basis. Its parameters are derived from the LEP and census data.

The contact status misclassification rate is defined as the proportion of all individuals in contact with at least 1 index case in the simulations who were initially classified as non-contacts. The model results show high contact status

misclassification in particular among the 15-29 year olds (largely a reflection of active household change).

Improved estimates of contact associated leprosy risks showed higher rate ratios in young children than adults. Apart from attributing such results to intensity of contact and sharing of environmental factors with source cases, they are also consistent with (but do not confirm) genetic susceptibility.

Apart from investigating household contact-associated risks of disease, the analysis of household dynamics in this thesis provides methods and baseline measures for understanding demographic and social pattern changes, of particular importance in this era of HIV.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Professor Paul Fine for his excellent ideas and supervision, which have contributed to the compilation of this thesis. It was such an honour working with him and I have gained quite a lot from his experience and wisdom. His commitment and interest in my work has been unsurpassed.

I would like to thank WHO/TDR for their financial support throughout my studies at London School of Hygiene and Tropical Medicine, without whose support publication of this thesis would not have been possible.

Many thanks to Basia Zaba, James Carpenter and Sian Floyd for their continued advice and useful comments on several aspects of this project. I am indebted to Emily Grundy for pointing me to literature on household dynamics models, to Chris Frost for literature on measurement error and to Jerome Huillard for assistance in SAS programming. I particularly want to thank Sian for her encouragement during the times when I was down and seemed to make little progress.

I would also like to thank Lyn Bliss and Keith Branson for their support on extraction of relevant data for this work. Their valuable data management skills and experience gained from the Karonga Prevention Study (KPS) were extended to me. Other members of the KPS both in London (UK) and Karonga (Malawi) have also helped me understand the KPS better and provided a friendly working environment.

Last but not least, I would like to thank my wife Esnat and daughter Temweka for their moral support, understanding and patience during the period of my studies. This work has been developed at the expense of their time and I owe it to them. I love them dearly.

LIST OF ABBREVIATIONS

CEB/CS	Children Ever Born/Children Surviving
CIR	Community Infection Ratio
DSS	Demographic Surveillance System
ICDDR	International Centre for Diarrhoeal Disease Research
KAS	Kasyata area
KPS	Karonga Prevention Study
KPT	Karonga Prevention Trial
LEPRA	Leprosy Relief Association
LEP-1	LEPRA Evaluation Project, first survey (1979-84)
LEP-2	LEPRA Evaluation Project, follow-up survey (1986-89)
LSHTM	London School of Hygiene and Tropical Medicine
MB	Multi-bacillary leprosy
MDHS	Malawi Demographic and Health Survey
MH	Mantel Haenszel
PAF	Population Attributable Fraction
PB	Pauci-bacillary leprosy
PYAR	Person years at risk
RFLP	Restriction Fragment Length Polymorphism
RR	Rate ratio (or Relative risk)
SAR	Secondary Attack Rate
SS	Sample Survey
SS/KAS	Sample Survey or Kasyata areas
SS/KKL	Sample Survey, Kasowa, Kasyata and Lower Songwe areas
TFR	Total Fertility Rate
UN	United Nations
WHO	World Health Organisation
WHO/TDR	Special Programme for Research and Training in Tropical Diseases

CHAPTER 1 INTRODUCTION

1.1 Rationale for the study

This research stems from a large epidemiological study of leprosy (and tuberculosis) in Karonga District, Northern Malawi (1). The epidemiological study began as a total population survey, called the LEPRA Evaluation Project (LEP). The study was started in 1979 and in later years it developed into the largest vaccine trial in Africa (2). It initiated several studies to investigate demographic, genetic, environmental and immunological factors underlying the patterns of mycobacterial disease and the behaviour of mycobacterial vaccines in this population. This particular project is an investigation of demographic issues (in particular, household dynamics) that relate to disease.

Household and dwelling contact are known risk factors for many contagious infections. Several studies have attempted to measure this contact-associated risk for various infections by following up individuals observed to be in household contact with cases. This is a relatively straightforward exercise for acute infections, with incubation periods of days or weeks, but special problems arise in studies of infections with long incubation periods. The risk associated with household contact may be underestimated due to misclassification of contact status of some individuals with source cases, particularly in cohort studies of diseases with long incubation periods. Such misclassification is likely to be most pronounced if the follow-up period is long. This is due to the dynamic nature of populations, which involves people changing households, forming new households, dying or migrating. Since studies to measure risks associated with household contact generally begin with contact defined at a single point in time, thus a "snap shot" of dynamic households, many individuals who experience household contact with cases at some point in time may be misclassified as non-contacts. This increases the risk in the reference (apparent non-contact) group, and results in under-estimation of the relative risk of disease associated with household or dwelling contact.

Of particular interest in the proposed study are the implications of household dynamics on leprosy risk in households. This is because detailed data on contact with leprosy cases and household membership, as well as on leprosy incidence, were available from the work in Karonga District over a long period of time. Although there is particular reference to leprosy, the concepts developed in this study are also applicable in general to other chronic diseases with long incubation periods, such as tuberculosis.

The long incubation period of leprosy disease has additional implications for this project. Because of this long and varied incubation period, some incident cases that arise during a follow-up period of study may be attributable to contact which occurred prior to the period of study rather than to contact recognised at the start of the study or during follow-up. This is unlikely to occur if the incubation period of a disease is short, in which case most contacts that lead to disease will occur during a study period. It is difficult to trace contact history for a population over a long period, and this can lead to both low sensitivity and low specificity of contact classification. These problems raise questions of how to take prior contact into account when investigating risk of disease associated with household contact. Obviously, prior contact will be most important when considering infections/diseases with long incubation periods.

Since the main epidemiological study in Karonga began prior to the HIV pandemic, but the project has now expanded to include work on HIV, an analysis of household dynamics has important implications beyond the study of leprosy, for which it was first designed. These early data provide baseline measurements against what is happening now in the population as a consequence of this new, now widespread, devastating and socially disruptive infection.

1.2 Objectives of the study

1. To describe demographic characteristics of the population and the dynamic nature of households in Karonga District, northern Malawi.

2. To develop a simulation model for household dynamics in a large population and use this to explore the relationship between "true" and observed risks associated with household contact.

3. To quantify misclassification of household contact with leprosy cases and investigate the relationship between age and sex patterns of household change with corresponding patterns of contact status misclassification.

4. To estimate relative risk of leprosy among household contacts versus non-contacts after adjustment for contact status misclassification and to compare these estimates with relative risks observed in the population under study.

5. To explore the implications of the long and varied incubation period and the declining prevalence of leprosy on both the magnitude and trend of contact status misclassification and rate ratios.

6. To discuss these findings with reference to the published literature on household contact as a risk factor for various diseases (in particular leprosy and tuberculosis) and consider some of the broader implications of household dynamics for epidemiological studies.

The approach to this project is in three steps: first, a descriptive analysis of the data, then development of a household dynamics simulation model to quantify misclassification of leprosy contact status, and finally estimation of the "true" underlying risk of leprosy adjusting for misclassification. The descriptive analyses not only provide baseline demographic characteristics of the population but also

parameters for the design and calibration of a stochastic household dynamics model for a developing country setting. The simulation model is used to investigate the age and sex patterns of contact misclassification. Finally, the "true" underlying relative risk of leprosy among household contacts versus non-contacts is estimated after adjusting the observed relative risks in the population under study for leprosy contact status misclassification using results obtained from the model.

Chapter 2 reviews knowledge on leprosy and tuberculosis in households and problems associated with measuring infection transmission. Chapter 3 gives a description of the LEP data used in this thesis.

Chapter 4 presents a demographic analysis of the LEP data. Section 4.1 covers age and sex structure of the population, Section 4.2 presents mortality estimation and Section 4.3 looks at household structures and movements.

Chapter 5 gives a description of the stochastic micro-simulation model and methods used for adjustment of rate ratios for contact status misclassification. Chapter 6 gives results from the simulation model with Section 6.1 presenting trends of sensitivity of contact status. Section 6.2 compares observed and adjusted rate ratios.

Chapter 7 outlines limitations of the simulation model, Chapters 8 and 9 present a general discussion of the results and recommendations for future work respectively.

The appendices have been grouped into several categories. Appendix A presents a list of variables analysed in this work and also describes a method for age re-distribution of individuals with birth estimates. Appendices B and C respectively outline methods used to estimate level of infant mortality in the first year of life and the under-ascertainment of infants. Appendix D outlines a method used to adjust relative risks of disease for contact status misclassification. Appendix E provides selected schematic diagrams for procedures in the stochastic simulation model of

household dynamics. Appendix F contains list of tables and figures from literature reviewed and from analyses of LEP data. Finally Appendix G is a presentation of selected tables and figures from the simulation model results and estimation of rate ratios adjusted for contact status misclassification.

CHAPTER 2 LITERATURE REVIEW

2.1 Household contact associated risk

When investigating or comparing the transmission potential of infectious agents, it is appropriate to standardize or control for factors related to environment or human behaviour. In this regard, households have proved to be useful communities for the measurement and comparison of directly transmissible human infections (3, 4). Household members live in close contact and share common factors such as environment, genetics and diet. They tend to be at high risk if an infected person is present in the household because closeness of contact is likely to be related to exposure intensity, which in turn, is likely to be related to the infection risk and occurrence of disease. Importantly, though households differ in different societies and parts of the world, they everywhere entail close contact and shared environment. Besides this, contact within households is relatively easily identified (5).

Much knowledge of communicability of various infections and their relation to exposure, susceptibility and immunity has been derived from observations on their spread in households or families. This is mainly in reference to acute communicable infections, however, as it is difficult to derive such knowledge for the so-called chronic infections, for reasons discussed in this thesis.

2.1.1 Acute infections in households

Charles Chapin, in a series of reports 1884-1905 (3), was one of the first to study the spread of acute infections, in particular measles, diphtheria and scarlet fever in households. He developed the secondary attack rate (SAR) as a measure of transmissibility in this context. The secondary attack rate is conventionally defined as the proportion of contacts of a primary case in a household who develop disease

as a consequence of this contact. To compute an SAR one requires date of onset of the first (primary) case in the family, enumeration of other persons (susceptibles) in the household at that time, and the dates of onset of cases that occur within a specified time from onset of the first case. One also requires knowledge of the incubation and serial intervals, in order to allow identification of the secondary cases. Ideally one needs also to know the proportion of all infections that lead to recognizable disease.

For communicable diseases with short infectious periods, measurement of morbidity risk among household contacts is relatively easy because excess risk is concentrated within a short period following invasion of household (3, 6). If one can ascertain the numbers of exposed contacts, and of secondary cases, the calculation and use of SAR are appropriate.

After Chapin, classical studies of the transmission of measles, chicken pox and mumps viruses in households were carried out by Hope-Simpson (4) using a variant of the SAR called the "susceptible exposure attack rate". As expected, many studies investigating transmissibility of measles have shown higher risks among household contacts than among individuals with no known household contact (7-9). Table 2.1 below gives a range of estimates of SAR from the published literature (10), showing high transmissibility of measles in households.

Table 2.1 Range of estimates of household secondary attack rates (SAR) from the published literature (10)

Infection	SAR
Measles	50-80%
Small pox	40-60%
Mumps	30-45%

The SAR measure was developed primarily for acute infections. Uncritical application of the SAR concept can be misleading (11). In many studies, the effect of "silent" infections in household and the influence of risk of infection outside of the household have not been taken into account. The latter error can lead to

overestimation of SAR. In addition, variable susceptibility to infection (significant differences in age and sex specific levels of immunity) and the failure to distinguish between co-primary and secondary cases also need to be considered in the determination and interpretation of SAR.

These problems reflect the need for caution in use and interpretation of the SAR even for acute infections. When considering "slow" infections with long incubation periods, we have the additional problem of tracing household membership and contact status over time. To keep a large group of people under close enough observation to record all "contacts", for a long duration, is in practice not possible.

2.1.2 Long-term (chronic) infections

Measurement of the transmissibility of chronic infections, in particular those with long incubation periods raises a variety of methodological problems, depending on the natural history of the infection in question. For example, studies of the transmissibility of vector borne infections involve consideration of vector population dynamics (e.g. studies of filarial transmission (12)) and studies of sexually transmitted infections require estimates of sexual contact frequencies (13-15).

This project deals with chronic infections transmitted by direct or respiratory contact, the two classic examples being leprosy and tuberculosis. These infections have very long incubation periods, making it difficult to identify the source and hence to ascertain the appropriate denominators (all exposed individuals) and numerators (secondary cases). In part because of this, studies of transmission of the leprosy and tuberculosis agents have emphasized *relative*, rather than absolute risks of disease among known contacts compared to (apparent) non-contacts.

2.1.2.1 Leprosy

Leprosy is a chronic disease of man resulting from infection with *Mycobacterium leprae*. The clinical forms of the disease range along a spectrum from tuberculoid, with low bacillary (paucibacillary) load to lepromatous, with high bacillary (multibacillary) load. Although the mode of transmission remains a controversial issue, skin-to-skin contact or the respiratory route have received most attention (16).

Studies of transmission of infection with *M. leprae* are difficult because of the absence of any test for infection and the long and variable incubation period of disease (5, 16). Given that the incubation period may last for many years, even decades, the original source of infection of any particular case is likely to be unknown or forgotten by the time of its clinical onset. In addition, some leprosy cases may remain undetected but yet be continuous sources of infection in the community (17).

Different forms of leprosy differ in transmission potential depending on their bacillary load. There are different classifications for these forms. Early studies used categories such as "neural" and "cutaneous" (18) but more recent studies have favoured the terms tuberculoid or paucibacillary and lepromatous or multibacillary (18-20). Multibacillary forms of leprosy are more infectious than paucibacillary and there are strong arguments indicating that they are responsible for most of the transmission.

The classical study of leprosy epidemiology was carried out in the Philippines by Doull *et al.* (18) who applied a historical cohort method originally developed by Frost for studies of tuberculosis. It was shown that individuals living in household contact with a paucibacillary (called "neural" by the authors) case were at approximately twice the risk of leprosy compared to individuals with no known household contact. Household contact with a multibacillary (called "cutaneous" by the authors) case increased the risk eightfold. The high relative risk of contacts of multibacillary cases

is consistent with the fact that these cases have a much higher bacillary load than do paucibacillary cases. Similar observations of risk of disease for contacts of multibacillary and paucibacillary cases were made in several other studies (19-23). In BCG trials carried out in Uganda (24) and New Guinea (25), household contacts were also observed to be at increased risk of leprosy compared to those with no known household contact. Variations in risk estimates from different studies may be due to differences in methodology and duration of follow-up and to the proportion multibacillary among source cases. In their study in Uganda, White *et al.* (24) noted a gradation in risk of disease when contact was at dwelling, compound or visitor level.

Table 2.2 Incidence rates (per 1000 person years at risk) of leprosy for household contacts of known leprosy prevalent cases in households.

Study location	Incidence rates (per 1000 pyar)		
	Household contacts	No known contact	Relative risk
Cordova and Talisay by Doull (18)	5.33	0.83	6.42
BCG trial, New Guinea by Hausfeld (25)	55.0*	5.0*	11
BCG trial, Uganda (24)	24.6	18.7	1.32

Note: * Estimates not adjusted for age

Among household contacts, studies have shown that the risk is particularly high in young contacts (under the age of 15 years) (18, 19, 21-23), contacts of multiple index cases (21) and contacts who are closely related to primary cases (19, 23). In general, there has been little evidence of sex differences in the incidence rates among contacts (21, 22, 26). Very few studies have tried to separate environmental and genetic factors on investigation of transmission. The only such studies (27, 28) failed to separate family and household members and, more importantly, did not collect data on contacts of single cases or standardise properly for age, sex, relationships and nature of contact. There is need for carefully planned longitudinal studies necessary to draw inferences on familial aggregation of disease.

The most detailed analysis of household contact as a risk factor for leprosy was carried out in Northern Malawi on over 80,000 initially disease-free individuals (19).

The analysis defined the risk group in terms of household and dwelling contact. The analysis by dwelling contact was similar to the close contact category as used in other studies (23, 24). This investigation found that individuals living in households or dwellings with multibacillary patients at the start of follow-up were at five- to eightfold increased risk of acquiring leprosy over the subsequent five years respectively compared to individuals not living in such households or dwellings. Household or dwelling contact with a paucibacillary case approximately doubled the risk. The findings were thus strikingly similar to those obtained by Doull (18) although these studies were done in different societies, with different designs and at different times. The Malawi study (19) showed no statistically significant difference in risk of disease between dwelling and household contacts of paucibacillary cases of leprosy. This is consistent with evidence that multibacillary cases are the most important sources of infection transmission in leprosy endemic areas. The increased risk in contacts of PB cases could reflect undetected contact with MB cases who were the sources of the PB cases.

In several published studies (17, 19, 29), it has been observed that only a small proportion (15-30%) of all incident leprosy cases have a history of identifiable household contact. The majority of new cases thus appear to arise from the non-contact group. The proportion of cases associated with household contact is a function of the level of incidence in the community and is influenced by household dynamics (given that transient contact with source cases may be important but go unnoticed). The percentage of cases with history of contact may increase as incidence declines, reflecting the concentration of risk factors for disease into households and families (e.g. genetics, habits/behaviour as well as contact) (30).

Beers et al (17) broadened the definition of contact to include neighbour, family and social contact in a retrospective study carried out in Indonesia. This study attempted to reconstruct leprosy incidence over 25 years (1971-96), through interviews and a house-to-house survey in a highly-endemic village. The authors found that 78% of 101 incidence cases reported having been in contact with a previously diagnosed

leprosy case at one point or another. By broadening the definition, this study might have captured "transient" contacts since most of such contacts were with family and social members. However, being a retrospective study, there was great potential for reporting bias and hence underestimation, or perhaps overestimation, of history of contacts because of the long incubation period of this disease.

2.1.2.2 Tuberculosis

Tuberculosis is a chronic infection that has been a major public health problem for centuries. Its agent, *M. tuberculosis*, is transmitted through respiratory contact with an infectious ("open") pulmonary case. Tuberculosis is similar to leprosy in many ways: the agents responsible for transmission are related, the disease has long incubation period and transmission of infection is attributable to a certain sub group of cases.

Several studies investigating risk of disease for household contacts have varied in scope and details of methodology, and comparison across studies is difficult.

A classic retrospective investigation, based on tracing historical contact of 132 families, was conducted by Frost between 1930-1 in a black community of Kingsport, Tennessee (6) (Doull borrowed this historical method for his study of leprosy in Cebu (18)). The investigation provided data on 794 present and former members of these families. Persons with history of household contact had twice the risk of pulmonary tuberculosis morbidity compared to those with no known household contact over the period investigated (6). Several subsequent studies have shown similar results (31-34). Although the studies measured increased risk in household contacts of cases, these are not simple secondary attack rates. The increased risk reflects clustering of risk factors (poverty, genetics, etc) in households as well as contact itself.

Tuberculosis differs from leprosy in having a test for infection (the tuberculin test), albeit an imperfect one. Sensitivity is thought to be more than 90%, but specificity varies as a function of exposure to cross-reacting (e.g. environmental) mycobacteria, and the test cannot distinguish between new and old infection (35, 36).

Several studies of household contacts have shown that prevalence of tuberculin reactivity is highest among contacts of smear positive cases, lower among contacts of smear-culture positive cases, and lower yet among contacts of smear culture negative cases. This trend reflects the relative transmissibility of these forms of disease. The prevalence of tuberculin "positivity" among household contacts of known cases gives some measure of the transmissibility of *M. tuberculosis* infection in households and is generally similar between males and females up to adolescence after which it is higher among males and rises with age. This age difference in adults may be a reflection of extra-household contact.

These studies of tuberculin sensitivity have in general been carried out in current contacts of cases and, as such, the issue of household dynamics does not arise.

Measurement of transmissibility of infection based upon a disease outcome is difficult for tuberculosis because of the long and variable incubation period (from infection to disease) and serial interval (between onsets of successive cases in a chain of transmission) and the fact that only a small proportion of infections lead to disease. According to the standard textbook view, about 10% of infections lead to disease, half within 5 years, and half after long intervals, many years or decades (37, 38). In fact these risks are age dependent, and primary disease (within 5 years) appears to be more common after infection of adults than after infection in childhood (39).

Only one study (39) has attempted to describe the full incubation period and serial intervals of tuberculosis. Given the long period from infection to disease for most

cases, tuberculosis cases generally do not know, let alone no longer live with or near the case from whom they contracted the infection.

Given these difficulties, most of the research on infectiousness and transmissibility of *M. tuberculosis* has been based upon measurement of infection, using tuberculin reactivity as an indicator.

Another effort to measure the contribution of intra versus extra household transmission was in terms of a "community infection ratio (CIR)", defined as the odds ratio of tuberculin positivity among non-contacts compared to contacts. The relative prevalence of infection (tuberculin reactivity) *observed* among current contacts of cases compared to that among current non-contacts (the community) would reflect the prevalence in the community. If CIR=1 this implies similar tuberculin reactivity in current non-contacts and current contacts of cases in households. A CIR measure of more than 1 would imply a higher prevalence of infection in the community than in households with cases. A study conducted in Peru (40) used the CIR measure as an index to investigate the relative importance of intra and extra household transmission of *M. tuberculosis* among children aged 6 months to 14 years by estimating the prevalence of *M. tuberculosis* infection in both groups. The study computed a CIR estimate of 0.40 (95% CI: 0.26, 0.64) and concluded that there was relatively higher frequent transmission of infection in households with cases than in the community. Although this was the case, they also concluded that there was substantial transmission of infection within the community. The more rapid the change in household membership, the more widely the infection would be distributed in the community. Although this study had design problems (e.g. selection of controls), it provides another example of the importance of household dynamics for patterns of infection and disease.

There has been considerable scope for misclassification of contact status in the studies of leprosy and tuberculosis cited above – a reflection of the dynamic nature of household membership coupled with the long and variable incubation periods of

these diseases. While some studies (19) have acknowledged the problem of misclassification, none have investigated the dynamic nature of households and how misclassification of contact status could affect the observed relative risk. This is part of the task that was undertaken in this study.

2.2 Households and their dynamics

Household dynamics, or changes in household membership over time, obviously complicate the study of transmission of infections with long incubation periods. The concepts of households and families are often confused because of their close relationship to each other and the lack of unambiguous definitions for either of them.

According to the United Nations (41) and Burch (42), a *household* is based on the living arrangements of persons, individually or groups, for providing themselves food and other essentials. Household is a comprehensive term used in many studies because it does not necessarily depend on relationships between the persons involved.

A *family* may be defined as those individuals who are related to a specified degree through blood, adoption or marriage (or cohabitation as parents). A *family household* consists of a married couple with or without children, man or woman with at least one child or any other combination of relatives living together. A *non-family household* consists of an individual living alone or sharing living quarters with one or more unrelated persons (41, 42).

Although several categories of families such as "nuclear" and "extended" exist, *family households* form the bulk of households today in most societies. A household in this study of Karonga district, northern Malawi, was defined as a group of people living together recognising one person as their head (19).

Understanding the dynamics of relationships between people and their membership in households is important for household dynamics modelling in relation to contact with source cases of disease in epidemiological studies. The presence of a relationship between two people implies that their activities are causally interconnected. Status change in one person frequently causes a change in the

other. To describe relationships and membership in households, one household member is usually selected as a reference person and the relationship between household members is viewed from his/her perspective (43).

Despite impressions of a wide variety of household or family forms, their structures and general behaviour look fairly similar in most societies, including those in Sub-Saharan Africa. Because of these similarities, many of the methods developed can be applied to any society conditional on data availability.

In the sections that follow we review some results from studies of households, and early methods developed for the study of household dynamics. We then discuss some macro- and micro-simulation models which evolved from the early methods.

2.2.1 Households in Malawi and neighbouring countries

Since problems associated with investigation of risk of disease in household contacts may be specific to certain societies, it is important to investigate the similarities in structures of households in several Sub-Saharan African countries, in particular those neighbouring to Malawi. This is important because although the modelling work undertaken here is based on data from Malawi, the methods should be applicable to any such investigation in developing countries especially in Sub-Saharan Africa where population and household characteristics are similar to those of Malawi.

There is no literature on household formation in Malawi, although descriptive analyses of household structures are available from family formation surveys (FFS) (44), Malawi Demographic and Health surveys (MDHS) (45) and censuses.

Malawi is divided into Northern, Central and Southern regions and within these regions, there are districts. The common household types (in Malawi) as reported in the 1992 MDHS (45) include nuclear households (couple with children only), single-

parent households and extended households. New households in Malawi, like elsewhere, are mainly formed through marriage with the husband assuming the headship except in a few cultures which are matriarchal. The role of extended families has weakened due to changes in labour migration, formal education (46) and increasing landlessness (47).

As a consequence of patterns and trends of marriage, fertility, mortality and migration in "western" countries, there has been a continuous decline in average household size, a growing number of non-family households and an upsurge of one-person households (48). Although there is a much smaller literature on household dynamics in rural African settings, several trends have emerged.

Analyses of census and survey data collected in Sub-Saharan Africa over the past 20 years have shown an increase in age at first marriage and at first birth, an increase in the incidence of divorce and separation and a decline in proportion married at a given point in time (49). Much of the recent research on households in Africa has been driven by the need for demographic projections or in the context of epidemiological studies of HIV.

According to the 1988-92 MDHS (45), respectively 74% and 85% of rural and urban households in Malawi were headed by men. Similar patterns are observed in households in different population in Sub-Saharan Africa, as shown in Appendix Table F.2. Overall, households are more likely to be headed by men than women. However, the head is not necessarily the economic provider or decision-maker in a household. Most respondents designate the oldest man or woman as head even when there is no clear hierarchy in authority and decision-making in the household (49). A comparison between rural and urban households showed that the proportion of households headed by women is relatively higher in the rural than urban areas. Due to the polygamous nature of most of these societies, a man can be a head of more than one household.

Data from six consecutive annual surveys were used to examine household structures during the HIV epidemic in a rural Ugandan population (50). On average, 26% of households were female-headed and this compared well with the rural Ugandan population in 1991 population and housing census (27.8%). Less than 1% of households were headed by individuals less than 19 years of age. The recent increase of deaths of young adults in communities affected with HIV/AIDS in sub-Saharan Africa (46) has led to an increase in single parent households and in orphaned children (49).

There is very little difference in size between rural and urban households in Malawi. The average number of persons living in a household was 4.5 during the 1988-92 period. These results are similar to those obtained in the 1987 census, when the average number of persons per household was 4.3. Table F.3 in Appendix shows the mean household size in several Sub-Saharan African countries as reported in Demographic and Health Surveys (DHS) by rural or urban area. The overall mean household sizes are similar, ranging from 4.3 in Malawi to about 5.6 in Zambia. The northern region of Malawi had the highest number of persons per household (4.8), with Karonga District having the highest in Malawi (5.3). The number of persons per room is often used as a measure of crowding, which may influence the spread of infection among household members. There was little variation in crowding between rural and urban households in Malawi as reported in the 1992 MDHS. The number of persons per room (size of room not known) was 2.8 in rural areas and 2.7 in urban areas. Overall, 56% of households had 1 or 2 persons per room and 34% of households had 3 or 4 persons. The percentages of households with more than 5 persons per room in rural and urban areas were 11% and 8% respectively (45).

The similarities between structure of households in Malawi and countries neighbouring to Malawi suggest the modelling work of household dynamics, based on data from Malawi, described in this thesis can be applicable to any such countries, especially the rural societies.

2.2.2 Modelling

While there has been steady development in household dynamics research and modelling in the "West", little work has been devoted to household dynamics in Sub-Saharan Africa (51). Thus, most of the work reviewed in this section is on models of household dynamics based on data from developed countries.

The study of household dynamics implies an understanding of the relationship between household members (49). Apart from looking at the importance of households from an epidemiological point of view, the study of household dynamics is important in many other disciplines such as housing, marketing and social security. For example, knowledge of size and composition of households and of their trends over time is essential for policy makers in evaluating housing needs.

Factors which have contributed to the slow progress in household dynamics research include the scarcity of data, the slow development of methods of analysis (48) and the increase in the number of categories when a detailed household breakdown is considered (52, 53).

Population and household dynamics (changes in their size and structure) and their projections (probable or imaginable future structures) are best modelled on the basis of an individual as a unit of analysis (48). In order to pursue modelling on this basis, we need to look at individuals, categorised by both their "pure" individual attributes (e.g. age and sex) and by relational attributes. Relational attributes refer to membership of and position within groups such as families and households (e.g. marital status and household position). Pure individual characteristics are deterministic and easy to update. The complexity of the modelling derives from the need to update the relational attributes over time.

This section will briefly review early methods and macro-simulation models of household dynamics but will concentrate on micro-simulation models as these are more relevant to this work.

2.2.2.1 Early methods

Among the first attempts to describe household dynamics was a research effort at forecasting growth of household numbers (41). These methods were important steps in early research on household dynamics modelling. They were devised to look at population and household projections without considering changes in household composition and movements between households over time. According to the 1973 UN Manual VII (41), one of the earliest techniques developed for forecasting households was based on the "*household-to-population ratio*". The method assumes that the growth rates of the population and of the number of households are equal, and implies that the household size remains constant. The household-to-population ratio based upon observed data was applied to future population projections to obtain the future number of households. However, given that household structures evolve with social change, they may grow at a different rate from the total population, which may invalidate the method for long term forecasting. This crude approach is appropriate when only elementary data are available. It was superceded by several other approaches, known as "*life table*", "*vital statistics*" and the "*headship rate*" methods.

The "*life table*" method was developed in 1951 for the United Kingdom by Brown (54) to model the distribution of families in a hypothetical stationary population broken down by age, sex and marital status, based on the 1947 British social survey data. This method equated families to households and further broke down the population distribution of the number of married couples, widows and widowers by number of children. The method gives a projection of families (households) by size of membership assuming a hypothetical stationery population. However, its validity was weakened by the stationary population assumption. The equating of

families to households may be invalid as not all unmarried adults or widowed people live on their own as a family but may join other families and form households.

The "*vital statistics*" method developed by Illing (55) in 1967 deals with projections of families of married couples and then transforms them into households. The transformation to number of households is a function of two ratios: the ratio of the number of family households to the number of families, and the ratio of non-family households to total households in the population. The advantage of this method is that family formation and dissolution can be related to demographic events such as marriage, divorce and death affecting family members, over time. As such, future trends of marriage, divorce and death by households can be estimated. However, the transformation from families to households does not take into account age and sex differentials in the relationship between families to households. The method is difficult to apply in developing countries due to lack of accurate data on marriage and divorce.

The "*headship rate*" method has been used extensively and is widely applicable due to the availability of appropriate data. The headship rate is defined as the proportion of individuals in each age and sex strata who are household heads. The method assumes that the number of heads of households directly represent the number of households in the population under study. The future number of households in each of age and sex headship category is obtained as a product of the number of persons in that category and their projected age-sex specific headship rate. The total future number of households is the sum of the age-sex specific number of households. Extensions of the headship rate approach have been developed (56, 57). The main methodological problem has been the estimation of future headship rates from those observed. It may not be appropriate to apply a constant headship rate to varying evolving population structures (58).

These "early" techniques are not appropriate for the problem under investigation, in that one cannot use them to quantify contact with source cases as was required for

this project. Although these methods are quick, easy to apply and require only simple data, the dynamic processes themselves remain unexplored (48). Aggregate cross-sectional data catch households at different stages of their life cycle and may obscure temporal variations. Macro-simulation models were thus developed to provide a more detailed picture of the behaviour of households over time.

2.2.2.2 Macro-simulation models

Macro-models of household dynamics are models that treat populations by strata e.g. groups defined by age and sex, and not as individuals. Thus, no consideration is made for occurrence of demographic events at an individual level. For example, considering mortality, the simulated number of deaths would be the product of the number of individuals and the age-specific mortality rate.

We briefly outline some of these models. Most of such work has been from institutions in developed countries in particular Scandinavia, the Netherlands and Germany.

The *ISP model*, described by Moller K.P. (59) in 1979, was developed by the "Institut fur angewandte Systemforschung und Prognose" (ISP) in Hanover, Germany for household projections within the framework of an economic model. This model produces household distribution by type and size. Events such as childbearing, mortality and external migration are simulated. The number of households is calculated by considering male adults and non-married female adults as heads of households using assumed headship rates (see Section 2.2.2).

The S*wedish model* is a dynamic household model constructed by Harsman *et al.* (57) during the late 1970s and early 1980s. It was aimed initially at projecting regional housing requirements for Stockholm. The model was based on the distribution of the population by household type at two consecutive time points and the implicit change of household status which individuals experience during the

intervening period. The model describes transitions that individuals in specific strata make and can be extended to account for births, deaths and migrants to produce complete population projections.

The *Primos housing model* was developed by Heida and Gordijn (59) in 1985 at the Netherlands Centre for Study of Physical Planning to assist in formulating housing policy. Individuals were distinguished according to age, sex, household status, marital status and whether one was a dependent child or not and living in an institution. Lack of data on changes in household status created difficulties and the authors relied on simplifying assumptions to allow modelling.

The *NIDI household model* was developed by Netherlands Inter-University Demographic Institute in 1987 and described dynamics of households in detail (59). The model generates, among other outcomes, household history for the cohort under study. Parameters for transitions between household types were obtained from the 1984 ORIN retrospective survey of household biography for the previous seven years (1977-84).

Due to the complex nature and states in which households may exist and the lack of substantial data, it is difficult to build macro-models for household projections. A higher level of detail (e.g. increase in number of subcategories of events) poses severe problems as the increase in explanatory variables results in an unmanageable number of groups (59). Methodologically, such models present consistency problems between household and population structures (52, 58). Updates are only made to one of the two structures. For example, mortality may be simulated in the population and yet there is no way of updating the household structure in terms of size and composition. Demographic events such as marriage, which involve more than one person, are not satisfactorily solved in these models. For example, it has been shown empirically that the process of leaving a parental household is closely related to marriage decisions, educational level and labour

force participation (58). But this cannot always be deduced from aggregate observed data.

These problems demonstrate the need for an approach allowing simultaneous modelling of population and household structures and their dynamics. This can only be achieved by modelling the behaviour of individuals (53, 58), emphasising the need for a micro-simulation modelling approach. With the availability of increasingly powerful computers, there has been increased interest in quantitative (descriptive or through computer simulations) analysis of households.

This thesis describes the development and application of a household model to investigate risk of disease associated with contact in households. Our interest was not in household or population projections but in contact misclassification with source cases of disease derived through incorporating, in a micro-simulation model, important selected demographic events affecting a population over time. It is appropriate to look at these events at a micro-level, hence, the need for a micro-simulation approach. Any grouping can be done from these micro-level data.

2.2.2.3 Micro-simulation models

Micro-simulation models attempt to model the various events, at an individual level, that together account for, in this context, the dynamics of households over time. Such an individual level analysis allows for updates of individual and household characteristics in successive time periods (60). The main difference between micro- and macro-models is that in micro-models the unit of analysis is the individual whereas in macro-models, we tend to look at groups e.g. females in the households.

Micro-simulation approaches to modelling households have been developed by several researchers and are dispersed over several disciplines. Earlier work on micro-simulation of household dynamics is traced back to Orcutt *et al* (61). Although their work concentrated on household dynamics, the intention was to model the economic sector of the Unites States. Orcutt *et al* (62) extended their work to produce a "Dynamic Simulation model of Income" (DYNASIM) which was used to assess the effects of national policy options such as female participation in the labour market. The National Bureau of Economic Research (NBER) housing model, designed to study the housing systems of Detroit and Pittsburgh, incorporated a micro-simulation module to examine household dynamics as reported by Clarke (60).

Household dynamics models have been incorporated into studies in several different disciplines. Examples include the work of Wilson and Pownall (63) on activity analysis; Bonsall (64) and Kreibich (65) on individual choice in transport systems and Wegener (66) on residential location.

Microsimulation techniques specific to household dynamics have currently received considerable attention. The basic structure of demographic micro-simulation modelling, developed initially by Hecheltjen (1974) (58), contained a description of demographic processes on the individual level but was restricted only to nuclear

families. Steger (1980) (58) extended Hecheltjen's model to families, households or institutions and introduced assumptions for transitions between these entities. The model was used for projections of the population and of family and household structures. Nakamura and Nakamura (67) developed a micro-analytic model devoted to population dynamics based on data for Alberta, Canada.

Keilman and van-Dam (68) developed a projection model for the simulation of household events in the Netherlands. Nelissen and Vossen (69) developed the Netherlands Dynamics Micro-Analytic Simulation Model (NEDYMAS). This is a broad-based model encompassing demographic, social security, labour market and income formation modules.

Clarke (60) used a micro-simulation approach to model household dynamics and produce forecasts of the number of different types of households over a 5-10 year period based on data from Yorkshire and Humberside in England. He described model specifications for death, migration and leaving home among other events.

Some simulation models concentrate on particular components of household dynamics. Willekens and Baydar (70) forecasted place-to-place migration using generalised linear models based on migration data collected by the Central Bureau of Statistics of the Netherlands from 1958 to 1982. Menken (71) and Teachman (72) used the proportional hazards model for modelling family formation and dissolution (time to event, in general).

The method developed by Keilman (73) was applied in several studies: for example a study by Kotowska (74) for population projections by age, marital status and region of residence in Poland; a study of the impact of household structure on social security in Netherlands (75); and studies of the dynamics of living arrangements in Netherlands, which showed increases in numbers of persons living alone (75) and cohabiting with time (76).

Several micro-simulation models more relevant to the work undertaken in this project have been developed and the methodology used described in detail (43, 58). One of the most popular micro-simulation models for household dynamics was developed at University of Frankfurt in Germany (58). Coded "*sfb 3*", the model was part of a general modelling effort aimed at analysing income distribution and social policy. The initial micro-data file used in the model had a record for each household containing among other variables size and head of household, one record for each person in each household with their personal characteristics, their relationship to the head of household and some socio-economic variables. The records of married persons contain pointers to their spouses and those of children point to their parents.

Menken (77) and Clarke (60) provided an excellent overview of micro-simulation modelling of household dynamics and Murphy (78) discusses logical and practical issues that arise when analysing household changes. The problem of sequencing of events (like births, deaths, migration and movements) in simulations has been much discussed (60, 79). However, whether one event occurs earlier than the other has been found to have a negligible effect when working with intervals of 6 months or even a year (80).

In micro-simulation models many explanatory variables can be introduced allowing a detailed hypothesis to be investigated. This is because the size of micro-simulation model is not determined by the number of explanatory variables but by population size. However, the modelling efforts are constrained by limitations of empirical data.

Consistency is maintained between population structures and family and household structures in a micro-simulation model. This is because changes at the individual level also initiate changes at family and household level due to the linkage of each individual to a specific family or household.

There is no standard modelling structure in the published micro-simulation approaches and, hence, standardised comparisons across population structures are not catered for. Micro-simulation requires more modelling resources and poses higher demands on empirical data (dynamic models of individual behaviour should be based on longitudinal data, as opposed to cross-sectional data, which may be scarce).

2.2.3 Developing countries

The micro- and macro-models reviewed above were based on data from developed countries but most of their methods should be generally applicable anywhere. Turning to developing countries, there has been very little work on household dynamics. An important and interesting exception is work carried out at the International Centre for Diarrhoeal Disease Research in Bangladesh (ICDDR,B), one of the most productive health research field stations in a developing country. Originally developed in 1960 as a Cholera Research Laboratory to study the epidemiology and prevention of cholera, it now conducts studies on a defined population (Matlab) and monitors important components of population dynamics (births, deaths, marriages and divorces, and in- and out-migration) over time. The main source of the routine data in Matlab, apart from periodic censuses and the Record Keeping System (RKS) (which has recorded the reproductive status and morbidity in women of reproductive age since 1978) is a Demographic Surveillance System (DSS). This has provided longitudinal registration of vital events since 1966 (81, 82). Individuals are registered in a computerised database with field surveillance designed to record all events and changes in household relationships over time. There were long delays in developing the system, due to expensive computer hardware, which diminished the usefulness of the data for policy and by other field stations. As such, the DSS was criticized as too costly and complex for practical applications. Even with a highly sophisticated DSS, the ICDDR(B) is unable to yield periodically updated information on formation and dissolution of households.

As an alternative to the Matlab system, the Household Registration System (HRS) (83) was designed at the Max-Planck Institute for Demographic research in Germany, in 2000, to resolve limitations of the Matlab DSS. HRS is a computer software system developed for diverse data collection from longitudinal household studies. This system monitors population and household dynamics with demographic surveillance and produces reports on cross-sectional and longitudinal data associated with studies of households and their members. Although appropriate for developed countries, the HRS is a form of technology transfer to developing countries. The system is currently in use in research sites in Africa (Burkina Faso, The Gambia, Ghana, Kenya, Mozambique and Uganda) and one site in Asia (Indonesia). It maintains consistent records of significant demographic events (such as births, deaths, marriages and migration) that occur to a population in a fixed geographical area over time and computes basic demographic information (age-specific birth, death and migration rates; age/sex distribution of the population and life table functions). The HRS can be easily modified to suit particular projects. The core structure of the HRS consists of characteristics of household members, their relationships and demographic events common to all longitudinal studies. Some requirements can be relaxed depending on population studied.

From this review, we see that there have been many published studies on household dynamics methodology based on data from developed countries but few have been reported from developing countries. Because of scarcity of data on households in developing countries, it has not been possible to make elaborate description of households that take into account various factors affecting their future growth and structural changes. This project will make use of a large longitudinal data set on households from Northern Malawi to discuss the effects of household dynamics on misclassification of contact status.

2.3 Exposure misclassification and relative risk

As discussed in previous sections, household dynamics play a crucial role in determining contact status. If household dynamics are not taken into account in an analysis, misclassification of contact status may occur. Misclassification is a well-known source of error in epidemiological studies (84-91). Subjects can be misclassified as to their exposure or disease status. There are two types of misclassification. *Non-differential misclassification* occurs when the probability of exposure (or disease) status being misclassified is independent of disease (or exposure) status. *Differential misclassification* occurs when the probability of exposure (or disease) status being misclassified depends upon disease (or exposure) status.

Misclassification has often been discussed in the literature in terms of disease outcome in cohort studies and in terms of exposure in case-control studies. However, exposure can also be misclassified in cohort studies. Subjects may either become non-exposed just prior to registration for the study or change from non-exposed to exposed after registration. Such situations are particularly common in studies with long duration of follow-up. This study will be looking at an example of exposure misclassification in cohort studies. Our exposure of interest is household contact with an infectious case.

Early discussions of the consequences of misclassification of study subjects have been presented in the literature by Bross (92), Diamond and Lilienfeld (93, 94), Newell (95), Harper (96), Gullen *et al.* (97) and Goldberg (85). The early and current discussions are based on results from analyses using methods developed to quantify misclassification and adjust observed estimates for misclassification. Most of these methods assume observed data from 2x2 tables together with known values of sensitivity and specificity (and hence misclassification rates) of the classification procedure.

There has been controversy on the extrapolation of known values of sensitivity and specificity (from validation studies or otherwise (98, 99)) obtained outside the study population. In a well-known example, Diamond and Lilienfeld (94) applied misclassification rates from an earlier study of the validity of statements concerning circumcision status and physical examination findings to a different study by Wynder of the relationship of circumcision status of husbands to cancer of the cervix (100). Their findings were challenged by Newell (95) on the basis of the applicability of misclassification rates from a different study.

Several methods have been developed to quantify and adjust estimates for measures of effect for misclassification (90). Copeland *et al.* (86) and Barron (101) developed formulae to estimate the "true" value of the relative risk (or odds ratio) as a function of misclassification. Barron derived the "true" state of nature (of the association between exposure and disease outcome) as a non-linear function of observed state of nature and the probability of misclassifiying with respect to each of the two variables. Farrington (102) developed a mathematical model to quantify misclassification bias in cohort studies. More recently, Reade-Christopher *et al.* (103) developed a regression model which helps assess the potential for bias in a follow-up study with categorical data when misclassification is ignored and corrects for misclassification bias in the estimates when reliable information is available. Sosenko and Gardner (104) extended the method of Copeland (86) to show the relationship between attributable frequency and misclassification bias.

The above methods assume the misclassified variable is dichotomous, that there is no bias due to confounding and that misclassification only affects the comparison of the two groups eg. disease occurrence in cohort studies. These methods also implicitly assume a variance of zero. In view of this, Espeland and Hui (105) developed a log-linear model, which, apart from incorporating information on error rates and biases, also enables variance estimation and can be generalised to larger contingency tables. Greenland (106) extended the method by Selen (107) for

variance of group means to estimate the variance of the error rates. Greenland (108) and Armstrong (89) also looked at the effect of misclassification in the presence of a covariate or confounder (3-way table). Inclusion of a third variable may be dictated by the need to control for confounding and to study heterogeneity in the measure of effect of exposure across levels of third variable (effect modification).

Applications of these methods have shown that misclassification is a function of exposure frequency, disease frequency, sensitivity and specificity (86, 101, 104, 105, 108). Use of the same specificity and sensitivity values on several misclassified data showed that biases in differential misclassification can vary. This variation depends on the true attributable frequency of exposure or disease (104).

The direction of bias in the estimates will depend on the type of misclassification. Various investigators have looked at the effect of differential versus non-differential exposure misclassification on the relative risk of disease. In non-differential misclassification, the bias is usually towards the null i.e. the strength of association is weakened. When differential misclassification is present, the bias in estimates can go either way (86, 104, 108). Barron (101) found that the relative risk of endometrial cancer between hypertensives and non-hypertensives adjusted for misclassification was greater than that estimated when errors of classification were ignored. Copeland *et al.* (86) convey in clear graphic and tabular form the direction and magnitude of bias in estimates by varying the error rates. These findings usually have involved different types of observed or computer-generated data.

The exposure of interest varies depending on the study undertaken. For example, in determining risk of lung cancer, smoking would be an exposure of interest (109). At times, surrogate exposures have been used as it is difficult to measure actual exposure. For example, coffee has often been used as surrogate for caffeine intake (110), and maternal residence at birth as surrogate to environmental exposure (111, 112). In determining the transmissibility of infection in household, household contact

can be our exposure of interest. Studies to investigate household contact as a risk factor for infection in cohort studies should take into account misclassification of household contact status because of the dynamic nature of households. This study will make use of a large data set on leprosy to obtain misclassification rates associated with household contact and will discuss the adjustment of relative risk estimates in this context.

2.4 Mortality, fertility and migration (in Malawi)

In order to understand transmission of infections in households through contact with source cases, there is need for information on changes in household composition and structures over time and patterns and trends of marriage, fertility, mortality and migration in populations. Demographic changes reflect these patterns.

Before studying household dynamics and its demographic effects, it is important to look at agreement of the observed data with previous results of studies conducted on the same or similar populations. The data used in this project for modelling household dynamics were obtained from a large epidemiological study in a rural population in Karonga District, Northern Malawi. We will wish to compare these data with results from censuses and surveys conducted in Malawi.

Sources of data on demographic events are limited. This is a problem in many developing countries. The two main sources of household data are population censuses and sample surveys. Although sample surveys furnish data on individual status transitions, they are in most cases set up for a different purpose. For example, the LEP data (1) employed in this project arose from an investigation of the epidemiology of leprosy in a rural population of Northern Malawi. However, observed transition rates of events between two surveys can be entered as parameters to run a micro-simulation model of household dynamics, as in this thesis.

Longitudinal studies also provide potential for household modelling if the individual and household data are linked between consecutive surveys (as they were in the LEP studies).

Although transition rates (probabilities) may also be derived from vital statistics (113), a disadvantage of such data is that they refer to legal situations as opposed

to factual circumstances e.g. marriage as opposed to cohabitation. They may also lack contextual data on the relationship between social economic status and demographic circumstances of individuals.

In Malawi, like in many Sub-Saharan African countries, the main sources of information are national censuses and household surveys e.g. Demographic Health Surveys (DHS) and Family Formation Surveys (FFS). The Malawi 1992 DHS was a nationally representative sample survey aimed at providing information on mortality, fertility and morbidity levels, among other health issues. Some information may also be available from research studies carried out in particular populations.

The common events affecting household structures over time are mortality, fertility, migration (49), increasing landlessness and marriage (47). These have important implications for future composition of households and for infection transmission. Changes in these demographic rates may affect household size and composition and hence the proportion of source cases and their contacts. As this may lead to bias in estimates of the risk of disease under investigation, an understanding of the demographic changes is important for understanding the implications of contact on infection transmission in households.

This section reviews published results on fertility, mortality and migration obtained from census and Demographic and Health Surveys (DHS) carried out in Malawi. The results will be compared to those from descriptive analyses of the LEP data.

2.4.1 Fertility

The populations in Sub-Saharan Africa are predominantly young, reflecting highly fertile, rapidly growing populations. Slightly over 45% of the population in these countries is under 15 years of age as shown in Table F.1 of Appendix. This has implications for household dynamics and for contact-associated risk of disease.

The crude birth rates for Malawi according to the 1977 and 1987 census results were 48 and 41 births per 1000 persons (114). Comparative figures for Karonga district were 50 and 37 births respectively, suggestive of a decline in fertility between 1977 and 1987.

The most widely used measures of fertility are the "total fertility" rate (TFR) and the "age-specific fertility" rate. Total fertility rate is defined as the average number of children a woman would have by the time she completes her reproductive period, assuming *current* levels of fertility.

The TFR, as reported in the 1987 census and MDHS (45), were 5.7 and 6.7 children respectively. These values were lower than that of 7.6 estimated in the Family Formation Survey (FFS) (44). Fertility among rural women was higher than that of urban women. For example, in the northern region it was 6.0 and 5.3 children for the rural and urban areas respectively.

Early childbearing leads to larger family size and is associated with increased health risks. Over one-quarter of women aged 15-19 years were reported in the MDHS in 1992 as having at least one child (45).

2.4.2 Mortality

The crude death rates for all Malawi as reported in the 1977 and 1987 censuses were 25 deaths and 14 deaths per 1000 persons respectively. The corresponding figures for Karonga District were 16 and 17 deaths per 1000 persons respectively. Although the national crude death rates for the 1977 and 1987 census suggest mortality decline, the crude rate for the 1998 census was 21.1 deaths per 1000 population. This increase in mortality is probably attributable to HIV/AIDS.

Infant mortality as reported in the Malawi Demographic and Health Survey (MDHS) report (45, 79) during the 1988-92 period was 134 deaths per 1000 live births. The under-five mortality appeared to decline from 258 per 1000 live births in 1978-82 period down to 234 per 1000 live births in the 1988-92 period (45, 79). Under-five mortality is higher in the rural than urban areas but varies by region. In the northern region, it was estimated at 202 deaths per 1000 in the 1988-92 MDHS report. These under 5 mortality rate estimates were measured in retrospect based on maternal interviews and are less reliable than prospective methods.

Deaths in childhood are highest for first births. Children are at high risk of mortality due to biological and socio-economic factors related to poverty, especially evident for children whose mothers started child bearing at a young age. Short birth intervals are also associated with higher mortality. There is evidence that the slowing of the decline in (under-five) mortality is partly due to the HIV epidemic. Most infants who acquire HIV from their mothers die in the first five years of life.

Investigations by Timaeus (115), using 1977 and 1987 Malawi census data, showed that mortality was higher for adult women than adult men. A series of enquiries in the early 1970s found similar results (116).

2.4.3 Migration

Migration rates are estimated from census data by comparing the district at birth and the current district of residence. Results from census data have concentrated mainly on internal migration (migration within Malawi).

In the 1966 census (117, 118), it was reported that 11% (453,782) of the total population were enumerated outside their district of birth. Of those enumerated outside their birth district, 74% were within the same region. Migration within a region was higher than from outside the region. For example, of those who migrated

to Karonga, 83% were from within the northern region and the remainder from the Central and Southern regions. Overall, the northern region has in the past experienced high net emigration to other regions of the country. In the early seventies, a large number of male labourers emigrated to work in the mines in South Africa (117, 118). This must have contributed to an increase in the proportion of female-headed households in the rural areas.

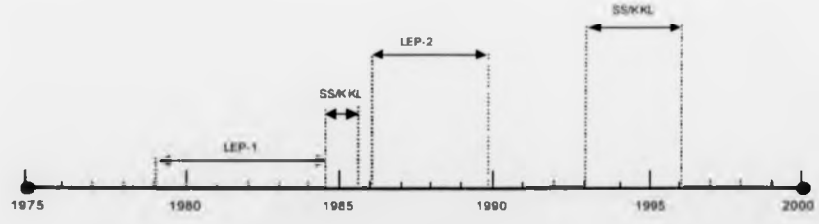
There was an increase in internal migration from the 1966 to 1977 census. In the 1977 census (117, 118) an overall higher number of male than female internal migrants was reported. The census also reported a higher number of male than female out-migrants from districts. These results are consistent with those obtained in 1987 census (114). The male dominance among migrants was greatest at ages 20-39, the most economically active age range. The age distribution of male migrants was older than female (median age of 26.6 and 24.5 years respectively). The overall modal age group for migrants was 20-29.

CHAPTER 3 DATA DESCRIPTION

3.1 Study area

Most of the data for this study were collected as part of two population surveys, carried out 5 years apart (on average), in Karonga District, Northern Malawi. The details of the fieldwork methods have been described elsewhere (1). The first survey (called LEP-1) was carried out from 1979 to 1984 and the second survey (LEP-2) from 1986 to 1989. The second survey coincided with the recruitment phase of a leprosy and tuberculosis vaccine trial (2). Over 112,000 and 146,000 individuals were interviewed in LEP-1 and LEP-2 respectively. In addition special surveys (Sample Surveys, Kasowa, Kasyata and Lower Songwe) coded SS/KKL for convenience were carried out in 1984, to estimate leprosy incidence prior to the vaccine trial, and in the 1990s, as part of the vaccine trial follow up. The temporal relationship of all these surveys is shown in Figure 3.1.

Figure 3.1 A time chart showing periods when surveys were conducted in Karonga District, northern Malawi.



The district was divided into 5 ecological "zones" on the basis of general ecological features (see Figure 3.2). These were the northern hills (Zone A), the northern lake

shore area (Zone B), the southern hills (Zone C), the semi-urban area around the district capital (Zone D) and the southern lake shore area (Zone E). Sometimes, for analysis purposes, these zones were grouped further into Northern (zones A and B) and Southern (zones C, D and E) Karonga.

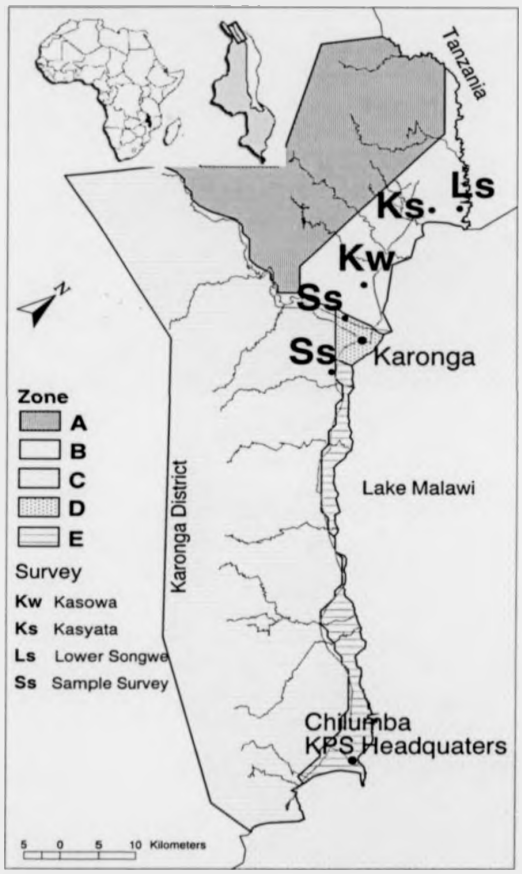
To achieve high coverage, the surveys were carried out by house-to-house visits by field teams. The areas covered in the two surveys were not quite identical (e.g. the LEP-1 survey omitted a small area in the southern tip of the district and sparsely populated areas on the west of the district were omitted from LEP-2).

3.2 Individual information

For each individual, interviewers collected information on birth year, sex, mother and father's names and identifiers, level of education, main occupation, village and household in which the individual was resident, position in household and year of joining household. The LEP data did not directly record spouse relationships. Where applicable, this has been inferred based on co-parent status (i.e. if a man and woman were recorded as father and mother of one or more individuals, it could be assumed that they were "spouses"). Not all co-parents were living together as couples.

Each individual was permanently identified in the project by an identification ("IDENT") number (six digits plus an algebraically determined check digit) and was assigned initially to that household in which they were first found and interviewed. Each household was assigned a unique (5-digit) "household number". An effort was made in LEP-2 to trace the whereabouts and status of all individuals seen in LEP-1. With such information, it has been possible to follow individuals from LEP-1 to LEP-2 and to observe whether they had changed households, died or migrated to other areas within or outside Karonga. A list of variables collected in the study, which are relevant for this research, is provided in Appendix A.1.

Figure 3.2 Map showing Karonga District (the study area), divisions of ecological zones and special survey (collectively termed SS/KAS) areas.



Information on education status was recorded as a categorical variable: never attended school, attended 1 through 8 years of primary school; or received at least some secondary education. Where there was contradictory information on education status recorded at two different interviews of the same individual, the more plausible information was accepted. If the two records appeared equally reasonable the information was considered missing (119).

The occupation of each adult was recorded. For purposes of analysis, the occupations were grouped broadly into fisherman or farmer, trader, salaried worker, "none" or casual worker, and student (those still attending school).

People were also categorised according to housing quality, defined in terms of construction material of dwellings (a dwelling was any structure in which people slept) (119). The four categories were: 1) houses constructed with locally made burnt bricks, 2) houses made with sun-dried bricks or pounded mud, 3) houses constructed using wood or bamboo poles and interlaced twigs and rods for the walls which are later plastered with mud and 4) temporary shelters made out of grass and other material.

3.2.1 Dates and local events calendar

Precise dates for births (or deaths) were rarely known or recorded. Birth month was recorded for only 10% (9130/89419) of individuals during LEP-1. The percentage varied with age, being 6% for children under 5, 12% for those aged between 5 and 30 years and 10% for those aged over 30.

For individuals who did not know their precise *years* of birth, a local events calendar was used to assign periods of time to births (1). These periods were chosen with reference to well-known local events. The main events and dates used in LEP-1 are shown in Table 3.1.

Table 3.1 Local events calendar used to estimate period of time to births in the LEP studies, Karonga District, northern Malawi 1979-89.

Year	Event
1914	Battle of Karonga (World War I)
1934	Major crop damage by locusts
1946	Passenger ship Viphya sank
1958	His Excellency Dr Hastings Kamuzu Banda returned to Malawi
1964	Malawi gained independence

Individuals who did not know their year of birth were assigned the mid-point year between the date of the earliest event they recalled and the previous event as their birth estimate.

Age was calculated by subtracting the assigned or precise year of birth from the year an individual was interviewed.

3.3 Household information

In this study, *a household was defined as a group of people living together and acknowledging one person as the head* (19). This definition was problematic in that it was not always unambiguous whether a household was the "same" the second time it was seen, if the composition had changed or a new head had taken over. For example, if a recently married son continued to stay with his father, it was not always clear whether the son acknowledged his father or himself as head. Because of the number and complexity of different circumstances, rather than formulate fixed rules, the decision as to whether a household was effectively the same (and hence deserved to preserve its old number) was left to the discretion of the interviewers. Due to the constantly changing composition of households in any society the definition of what makes up any particular household is to some extent arbitrary. It should be noted that the interviewers were from the local population, they underwent a standardised training experience and they were routinely supervised.

In addition, they were encouraged to write additional relevant information on questionnaires and to discuss unusual circumstances which arose with their supervisors and with the Project leader Dr Jorg Ponnighaus. It is thus considered that their decisions on household status were well informed and consistent.

Information collected on households included household serial number, geographical location, identification of the head of household, village in which the household was located, number of dwellings in the household, year of arrival of the household in the current place and previous location of the household where applicable. Sometimes position on household and individual records forms for heads did not match due to errors in coding and data entry.

Most households consisted of nuclear or extended families but sometimes included more distant relatives, visitors, renters and workers. The positions of individuals in a household were categorised as: head of household, "member", visitor for less than 6 months, visitor for 6 months or more, employed worker or servant, renter, patient, and relative of employed worker or renter. A "member" could be a spouse, child or (sometimes distant-) relative of the head of household. For most analyses described in this report, position in household was categorised as head, "member" and "other" (which included all the other positions).

Atypical "households", which contained unusual groupings of individuals, such as traditional healers' camps, boarding schools, fishermen's camps, young pioneers camps, construction camps, Irrigation schemes and the Karonga District prison were excluded from the household structure analyses presented here, though the individuals involved are included in the population analyses.

Due to the polygamous nature of society, one man could at one time be a head of one household with several wives in one place or a head of several households in different places. Records for personal and household information were kept separate. If an individual was a head of two or more households, his personal

information was kept and recorded against only one of the households of which he was head i.e. the one in which he was first interviewed. This convention was instituted in order to avoid repeated personal records and multiple counting of heads of households.

Geographical co-ordinates, precise to 10x10 metres, were used to locate households within a square kilometre grid drawn on aerial photographs. Villages were assigned unique 4-digit numbers that could be used as prefixes to all the household identification numbers within that village. These codes identified a village as being within Karonga, outside Karonga but within Malawi or outside Malawi and are useful for tracing migration between LEP-1 and LEP-2.

Population mobility has been defined as the movement of individuals in space and time and includes all movements regardless of distance, duration or intentions (120). In this thesis mobility has been categorised into migration (in or out of Karonga District) and movements within the district. Migration involves change of place of residence and in some contexts the term implies that there is little or no expectation of return. We are concerned with migration in relation to Karonga District and repetitive movements within Karonga. In the analysis of movements of individuals between households, only individuals who were seen at both LEP-1 and LEP-2 were used. If the LEP-2 household for an individual was different from that recorded in LEP-1 then that implied a change in household.

3.4 Leprosy cases

Trained Paramedical Leprosy Control Assistants (LCAs) examined all individuals thoroughly for leprosy (1). For the household contact analyses, index cases were defined as individuals diagnosed as having leprosy prior to (e.g. known from the LEPRA Control Programme) or at their first examination in the LEP-1 survey. Incident cases were those cases diagnosed only *after* the first examination. All diagnoses were categorised according to a diagnostic certainty scale (certain,

probable and possible), type of leprosy (pauci- or multi-bacillary) and exposure-reference time (time elapsed between first LEP-1 household examination date and first registration date for leprosy).

A case was classified as "New" if the first registration date was within six months prior to LEP-1 date (i.e. the reference time was within 6 months). If the registration date was more than six months before the LEP-1 examination date but after 1/1/75 a case was classified as "Old". If it was registered before 1/1/75, it was classified as "Ancient". This allows separation of new and recent cases, which were likely to be active and still infectious, from the older cases many of whom had been cured and were thus less likely to be still infectious. This distinction becomes important when we consider risks of disease among people who have been in contact with certain patients recently, or in the past.

Diagnostic certainty was determined by an algorithm which combined historical, clinical, microbiological and histopathological data (121). There were three categories: N ("narrow"), M ("middle") and O ("out"), which may be interpreted as the diagnosis being considered certain, probable or only possible, respectively. Individuals whose leprosy diagnosis was doubtful (i.e. "possible" according to the project's terminology) were neither included as index cases nor were they considered at risk of disease in the analyses presented here.

A case household was defined as that in which at least one (index) case was resident at the first LEP-1 examination. All other individuals living in that household at the time were considered to be in contact with the index case(s) and at contact-associated risk of disease.

Incidence analyses presented here exclude incident cases which occurred in the vaccine trial.

3.5 Special surveys

Some areas were visited more than twice by field teams in order actively to ascertain leprosy cases. These areas included Sample Survey, Kasowa, Kasyata and Lower Songwe as shown in Figure 3.2 and are coded SS/KKL in Figure 3.1. These special surveys were conducted between LEP-1 and LEP-2 in 1984 and also after LEP-2 between 1993 and 1996. Apart from active leprosy case finding, women in some of these areas (Sample Survey or Kasyata (SS/KAS)) were asked about the number of children they ever had and how many were still alive. These surveys were helpful not only for indirect estimation of child mortality but also for tracing of individuals who changed households more than once between LEP-1 and LEP-2.

The data used for this tracing exercise pertained to individuals who were resident in the Sample Survey or Kasyata (SS/KAS) geographical areas at LEP-1 (1980-84) and LEP-2 (1986-89) as determined by the geographical coordinates of their household (122). Using these geographical co-ordinates of the households, boundaries were demarcated for the areas included in the sample surveys. All households, and their residents, located within these boundaries during LEP-1 and LEP-2 surveys were identified providing data for estimation of the "sensitivity" of initial observed contact status as a measure of an individual ever having had actual household contact with a case.

This analysis included people seen during LEP-1, SS/KAS and LEP-2 surveys. The data extracted consisted of individuals who were identified as index cases of leprosy during the LEP-1 survey and incident cases identified during the follow-up (either in SS/KAS or during LEP-2 survey). Using such data, we calculated the risk of disease, sensitivity of observed household contact status, observed risk ratios and risk ratios adjusted for observed contact status misclassification. Index and incident cases were those with a diagnostic certainty of "certain" or "probable".

Case households were defined at LEP-1 as well as at LEP-2 and during the SS/KAS because an index case, just like any other individual in the study area, could change households over time. Contact status was determined at LEP-1, SS/KAS and LEP-2 depending on the household in which an individual was resident at that time.

Some individuals changed household more than once between LEP-1 and LEP-2 survey. Such individuals could either change to a different household each time they moved ("forward" move) or change back to their previous, or original, household ("return" move) the second time they moved.

3.6 Pedigree analysis

Because of the linking of individual records to their parents' records, it is possible to reconstruct family pedigrees. This allows identification of the familial relationship between individuals within households. Our interest in this analysis was mainly to identify relationships between index cases and other household members and to investigate whether the patterns of household change were influenced by the nature of the relationship to the head of household.

To identify relationships, we examined individual IDENTS linked to the IDENTS of their parents. By using a simple set of rules (full siblings share both parents, half siblings share one parent, etc) we were able to identify relationships which included children, siblings, half siblings, cousins, nephews, nieces and grand-children of the index case (or head) in household.

3.7 Data quality and management

Considerable attention has been given to data collection and processing throughout the history of the KPS to ensure high quality. Apart from frequent supervision of the field teams during data collection and occasional independent data collection to assess inter-observer variation, a series of manual and computer-assisted checks are employed to detect and minimize errors.

Field questionnaires were (and still are) sent to the coding office at the project headquarters. The forms are there checked for completeness and consistency including visual comparison with information collected in related households. Requests for collection of missing information and correction of errors or checking of apparently inconsistent data were sent to field teams on a weekly basis.

Data from the field questionnaires were at first coded onto coding forms and sent to Malawi Government Ministry of Finance Data Processing Unit (DPU) in Blantyre for key-punching onto tapes. Since 1985, data have been entered directly onto micro-computers at the project headquarters in Chilumba.

Tapes from the DPU and back up discs or tapes from Chilumba have been sent to the London School of Hygiene and Tropical Medicine where data are transferred onto micro-computers and subjected to further series of consistency checks. This initiated a further round of queries to Chilumba and to the field to correct errors identified. At regular intervals the cleaned data were and are archived and used for analysis. Currently, virtually the entire process of checking, (double) entry, correcting and archiving takes place at project headquarters.

Over the more than 20 years of the KPS, information technology has changed greatly, and the project has had repeatedly to convert all data to suit new hardware

and software systems. Currently the data are held on a combination of Oracle and Foxpro databases from which files are extracted for analysis in SAS or STATA.

The accumulated data on approximately 250,000 individuals provide a large unique resource with which to investigate household dynamics and their implications on household contact with leprosy cases over a period of 10 years between 1980 and 1989.

CHAPTER 4 DESCRIPTIVE ANALYSIS

This first part of the analysis is a detailed description of the demographic data. This is essential background information to set what follows in context. Since the study in Karonga, Northern Malawi is on-going, the descriptive analysis provides baseline measurements against which current developments can be compared. Further, these descriptive analyses provide estimates of input parameters for the stochastic micro-simulation model of household dynamics that has been developed and is described in Chapter 5.

Findings presented include the age and sex structure of the population, mortality, household structures and individual movements. Wherever possible, the results are compared to Malawi census data.

4.1 Age-sex structure of the population

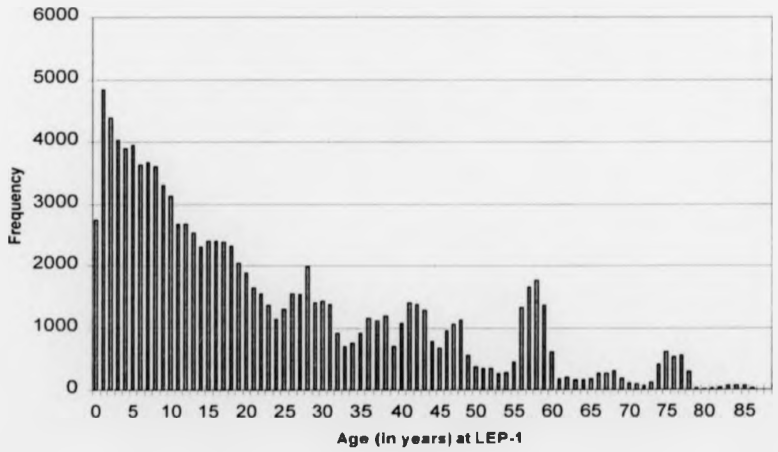
The total number of individuals seen during LEP-1 was 112886, of whom 59370 (52.6%) were female; in LEP-2 146129 individuals were seen, of whom 75575 (51.7%) were female. The sex compositions, observed in LEP-1 and LEP-2, were virtually identical to those observed in the 1977 and 1987 National Censuses for Karonga District (114).

The age range of individuals (with precise or assigned years of birth) in LEP-1 and LEP-2 was 0-89 and 0-93 years, with mean ages of 22.7 and 22.5 years respectively. These mean ages are virtually identical to the estimate derived from the 1987 National Census for Karonga. The median ages as in LEP-1 and LEP-2 were 17 and 16 respectively. The mean and median ages for females (23.5 and 18 years respectively) in LEP-1 were higher than those for males (21.9 and 15 years). A similar pattern was seen at LEP-2.

The population of Karonga was predominantly young, with 45.6% of the population under 15 years at both LEP-1 and LEP-2, reflecting a high fertility population. The percentage aged under 15 years was 46% for Karonga in the 1987 census.

Figure 4.1 shows the age distribution of the population of Karonga in LEP-1. The apparent age heaping especially for adults is a direct result of birth estimates based on the local events calendar as shown in Table 3.1, Section 3.2.1. A smoothed age distribution, obtained by giving those with estimated birth dates a random age in the "plausible" window based on the distribution of those with precise years of birth, is shown in Figure 4.2. This distribution shows a smooth decline with age with very little heaping. The procedure for smoothing is explained in Appendix A.2.

Figure 4.1 Age distribution of population of Karonga District, northern Malawi identified in the 1979-84 survey (LEP-1)



Figures 4.1 and 4.2 also reveal a deficit of infants aged under 1 year old. This is an artifact attributable to using years of birth to compute ages. To understand this, note that any child born in December of the year preceding a household's inclusion in the survey would be classified as being 1 year old if the interview took place in January and the child was in fact only 1 month old. The underestimation of infants based on this computational approximation is discussed in Appendices B and C, where it is shown that this method of assigning age leads to an underestimate of the number of infants by approximately 50% and that about 5480 infants under one year of age actually were seen during LEP-1.

Figure 4.2 Age distribution after redistributing those with birth estimates proportional to those with precise years of birth in LEP-1 (1979-84) in Karonga, Malawi.

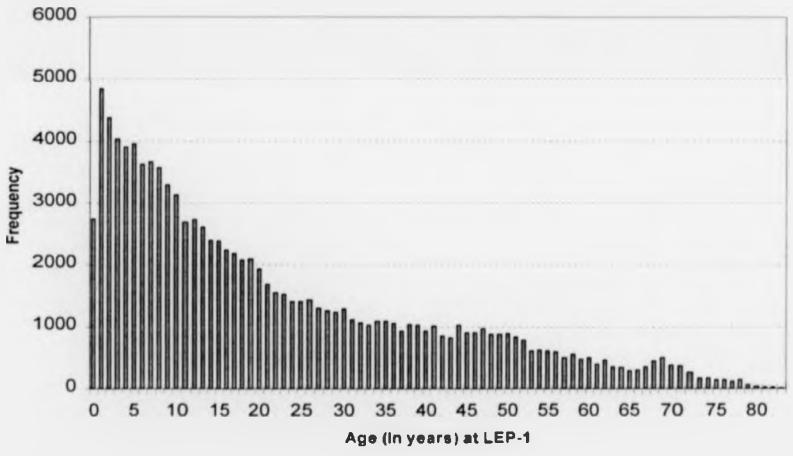
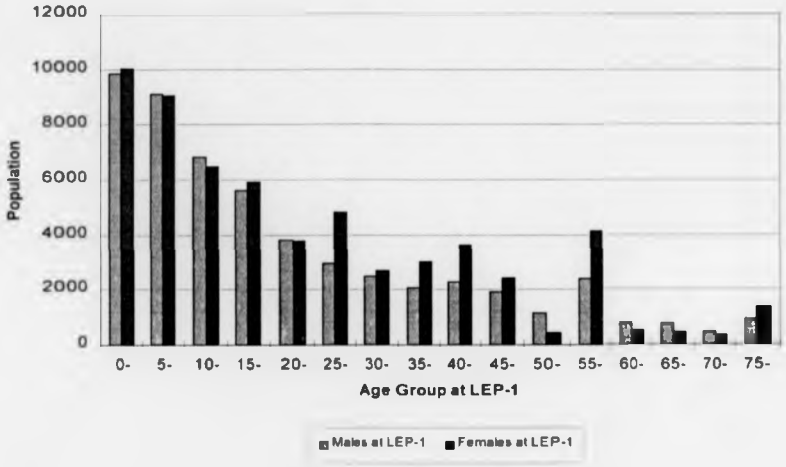


Figure 4.3 shows that the age distributions (not smoothed) of males and females under the age of 25 years in LEP-1 were similar. Females notably exceeded males in all middle age groups (25-50 years) probably because many males leave the district to find employment. LEP-2 revealed similar patterns.

The odd patterns for 50-54 and 55-59 year age groups in Figure 4.3 (in terms of frequency, with more females in 55-59 age group) are attributable to use of birth estimates (see Table 3.1), and the fact that women were less likely to give a precise birth year than men (Figure 4.4). This pattern was not observed in the smoothed distribution.

Figure 4.3 Population by age and sex, Karonga District, northern Malawi: LEP-1, 1979-84.

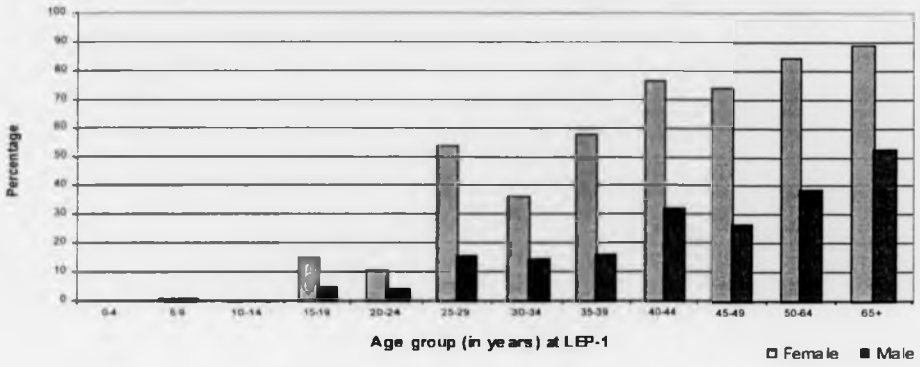


4.1.1 Birth year estimates: age and sex distributions

Birth year estimates were used in LEP-1 and LEP-2 surveys for individuals who did not give or know their precise year of birth. This led to an uneven age distribution, as seen in Figure 4.1. Seventy-nine percent and 85% of the population reported their precise year of birth in LEP-1 and LEP-2 respectively (the increase is likely to reflect increasing literacy in the population over time).

From Figure 4.4, we observe that the proportion of individuals who did not give their precise years of birth increased with age, and was consistently higher for females than males, with an overall figure of 30% (17479/58264) for females and 11% (5716/51964) for males.

Fig 4.4 Age-specific percentage of individuals with birth estimates by sex. Karonga District, northern Malawi (1979-84)



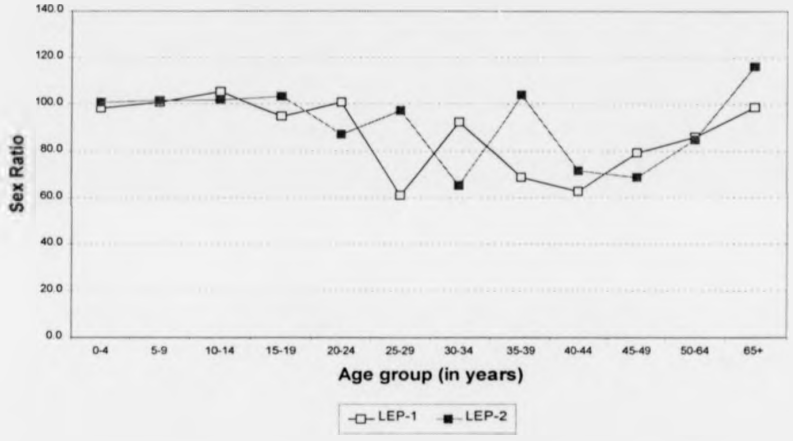
4.1.2 Sex ratios

The sex ratio is the ratio of males to females in a population, usually expressed as the number of males per 100 females. The sex ratios for all Karonga were 90.4 and 93.6 during LEP-1 and LEP-2 respectively. These were within one percent of the 1977 and 1987 National Census sex ratios for Karonga District.

Figure 4.5 shows an excess of females over males for those aged over 25 years in both LEP-1 and LEP-2 surveys. This is probably due to out-migration of males, especially young adults, to urban areas outside Karonga in search of employment and better economic opportunities. The figure shows the sex ratios with wider age groups after 50 to reduce the effect of age misplacement due to birth estimates used. The sex ratios at older age groups increase with age. The increase is

unexpected because women live longer than men in many populations (79) and one might expect the sex ratios to fall for older age groups. The observed trend is probably in part a reflection of men returning to their homes after having retired from employment outside the district, but it may not be precise because the analysis combined individuals with birth estimates and those with precise years of birth.

Figure 4.5 Age-specific sex ratios during LEP-1 and LEP-2 surveys. Karonga District, Northern Malawi, 1979-89.



The sex ratios for LEP-1 and LEP-2 revealed the same patterns except that the ratios at LEP-2 were shifted forward by approximately 5 years, the time between LEP-1 and LEP-2. This is reassuring as it confirms that we are dealing with the same population in both studies and appears to reveal an interesting birth cohort effect (e.g. the low sex ratio for 25-29 year olds in LEP-1 may reflect that males born around 1955 were especially likely to work outside the district).

4.1.2 Birth rate estimate

There were 2740 infants out of 112886 individuals recorded at LEP-1. This translates to a crude birth rate of 24 births per 1000 persons which is low compared to 1977 and 1987 Malawi Census figures. This is partly attributed to the way ages were computed using birth years only leading to an underestimation of infants.

After adjusting for this underestimation, the crude birth rate for LEP-1 and LEP-2 were 49 and 43 births per 1000 persons respectively. These are still underestimates as total infants were obtained from cross-section studies and such numbers were not corrected for actual infant (or neonatal) mortality in that population.

4.2 Mortality estimation

Deaths were ascertained prospectively through follow-up from LEP-1 to LEP-2. Using that information, direct estimates of overall and age-sex specific mortality were obtained using the person years at risk approach. Out of 112886 individuals recorded at LEP-1, 5664 were recorded dead during the 5-year follow-up period between LEP-1 and LEP-2. This translates to a crude death rate of 10 deaths per 1000 persons per year.

Mortality estimation for individuals with precise years of birth was done separately from those with assigned years of birth in order to explore if their mortality distributions were different.

Indirect estimates of child mortality were obtained using the indirect CEB/CS demographic technique developed by Brass (123) based on information from women surveyed in the Kasowa and Lower Songwe areas in 1993 and 1995.

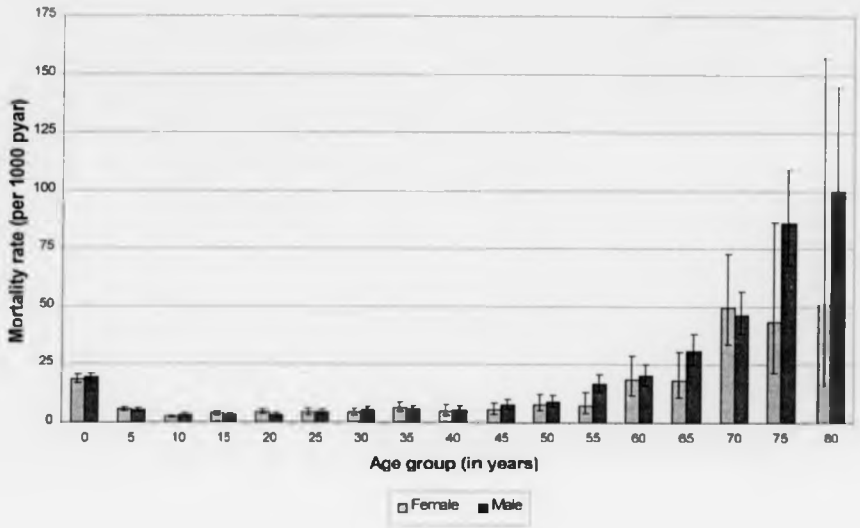
4.2.1 Direct mortality estimation

4.2.1.1. Person years at risk (PYAR) analysis

Age-specific mortality rates could be calculated directly from follow-up data on the LEP-1 population as obtained in the LEP-2 survey. Person years at risk (PYAR) were computed by subtracting year of interview at LEP-1 from the earlier of the date of death, date of leaving household (when this implied lost to follow-up) or date of LEP-2 interview. The overall mortality rate was estimated as 7.3 (95% CI: 7.1, 7.6) per 1000 person-years at risk which is low. The results shown in Figure 4.6, reveal high mortality for children under 5 years of age and in older age groups, with no systematic differences between males and females for children and middle-aged adults. Extreme values for mortality in the oldest age groups, aged over 85, were unreliable as they were based on small numbers (2.3 PYAR and only 1 death for females, 33.4 PYAR and 5 deaths for males) and are not shown on the graph.

The observed mortality among under fives, approximately 20 per 1000, appears low, and is discussed further in Section 4.2.1.3. The lowest mortality is observed in children aged 10-14 years. For older people, the estimates are low because we only dealt with individuals with precise years of birth (see next section). Over age 45, mortality rates were generally higher for males than females with an adjusted rate ratio of 1.35 (95% CI: 1.11, 1.65).

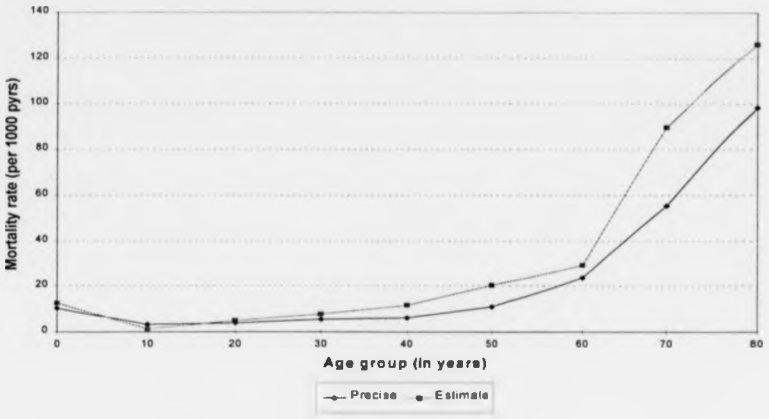
Figure 4.6 Mortality rates (per 1000 PYAR) between LEP-1 and LEP-2 surveys by sex for individuals with precise years of birth. Karonga District, northern Malawi 1979-89



4.2.1.2 Individuals with precise or assigned birth years

In Figure 4.7, we observe higher mortality in individuals who could only estimate their birth years using the local events calendar, compared to those with precise years of birth. This difference widens with increasing age and provides justification for separate mortality estimation for the two groups.

Figure 4.7 Mortality rates (per 1000 pyar) for individuals with precise or estimated years of birth: Karonga District, Malawi 1979-89.

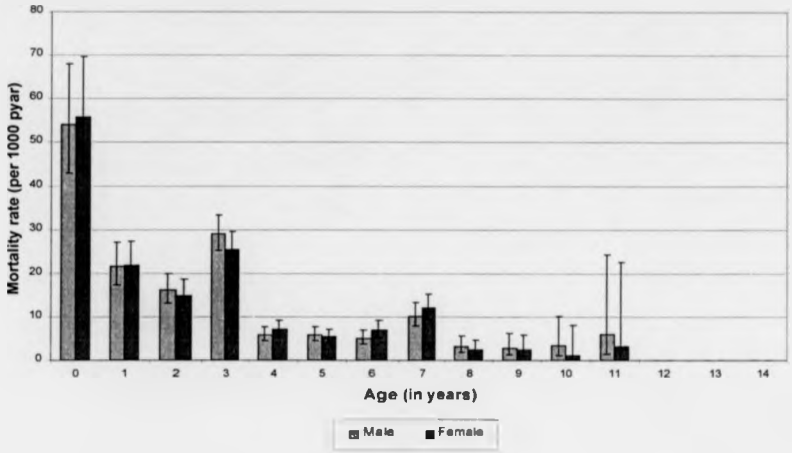


4.2.1.3 Children under 5 years of age

We examined mortality by single years for children who were under 5 years of age at LEP-1. A child from this age group at LEP-1 would be at most 14 years old by the end of LEP-2 (if first interviewed at age four at the start of LEP-1, in 1979 and again seen at the end of LEP-2, in 1989).

Figure 4.8 shows no evidence of sex differences in age-specific child mortality. Although there was a general decrease in mortality with increasing age of children, there was a sharp increase for children about 3 years old and to a lesser extent at ages 7 and 11. Similar patterns in males and females make it look less like chance and more like an artifact, perhaps attributable to some subtle digit bias associated with a recollection of approximately how old the children were when they died rather than the actual year of death. It is possible that the patterns could be real: at age 3, deaths might be associated with weaning and at age 7, they might reflect an increased exposure to infection as it coincides with start of school for most children.

Figure 4.8 Age-specific mortality rates (per 1000 pyar) between LEP-1 and LEP-2 surveys for children under 5 years of age at LEP-1. Karonga District, northern Malawi 1979-89.



Most importantly, these figures greatly underestimate true child mortality rates in Karonga, especially for infants. This is evident when we compare them with census results discussed in Section 2.4.2. These mortality estimates are averaged over 5 years. Under-ascertainment of very young infants at LEP-1 could have led to selective exclusion of the group at highest mortality risk. Mortality in the first year of life is not uniform – it is much higher in the first few months than later months of the year, whereas children under 1 year of age interviewed in the LEP surveys are an over-representation of infants in the later part of the first year of life after which period relatively low mortality is expected.

4.2.2 Indirect mortality estimation

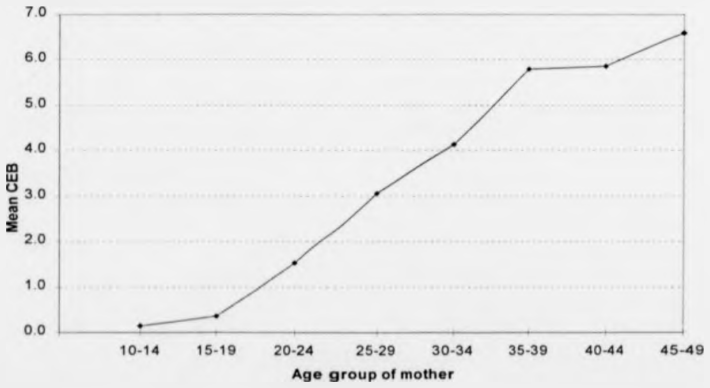
Information was collected from women of Kasowa and Lower Songwe areas regarding number of live born children they had had and how many were still alive. Based on this information, indirect estimates of child mortality were obtained using

the children ever born/children surviving (CEB/CS) technique developed by Brass (124).

Of 3106 women interviewed in these surveys, 2983 (96%) gave precise years of birth and only these women were used in the analysis.

Figure 4.9 shows the expected increase in mean number of children ever born (CEB) with age of mother up to age group 45-49. The mean CEB to women aged 45-49 (6.60) compares well with the Malawi Demographic and Health Survey (MDHS) figure of 6.7 estimated in 1992 and with the rural total fertility rate of 6.0 estimated during the 1988-92 period in the northern region of Malawi (45).

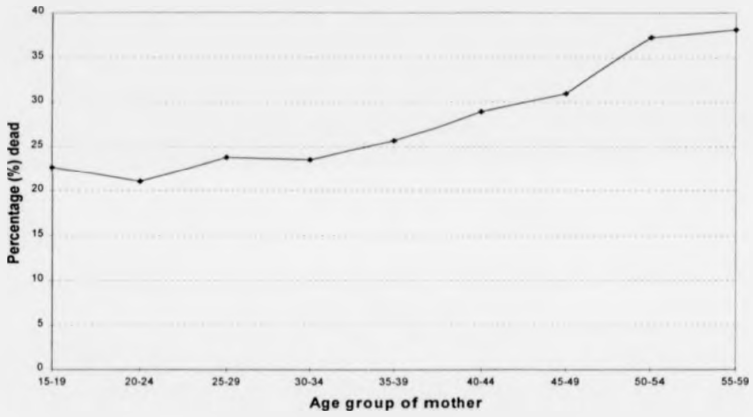
Figure 4.9 Mean Children Ever Born (CEB) by age of mothers when interviewed (sample survey data, Karonga District, Malawi 1993-6)



The proportion dead of CEB to women aged between 15-49 years increased with age of the women, from 23% in those aged 15-19 up to 31% in those aged 45-49 years as shown in Figure 4.10. The increase reflects in part the longer follow-up of those born longer ago in the past.

In order to obtain indirect estimates of under-5 mortality, the CEB/CS Brass technique (124) was used. The proportion dead of CEB by age of mother is the required input for this method. The Brass technique allows a choice of several standard models of mortality to compare with the mortality observed in the population under study. The "South" model was used as the standard against which under 5 mortality in Karonga was compared (123) since this model is characterised by high child mortality relative to adult mortality.

Figure 4.10 Proportion dead of CEB by age group of mother when interviewed (sample survey data, Karonga District, Malawi 1993-6).

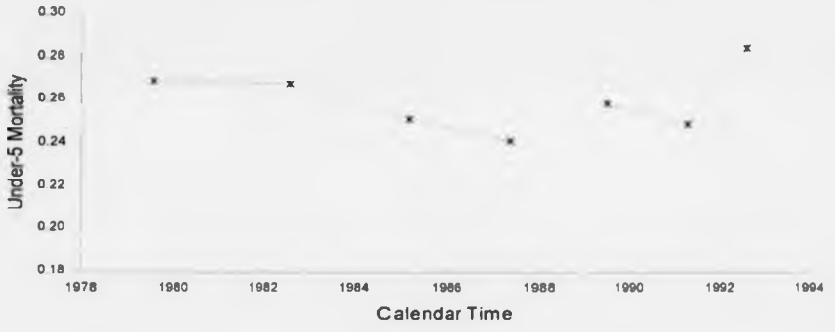


Mortality levels obtained from each age group of mothers correspond to different periods of time. They can be time-located over a calendar period using an extension of the original Brass method (125). They are then converted to under-five mortality for easy comparability across different age groups of women.

The date chosen as reference point to time-locate mortality levels relating to each age group of mothers was 1994 because 99% (3082) of the women had been interviewed by 1994 and it was also the mode and median year of interview.

After time-locating mortality, Figure 4.11 shows that under-5 mortality ranged between 24-28% in the period 2-15 years prior to the survey (started in 1993). Under-5 mortality was 269, 252 and 259 per 1000 live births in 1979, 1985 and 1989 respectively. These under-5 mortality rates are higher than the overall under-5 mortality value in all Malawi of 234 deaths per 1000 live births reported in MDHS (45) for the period 1988-92. It is interesting that these estimates of under-5 mortality were obtained from SS/KAS areas located in north Karonga which we show in the next section has higher mortality rates than the southern part of Karonga.

Figure 4.11 Trend in under-5 mortality for periods before 1993-6 survey: Kasowa area, Karonga District, northern Malawi.



Ignoring the most recent point, the plot of mortality trend shows a suggestion of slow decline in under 5 mortality up to 1988 but the trend is not clear the closer we get to the survey period. The most recent point is often ignored in such analyses because it is highly erratic and usually shows higher mortality than the general trend (45). The method of estimation assumes that mortality of children is related solely to age of women and this is not true for younger mothers. This method does not take into account the complex relationship between HIV/AIDS and infant and child

mortality in populations affected by HIV/AIDS. Children of HIV-infected mothers are at greater risk of mortality if they have been infected vertically or are left as orphans.

4.2.3 Risk factor analysis for mortality

Risk factors for mortality using a Poisson regression model are outlined in Table 4.1. There was a relationship between mortality and geographical residence zone, with the highest mortality in the northern hills. Mortality decreased from the northern part down to the southern part of the district. Categorising zones into northern (zones A and B) versus southern Karonga (zones C, D and E), shows that the individuals living in the northern part had 1.30 (95% CI: 1.19,1.39) times higher mortality than those in the southern part. The inclusion of the town centre within the south was justified as the mortality rates are almost the same as for the southern part.

After controlling for age and sex, there was no evidence that leprosy cases were at an elevated risk of dying with a rate ratio of 1.14 (95% CI: 0.84, 1.54) after adjustment for education, occupation, position in household, geographical zone and BCG scar status. Not surprisingly tuberculosis was associated with an appreciable increase in mortality (rate ratio 2.30, 95% CI: 1.64, 3.22) after adjustment for age, sex, education, occupation, position in household, geographical zone and BCG scar status.

The estimates adjusted for age and sex also show that individuals with no BCG scar were 1.43 (95% CI: 1.31, 1.57) times at risk of death compared to those with a BCG scar recorded in LEP-1.

Table 4.1 Assessment of risk factors for mortality. Karonga District, northern Malawi 1979-89.

Risk Factor		Crude estimates			Estimates adjusted for age and sex only.	Estimates adjusted for other risk factors (including age, sex, education level and occupation)	
		Number of deaths /Person years (PYR)	Crude mortality rate (per 1000 PYR)	Rate Ratio	Rate Ratio (95% CI)	Rate Ratio (95% CI)	P-value
Geographical zone	Northern Hills (zone A)	210/17796	11.8	1	1	1	
	Northern lake shore (zone B)	1261/155766	8.1	0.69	0.68 (0.59,0.79)	0.68 (0.58,0.79)	<0.001
	Semi-urban area (zone D)	286/43413	6.6	0.56	0.55 (0.46,0.66)	0.60 (0.49,0.74)	<0.001
	Southern Hills (zone C)	723/106500	6.8	0.58	0.56 (0.48,0.66)	0.56 (0.48,0.66)	<0.001
	Southern lake shore (zone E)	511/83547	6.1	0.52	0.49 (0.42,0.58)	0.51 (0.43,0.61)	<0.001
North versus South*	Zones A, B	1471/173562	8.2	1	1	1	
	Zones C, D, E	1520/233461	6.5	0.77	0.75 (0.70,0.81)	0.77 (0.72,0.84)	<0.001
Leprosy case	Non-case	2939/401576	7.3	1	1		
	Case (includes MB & PB cases)	52/5458	9.5	1.30	1.25 (0.94,1.65)	1.14 (0.84,1.54)	0.401
Tuberculosis case	Non-case	2945/404884	7.3	1	1		
	Case	46/2150	21.4	2.94	2.69 (2.00,3.61)	2.30 (1.64,3.22)	<0.001
BCG Scar	Yes	1127/189458	6.0	1	1	1	
	No scar	1725/172545	10.0	1.68	1.48 (1.36,1.62)	1.43 (1.31,1.57)	<0.001

* Fitted in a separate poisson regression model with all risk factors, except geographical zone

Mortality was 1.14 (95% CI: 1.03, 1.25) times higher in individuals who could only estimate their birth years using the local events calendar than those with precise years of birth after adjusting for age, sex, education, occupation and housing. This finding of borderline significant difference probably reflects residual confounding of socio-economic factors in two groups, which in turn are related to poverty.

4.3 Summary

The population (52% female) of Karonga during the 1980s was predominantly young, typical of rural African communities, with 46% of the population under 15 years of age. The crude sex ratios (number of males per 100 females) at LEP-1 and LEP-2 were 90 and 94 with an excess of females over males for those aged between 25 and 40 years. This is attributable to migration of males to urban areas outside Karonga in search of better employment opportunities. The age distribution reflects a high fertility population, with TFR of 6.6 children.

The crude birth rates for Karonga District during LEP-1 and LEP-2 were estimated to be 49 and 43 births per 1000 persons respectively. The respective figures for the 1977 and 1987 census were 50 and 37 births per 1000 persons (114). The total fertility rate of 6.6 from the LEP data was virtually identical to the MDHS (45) figure of 6.7 but slightly higher than the 1987 census figure of 5.7.

We noted a shortfall of infants, due to the method of computing age as (interview/examination year - birth year). Taking into account mortality in the first year of life, we estimate that this method leads to 47% of infants being misclassified as one year olds. This underestimation of infants leads to inaccurate estimates of infant mortality and morbidity, especially in studies in which reporting of birth dates is poor.

We note high mortality in young children (under 5 years of age) and older adults, with mortality higher for males than females aged over 45. We also noted significantly higher mortality in TB cases than non-cases, rate ratio of 2.30 (95% CI: 1.64, 3.22), but no significant difference in mortality between leprosy cases and non-cases, rate ratio of 1.14 (95% CI: 0.84, 1.54), after adjusting for age, sex and socio-economic status. This is consistent with data from other populations showing that leprosy is not a "fatal" disease (16).

The finding of significantly lower mortality among those with compared to those without a BCG scar, rate ratio of 1.43 (95% CI: 1.31, 1.57), is interesting. Though it persists after adjusting for education and occupation, we suspect that it reflects residual socio-economic confounding rather than a biological effect of the vaccine. No evidence of a mortality effect associated with BCG was seen in the KPT vaccine trial (16, 126). Recently, studies have been conducted that investigate association between vaccine and survival (127, 128) showing that BCG and measles vaccines halved child mortality. Although several problems have been pointed out with such studies, in areas of high mortality, vaccines may have substantial non-specific effects on mortality from all causes.

Mortality in individuals with birth estimates was 1.14 (95% CI: 1.03, 1.25) times higher than those with precise years of birth. We also note higher mortality in the north than south Karonga, with a rate ratio of 1.30 (95% CI: 1.19, 1.39) probably reflecting differential environmental factors.

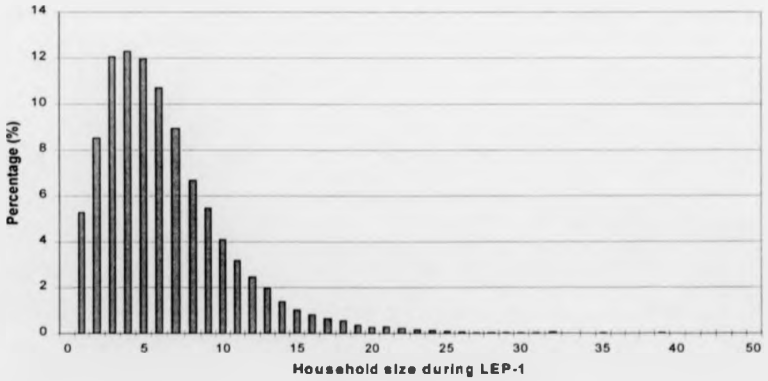
4.4 Household structures

The household structure analyses included consideration of household size, age and sex structure of heads of households, changes in headship and membership of households. To determine change of household by individuals between LEP-1 and LEP-2, household serial numbers were compared (see Section 3.3 on definition of new households and problems raised).

4.4.1 Household size

There were 17905 and 23665 households identified in LEP-1 and LEP-2 respectively. Figure 4.12 shows the frequency distribution of size of households at LEP-1 (similar distribution at LEP-2 in Figure F.2 of Appendix). The mean and median household sizes were 6.42 and 5 in LEP-1 and 6.37 and 5 in LEP-2 respectively.

Figure 4.12 Relative frequency distribution of household size for Karonga District, Malawi during LEP-1 (1979-84).



4.4.2 Position in household

Table 4.2 shows the distribution of LEP-1 households by size and membership status. It will be noted that not all households were recorded as having a *resident* head, reflecting the fact that some individuals were heads of more than one household but were recorded only within the household in which they resided (see Section 3.3). Thus, it is possible that some single individual households were really two-person households with an absent head. By definition, if a household had only 1 individual, that one is the head. However, this was not always the case not only because of absentee heads but also in part due to errors in recording positions on household and personal forms. This was observed in 5 of the 23 single individual households, in which the recorded single resident "member" based on personal records form was actually recorded as a head on the household form. Nineteen of these 23 individuals were aged over 30 and mostly women (20/23).

The majority (600, 63%) of the 947 single individual households consisted of a single male. Most of these individuals were over 40 years of age (see Appendix Table F.10) and were the recorded heads (99% of the males and 94% of the females). A similar distribution of position in household of varying sizes was observed at LEP-2.

Table 4.2 Frequency distribution of the population by position and household size at LEP-1

	Position in Household at LEP-1				
Household Size	Frequency (row%)			Total individuals (col %)	Total Households (col %)
	Head	Member	Others		
1	924 (97.6)	23 (2.4)	0 (0.0)	947 (0.8)	947 (5.3)
2	1484 (48.6)	1537 (50.4)	29 (1.0)	3050 (2.7)	1525 (8.5)
3-5	6356 (24.5)	19294 (74.2)	344 (1.3)	25994 (22.6)	6502 (36.3)
6-10	6308 (13.0)	40839 (84.3)	1299 (2.7)	48446 (42.2)	6430 (35.9)
11+	2446 (6.7)	32026 (87.9)	1951 (5.4)	36423 (31.7)	2497 (14)
Total	17518	93719	3623	114860	17901

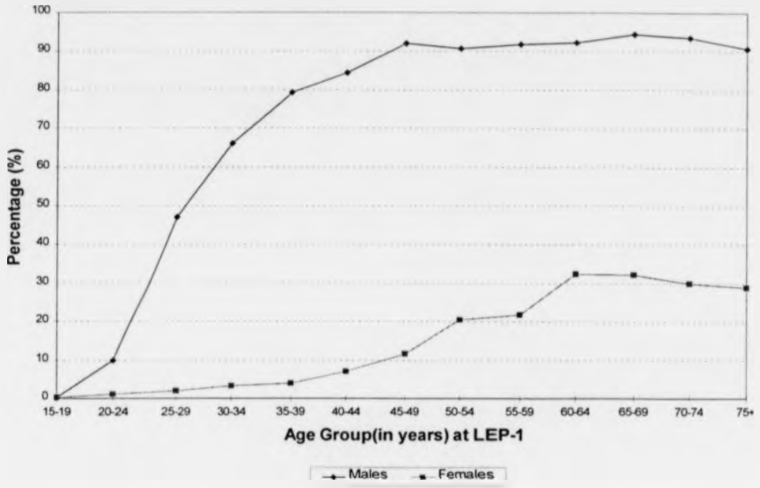
More than 99% (17828) of the numbered households recorded at LEP-1 were seen only once during the LEP-1 period, but 100 households were seen twice and one was seen three times. The total number of individuals in Table 4.2 is higher than actually observed at LEP-1 (112886) because some individuals were actually identified in more than 1 household in LEP-1. Keeping only one record for each individual would lead to underestimate household size. It should be noted that heads of multiple households were not necessarily interviewed more than once in LEP-1. Two of the 101 households seen more than once were recorded as having a different head at the second visit. The two previous heads were not seen again in LEP-1 (both were last seen alive, after LEP-1, in 1986, one in a different household and the other with no record of household). Their household changes were probably a reflection of remarrying or polygamy within or outside Karonga District. 14780 (85%) heads of LEP-1 households were males and 11168 (64%) gave precise years of births.

Of the 23,665 households recorded during LEP-2, 1603 were seen more than once, and of these 48 were recorded as having two heads within the study period. About 84% (19408) of heads of LEP-2 households were males and 70% of heads gave precise years of birth.

Figure 4.13 shows the percentage of males and females who were heads of households in LEP-1, by age. The proportion of males who were heads rose to 90% at age group 45-49. Approximately, 92% of male adults over age 45 were household heads.

The proportion of females who were heads during LEP-1 increased slowly with age to approximately 30% by age 60. This pattern was similar in LEP-2. As a consequence, female heads of households were on average older than male (mean age of 55 and 47 years respectively), see Table F.11 in Appendix.

Figure 4.13 Percentage of males and females who were heads of household, by age at LEP-1, Karonga District, Malawi 1979-84.



4.4.3 Changes in heads of households

There were 15015 households identified at both LEP-1 and LEP-2. In 9.3% (1390) of these households the head changed between the surveys. Table 4.3 presents a breakdown of these households by sex of initial and subsequent head.

Table 4.3 Cross-tabulation of previous and new heads for households which changed heads between LEP-1 and LEP-2, by sex.

		New head at LEP-2		
		Female	Male	Total
Head at LEP-1	Female	103	287 (74%)*	390 [28%]**
	Male	824 (82%)*	176	1000
	Total	927 (67%)*	463 (33%)*	1390

Note: * =row percentages, **=column percentage.

Of these households, 1000 (72%) had a male head in LEP-1 but only 463 (33%) had a male head in LEP-2. In the majority of the households, (824+287=)

1111/1390 (80%), the headship was assumed by an individual of the opposite sex, i.e. 82% of households originally headed by men were taken over by women, and 74% of those originally headed by women were taken over by men. The changes were due to death of the original head in 973 (70%) of these households (831 males, 142 females), 97% of whom were aged over 40 years. In another 160 (12%), the original head was recorded as having left household (119 males, 41 females). The LEP-2 status was unknown or missing for 258 (18%) of the original heads who had been supplanted (51 males, 207 females). This explains the age trend difference between male and female heads in Figure 4.13, a reflection of women taking over headship of households after being widowed.

4.4.4 Position changes in household between the surveys

Tables 4.4a and b summarise position changes by age and sex between the LEP-1 and LEP-2 surveys. The tables show that in general people remain in the same household and same status. This is particularly so for male and older (over age 30) female heads. An appreciable proportion of male members over age 20 become heads of new households. The majority of "Others" are recorded as members of other households, with older males (over age 20) becoming heads when seen in LEP-2.

Table 4.4a Frequency table for changes in position between LEP-1 and LEP-2 for surviving resident males by age, Karonga District, Northern Malawi 1979-89. The percentages of the total by age who stayed in the same household or changed are given in parenthes

			LEP-2 Household and position								
			Same Household				New Household				TOTAL
		Age	Head	Member	Others	Total	Head	Member	Others	Total	
LEP-1 Position	Head	0-9	-	-	-	-	-	-	-	-	-
		10-19	7 (77 8%)	-	-	7	0	2 (22 2%)	0	2	9
		20-29	1333 (93 8%)	2 (0 14%)	0	1335	12 (0 8%)	54 (3 8%)	20 (1 4%)	86	1421
		30-	10109 (95 5%)	2 (0 02%)	0	10111	244 (2 3%)	144 (1 4%)	90 (0 9%)	478	10589
		All	**11449**	**4**	**0**	**11453**	**256**	**200**	**110**	**566**	**12019**
	Member	0-9	-	11998 (79 7%)	9 (0 06%)	12007	-	2830 (18 8%)	209 (1 4%)	3039	15046
		10-19	15 (0 2%)	7043 (72 9%)	11 (0 11%)	7069	632 (6 5%)	1777 (18 4%)	183 (1 9%)	2592	9661
		20-29	78 (2 5%)	1261 (40 8%)	11 (0 36%)	1350	1343 (43 5%)	322 (10 4%)	72 (2 3%)	1737	3087
		30-	206 (15 8%)	445 (34 1%)	3 (0 23%)	654	519 (39 8%)	93 (7 1%)	38 (2 9%)	650	1304
		All	**299**	**20747**	**34**	**21080**	**2494**	**5022**	**502**	**8018**	**29098**
	Others	0-9	-	52 (17 7%)	5 (1 7%)	57	-	223 (76 1%)	13 (4 4%)	236	293
		10-19	0	21 (18 9%)	1 (0 9%)	22	9 (8 1%)	75 (67 6%)	5 (4 5%)	89	111
		20-29	0	15 (8 5%)	10 (5 6%)	25	83 (46 9%)	49 (27 7%)	20 (11 3%)	152	177
		30-	4 (1 5%)	12 (4 6%)	23 (8 8%)	39	142 (54 6%)	36 (13 8%)	43 (16 5%)	221	260
		All	**4**	**100**	**39**	**143**	**234**	**383**	**81**	**698**	**841**
TOTAL			11752	20851	73	32676	2984	5805	693	9282	41958

Table 4.4b Frequency table for changes in position between LEP-1 and LEP-2 for surviving resident females by age, Karonga District northern Malawi 1979-89. The percentages of the total by age who stayed in the same household or changed are given in parenthes

			LEP-2 Household and position							
			Same Household				New Household			TOTAL
		Age	Head	Member	Others	Total	Head	Member	Others	Total	
LEP-1 Position	Head	0-9	-	-	-	-	-	-	-		
		10-19	0	4 (66 7%)	0	4	0	1 (16 7%)	1 (16 7%)	2	6
		20-29	20 (23 5%)	35 (41 2%)	0	55	2 (2 4%)	25 (30 5%)	3 (3 5%)	30	85
		30-	1810 (79 9%)	133 (6 6%)	2 (0 1%)	1745	14 (0 7%)	241 (12%)	16 (0 8%)	271	2016
		All	**1830**	**172**	**2**	**1804**	**18**	**267**	**20**	**303**	**2107**
	Member	0-9	-	11148 (76 6%)	14 (0 1%)	11162	-	3184 (21 9%)	201 (1 4%)	3385	14547
		10-19	6 (0 07%)	4252 (46 4%)	93 (1 0%)	4351	25 (0 27%)	4538 (49 5%)	256 (2 8%)	4819	9170
		20-29	43 (0 68%)	4205 (66 3%)	21 (0 3%)	4269	62 (1%)	1884 (29 4%)	146 (2 3%)	2072	6341
		30-	830 (6 2%)	10251 (76 4%)	6 (0 04%)	11087	475 (3 5%)	1700 (12 7%)	153 (1 1%)	2328	13415
		All	**878**	**29856**	**134**	**30868**	**562**	**11288**	**756**	**12604**	**43473**
	Others	0-9	-	58 (19 9%)	6 (2 1%)	64	-	214 (73 3%)	14 (4 8%)	228	292
		10-19	0	20 (10 1%)	2 (1 0%)	22	0	171 (86 4%)	5 (2 5%)	176	198
		20-29	1 (0 4%)	28 (11 7%)	7 (2 9%)	36	6 (3%)	181 (75 4%)	17 (7 1%)	204	240
		30-	1 (0 4%)	29 (11 8%)	8 (3 3%)	38	35 (16 9%)	165 (67 3%)	7 (2 9%)	207	245
		All	**2**	**136**	**23**	**160**	**41**	**731**	**43**	**815**	**975**
TOTAL			2511	30163	159	32833	619	12264	819	13722	46555

4.4.5 Summary

These data give an insight into the patterns of household structure and position change in this rural African population. Households contained on average 6.4 individuals in LEP-1 and LEP-2, and only 18% and 4% of households had more than 10 or 15 individuals respectively.

About 85% of heads of households during both LEP-1 and LEP-2 were males. This was expected in this population due to the patrilineal nature of the northern region society. Women often assumed headship after the male heads had either died or left the household. Thus, female heads were, on average, older than male heads.

As in most societies, new households are in general formed by young adult males and females, with the male assuming the headship role. The observed changes in household membership status between surveys are age and sex dependent. Most individuals remained in the same household. This was particularly true for male and old female heads. Young female heads (under 30 years) became members of the same or other household over time. For male members, the probability of becoming head of a new household in the 5 years increased with age, being 40% for males over age 30. We will explore these changes in greater detail in the next section.

Approximately 5% of households were recorded as having only a single resident individual. A high proportion (96%) of the single person households in LEP-1 dissolved over the following 5 years, as individuals became members of other households by LEP-2.

The majority of older adults (aged over 40 years) in one-person households were females – a reflection of more men than women re-marrying after divorce or death of a spouse and also of husbands (absentee heads) working out of Karonga.

4.5 Change of households

In this section, we present results of analyses of movements of individuals between household, in- and out-migration by age, sex and position in households during the 5-year period between the LEP-1 and LEP-2 surveys.

4.5.1 Movements between households

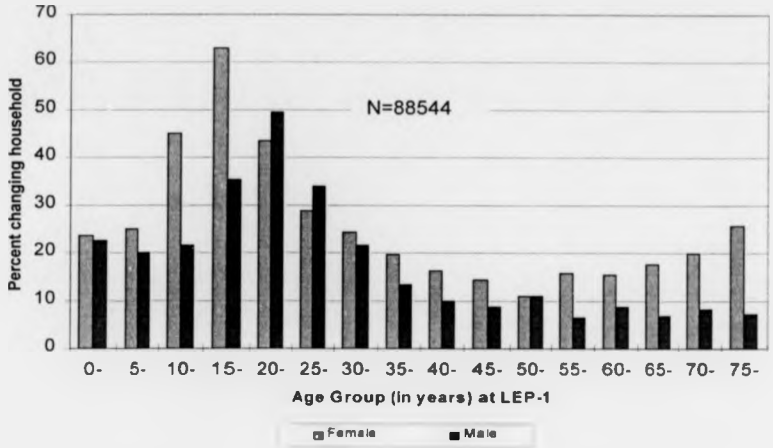
There were 88544 individuals identified and recorded at both LEP-1 and LEP-2 of whom 53% were female. It was observed that 30% (13720) and 22% (9306) of females and males respectively changed households between LEP-1 and LEP-2. Of the 23026 individuals who changed households, 60% were women.

Figure 4.14 shows important age and sex differences in individual propensity to change household during the inter-survey interval (five years for most individuals). We note high rates of household change (20%), but no apparent sex differences, in children under 5 years of age.

There was a very active change of households among young adults aged 10-30 years – with more than 30% changing households in most age groups. There is an earlier age peak for females (over 60% in the 15-19 year age group) compared to males (slightly below 50% in the 20-24 year age group).

The figure also shows a relatively low rate of household change for persons over 40 years of age, with a higher rate of change for older females (over 15%) than males (under 10%) in all age groups. The rate of change for older females increases with age.

Figure 4.14 Percentage of individuals changing household by age and sex between LEP-1 and LEP-2 surveys: Karonga District, northern Malawi, 1980-89.



4.5.2 Forward and return moves (three survey population)

A more detailed analysis of movements of individuals between households was possible for a subset of individuals observed in three surveys: LEP-1 survey, sample survey (SS) and LEP-2 survey, see Figures 3.1 and 3.2. This gave an opportunity to measure the proportions of individuals who made forward and return moves (see Section 3.5 for definitions), by age and sex.

The following analysis includes 3902 individuals who were seen in all three surveys (LEP-1 survey, SS and LEP-2 survey). Movements were classified by comparing households in which an individual was resident at any of these points.

Ninety-five percent of individuals recorded in the three surveys were first interviewed between 1979-81 during LEP-1. The SS were done in 1984 and 91% were also interviewed between 1988-89 during LEP-2. Thus, on average, these

surveys are five years apart and span changes in households over a ten-year period. Movement patterns are tabulated by sex in Table 4.5.

Table 4.5 Frequency distribution of movement patterns by sex.

Type of move	Female	Male	Total (col %)
AAA	1341	1316	2657 (68.0)
AAB	334	223	557 (14.3)
ABA	65	62	127 (3.3)
ABB	170	154	324 ((8.3)
ABC	161	76	237 (6.1)
Total			3902

Note **AAA** - no move. **AAB** - a move after the SS. **ABA** - return move to previous household at LEP-1; **ABB** - a move within or after LEP-1 but no move after SS and **ABC** – two forward moves to different households.

We observe that (3902-2657=) 1249 (32%) individuals changed households when all the three surveys were considered. This compares to (127+324+237=) 688 (18%) between LEP-1 and SS and to (557+127+237=) 921 (24%) between the SS and LEP-2.

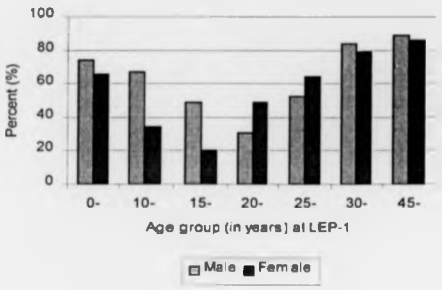
Of the 88544 total individuals seen at both LEP-1 and LEP-2, 23026 (26%) changed households between the surveys. This is similar to the rate of household change between the SS and LEP-2 observed in the population surveyed three times.

Altogether, 127 (3.3%) and 237 (6%) made two moves, either return (ABA) or forward (ABC) moves respectively. Of these, women were more likely to make forward moves than men (161/226=71% compared to 76/138=55%, p=0.002).

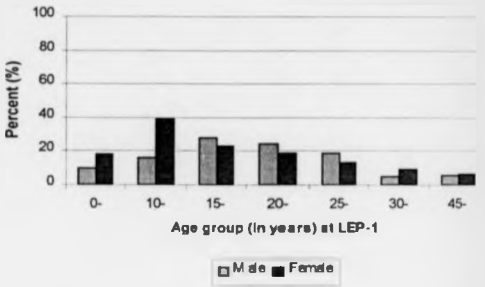
These several populations are broken down further by age in Figure 4.15. We see the age distribution of those who moved between SS and LEP-2 (AAB, Fig 4.15b) shifted 5 years to the right of those who moved between LEP-1 and the SS, (ABB, Fig 4.15c), as expected. The proportion of individuals who made forward moves (ABC), between the three surveys, peaks in the 15-19 year age group (about 30% of all females and 15% of all males) and steadily decreases with age thereafter to less than 4% in those aged over 45, see Figure 4.15d. The percentages are small for return moves (ABA).

Figure 4.15 Distribution of movement patterns between the three surveys (LEP-1, SS and LEP-2) by age when first seen in LEP-1 and sex. Karonga District, northern Malawi, 1979-89. Percentages refer to all individuals seen in the three surveys.

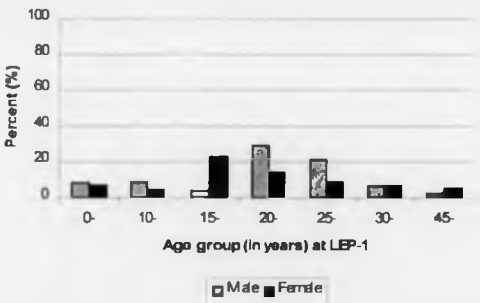
a) AAA



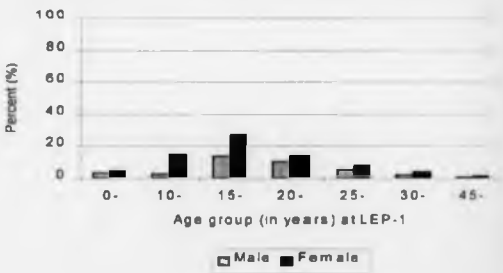
b) AAB



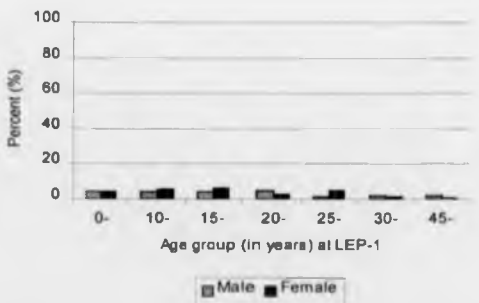
c) ABB



d) ABC



e) ABA



f) Proportion of return moves (ABA) among those who moved twice (ABA+ABC)

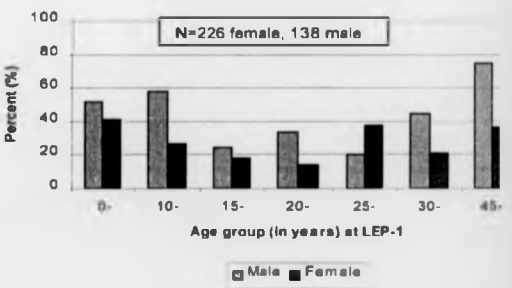


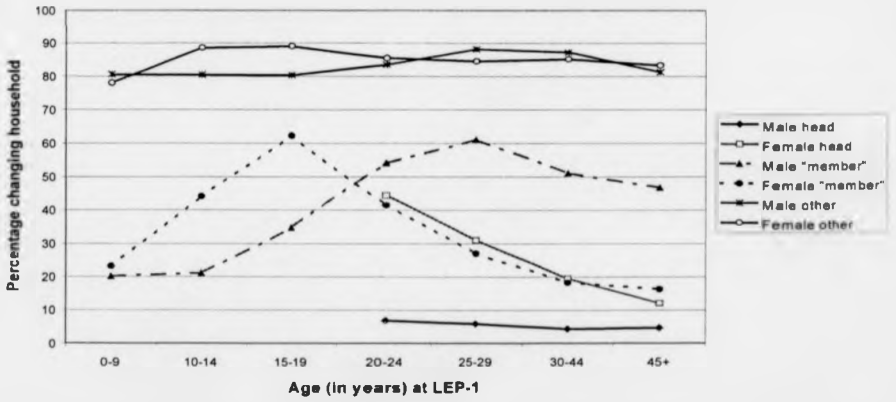
Figure 4.15(f) shows that the percentage of individuals making return moves is relatively low for males between ages 15-29 and for females between 15-24. After adjusting for age and position in household, we show that males were 2.26 (1.38, 3.74) times more likely to make return moves than females (p=0.001). There was no evidence of significant age and sex interaction (p=0.482). There is a higher proportion of return moves by females than males in only one age group, 25-29, perhaps due to early divorce. The age distribution for individuals making return moves shows that those aged 15-19 and 20-24 are less likely to make return moves compared to those aged 0-9 (OR of 0.34 (0.17, 0.68) and 0.36 (0.14, 0.92) with p-values of 0.002 and 0.032 respectively. The pattern for those over 30 years suggests an increase in return moves with age (with fewer females making return moves than males). There were no significant differences in making return moves for members and visitors compared to heads of households; OR 1.87 (0.74, 4.75) and 1.08 (0.30, 3.84) and (p=0.188 and 0.910) respectively after adjusting for age and sex. Family relationship is important in this analysis as children make return moves back to their parents. During simulations of household dynamics, the proportions from this analysis were used to determine the probability of a return or forward move among individuals considered as having changed household twice.

4.5.3 Movements by age and position in household

The propensity of an individual to change household depends on his or her position in the household (membership status and relationship to head). Figure 4.17 shows the same data as in Figure 4.14 broken down by position in household. It shows that individuals identified as visitors or renters in households were far more likely to change households than members who, in turn, were more likely to move than heads. Over 80% of individuals coded initially as visitors, renters or their relatives changed households.

The high proportion changing household among female members below age 25 reflects the pattern seen in Figure 4.14. The peak age for female "members" changing households (62% in 15-19 age group) was earlier than of males (in 25-29 age group). The proportions of female heads of households changing households were higher than those of male "heads" across all ages. Less than 10% of male heads aged over 20 years changed household compared to between 45% and 12% for female heads. We note a rapid decline in household change with age for female compared to male heads. There were too few heads in the age group 15-19 (2 changed households respectively out of 9 and 6 male and female heads) for reliable consideration, and thus they have been omitted from Figure 4.17.

Figure 4.17 Percentage of individuals who changed households between LEP-1 and LEP-2 by age, sex and position in household at LEP-1, Karonga District, northern Malawi 1979-89.



Although stratifying these positions further by relationships to heads of household led to small numbers in some categories, this provided expected results. Breaking down the "other" group of visitors or renters further as in Table 4.6 we show that 60% (676) of them were recorded initially as short-term visitors, i.e. in households for less than 6 months. We note no differences in household change between those

recorded initially as short- versus long-term visitors and a higher probability of change by workers or renters.

Table 4.6 Breakdown of position of visitors or renters ("others"), for household change, Karonga District, northern Malawi, 1980-89.

Category under this position	Percentage who moved	Total
Visitor for less than 6 months	517 (76.5%)	676
Visitor for more than 6 months	162 (76.1%)	213
Itinerant	12 (85.7%)	14
Employed worker or renter	198 (85.0%)	233
Total	889 (78.3%)	1136

The group recorded as "others" included children, siblings and grandchildren of heads of households. In Tables 4.7a and b, we note a relatively low percentage of household change among children of heads, though the numbers were very small in particular for males.

Table 4.7a Percentage household change for male visitors or renters broken down by relationship to head of household. The table shows the number (and percent) who changed household in each category and total.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Unrelated	Other relations	Total
0-14	2 (33.3%)	-	4 (66.7%)	41 (67.2%)	162 (83.9%)	10 (100%)	219 (79.4%)
Total	6		6	61	193	10	276
15-29	6 (40%)	5 (83.3%)	5 (100%)	2 (66.7%)	152 (84.9%)	2 (100%)	172 (81.9%)
Total	15	6	5	3	179	2	210
30-44	-	-	2 (100%)	-	109 (88.6%)	2 (100%)	113 (89.0%)
Total			2		123	2	127
45+	0 (0%)	-	-	-	68 (81%)	-	68 (80.0%)
Total	1				84		85

Table 4.7b Percentage household change for female visitors or renters broken down by relationship to head of household. The table shows the number (and percent) who changed household in each category and total.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Unrelated	Other relations	Total
0-14	1	1	10	44	177	9	242
	(16.7%)	(100%)	(90.1%)	(65.7%)	(82.7%)	(90%)	(78.3%)
Total	6	1	11	67	214	10	309
15-29	42	4	1	2	177	4	230
	(70%)	(100%)	(100%)	(100%)	(87.6%)	(90%)	(83.9%)
Total	60	4	1	2	202	5	274
30-44	10	3	-	-	73	5	91
	(71.4%)	(100%)			(83.9%)	(83.3%)	(82.7%)
Total	14	3			87	6	110
45+	3	2	-	-	37	7	49
	(60%)	(100%)			(82.2%)	(70%)	(79.0%)
Total	5	2			45	10	62

"Members" included children, siblings, spouses and other relatives of heads of households as shown in Appendix Tables F.12a and b. Breaking down this group, we note that parents and children of heads of households were less likely to change household compared to other relatives. Those unrelated to the head of household [52% (6064/11763) of whom are spouses] changed households more than parents and children of heads but were less likely to change households compared to other relatives (see Tables F.13a and b in Appendix).

4.5.4 Risk factor analysis for household change

Table F.6 in Appendix shows results from a multivariate regression to assess risk factors for household change. It presents crude estimates of household change for each risk factor and also estimates adjusted for age, sex and other risk factors. The risk factors considered included age, sex, BCG scar status, position in household, size of household and geographical zone.

Households were categorised into sizes 1, 2, 3-5, 6-10 and those with more than 10 individuals resident. An analysis of movements by size of household at LEP-1 was

done to investigate whether changing household was related to size of household. Movements of individuals by size of household revealed that the probability of moving from middle-sized households (3-10) was lower than from small (1 and 2) and larger (11+) households (see Figure F.1 in Appendix). The peak moving age was directly related to household size i.e. it was older for larger than smaller households.

The propensity to change household is determined mainly by age and sex, and position in household. Smaller households are unstable and more likely to dissolve through members marrying or being absorbed into existing households.

The rate of household change varied by geographical location of household. We observed that 22% (1299/5879) of individuals located in geographical zone A at LEP-1 changed households compared to 24% (8740/36038), 27% (5786/21816), 27% (2315/8629) and 30% (4864/16163) of individuals in zones B, C, D and E respectively. The OR increases from 1.27 (95% CI: 1.18, 1.37) for zone B compared to zone A (reference group) to 1.56 (95% CI: 1.45, 1.69) for zone E after adjusting for age, sex and risk factors listed in the preceding paragraph and shown in Table F.6 of Appendix. Earlier analysis (122) showed that there was more leprosy in north than south Karonga. In terms of movements, people in the south were more likely to change households than those in the north.

The estimate of household change adjusted for age and sex does not show any significant differences between individuals with no BCG scar and those with a BCG scar recorded in LEP-1, with an OR of 1.02 (95% CI: 0.98, 1.06).

4.5.5 Migration out of the study area

This analysis was carried out using records of individuals who were seen in LEP-1 but not in LEP-2. An indicator variable showed whether these individuals were just absent from the household, had left their previous household or had died by the

time the household was interviewed at LEP-2. For individuals recorded as having left their previous household, it was possible to find out whether they had moved within Karonga District, within Malawi but outside Karonga or outside Malawi (by checking the code of village in which their new household was located). With this information, it was possible to look at the characteristics of out-migrants between LEP-1 and LEP-2.

Age and sex breakdown

Altogether 12% (13275) of individuals recorded at LEP-1 had out-migrated over a 5-year period by the time the LEP-2 survey was done. The percentage for males out-migrating (12%) was similar to that for females (11%). The data are presented by age and sex in Table 4.8.

For both sexes, the rates of out-migration increased with age and peaked in the age group 20-24 (21% for females, 24% for males) and declined to very low levels for individuals aged over 45. The rates are higher for girls than boys under 15 years of age, but consistently higher for males above that age. Whereas young girls appear more likely than boys to leave the district to marry, the higher rates for male adults probably reflect the search for employment outside the district. The higher rates of (young adult) males out-migrating are in contrast to the household change within the district where (young adult) females change households more frequent than males.

Table 4.8 Out-migrating from Karonga District between LEP-1 and LEP-2. The table shows the number (and percent) out-migrating in each category and denominator in bold.

Age Group	0-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	Total
Females	1202	1183	976	1062	792	529	786	239	6769
	(12.0%)	(13.1%)	(15.1%)	(17.9%)	(20.9%)	(10.9%)	(8.4%)	(2.4%)	(11.4%)
	10018	**9048**	**6484**	**5925**	**3780**	**4861**	**9399**	**9847**	**59362**
Males	1075	893	803	1041	914	550	917	313	6506
	(10.9%)	(9.8%)	(11.8%)	(18.5%)	(24.0%)	(18.6%)	(13.4%)	(3.7%)	(12.2%)
	9842	**9118**	**6825**	**5619**	**3802**	**2957**	**6854**	**8480**	**53497**

Out-migration by position in household

Overall, 7% (1233) of heads of households left the district, whereas 12% (10754) of "members" and 33% (1282) of visitors or renters or their relatives out-migrated. The details by age and sex are shown in Table 4.9.

Though male heads of household were less likely to leave the district than were male household members, the opposite was true of female heads, who were more likely to out-migrate than either male heads or male or female members. The propensity to change household was particularly high for female heads below 30 years of age, of whom 44 out of 145 left the district. It is likely that many of these female heads left Karonga either to remarry or to seek employment elsewhere.

Table 4.9 shows relatively high proportions of visitors, renters or their relatives out-migrating across all ages. The distribution peaks at 40% for females and 52% for males aged 20-24.

Table 4.9 Probability of out-migrating by age, sex and household status. The table shows the number (and percent) out-migrating in each category and denominator in bold.

Position in Household	Sex	Age group (in years)								Total
		0-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	
Head	Female	-	-	-	3 (33.3%) **9**	15 (34.1%) **44**	26 (28.3%) **92**	42 (9.1%) **461**	46 (2.2%) **2068**	132 (4.9%) **2674**
	Male	-	-	-	2 (15.4%) **13**	53 (14.3%) **370**	216 (15.7%) **1377**	603 (11.7%) **5174**	227 (2.9%) **7725**	1101 (7.5%) **14659**
Member	Female	1084 (11.3%) **9621**	1110 (12.6%) **8813**	922 (14.6%) **6316**	1005 (17.6%) **5724**	669 (19.3%) **3466**	448 (9.8%) **4553**	640 (7.4%) **8613**	160 (2.1%) **7596**	6038 (11.0%) **54702**
	Male	983 (10.4%) **9481**	844 (9.4%) **8944**	786 (11.6%) **6750**	970 (17.8%) **5441**	709 (22.6%) **3137**	235 (17.5%) **1340**	159 (12.6%) **1258**	30 (5.7%) **523**	4716 (12.8%) **36874**
Others	Female	118 (29.7%) **397**	73 (31.1%) **235**	54 (32.1%) **168**	54 (28.1%) **192**	108 (40.0%) **270**	55 (25.5%) **216**	104 (32.1%) **324**	33 (18.0%) **183**	599 (30.2%) **1985**
	Male	92 (25.5%) **361**	49 (28.2%) **174**	17 (22.7%) **75**	69 (42.3%) **163**	151 (52.2%) **289**	97 (42.2%) **230**	152 (38.4%) **396**	56 (25.5%) **220**	683 (35.8%) **1908**

4.5.6 Migration into the study area

This analysis was based on individuals resident in the study area during LEP-2 but who were not seen or interviewed during LEP-1. There were differences in coverage of LEP-1 and LEP-2 surveys, but this had little effect in these estimates, as the areas concerned were small.

Each new individual identified in LEP-2 had the name of his or her previous village recorded, wherever possible, which indicated whether they had been staying outside the district before they joined their current (LEP-2) household. This analysis was restricted to individuals whose year of joining current household was within ten years of LEP-2 interview date. Ten years was chosen because it was the maximum time between last LEP-2 interview year (1989) and start date of LEP-1 (1979). This was done so that we could only consider individuals who in-migrated after LEP-1. Any individual who had been in the household for more than 10 years "must" have been resident in the household during LEP-1 survey. Figure 4.18 shows that 96% of all new household members identified in LEP-2 were recorded to have joined the household within the previous 10 years.

Age and sex breakdown

Of the 146,115 individuals recorded at LEP-2, 10% (13969) had in-migrated into Karonga District after LEP-1. The percentage of females (10%) in-migrating was similar to that of males (9%). Table 4.10 shows that the percentages of in-migrants was higher for females than for males between the ages 10-29. This is likely to reflect the tradition in northern Malawi for women to join their husbands' households upon marriage.

Figure 4.18 Relative frequency distribution of time since joining LEP-2 household, Karonga District, northern Malawi, 1986-9. Analysis is restricted to individuals who moved into households between LEP-1 and LEP-2.

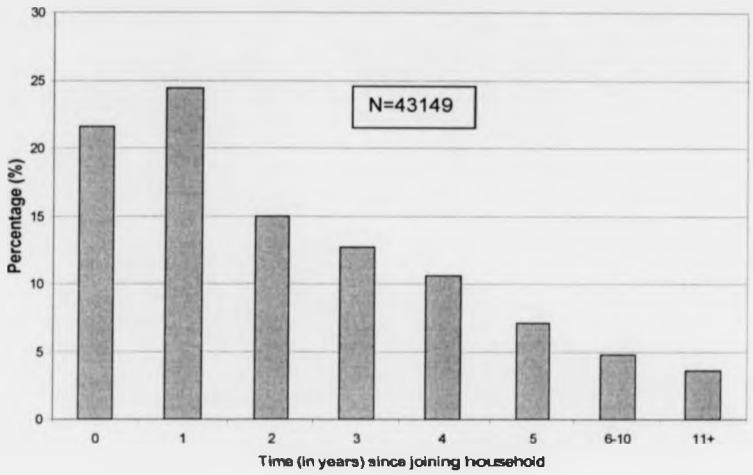


Table 4.10 Frequency distribution of individuals in-migrating into Karonga District between LEP-1 and LEP-2, by age and sex. The table shows the number (and percent) in-migrating in each category and denominator in bold.

Age Group	0-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	Total
Females	673	981	1175	1187	1267	858	1033	517	7691
	(5.6%)	(8.9%)	(11.7%)	(16.3%)	(19.2%)	(20.0%)	(8.9%)	(4.1%)	(10.2%)
	11990	**11029**	**10038**	**7298**	**6604**	**4296**	**11626**	**12693**	**75574**
Males	708	1033	924	865	850	680	859	359	6278
	(5.9%)	(9.3%)	(9.0%)	(11.5%)	(14.8%)	(16.3%)	(9.7%)	(3.3%)	(8.9%)
	12087	**11171**	**10223**	**7522**	**5754**	**4175**	**8840**	**10769**	**70541**

In-migration by position in household

Among heads of households, members and visitors recorded in LEP-2, 0.1%, 10% and 43% respectively had recently in-migrated into Karonga District (see Table 4.11).

There was no clear pattern by age and sex for the few, (26, 0.2%) immigrants who became heads of households. Concerning members, there were more in-migrant females (6508) than in-migrant males (4774). This relative excess increased with age.

Not surprisingly, a high proportion of all visitors and renters were immigrants, in particular among the young adults (more than 50% and 60% among females and males, respectively, in the age group 25-29).

Table 4.11 In-migration statistics between LEP-1 and LEP-2, by age, sex and household status. The table shows the number (and percent) in-migrating in each category in LEP-2, Denominators (Total LEP-2 population) are in bold.

Position in Household	Sex	Age group (in years)								Total
		0-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	
Head	Female	-	-	-	0	1	0	1	1	3
					(0.0)	(2.0%)	(0.0)	(0.2%)	(0.0)	(0.1%)
					7	**49**	**70**	**550**	**2984**	3660
	Male	-	-	-	0	1	0	11	11	23
					(0.0)	(0.1%)	(0.0)	(0.2%)	(0.1%)	(0.1%)
					19	**684**	**1936**	**6631**	**9797**	19067
Member	Female	494	833	1082	1078	1064	685	828	444	6508
		(4.3%)	(7.8%)	(11.1%)	(15.4%)	(17.5%)	(17.6%)	(7.8%)	(4.7%)	(9.5%)
		11382	**10691**	**9773**	**6989**	**6083**	**3888**	**10574**	**9475**	68855
	Male	501	886	838	775	653	463	463	195	4774
		(4.4%)	(8.2%)	(8.4%)	(10.6%)	(13.9%)	(24.6%)	(30.1%)	(32.6%)	(9.9%)
		11480	**10844**	**10006**	**7312**	**4682**	**1880**	**1537**	**599**	48340
Others	Female	179	148	93	109	202	173	204	72	1180
		(29.4%)	(43.8%)	(35.1%)	(36.1%)	(42.8%)	(51.2%)	(40.7%)	(30.9%)	(38.6%)
		608	**338**	**265**	**302**	**472**	**338**	**501**	**233**	3057
	Male	207	147	86	90	189	211	375	148	1453
		(34.1%)	(45%)	(39.8%)	(47.1%)	(50.3%)	(61.2%)	(57.6%)	(40.9%)	(47.3%)
		607	**327**	**216**	**191**	**376**	**345**	**651**	**361**	3074

The descriptive analyses presented in this chapter provide an understanding of the general population and dynamics of households in the population. It also provides parameter values necessary for the development of a stochastic simulation model of household dynamics described in the next chapter.

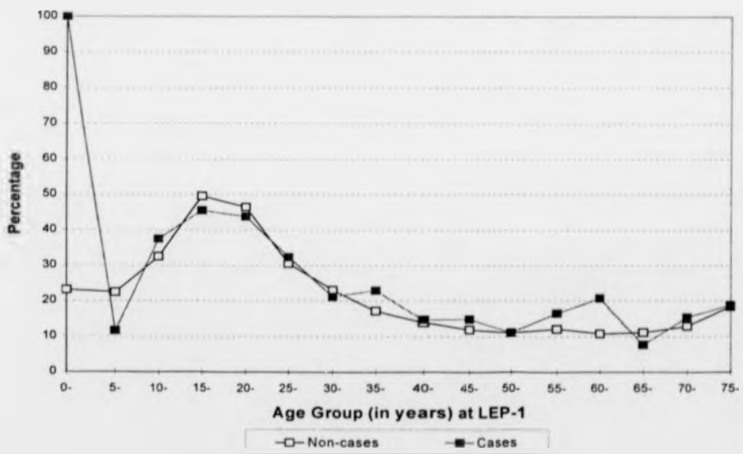
4.5.7 Leprosy cases and their movements

Leprosy is known to be associated with stigma in some populations, thus, it was of interest to investigate whether the presence of a leprosy case in a household affected tendency of people to change household.

Altogether there were 2176 leprosy cases (296 MB and 1880 PB cases) identified at LEP-1. These are called "index cases" in the contact analysis paper. An investigation was carried out to see whether there were any differences in household change between these individuals and persons who had never been diagnosed with leprosy.

Of individuals observed at both LEP-1 and LEP-2, 13% (11627) were living in households with at least one case in LEP-1. These are the LEP-1 case households. Of all the households seen in LEP-1 and LEP-2, 10.3% (1845) and 6.9% (1627) respectively had at least one known LEP-1 leprosy case resident.

Figure 4.19 Age-specific percentage of changing households for non-cases and leprosy cases, between LEP-1 and LEP-2, Karonga District, Malawi 1979-89.



Out of 1704 index cases of leprosy (233 MB and 1471 PB cases) seen at both LEP-1 and LEP-2, 24.3% (47 MB and 367 PB cases) had changed households by LEP-2. There were no apparent differences in changing households between leprosy cases (MB or PB cases combined) and non-cases as shown in Figure 4.19. The OR for leprosy cases changing household was 1.08 (0.96,1.22) after adjusting for age and sex (Appendix Table F.6). There was only 1 case in age group 0-4 (and this individual changed household by LEP-2). The OR does not change appreciably after further adjustment for size of household, position in household and geographical zone in which a household is located, 1.12 (0.99, 1.22). There were no significant differences in the tendency to change households between "New", "Old" and "Ancient" cases of leprosy relative to non-cases after adjusting for age and sex, with ORs of 0.95 (0.77,1.18), 0.99 (0.80, 1.22) and 1.18 (0.95, 1.45) respectively.

Of the 76927 individuals recorded as living in non-case households at LEP-1, and who were seen in LEP-2, 25.3% (19482) had changed households by LEP-2 whereas 30.5% (3547) of individuals recorded as living in case households at LEP-1 had changed households by LEP-2 ($p<0.001$). This difference reflects geographical zone and socio-economic differences. There was more leprosy and less movement in the north rather than south Karonga.

An analysis of whether or not the presence of a leprosy case in a household affects the tendency of people to move into that household is difficult as it could consider precise dates of household change and data on the membership of households at the time of the move. This was limited in that precise dates were missing for most individuals. However, in simplest terms of the individuals who changed households, 6.2% (1431/23029) and 9.3% (2146/23029) moved into case households as defined at LEP-1 and at LEP-2 respectively.

Return moves

There were only 92 "index cases" of leprosy (individuals who had ever been diagnosed with leprosy) in the villages seen during three surveys (LEP-1, SS and LEP-2). Of these index cases, 60% were female and 34% changed household. Only 2 (2%) individuals made return moves and 5 (5%) made forward moves, which was similar to the 3.3% and 6% respectively for all individuals seen at all the three sample surveys.

Of the 7 index cases of leprosy recognised as having moved twice (ABA and ABC moves), 2 (29%) made return moves. A breakdown of movements of index cases further by age and sex yielded numbers too small for meaningful analysis.

4.5.8 Summary

Approximately 26% of individuals seen in both LEP-1 and LEP-2 were found to have changed households over the five year inter-survey interval.

More than 20% of children under 10 years of age in LEP-1 were found in a different household in LEP-2. Household change (including return or forward moves) in children is likely to be dependent largely on movements of adults. Children may move with their mothers, but are occasionally sent away to live with other relatives. Some of the children may have been orphans, with relatives taking responsibility of such children. Given the recent increase in orphanhood as a consequence of AIDS-related death of parents, it is likely that the rate of household change of children will have increased in recent years. This aspect of the data deserves further analysis.

The highest rates of household change were in adolescents and young adults peaking at greater than 60% among girls aged 15-19 and 50% among males 20-24. Active household change in young adults is attributable primarily to marriage and

search for employment opportunities. The earlier age peak in household change for young women, compared to men, is a reflection of age at marriage and setting up new households in this society.

Adults over age 30 changed household relatively infrequently, in particular the males, the majority of whom are household heads. The higher rates of household change for older women compared to older men are likely to reflect separation and widowhood – custom dictates that women, not men, leave the marital home if the marriage is terminated. Thus, older women are likely to leave and join their children's or original parental households. Widowhood or separation was not investigated from the data because marriage was not recorded explicitly.

The propensity of an individual to change household depends on his or her position in household and relationship to head of household. "Members" of households included children, spouses and other relatives of heads of households, the majority of whom were dependents aged under 15 years and, hence, relatively unlikely to change household. Not surprisingly, visitors and renters were a highly mobile group relative to heads and members of households.

The proportion of individuals observed to change household twice over 10 years was relatively high in adolescence and young adulthood, with the proportion higher for females than males. However, the proportion of return moves to previous households, among those who moved twice was higher for males than females across all ages except in 25-29 (perhaps due to early divorce). The proportion is highest in children aged under 10 years and adults over 30 years. Adult males in this age group were probably returning from employment (e.g. rice plantation schemes), and children from visiting relatives within Karonga District.

Approximately 12% of the population moved out of the district over the 5 years between the surveys. The probability of out-migration was highest among young adults, peaking at 21% and 24% among females and males aged 20-24,

respectively. These high rates reflect the search for employment in urban areas and marriage outside the district. Between 10 and 15% of children also left the district, probably to visit relatives or to attend primary boarding or secondary schools outside the district. Young female household heads aged below 30 years were particularly likely to leave the district (approximately 30%). Many of these women were probably divorced or widowed, and left the district to remarry.

Approximately 10% of individuals identified in LEP-2 were in-migrants to the district. Thus, there is a net out-flow of migrants from the district. The percentages of in-migrants in young adulthood were higher for females than males, probably due to marriage patterns. In the northern region of Malawi, women join their husbands' households.

There were no apparent differences in changing households between non-cases and leprosy cases. In this context it should be noted that clinical leprosy is not "permanent". Many cases heal or recover totally. These analyses are consistent with the impression of field staff on the KPS that there is little if any stigma attached to leprosy in this population (personal communication).

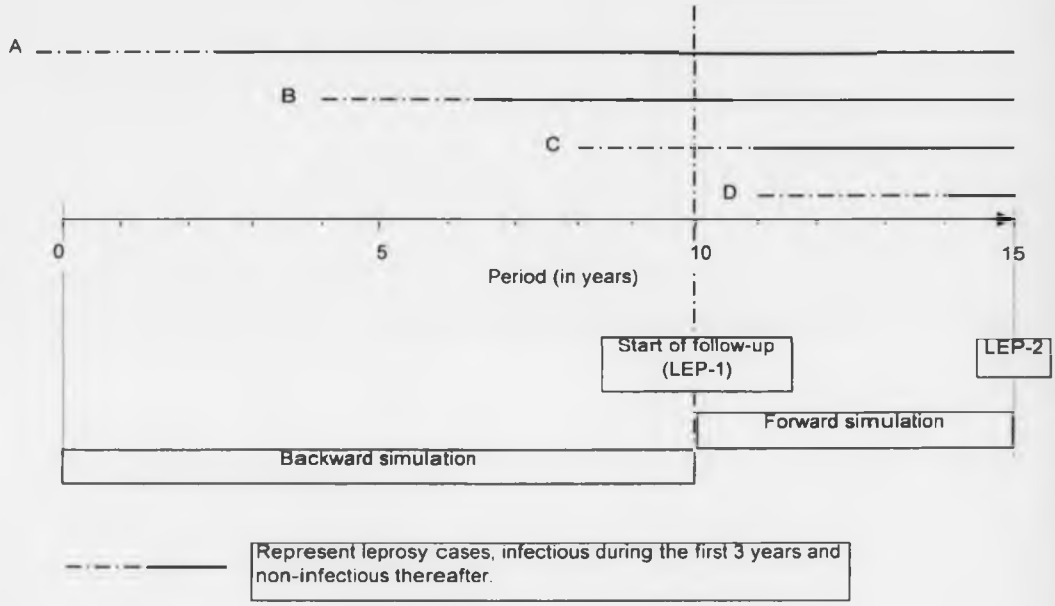
CHAPTER 5 SIMULATION METHODOLOGY

The next part of the project was the development of a stochastic micro-simulation model of household dynamics in Karonga District with particular reference to demographic events which affect household contact status over a period of time. The following terms and definitions will be used in describing the model and its results. Figure 5.1.1 is used as an aid in defining these terms. The follow-up refers to the time between the LEP-1 and LEP-2 surveys and to an analogous simulated situation. We have assumed that an incident case (or index case) is infectious for a period of 3 years from year of onset of disease.

Term	Definition
Leprosy case	Anyone ever diagnosed with leprosy, regardless of whether still infectious at LEP-1.
Index case	Any leprosy case in the population who may have been responsible for transmission of infection up to the start of a follow up study. These cases were used to define contact status at start of the observed (LEP-1 to LEP-2) or simulated follow-up study.
Incident case	A leprosy case in the population with onset during the follow-up period. For relatively long incubation period simulations, this is also considered an index case.
Case household	A household in which at least one (index) case was resident.
"Old" index case	A leprosy case with onset more than 3 years prior to start of follow-up.
"New" index case	A leprosy case with onset less than 3 years prior to the start of follow-up.
Incubation period	The time from initial infection to the onset of clinical signs of disease.

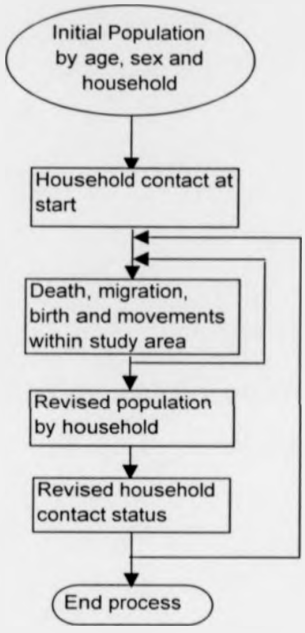
Initial contact	An individual without leprosy living in a case household at the start of a follow-up study.
Sensitivity of initial contact status	The proportion of actual contacts which were correctly recognised as contacts at start of a follow-up study.
"Forward" sensitivity of initial contact status	The proportion of individuals in contact with at least 1 (index) case at any time during a follow-up study, who were also recognised to be contacts at the start of follow-up study. This concept is particularly appropriate for relatively short incubation periods.
"Backward" sensitivity of initial contact status	The proportion of individuals who were contacts prior to the onset of a follow up study who were observed to be contacts at the start of follow-up study. This concept is particularly appropriate for relatively long incubation periods. The relevant time period for such contacts in the past is that period during which transmission could lead to clinical onset during the follow-up.
Specificity of initial contact status	The proportion of individuals with no relevant contact, who were correctly identified as non-contacts. This was initially set to 1.
Starting population	All individuals recorded in 1854 households sampled from LEP-1 data, who formed the starting population for all the simulations.
Three-survey sample	3902 individuals who were seen three times, in the LEP-1, SS/KAS and LEP-2 surveys (see Figures 3.1 and 3.2).

Figure 5.1.1 Diagram of time period under consideration in these studies. The prospective follow-up study is considered to take place between years 10 and 15. This is analogous to the observed follow-up between LEP-1 and LEP-2, and is modeled in the "forward" simulation. Events which took place during the previous 10 years (0-10) are modeled in the "backward" simulation. As per definition, A and B are "old" index cases, C is a "new" index case and D is an incident case.



The most important demographic events modelled were death, birth, marriage, migration and movement of individuals between households. Flow diagrams of these demographic processes are provided in Appendix E. Figure 5.1.2 shows an overview of the model.

Figure 5.1.2 Overview of the stochastic micro-simulation model for household dynamics.



Briefly, the model has an initial sample drawn from the actual population and thus defined by age, sex and household. Individuals have initial (observed) contact status set depending on whether they were in case or non-case households. Events such as birth, death, in-migration, out-migration, marriage and movement between households are simulated on an individual and annual basis. This is done for a period of 5 or 10 years in order to explore shorter and longer incubation periods respectively, as explained in section 5.3.1 and 5.3.2. Contact status histories are tracked through the simulation. "Sensitivity" of initial contact status values (and contact status misclassification rates) are calculated, based on these histories.

For each simulation run, the records of individuals generated are interpreted as a random sample from the population that would have been observed if the events modelled had been taking place in the real world. Due to the stochastic approach, the predictions differ between simulations, but the mean and variance of the results can be used to provide an estimate of the expected sensitivity of initial contact status and confidence intervals for it.

Simulations on this scale take considerable computational space and time. These problems were taken into consideration when deciding on the initial sample size. The simulations were run using the SAS statistical package.

5.1 Starting population

Individuals were the units of analysis, and the starting point of the simulation process was a sample of individuals (with their associated attributes), with each individual linked to one household.

There are two ways to obtain a starting population and their attributes (60). One way is to generate the sample assuming the attribute distributions are similar to that of the population under study. This is appropriate when data are in aggregated form as would be the case if an age and sex structure of a population were provided. Multiplying the total number of individuals by the proportions in each age and sex category will generate a sample with the same cross-sectional characteristics as the actual population.

An alternative approach, used in this project, is to use a real (random or stratified) sample of the population as the starting point of the simulation. This is preferable if the individual data are available. It has not been possible in many situations to use a real sample because data have only been available in aggregated form. However,

the LEP data used here are unique in that they are in disaggregated form and include records of individuals linked to their households.

For this simulation, a random sample of households was drawn from all 17905 LEP-1 households, based on Monte-Carlo techniques (see Section 5.2), and all individuals resident in the sampled households at LEP-1 were included in the starting population. The chosen sample contained 1854 households (approximately 10%) which, in turn, contained 11401 individuals (52% female) of whom 224 (2%) were index cases of leprosy (52% female). It was found that 10% of LEP-1 households had at least one index case resident and approximately the same proportion was observed in the starting population [196 (11%) were index case households].

The analysis of contact-associated risk (19) was restricted to individuals seen during both LEP-1 and LEP-2. New households entered or formed by these individuals during the follow-up period have implications for contact misclassification. A procedure for formation of new households was thus included in the model.

Of the households recorded in the LEP-2 survey, 29.9% were new (i.e. not seen during LEP-1). However, it was possible for households existing at LEP-1 but outside the geographical area under consideration for LEP-2 to move into LEP-2 geographical area. Such households may have appeared "new" while in fact they were not. For the "formation of new households" procedure, an extra 792 households were required to represent households newly formed throughout the period of simulation.

5.1.1 Characteristics of the starting population

All the non-case individuals in case households were recognised as initial contacts. About 10% (1177) of non-case individuals were recognised as initial contacts (of

whom 54% were females compared to 52% (5159) in the non-contact group). Table F.5 in Appendix presents initial individual and household attributes.

The initial household size shows a similar distribution to that observed at LEP-1 with about 6% of households of size 1. The mean and median household sizes were 6.1 and 5 (IQR of 3-8) respectively.

Table 5.1.1 shows the age and sex distribution of individuals in the starting population. The age groupings were chosen because they were used in the published contact analysis (19).

Table 5.1.1 Frequency tabulation (and age-specific percentages) of initial age distribution for males and females in the starting population.

Sex	Age groups (in years)									Total
	0	1-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	
Male	129 (1.1%)	868 (7.6%)	936 (8.2%)	721 (6.3%)	612 (5.4%)	353 (3.1%)	310 (2.7%)	702 (6.2%)	858 (7.5%)	5489
Female	130 (1.1%)	840 (7.4%)	894 (7.8%)	617 (5.4%)	621 (5.4%)	376 (3.3%)	485 (4.3%)	897 (7.9%)	1049 (9.2%)	5909
Total	259	1708	1830	1338	1233	729	795	1599	1907	11398

5.1.2 Input parameters (and their confidence intervals) for event occurrence

The probabilities of an individual experiencing each demographic event were derived from either the LEP or census data. The LEP data provided average 5-year risks of death, out-migration and household change within the study area. The annual risks of events occurring (shown in Table 5.1.2) were calculated using the formula

$$R_1=1-\{1-R_5\}^{0.2}$$

where R_1 and R_5 are the annual and 5-year risks respectively.

Table 5.1.2 Annual probabilities of events occurring by age and sex.

EVENT	SEX	Age group (in years) 0	1-4	5-9	10-14	15-19	20-24	25-29	30-44	45+
Death	Male	0.1523	0.0392	0.0077	0.0047	0.0033	0.0038	0.0069	0.0063	0.0254
	Female	0.1259	0.0297	0.0071	0.0045	0.0024	0.0062	0.0051	0.0074	0.0226
Change of household	Male	0.0500	0.0500	0.0439	0.0476	0.0836	0.1274	0.0799	0.0321	0.0168
	Female	0.0528	0.0528	0.0558	0.1125	0.1801	0.1079	0.0658	0.0428	0.0347
Out-migration	Male	0.0229	0.0229	0.0204	0.0247	0.0402	0.0535	0.0403	0.0283	0.0075
	Female	0.0252	0.0252	0.0276	0.0321	0.0387	0.0459	0.0228	0.0173	0.0049

Some parameter values show substantial variation by age, whereas others are similar across consecutive age groups.

The most important input parameter for determining household contact status misclassification is the annual rate of change of household. Confidence intervals on the rates of change of household (shown in Table 5.1.3) were calculated to investigate the effect of uncertainty in point estimates on sensitivity of initial contact status. In strict sense, the annual rates of household change are actually annual risks (proportions).

Table 5.1.3 95% confidence interval for annual risk of change of household by age and sex.

SEX		AGE CATEGORIES 0	1-4	5-9	10-14	15-19	20-24	25-29	30-44	45+
Male	Lower limit	0.0477	0.0477	0.0417	0.0450	0.0796	0.1208	0.0742	0.0284	0.0127
	Upper limit	0.0523	0.0523	0.0460	0.0502	0.0878	0.1342	0.0853	0.0357	0.0212
Female	Lower limit	0.0504	0.0504	0.0533	0.1081	0.1739	0.1020	0.0621	0.0403	0.0297
	Upper limit	0.0552	0.0552	0.0584	0.1170	0.1866	0.1140	0.0694	0.0478	0.0442

Such narrow confidence intervals suggest that varying parameter values within these limits will have only a small effect on the contact status misclassification rates observed in the simulations. Although the LEP study was a total population survey, the confidence intervals are still useful if we consider another population in Malawi or elsewhere with similar behaviour and characteristics as that of the LEP study area, or the same LEP study area at a different time.

Apart from using 95% confidence limits as input parameters in the model, we shall also investigate the variation in sensitivity of initial contact status by varying the mean annual risk of household change by +/- 20% (to investigate a population reasonably similar to that in Karonga).

5.2 Micro-simulation model for household dynamics

The major assumptions made were that 1) rates of movement between households, deaths, and migration are constant during the period of the simulation and 2) movements of individuals between households are random according to these rates. In fact, random movement is modified by a proximity function, which incorporates a non-random function in the process. The latter assumption is, however, unlikely to capture the social realities of household change in the real world, and we discuss its implications in Section 5.2.3.1.

A simple model using appropriate age-specific mortality rates and risks of movements between households was first constructed and checked against the data. Elaborations were introduced gradually to produce a comprehensive model closer to reality.

Disease, migration and contact with index cases of disease were later incorporated into the model. Further assumptions included the following:

1) Disease status in index cases is not misclassified.
2) Distance between (i.e. proximity of) households does not affect infection risk because we only consider within-household infection.

In simulating the implications of long incubation periods, the incident cases that were generated were assumed to become infectious at onset of disease and to remain infectious for 3 years. Infection among contacts was assumed to start at onset of disease.

Monte-Carlo methods

Some characteristics such as ageing are deterministic and easy to update in simulations while others are updated stochastically using conditional probabilities (based on Monte-Carlo techniques). A number is drawn from a uniform distribution of the interval (0,1). This number is compared to the appropriate conditional probability. A process (or event) is assumed to have occurred when the number drawn does not exceed that probability. For example, for every year a woman spends in a specific age group, a random number, x, in the interval 0-1 is drawn and compared to the age-specific fertility rate. If x is less than the age-specific fertility rate then a birth is considered to have occurred to that woman, otherwise she does not experience birth in that year. This technique has been extensively discussed in the literature (60, 69).

5.2.1 Death procedure

The parameters used for simulating death were the age and sex specific mortality risks obtained from LEP data (for infant and child mortality, we used rates from indirect estimation). The processes involved are shown in the flow chart (see Figure E.1 in Appendix).

If an individual dies, an indicator variable is set to mark the individual's record for deletion from their household. Death for each individual is simulated on an annual basis with risks of dying by age and sex given in Table 5.1.2.

If an individual in a single-person household dies, the household is dissolved and the number of households reduced by 1. If the individual was not in a one-person household, the household size is reduced by 1. The program checks to see whether

(s)he was the head of household. If this person who has died was the head of the household, a new head of household is allocated (as described in Section 5.2.4).

5.2.2 Out-migration procedure

This procedure models the out-migration of individuals from the study area (see flow diagram, Figure E.2 in Appendix) and works in a similar way to the "death" procedure described in Section 5.2.1. The input parameters are the annual out-migration rates by age and sex shown in Table 5.1.2. If out-migration has occurred, the variable recording number of out-migrants will be incremented accordingly and the number and size of households will be reduced. What happens if the head of a household out-migrates is described in Section 5.2.4.

5.2.3 Change of household procedure

This procedure traces movements of individuals between households within the study area. The flow diagram is given in Figure E.3 of Appendix. Individuals are initially classified by whether they were resident in a household with case or not. The "true" contact status (a binary variable) from each simulation is set to that of observed contact status at start and is successively updated on an annual basis during the simulation period. The issue of contact prior to a study is discussed in Section 5.3.2.

The parameters governing this procedure are annual age and sex specific rates of change of household obtained from the LEP study and shown in Tables 5.1.2 and 5.1.3. The confidence limits were also used as parameter values in the micro-simulation model to assess variability in sensitivity values.

An individual who has changed household is marked for allocation to a new household. For each year of the simulation, an indicator variable is set to show

whether an individual changed household in order to keep track of the number of times an individual has changed household in the course of simulation.

In the simulations, an individual can only change household once a year. The probability of an individual changing household is independent of whether or not they moved before (i.e. in the preceding simulation years).

If an index case moves into a household, the "true" contact status of all the individuals in that household not in contact with a case before is changed. However, if individuals in the new household were contacts before (from a different household or of the same household if it already had a case), their "true" contact status during that year of simulation does not change.

For all moves, the new and previous household sizes are adjusted accordingly.

5.2.3.1 Household allocation

Allocation of households to individuals marked as having either changed households, in-migrated or whose households have been dissolved during that year of simulation is done on an annual basis as described below.

Initially, 5% of households were of size 1. We fixed the allocation procedure to maintain this proportion approximately constant in successive years of simulations. This was done through introducing "marriage", (see Section 5.2.6), which is closer to reality.

Individuals who change household are randomly allocated (with probability 0.05) to a new household otherwise they are randomly allocated to existing households.

The allocation of individuals to existing households is influenced by "proximity" of households using the serial household numbers generated. It was assumed (and

this is in fact the case) that the closer these numbers, the smaller the physical distance between the households and the more likely a movement was to occur between them.

Generation of the destination household serial number for individuals who move is a "random" function of the previous household, based on a standard normal random value and pre-determined constant standard deviation. The standard deviation used was 100. Explicitly,

$$H_n = H_o + z^*\sigma$$

Where H_n and H_o are new and previous household serial numbers, z and σ are the standard normal random value and standard deviation respectively.

The choice of standard deviation (100) was such that a large proportion of individuals are allocated to nearby households but some may still move considerable distances. Using this type of random allocation, an individual has a higher chance of being allocated to a closer than to a distant household.

5.2.3.2 Forward and return moves

Records of an individual's previous households and the number of times they have changed household up to the current year in the simulation period help in assigning the individual to either a previous or new household. For those who have changed households before, instead of just randomly assigning them to different households each time they move, the program assigns them on an annual basis to either their previous household (return move) or to a new one (forward move).

Table 5.2.1a shows parameter values, by age and sex, for determination of a return or forward move. They were obtained from analysis of LEP data based on individuals who were seen at LEP-1, SS/KAS and LEP-2 (see Section 4.3.7.2).

Table 5.2.1a Conditional probabilities of individuals making return (as opposed to forward) moves by age and sex, Karonga District, northern Malawi 1979-89.

	Age Group (in years)								
SEX	**<1**	**1-4**	**5-9**	**10-14**	**15-19**	**20-24**	**25-29**	**30-44**	**45+**
Male	0.52	0.52	0.52	0.58	0.24	0.33	0.20	0.44	0.75
Female	0.41	0.41	0.41	0.27	0.18	0.14	0.38	0.21	0.36

Table 5.2.1b shows an example of the overall proportion of return or forward moves after one (5-year period) simulation run. Of the 458 individuals who moved more than once, 118 (26%) made return moves. Of the individuals who made 2 and 3 moves, 26% and 29% made return moves respectively. Very few individuals made more than 3 moves in the 5-year simulation period.

Table 5.2.1b Frequency distribution of individuals by number and direction of moves after a 5-year simulation period

	Number of moves made in this period					
	0	**1**	**2**	**3**	**4**	**Total**
Return move	0	0	106 (26%)	12 (29%)	0	118
Total	9284	2502	415	42	1	12244

5.2.4 Procedure for assigning new positions in households

5.2.4.1 Head of household

After the first year of a 5-year period simulation run, about 7% (60) of heads of households were found to have either died or out-migrated or changed households. If the head of household either dies or out-migrates or changes household within the district, a new head must be assigned. The program checks for members of households who may be eligible to assume headship.

The oldest "member" of the household is automatically assigned as the new head. Individuals in the household who are not members are not eligible to become heads. This is justified by the data in that of the new heads, a high proportion were previously members of households as opposed to visitors or renters, as shown in Tables 4.4a and 4.4b. From the data, the youngest head of a household was in the age group 15-19.

If the oldest "member" who has been assigned as the new head of the household is below the chosen cut-off age of 18 years, that household is marked for dissolution and its remaining residents are randomly allocated to other existing households using the neighbourhood preference described in Section 5.2.3.1.

5.2.4.2 Other positions in household

The probabilities of allocating new positions to previous heads of households or other individuals if they changed households were estimated from LEP data (see Tables G.2). Table 5.2.2 gives a sample cross-tabulation of position before and after a 5-year period simulation run for individuals who changed households at least once in the 5-year period.

Table 5.2.2 Overall frequency of position in household for individuals who changed households before and after the simulation process.

		Position in household after simulations			
		Head	Member	Others	Total
Position in household before simulations	Head	18 (36%)	27 (54%)	5 (10%)	50
	Member	17 (4%)	430 (88%)	39 (8%)	486
	Others	0 (0%)	15 (94%)	1 (6%)	16
	Total	35	472	45	552

If an individual who changed household was a head, allocation of new position was based on results from the LEP data. An individual remains a head with probability of 0.31. They become either a visitor or renter in the household they have moved to, with probability 0.22 otherwise they are assigned to position of member of household.

For those who were not heads of households, different proportions were used in assigning new positions in the household they had moved to. For members of household there was a 7% chance of being visitors or renters in the new household whereas the chance of becoming a "member" in the new household for visitors or renters was 90%.

For individuals who had not changed household, it was assumed that the same positions would be retained - based on LEP data (see Section 4.4.4).

5.2.5 Birth procedure

All women in the current sample at each year of the simulation period, who are in the reproductive age group 15-49, and who did not give birth the previous year are assumed to be at risk of giving birth.

Births are simulated on an annual basis based on age-specific fertility rates (number of births per 1000 women) for Karonga district obtained in the 1987 census and listed in Table 5.2.3. Birth rates are assumed to remain constant throughout the simulation period.

Table 5.2.3 Age-specific fertility rates for women in the reproductive age group (15-49) in Malawi.

	Age Groups						
Source	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**
***Census, 1987 (Karonga)**	0.196	0.317	0.302	0.251	0.186	0.091	0.013

Note: *Rates have been adjusted for apparent under-reporting of births

If a birth occurs, a child record is created and the woman is assigned as its mother, so the household number assigned to the child is the same as the mother's. Sex is assigned randomly on the basis of the observed proportion of males and females under 1 year of age (49% of children under 1 year of age were male in LEP-1 and LEP-2 surveys, and the 1987 census). Other variables initialised are current age (set to zero), number of moves made (zero), whether the child is an index or incident case of leprosy (non-case initially) and position in household. The infant is considered a "member" (if the mother was a head of household) or else has the same position as the mother.

The contact status assigned to the infant is the same as the mother's current annual contact status (for update of these to obtain "true" contact status, see Section 5.3). In the situation where the mother was the index case, the child is considered a contact. For simulations in which we assumed a long incubation period (see Section 5.3.2), assignment of contact status was slightly different. If the mother was in household contact with an index case, the child was assigned a contact status (subdivided into new/old contact categories) depending on whether the contact was with a "new" or "old" index case. These newly generated records of children are then added to the current sample and are used in subsequent years of the simulation period.

In order to achieve realistic birth spacing, a woman who delivers a birth in year *n* is not exposed to the risk of giving birth in year *n+1*.

5.2.6 Marriage procedure

Because of the way the household allocation procedure has been designed, we had an excess of single-person households compared to the data (about 5% of LEP-1 households constituted single-persons). The marriage procedure was included to be closer to reality as some people move out from households to set up households of their own.

After randomly allocating households to individuals who have changed households and computing household size, those individuals (movers) between 10 and 40 years of age who were originally assigned to single-person households are eligible for "marriage". Their age is categorised into 5-year age groups as 10-14, 15-19, 20-24, 25-29, 30-34 and 35-39. A sample output of individuals eligible for marriage by age and sex is given in Table 5.2.4. We note a consistently higher proportion, by age, of females who have changed households and are in single-person households. This is a reflection of the relatively high proportion of young female adults changing households compared to males, and high male out-migration over age 20.

Table 5.2.4 Frequency distribution for individuals eligible for "marriage" by age and sex.

	Age Groups (in years)					
	10-14	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**
Female	75	73	42	30	16	13
Male	27	30	37	22	7	9
Total	102	103	79	52	23	22

Here we have assumed the man to be older than the woman when matching them as "marriage" partners to occupy the same household. A man can only marry a woman who is, on average, five years younger. The only matching is by age. Some studies (43, 58, 60, 69) have looked at matching differently. This is discussed in Chapter 7.

Only one household is needed for the two matched individuals. This leads to the following situations.

(1) If we have n males and (n+x) females in an age group and its preceding age group respectively, then only the first n females and males would be matched. The extra (x) females will remain in single-person households. The same applies to males if the numbers are reversed.

(2) Due to the way "marriage" is being implemented, a man can only be matched to a woman in a lower age group. Thus, women who are 35-39 years and boys 10-14 years of age are not paired with anyone and, they remain in single-person households. In any case, all one-person households with individuals under 18 years of age are marked for dissolution.

5.2.7 In-migration procedure

The assumptions made in this procedure are that individuals migrating into the population have never been in contact with leprosy cases before and in-migration rates are constant over the simulation period. The initial contact status of in-migrants depends only on the households to which they are allocated upon arrival.

In each year of the simulation period, a file of the total number of individuals by age is created. This procedure uses in-migration rates by age and sex, shown in Table 5.2.5 and derived from LEP data for the period between LEP-1 to LEP-2 surveys. These annual rates were computed based on the knowledge that the LEP studies were, on average, 5 years apart.

Table 5.2.5 Annual in-migration rates (per person) by age and sex, in Karonga District, Malawi 1986-89.

Sex	Age Groups								
	0	1-4	5-9	10-14	15-19	20-24	25-29	30-44	45+
Male	0 0120	0 0120	0 0192	0 0188	0 0241	0 0315	0 0349	0 0202	0 0068
Female	0 0115	0 0115	0 0185	0 0246	0 0349	0 0417	0 0436	0 0184	0 0083

The frequency of in-migrants in each age group is computed as the product of the current number of individuals with a specific age and sex and the appropriate in-migration rate. Separate records equal to the number of in-migrants, by age and

sex, are created. For example, if there were 1000 males in the 15-19 age group in a particular year of the simulation period, then 24 new records would be created for male in-migrants in that age group (obtained as a product of 1000 and the in-migration rate, 0.0241).

The next step in the procedure is initialisation of attributes for in-migrants. The number of moves such individuals have made is set to zero. In-migrants are neither incident nor index cases. Within each age group, age is randomly assigned based on the outcome of a random uniform generated number (0,1).

Contact status is assigned based on the household to which an individual has been allocated (see household allocation, Sections 5.2.3.1 and 5.2.3.2). Taking in-migrants to be non-contacts initially assumes that these people have not been in contact with leprosy index cases before migrating into the study area. In reality, some of the in-migrants may have been in contact with cases before coming into the area. However, Karonga District was known to have a relatively high leprosy prevalence at the time and, as such, the chance of an in-migrant being in contact with an index case of leprosy elsewhere was relatively small.

Positions in households are allocated randomly to the in-migrants using the rules explained in Section 5.2.4. From the LEP data, 80.9% of in-migrants became "members" of households and 18.9% were either visitors or renters. A negligible proportion (0.2%) became heads of households. In-migrants could thus only be allocated to positions of either "member" or visitors/renters based on probabilities of 0.81 and 0.19 respectively. The new records of in-migrants are then appended to the file containing all individuals in the current sample used for simulation.

5.2.8 Contact histories and events

The simulation program keeps a record of the contact status of every individual during each year of the simulation period. Records of individual events and characteristics from the simulations are kept in a master file after each simulation run. This master file contains a track of events such as household change status, in-migration status and those born into this population for each year of the period of simulation.

A separate file of individuals who have died or out-migrated throughout the simulation period is also kept.

5.3 Quantification of misclassification of contact status

The simulation methodology described above provides a way of modelling household dynamics within a study area. The complexity of the model can be varied depending on time and the research question being addressed. We deliberately limit the complexity of these simulations because the objective was not to simulate reality precisely but to better understand the principles of household dynamics and how they affect household contact.

The main outputs of interest from the simulation model are the contact histories of individuals for each year of the simulation. Using these contact histories, together with initial observed contact status in the LEP data, we can calculate the "sensitivity" of initial (observed) contact status by age, sex and, for the long incubation period simulation model, time-reference of leprosy cases ("new" or "old"). The implications of time-reference of leprosy cases are discussed in Section 5.3.2.

The results generated through the simulations are then considered as the "true" underlying state of household contact status. The comparison between contact status at the start of follow-up (initial observed contact status in the LEP data) and "true" underlying contact status were used to derive estimates of *sensitivity* of initial contact status. Values of the sensitivity have been estimated for different age and sex categories (and time-reference of leprosy cases in households, for the long incubation period circumstance). The general approach to quantifying sensitivity of initial contact status is illustrated in Table 5.3.1.

Table 5.3.1 Frequency distribution of individuals according to "true" and observed household contact status with an index case of leprosy.

Initial "observed" contact status	"True" underlying contact status		Total
	Contact	Non-contact	
Contact	A	B=0	A+B
Non-contact	C	D	C+D
Total	A+C	B+D	A+B+C+D

Note that B=0 in Table 5.3.1 because all individuals who were at the start observed to be contacts are considered to remain so during the simulation period. However, of the C+D initial non-contacts, C became contacts in the course of the simulation. In other words, the simulation model enables us to estimate the number of individuals who were ever in contact with index cases (A+C). Then, the sensitivity of initial contact status is defined as the proportion of "true" contacts who were correctly observed as contacts at start of follow-up of a study, i.e. A/(A+C). Contact status misclassification rate is the complement of sensitivity of initial contact status (i.e. 1- sensitivity of initial contact status = 1-A/(A+C) = C/(A+C)). Specificity of initial contact status, the proportion of "true" non-contacts correctly observed as non-contacts initially, is D/(B+D), which is 100% with the initial assumption of B=0.

There are two refinements to the way in which sensitivity of initial contact status is calculated, depending on incubation periods of disease under study.

5.3.1 "Short" incubation period

If the incubation period of the disease is considered relatively short, then most contacts which lead to disease in the study period is likely to occur at the start or during the study period. To investigate contact status misclassification under this scenario, simulations have been run for 5-year periods with contact status updated yearly. In this scenario, we consider any contact at any time during the study period as relevant for transmission of infection. Thus, the *sensitivity of initial contact status* is defined as the proportion of individuals in contact with at least 1 index case in years 1-5 of the simulation, who were also contacts at the beginning of the study. This is termed, for convenience, "forward" sensitivity of initial contact status.

5.3.2 "Long" incubation period

Due to the long and varied incubation period of diseases such as leprosy and tuberculosis, some incident cases that arise during the follow-up period of a cohort study may be attributable to earlier contact outside the period of study rather than to contact recognised at the start of the study or during follow-up, as illustrated in Figure E.4 of appendix.

If the average incubation period is relatively long and follow-up period is relatively short, then most of the "relevant" contacts (i.e. relevant to cases who have onset during the study period) would have occurred prior to the study.

Another important factor to consider about a disease under study is its incidence trend over time (30, 129-131). Leprosy has declined in the Karonga population, as discussed in Section 4.5.8. Thus, a long incubation period, coupled with declining incidence, may mean that a large proportion of the incident cases that arise in a period of study would be due to unobserved contact preceding the study. It is difficult to trace contact histories for such cases and this probably leads to low *sensitivity* and low *specificity* of initial contact status classification.

When measuring prior contact (mainly with respect to infections with long incubation periods) the same model can be used to simulate the period prior to the initiation of a cohort study. Current contact status, in the last year of the simulation period, is equivalent to "observed" contact at the start of the cohort study whereas "true" contact status can be derived from household membership *prior* to the last year of the simulation period. However, depending upon the incubation period of the disease in question, not *all* earlier contact may be relevant for disease observed in the study period and should therefore not be considered when defining sensitivity of initial contact status. The length of simulated observation time prior to the initial survey, needed to identify the relevant "window of opportunity" for infection, needs to be carefully defined.

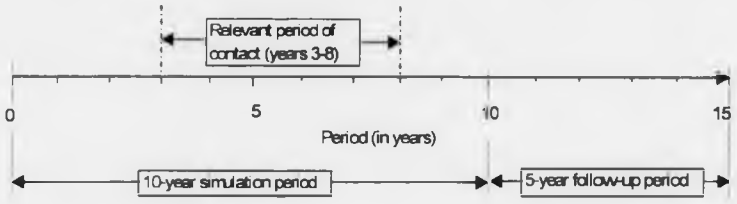
This was achieved by running a series of 10-year simulations, and investigating how varying the timing of relevant contact (with different assumptions about incubation period) affected contact status misclassification. The incubation periods used in the simulations were 5, 7, and 9 years.

Figure 5.3.1 illustrates how the relevant contact period for computing "backward" sensitivity of initial contact status is identified.

The illustration is of a 10-year period simulation run with an assumed incubation period of 7 years, and zero variance. A 5-year period of prospective follow-up for incident disease is assumed to start after simulation stops on year 10. For these assumptions, the relevant period for contact that might lead to disease during follow-up is between years 3 and 8 of the 10-year period simulation run. Contacts occurring within this period are thus considered "true" in so far as they could lead to disease onsets during the follow up period from year 10 to year 15. The initial contacts at year 10 of simulation, at the start of the prospective follow-up, are considered as "observed". Thus, in this example, "backward" sensitivity of initial contact status can be defined as the proportion of "true" underlying contacts in

years 3 to 8 of the simulation period that were correctly observed as contacts at the last year of simulation. Such an approach recognizes the relevant contact period necessary for transmission of infection in order for disease onset to occur during a follow-up.

Figure 5.3.1 Diagram of simulation over 10-year period to identify the relevant period of contact for computing "backward" sensitivity of initial contact status assuming a 7-year incubation period. It is assumed that a cohort study begins on year 10, and follows up the population for 5 years, until year 15. Thus, the "initial" contact status is that observed at year 10.



"Backward" sensitivity of initial contact status has been computed based on contact with all the index cases together as well as separately for contact with "old" and "new" index cases (explained in Section 5.3.2.1 below). Analogous definitions have been used for incubation periods of 5 and 9 years. In the simulation model, we have assumed that incubation period is a constant. A more complex model would incorporate variable incubation periods.

Fifty simulations of 10-year periods each were run with incubation periods of 5, 7 and 9 years separately and a fixed 5-year follow-up period. Table 5.3.2 shows the relevant period of contact, for each incubation period during a 10-year period simulation. To calculate sensitivity based on an n-year incubation period for a 10-year simulation period, we are interested in contacts arising between b-n and e-n where b and e are beginning and end of the cohort study period respectively.

Table 5.3.2 Period of contact during a 10-year simulation period, starting at year 0, for disease onset to occur during a five year follow up period from year 10-15.

Incubation period (in years)	Relevant period of contact
5	Between years 5 to 10.
7	Between years 3 to 8.
9	Between years 1 to 6.

5.3.2.1 Generation of incident cases

Due to the long period under simulation, it is inevitable that some initial index cases will be lost to migration and death. We have to consider generation of incident cases in our simulations to avoid depletion of "index cases" over time. Incident cases have been generated based on age and sex incidence rates (Table 5.3.3) obtained from previous studies in the same population (132). Incident cases are generated by age and sex but without considering household contacts as being at a higher risk of developing disease than non-contacts. Placing contacts at a higher risk when generating incident cases, by simulation was considered inappropriate for this study because we are investigating the risk of disease associated with household contact.

To incorporate the declining incidence of leprosy with time in the simulations, the input parameters for incidence generation were reduced each year of the simulation period. Over a 10-year period, we assumed a conservative estimate of 50% decline in incidence based on incidence trends observed in the population. The age and sex specific incidence rates in Table 5.3.3 were observed over a 10-year period. Assuming an exponential decline in the incidence rates over time, we obtained a parameter value of 0.077 as the constant annual rate of change in the incidence rate. The declining parameter estimates are shown in Appendix, Table F.7.

Table 5.3.3 Average annual incidence rates of leprosy in Karonga District by age and sex, 1980-89

	Age groups (in years)										
SEX	**0-4**	**5-9**	**10-14**	**15-19**	**20-24**	**25-29**	**30-34**	**35-44**	**45-54**	**55-64**	**65+**
Male	0	0.0003	0.0009	0.0016	0.0007	0.0010	0.0013	0.0011	0.0020	0.0013	0.0018
Female	0	0.0004	0.0012	0.0013	0.0014	0.0011	0.0021	0.0020	0.0022	0.0023	0.0018

The model also considers contact with incident cases because such contact might lead to disease over long simulation periods. In these simulations, we assume that the infectious period for incident cases is for 3 years from year of onset of disease. During this period, such cases are considered "new". Within the 3-year period, any individual who comes into contact with these cases is considered as having contact with a new case. After the 3-year period is over, such cases are considered as "old" cases. "Old" index cases are thus those that were present at the start of follow-up at year 10 with onset more than 3 years prior to start of follow-up. Since sensitivity is defined at start of follow-up, "new" cases are thus leprosy cases with onset less than 3 years prior to the start of follow-up at year 10. The aim of separating "new" and "old" cases is to look at contact-associated risks separately as was done in the published analysis where new cases were assumed to be more infectious than old or ancient cases.

5.3.3 Sensitivity of initial contact status

Each (5-year or 10-year period) simulation run produces sensitivity of initial contact status values by age and sex. The mean sensitivity of initial contact status is the average over all simulation runs.

The 95% confidence intervals of sensitivity of contact status are calculated assuming normality. The standard deviation is calculated from the generated sensitivity values. The crude 95% confidence intervals are computed by ordering the values of sensitivity of initial contact status and obtaining the 2.5^{th} and 97.5^{th} percentiles as outlined in (133).

5.4 Adjustment of rate ratios for misclassification of contact status

The relative risks (rate ratios) of disease among household contacts compared to non-contacts as estimated in cohort analyses (19) are adjusted for misclassification of household contact status using sensitivity of initial contact status values obtained

from the simulation model and plausible specificity values. Table 5.4.1 shows the observed frequencies of disease in the contact and non-contact groups and how the relative risks were computed.

The observed relative risks among household contacts compared to those with no known household contact will be $(n_{11} / n_1)/(n_{01} / n_0)$, where n_{11}, n_{10}, n_{01}, n_{00} are cell frequencies, n_1, n_0 are row totals and n_1, n_0 are column totals. Since we had observed contacts, the "true" contacts and incidence cases were obtained by adjusting for misclassification of household contact. The method used for adjusting for misclassification in this analysis is an adaptation of the formulae developed by Copeland *et al.* (86) for disease misclassification in cohort studies, details of which are shown in Appendix D.

Table 5.4.1 Initial contact status and incidence cases in a cohort study

Initial contact status	Disease status		
	Incidence cases	Non-cases	Total
Household contact	n_{11}	n_{10}	n_1
Non-contact	n_{01}	n_{00}	n_0
Total	n_1	n_0	n

The true underlying relative risk assuming no misclassification of household contact status will be $(n_{11}*/ n_1*)/(n_{01}*/ n_0*)$ where $n_{11}*$, $n_{01}*$ and n_1*, n_0* are "true" cell frequencies and row totals respectively after adjusting for contact status misclassification as estimated using formulae in Appendix D.

For rate ratios, the same principles for adjusting for contact status misclassification apply. We need to take into account incident cases and person-time at risk in the contact and non-contact groups of a cohort study. Thus, n_1 and n_0 in last column of Table 5.4.1 would represent person-times. However, the methods for adjusting the cell frequencies are the same.

We shall assume non-differential misclassification of contact status between the two groups i.e. the chance of incident cases being misclassified in the contact and non-contact groups is the same.

5.4.1 Variance estimation for observed and corrected rate ratios

Variation in the estimates of 'true' underlying (corrected) rate ratios above was based on values of sensitivity of contact status obtained from confidence limits (crude or otherwise). The confidence intervals of the 'true' underlying rate ratio should take into account the variation in the sensitivity of contact status as outlined by Greenland (106). The 95% confidence intervals for the observed rate ratios are obtained using the classical variance estimation of $(1/D_0 + 1/D_1)$ with an error factor of $\exp(1.96^* \sqrt{(1/D_0 + 1/D_1)})$ where D_0 and D_1 are number of cases in the unexposed and exposed groups respectively. The variance estimation procedure for the 'true' underlying (corrected) rate ratios has been outlined in Appendix D. The variance of sensitivity of contact status used here was obtained from simulations whereas Greenland (106) derives it from validation 2x2 tables based on the binomial distribution.

CHAPTER 6 RESULTS

This chapter presents estimates of ("forward" and "backward") sensitivity of initial contact status obtained from the LEP data and those based on the simulation model. It is also a presentation of rate ratios corrected for contact status misclassification based on sensitivity values obtained from the simulation model.

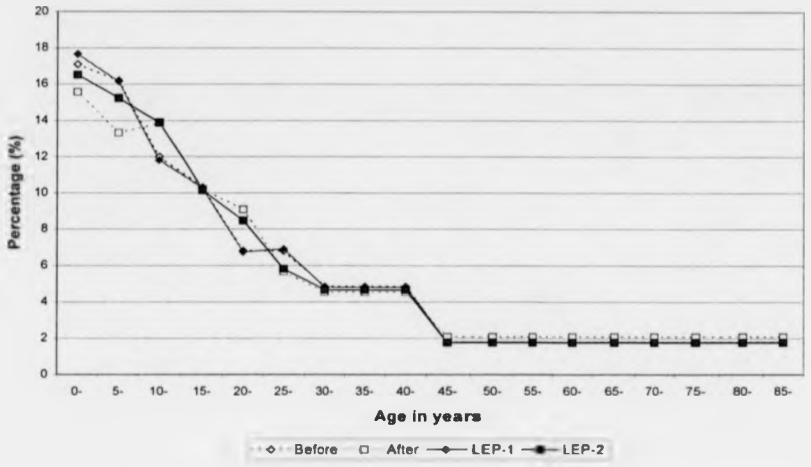
6.1 Stability of the simulation model

Stability of the model was investigated by comparing age/sex structure of the population and household size distribution before and after simulations.

Figure 6.1.1 shows that the age distribution before and after a five year simulation period are similar. The distribution before simulations was based on the "initial" sample drawn from the LEP-1 population, hence the similarity between age distributions at LEP-1 and before simulations as well as between LEP-2 and after simulations. This reflects how well the simulation captures age and sex population structure.

Compared to LEP data, there seem to be an overestimate of infants (less than 1 year old) after simulations. However, births in the model were simulated on an annual basis over a 5-year period, and should be closer to the true number of infants than observed in LEP data. The simulations led to underestimate of children between 1-9 years of age perhaps because the assumed mortality rates were too high.

Figure 6.1.1 Age distribution at start and after 5-year simulations (which is an average of 100 runs) compared to observed LEP-1 and LEP-2 distributions.

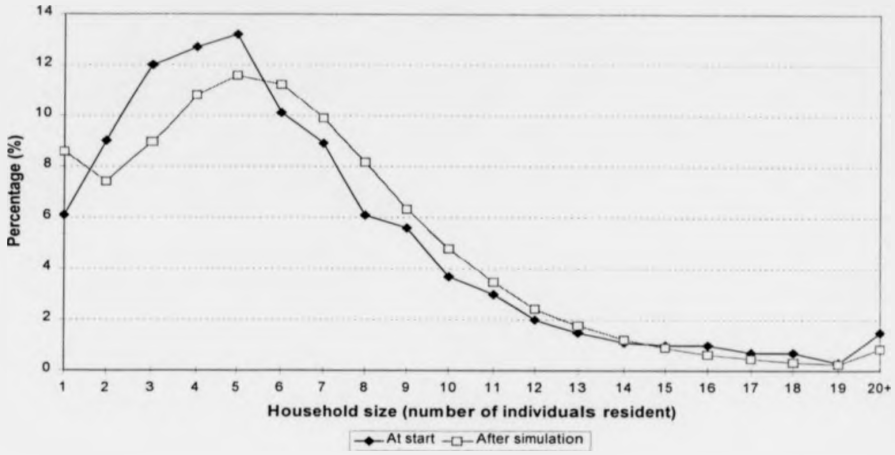


Note: The wider age groups (30-44) and 45+ have been spread over 3 and 9 year age groups respectively

Figure 6.1.2 shows household size distribution before and after the simulations were run. The proportion of single-person households after simulations was slightly higher than that obtained before simulations (8.5% compared to 6% before simulations). This difference was considered acceptable for the purpose of this project (early simulations without the marriage procedure produced 30% single person households).

Apart from this increase in single person households, the model led to some decrease in households of size 2-5 and an increase in households of size 6-12. Rather than attempt to reproduce the observed distributions precisely, it was considered that the age and household size distributions generated from the simulation model were sufficiently comparable to the observed LEP data for the main purposes of this project (the tracking of contact status).

Figure 6.1.2 Household distribution at start and after (5-year period) simulation runs. The "after" distribution is an average over 100 runs.



6.2 Sensitivity of contact status

6.2.1 Sensitivity of initial contact status based on LEP data

Apart from obtaining sensitivity of initial contact status from simulations, we were also able to estimate contact status misclassification directly from the LEP data.

This analysis was done to estimate contact status misclassification based on individuals recorded at LEP-1 and LEP-2 compared to contact status misclassification based on individuals recorded in the "three-survey" population, with an additional survey (SS/KAS) between LEP-1 and LEP-2.

3902 individuals were identified at 3 time points (at LEP-1, SS/KAS and LEP-2) as explained in Section 3.1. Thus, we could identify the proportion of all observed "true" contacts by LEP-2 (individuals who were ever classified as contacts on the basis of observations in LEP-1, SS/KAS and LEP-2). This gives an estimate of the forward "sensitivity" of initial contact status as observed in the initial survey. According to the definition, an individual who was once in contact with an index case remains a "contact" although they may move to a non-case household over time.

Of the 3902 individuals who were interviewed in all the three surveys (LEP-1, SS/KAS and LEP-2), 10% (367) were living in a case household in at least one survey and about 2% (92) were index cases of leprosy.

Of all the individuals who resided in the SS/KAS geographical area at all three surveys, 10% (367), 11% (415) and 7% (261) were observed to be living in case households at LEP-1, SS/KAS and LEP-2. The overall (forward) sensitivity of initial contact status was 77%. That is, of the 477 individuals who were recognised as

"true" contacts at either LEP-1 or SS/KAS or LEP-2, (367=) 77% were correctly observed as contacts at LEP-1.

A total of 80 incident cases were identified by the time the LEP-2 survey was completed among individuals who resided in the SS/KAS geographical area. Of these incident cases 13 (16%) and 10 (13%) had been living in household contact with at least one index case at LEP-1 or SS/KAS respectively previous to onset of disease.

The sensitivity of initial observed contact status at LEP-1 for the incident cases was about 93% (13/14). In other words, 13 of the 14 incident cases who were contacts (at either LEP-1 or SS/KAS or LEP-2) were observed as contacts at LEP-1. The sensitivity of initial contact status for individuals who did not develop disease (non-case group) was 76.5%. That is, 354, or 76.5% of the 463 individuals who were recognised as contacts at either LEP-1 or SS/KAS or LEP-2 were observed as contacts at LEP-1. Given the small numbers involved in particular of incident cases, these sensitivity values are not significantly different (p=0.27).

Table 6.2.1 shows sensitivity of initial contact status by age and sex for individuals who were recorded in all the three surveys (LEP-1, SS/KAS and LEP-2) and those recorded in LEP-1 and LEP-2 only.

Table 6.2.1 Age and sex breakdown of "sensitivity" of initial contact status (as observed in LEP-1), with numerator (in parentheses). "True" contact status was defined on the basis of individuals recorded at LEP-1, SS/KAS and LEP-2 (three-survey sample) or on the basis of individuals recorded at LEP-1 and LEP-2 only.

	Male		Female	
Age group (in years)	Three-survey sample	LEP-1 – LEP-2	Three-survey sample	LEP-1 – LEP-2
0-9	0.80 (66)	0.84 (1612)	0.78 (79)	0.83 (1578)
10-19	0.77 (50)	0.82 (1074)	0.71 (36)	0.75 (1004)
20-29	0.73 (16)	0.83 (420)	0.57 (17)	0.78 (534)
30-44	0.68 (13)	0.86 (353)	0.83 (35)	0.80 (653)
45+	0.92 (23)	0.85 (542)	0.82 (32)	0.81 (691)
Overall	0.79 (168)	0.83 (4003)	0.76 (199)	0.80 (4460)

The sensitivity values based on the "three-survey" sample data were lower than those based on LEP-1 and LEP-2 data (overall respective values of 0.79, 0.83 for males; 0.76, 0.8 for females). These differences are as expected because additional contacts initially missed at start of follow-up must be identified in 3 than in 2 examinations.

There is still need to conduct simulations because the population surveyed three times was relatively small, and the period between these surveys was more than 1 year.

6.2.2 "Forward" sensitivity of initial contact status estimated from simulations

What we define as "forward" sensitivity is appropriate for situations in which the incubation period of disease is relatively short (most contacts leading to disease occur during the study period, as explained in Section 5.3.1).

Table 6.2.2 shows the age-sex distribution of contacts in the "initial" sample at the start of simulations, along with average "forward" sensitivity of this initial contact status, defined at LEP-1, as an estimate of all contacts arising over the next five years (based on 100 five-year period simulation runs).

Of the 11401 individuals in the starting population, 1176 (10.3%) were observed to be contacts at the start (LEP-1). There are no apparent sex differences in sensitivity for children under 5 years of age. The sensitivity peaks at 0.73 and 0.68 in boys and girls aged 5-9 years respectively, and is lower among females than males in age groups 5-19. The values are lowest in the 25-29 age groups in both males and females (0.56 and 0.52 respectively). However, the distribution lowers earlier for females (in 15-19 age group) than males. Misclassification decreases (sensitivity increases) with age for individuals over 30 years of age, with misclassification greater for males than for females. However, these are based on small numbers and should be treated with caution.

Table 6.2.2 Initial contacts and simulation results of "forward" sensitivity by age and sex

	SEX	Age groups (in years)									
		0	1-4	5-9	10-14	15-19	20-24	25-29	30-44	45+	Total
*Initial contacts (%)	M	11 (8.5)	105 (12.1)	112 (12.0)	66 (9.2)	64 (10.6)	38 (11.0)	31 (10.4)	45 (6.8)	71 (8.6)	543
	F	15 (11.5)	83 (9.9)	100 (11.2)	74 (12.1)	87 (14.2)	33 (9.0)	45 (9.4)	82 (9.4)	114 (11.5)	633
											1176
Average forward sensitivity of contact status after simulations	M	0.62	0.66	0.73	0.72	0.62	0.59	0.56	0.62	0.70	
	F	0.60	0.66	0.68	0.65	0.58	0.61	0.52	0.67	0.74	

Note: *The denominator, on which the % are based, excludes index cases in the starting population. M-Male and F-Female.

6.2.2.1 Sensitivity based on lower and upper annual change of household

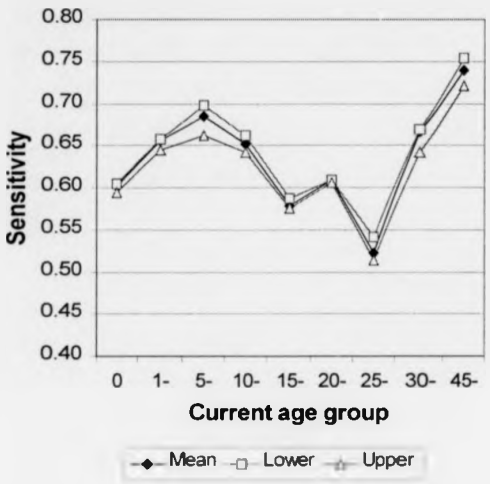
One hundred simulations, of 5-year periods each, were run separately using the 95% lower and upper confidence limits of the annual rate of change of household shown in Table 5.1.3. The values of sensitivity of initial contact status obtained from these simulations are compared to those obtained using the mean annual rate of household change.

The sources of uncertainty in sensitivity of initial contact status may be due not only to variation in change of household but also to stochastic variability in the simulations. The latter reflects our uncertainty about the relation between the contact status and demographic events of interests. Crude 95% confidence limits of sensitivity of initial contact status from the simulation model capture this stochastic variability.

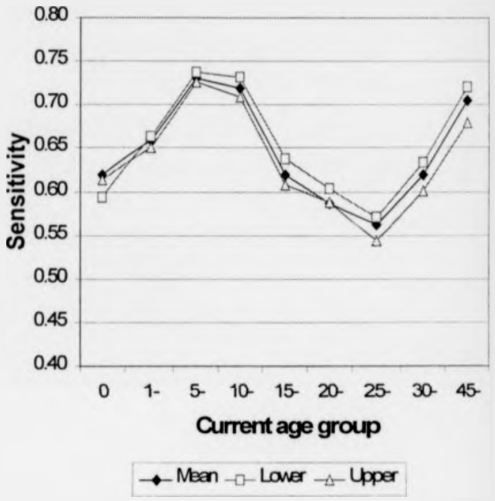
Figures 6.2.1a and b show "forward" sensitivity of initial contact status, for males and females, by different annual rates of household change (95% confidence limits). Each curve represents an average of 100 simulation runs. The slight increase in sensitivity for females aged 20-24 years is attributable to an increase in return rather than forward moves.

Figure 6.2.1 Average forward sensitivity of initial contact status for (a) females and (b) males over 100 runs of 5-year periods each using lower, upper and mean annual proportions changing household.

(a) Females



(b) Males



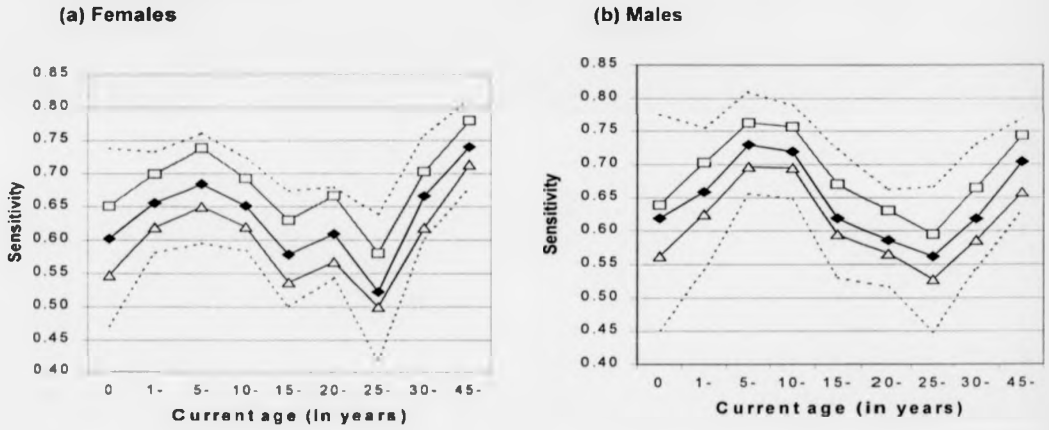
Note: "Mean", "Lower" and "Upper" present age-specific average sensitivity of initial contact status obtained using mean annual proportion changing household, lower and upper 95% confidence bounds respectively

We observe only small differences in the sensitivity of initial contact status for females or males regardless of whether one uses the lower, upper or mean annual rate of change of household. Thus, it is adequate to only investigate sensitivity values based on the mean annual rate of changing household.

6.2.2.2 Confidence intervals for sensitivity of contact status

Figures 6.2.2a and b show crude 95% confidence intervals for the "forward" sensitivity of initial contact status obtained from the simulations. The model was also run using the lower and upper values of mean annual percentage of individuals changing household based on an assumed relative precision of 20% (chosen arbitrarily) of the mean estimate. The results for relative precision are based on the average of 50 (5-year period) simulation runs.

Figure 6.2.2 Crude confidence intervals of sensitivity of initial contact status for (a) females and (b) males, by age.



Note: ♦= average sensitivity of initial contact status, = crude 95% confidence limits for sensitivity of initial contact status, Δ, = Lower and Upper average sensitivity of initial contact status values obtained using upper and lower bound (+/-20%) of annual household change respectively

We note that the crude (based on ordering) 95% confidence intervals for the "true" sensitivity of initial contact status are much wider compared to the ones obtained when we assume normality and also compared to the range of the sensitivity values obtained using a relative precision of 20%. This is more pronounced for those aged under 1 year and between 25-29 years.

The crude 95% confidence intervals give the bounds within which the true value of age/sex specific sensitivity of initial contact status is expected to lie without varying the input parameters. Varying the estimates of change of household by up to 20% still gives values of sensitivity of initial contact status that lie within the crude 95% confidence intervals across all age groups. The variation has to be more than 20% of the rate of change of household to have a marked effect on inferences.

Based on these results and Section 6.2.2.1, we concentrate on sensitivity estimates obtained using mean proportion of individuals changing household from the descriptive analyses and the crude upper and lower confidence limits from the simulations.

6.2.2.3 Trend of sensitivity with duration of follow-up

The stochastic simulations of household dynamics were carried out considering diseases with incubation periods of different lengths. The simulations considering relatively short incubation periods (a year) mainly investigated the age-sex trend of sensitivity of initial contact status for a fixed 5-year period of follow-up. This was because in the published cohort study in which rate ratios of disease associated with household contact were computed (19), the initial (LEP-1) and follow-up (LEP-2) studies were, on average, 5 years apart. To adjust the relative risk estimates for contact status misclassification, it was considered appropriate to obtain sensitivity of contact status from a 5-year period and with the same age and sex breakdown.

Published cohort studies investigating the effect of (household) contact with index cases on risk of disease have used different durations of follow-up. Ideally, the relative risk estimates from these studies should be adjusted based upon the sensitivity of initial contact status estimated for a similar duration of follow-up as in the study. Table 6.2.3 presents results from 50 simulation runs of the model for

follow-up periods from one to five years. Estimates of the "forward" sensitivity of initial contact status, by age and sex, are shown.

Table 6.2.3 Forward sensitivity of initial contact status by age and sex from 50 simulation runs with different duration of follow-up based on contact with all index cases

	DURATION OF FOLLOW-UP (IN YEARS)									
	1		2		3		4		5	
Age group (in years)	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0	0.89	0.87	0.79	0.80	0.71	0.71	0.66	0.65	0.62	0.60
1-4	0.93	0.91	0.86	0.83	0.78	0.78	0.70	0.70	0.66	0.66
5-9	0.92	0.92	0.84	0.82	0.81	0.78	0.77	0.72	0.73	0.68
10-14	0.90	0.89	0.85	0.81	0.79	0.73	0.74	0.69	0.72	0.65
15-19	0.87	0.87	0.80	0.78	0.72	0.72	0.64	0.64	0.62	0.58
20-24	0.87	0.83	0.77	0.75	0.69	0.69	0.66	0.66	0.59	0.61
25-29	0.86	0.89	0.77	0.79	0.69	0.67	0.60	0.54	0.56	0.52
30-44	0.87	0.90	0.76	0.81	0.69	0.76	0.64	0.71	0.62	0.67
45+	0.90	0.92	0.85	0.87	0.79	0.82	0.74	0.77	0.70	0.74

As expected, the sensitivity declines with duration of follow-up. Sensitivity values for a 1-year period of follow-up are over 85% whereas those for a 5-year period range between 50% and 74%.

6.2.3 "Backward" sensitivity of initial contact status

We now consider the sensitivity of initial contact status relevant to cohort studies of diseases of relatively long incubation periods. In such situations, some incident cases that arise during a follow-up period of study may be attributable to contact which occurred prior to the period of study. "Backward" sensitivity was defined in Section 5.3.2 and relates to the sensitivity of contact status observed at the start of a cohort study, as a measure of relevant contact which occurred in the past, before the study began.

6.2.3.1 Sensitivity of contact status with duration of incubation period

The actual incubation periods for diseases such as leprosy and tuberculosis can vary from several months to many years, or even decades, in length. To explore this variability, we made 50 (10-year period) simulation runs of the model assuming fixed incubation periods of 5, 7 and 9 years. This 10-year period precedes the start of the cohort study. In this situation (illustrated in Figure 5.3.1), the "initial" contact status refers to the situation *at the end of* the simulation period at which time the cohort study is assumed to begin. The incident cases occurring in the subsequent (five-year follow up) period are attributable to contacts and infections which occurred prior to the onset of the cohort study. Thus, some of the individuals who became incident cases would be wrongly classified as non-contacts - leading to reduced of specificity of contact status, as discussed in Section 5.3.2 and explored further in Section 6.3.6.

Table 6.2.4 presents estimates of "backward" sensitivity of initial contact status by age for male and female contacts of all index cases of leprosy (new or old cases) for different incubation periods. We see the expected trend of decline in sensitivity with increasing incubation period.

Table 6.2.4 Values of backward sensitivity of contact status, together with standard deviations in parenthesis, for contacts of all index cases (old or new) from simulations with long incubation period after 50 simulation runs of a 10-year period with fixed 5 years of follow-up.

	LENGTH OF INCUBATION PERIOD (in years)					
	5		7		9	
Age group (in years)	Male	Female	Male	Female	Male	Female
0-9	0.60 (0.0308)	0.60 (0.0295)	0.52 (0.0374)	0.52 (0.0310)	0.48 (0.0318)	0.48 (0.0308)
10-14	0.62 (0.0358)	0.59 (0.0394)	0.54 (0.0413)	0.50 (0.0438)	0.51 (0.0478)	0.45 (0.0386)
15-19	0.61 (0.0291)	0.57 (0.0405)	0.53 (0.0401)	0.46 (0.0391)	0.47 (0.0405)	0.43 (0.0456)
20-24	0.60 (0.0367)	0.60 (0.0482)	0.50 (0.0505)	0.50 (0.0424)	0.43 (0.0421)	0.46 (0.0586)
25-29	0.57 (0.0468)	0.67 (0.0373)	0.47 (0.0471)	0.57 (0.0434)	0.40 (0.0565)	0.54 (0.0475)
30-44	0.64 (0.0427)	0.78 (0.0390)	0.56 (0.0426)	0.71 (0.0484)	0.51 (0.0529)	0.70 (0.0448)
45+	0.68 (0.0320)	0.67 (0.0327)	0.60 (0.0449)	0.61 (0.0350)	0.56 (0.0402)	0.57 (0.0410)

6.2.3.2 Sensitivity of contact status by type of index case ("old" and "new")

The analysis was repeated to explore the implications of contact with "old" or "new" index cases – (see Section 5.3.2.1). The results are shown in Table 6.2.5.

Table 6.2.5 "Backward" sensitivity of initial contact status for males and females by age and whether index case was "old" or "new".

Age group (in years)	Duration of incubation period (in years)											
	5 years				7 years				9 years			
	Time reference of index case.				Time reference of index case.				Time reference of index case.			
	New		Old		New		Old		New		Old	
	M	F	M	F	M	F	M	F	M	F	M	F
0-9	0.35	0.34	0.53	0.53	0.20	0.19	0.45	0.45	0.07	0.07	0.41	0.40
10-14	0.32	0.32	0.56	0.54	0.18	0.16	0.45	0.44	0.05	0.05	0.43	0.38
15-19	0.34	0.34	0.55	0.51	0.18	0.19	0.47	0.38	0.04	0.11	0.41	0.34
20-24	0.33	0.36	0.56	0.52	0.16	0.20	0.45	0.41	0.04	0.10	0.37	0.37
25-29	0.33	0.39	0.52	0.58	0.18	0.23	0.41	0.47	0.04	0.14	0.34	0.43
30-44	0.35	0.39	0.57	0.68	0.20	0.26	0.48	0.62	0.05	0.15	0.44	0.59
45+	0.34	0.35	0.62	0.61	0.19	0.19	0.53	0.55	0.04	0.05	0.50	0.51

We see that the values of sensitivity of initial contact status for both male and female contacts of "new" cases are consistently lower than for contacts of "old" index cases for the different incubation periods.

In addition, we see that among those over 24 years of age the sensitivity of initial contact was generally higher for females than for males.

Sensitivity, in this situation, is defined as the proportion of contacts, within the relevant contact period, also observed as contacts at the last year of simulation (start of follow-up study). "New" cases were assumed to be infectious for only 3 years before they were considered as "old" cases (see Section 5.3.2.1). Thus, according to the definition of sensitivity, only individuals living in contact with "new" cases which arise within the last 3 years of the simulation period prior to onset of a follow-up study will actually be observed as contacts at the last year of simulation. The further the relevant period of contact (see Section 5.3.2) is from the onset of follow-up study, the fewer contacts of "new" cases (denominator) observed. The denominator increases with incubation period (since we have a much wider relevant period of contact, hence, more contacts observed).

Because of this, most contacts with "new" cases occurred early in simulations and are not actually observed at the last year of simulations leading to low sensitivity. Sensitivity based on contact with "old" cases, is relatively high because most contacts which occur throughout the simulation period are also observed at the last year of simulation.

The issue of "old" and "new" cases is also relevant for specificity i.e. considering someone a contact but the contact is with an "old" (no longer infectious) case – similar to dealing with a false positive. Risk of disease should thus be greater for observed contacts of "new" cases than for observed contacts of "old" cases.

6.2.4 Summary

This section began by demonstrating that the household dynamics model developed in Chapter 5 provided sufficiently stable age and household size distributions, over five years of simulation, for it to be used for tracing contact histories over time.

A direct estimate of the sensitivity of initial contact status (as observed at LEP-1) as a measure of all contact over the next five years, based on individuals examined three times (LEP-1, SS/KAS and LEP-2) was 77%.

There were no apparent sex differences in sensitivity values for children aged under 5 years as their movements are largely dependent on their parents or guardians.

A low sensitivity of initial contact status measures is a reflection of a high rate of household change. The earlier lower sensitivity among young adult females than males is related to earlier household change for females than males. Household change is relatively low for older adults leading to increased sensitivity values; higher for females than males. Although old females are more likely to change household within the study area than old males, there was higher male (who became contacts) in-migration and return moves to the study area at older age groups than females resulting in lower sensitivity in males.

The longer the duration of follow-up in cohort studies, the more difficult it is in practice to capture all new contacts who might go on to develop disease, hence, the lower the sensitivity of initial contact status. The methods for taking into account contact status misclassification to obtain precise estimates are in principle appropriate for any cohort analysis of household contact.

The longer the incubation period, the lower the sensitivity of initial (observed) contact status, and the greater the likelihood for an actual earlier contact to go unrecognised. It is important to appreciate that, when dealing with diseases such as leprosy and tuberculosis, disease onsets seen during follow-up may well be attributable to unobserved earlier contact.

The "forward" and "backward" perspectives describe an important distinction. We observed that, in general, values for "backward" sensitivity are lower than "forward" sensitivity of initial contact status from a 5-year period simulation model, with differences being more pronounced the longer the incubation period - reflecting the importance of missed earlier contact. "Forward" sensitivity considers all contacts during the follow-up period. This is relevant for infection transmission for diseases with relatively short incubation period. "Backward" sensitivity assumes that relevant contact occurred earlier before the study. This is mainly useful for the study of diseases with relatively long incubation period or with relatively shorter period of follow-up than its incubation period.

6.3 Adjusting estimates for contact status misclassification

6.3.1 Adjusted relative risks based on observed sensitivity

Age and sex-specific risk ratios (observed and corrected) were computed from the observed data based on incident cases observed between LEP-1 and LEP-2 among the 3902 individuals seen in the three-survey sample. Tables 6.3.1a and b show small differences between the observed and corrected risk ratios. The correction for contact status misclassification was based on sensitivity values from Table 6.2.1.

Table 6.3.1a Males

Age	Household contacts		Non-contacts		RISK RATIO (contacts relative to non-contacts)	
	Cases/ Total	Risk	Cases/ Total	Risk	Observed	Corrected
0-9	3/83	0.036	7/646	0.011	3.34 (0.86, 12.9)	3.63 (0.79, 16.59)
10-19	3/65	0.046	6/354	0.017	2.72 (0.68, 10.89)	3.03 (0.59, 15.47)
20-29	0/22	0.000	2/114	0.018	-	-
30-44	0/19	0.000	5/197	0.025	-	-
45+	1/25	0.040	3/269	0.011	3.59 (0.37, 34.48)	3.66 (0.36, 37.69)
CRUDE	7/214	32.710	23/1580	14.557	2.25 (0.96, 5.24)	2.29 (1.04, 5.02)
			MH- Estimate		2.22 (0.95, 5.22)	2.24 (1.01, 4.94)

Table 6.3.1b Females

Age	Household contacts		Non-contacts		RISK RATIO (contacts relative to non-contacts)	
	Cases/ Total	Risk	Cases/ Total	Risk	Observed	Corrected
0-9	3/101	0.030	10/568	0.018	1.69 (0.46, 6.13)	1.75 (0.43, 7.15)
10-19	0/51	0.000	5/286	0.017	-	-
20-29	1/30	0.033	6/243	0.025	1.35 (0.16, 11.21)	1.40 (0.12, 15.86)
30-44	2/42	0.048	9/294	0.031	1.56 (0.34, 7.20)	1.58 (0.32, 7.86)
45+	1/39	0.026	13/362	0.036	0.71 (0.09, 5.46)	0.71 (0.09, 5.52)
CRUDE	7/263	0.027	43/1753	0.025	1.09 (0.49, 2.41)	1.07 (0.52, 2.19)
			MH- Estimate		1.15 (0.51, 2.56)	1.16 (0.56, 2.38)

The Mantel-Haenszel estimate of the risk ratio for contacts versus non-contacts in this population sample adjusted for age and sex is 1.51 (0.85, 2.68). The relative

risk of disease for contacts compared to non-contacts was highest in young children and in old individuals, but the numbers are relatively small and the difference not significant.

6.3.2 Adjusted estimates based on sensitivity values from simulations

Having computed "forward" sensitivity of initial contact status values from the simulation model, they were used to adjust published estimates of relative risks of leprosy (household contacts versus non-contacts) for contact status misclassification.

The published rate ratios (19) were derived in an analysis stratified by age, sex and BCG status of individuals and also by type of disease in the index cases (MB or PB). Tables 6.3.2a and b show observed data aggregated over BCG status and type of disease in the index case, together with their age-specific incidence rates of disease (in household contacts and non-contacts) and rate ratios. Table G.1a and b in Appendix shows further adjustment by BCG status.

Table 6.3.2a Observed rate ratios of disease for household contacts compared to non-contacts for males, by age.

	Household contacts		Non-contacts		Observed rate ratio
Age group (in years)	Cases/Total person years	Incidence rate (per 1000pyr)	Cases/Total person years	Incidence rate (per 1000pyr)	(contacts vs non-contacts)
0-9	1/4984	0.201	6/54953	0.109	1.84 (0.22, 15.26)
10-19	13/4573	2.843	32/47230	0.678	4.20 (2.20, 7.99)
20-29	2/1816	1.101	11/20206	0.544	2.02 (0.45, 9.13)
30-44	1/1311	0.763	20/22771	0.878	0.87 (0.12, 6.47)
45+	5/2359	2.120	37/31843	1.162	1.82 (0.72, 4.64)
Total	22/15043	1.462	106/177003	0.599	2.44 (1.54, 3.86)
MH-rate ratio					2.57 (1.64, 4.01)

Table 6.3.2b Observed rate ratios of disease for household contacts compared to non-contacts for females, by age.

Age group (in years)	Household contacts		Non-contacts		Observed rate ratio (contacts v non-contacts)
	Cases/Total person years	Incidence rate (per 1000pyr)	Cases/Total person years	Incidence rate (per 1000pyr)	
0-9	4/5042	0.793	6/53570	0.112	7.08 (1.20, 25.10)
10-19	8/4415	1.812	27/45509	0.593	3.05 (1.39, 6.72)
20-29	3/2213	1.356	25/28496	0.877	1.55 (0.47, 5.12)
30-44	3/2705	1.109	52/35791	1.453	0.76 (0.24, 2.44)
45+	9/3080	2.922	50/37258	1.342	2.18 (1.07, 4.43)
Total	27/17455	1.547	160/200624	0.798	1.94 (1.29, 2.92)
MH-rate ratio					2.04 (1.35, 3.07)

Note These figures combine individuals with and without BCG scars and do not distinguish contacts by type of disease in the index case.

Although the rate ratios and absolute frequencies shown here are aggregated over BCG status, there is no evidence of differential misclassification of household contact status between the groups with and without BCG scars as there was no association between BCG scar status and observed contact status in the study population (p=0.656). The principles of adjusting for misclassification of contact status are applicable in tables broken down further by other variables but further breakdown of these tables would lead to very small numbers. We here assume that contact status misclassification differs by age and sex but is non-differential across other groups.

In Table 6.3.3 we show values of "forward" sensitivity of initial contact status obtained from the simulations (averaged over 100 runs of 5-year periods each) by sex, with age breakdown as in the contact analysis paper (19). To investigate the variation in rate ratios obtained, they were also adjusted using the crude 95% confidence limits of sensitivity of initial contact status.

Table 6.3.3 Forward sensitivity of initial contact status and their 95% crude confidence intervals, by age and sex

Sex	Age group (in years)				
	0-9	10-19	20-29	30-44	45+
Male	0.69 (0.62, 0.75)	0.68 (0.62, 0.75)	0.58 (0.51, 0.64)	0.62 (0.54, 0.73)	0.70 (0.63, 0.77)
Female	0.66 (0.60, 0.71)	0.62 (0.56, 0.69)	0.58 (0.53, 0.63)	0.67 (0.60, 0.76)	0.74 (0.68, 0.81)

"Corrected" underlying status of contacts and non-contacts and their corresponding incidence rates and rate ratios of leprosy for household contacts compared to non-contacts are shown in Tables 6.3.4a and b, based on the sensitivity of initial contact status values given in Table 6.3.3. For example, if we consider males in the 0-9 age group, there were 1 and 6 cases observed in the contact and non-contact groups respectively (from Table 6.3.2a. Using a sensitivity value of 0.69 obtained for males in 0-9 age group, the corrected number of cases in the contact group was 1/0.69=1.5. The corresponding corrected number of cases in the non-contact group was (0.69*6-(1-0.69)*1)/0.69=5.6 as shown in bold in Table 6.3.4a based on formulae given in Appendix D. Similar corrections were made to the person year denominators of the incidence rates.

Table 6.3.4a Corrected rate ratios of disease for household contacts relative to non-contacts for males, by age.

Age group	Household contacts Cases/Total person years	Incident rate (per 1000 pyr)	Non-contacts Cases/Total person years	Incident rate (per 1000 pyr)	Corrected rate ratios (contacts v non-contacts)
0-9	**1.5**/7248	0.201	**5.6**/52689	0.105	1.91
10-19	19.2/6741	2.843	25.8/45062	0.573	4.96
20-29	3.5/3155	1.101	9.5/18867	0.505	2.18
30-44	1.6/2119	0.763	19.4/21963	0.883	0.86
45+	7.1/3350	2.120	34.9/30852	1.131	1.87
Crude	32.8/22613	1.451	95.2/169433	0.562	2.58
MH-rate ratio					2.74

Table 6.3.4b Corrected rate ratios of disease for household contacts relative to non-contacts for females, by age.

Age group	Household contacts Cases/Total person years	Incident rate (per 1000pyr)	Non-contacts Cases/Total person years	Incident rate (per 1000pyr)	Corrected rate ratios (contacts v non-contacts)
0-9	6.1/7649	0.793	3.9/50963	0.077	10.28
10-19	13.0/7145	1.812	22.1/42779	0.516	3.52
20-29	5.2/3807	1.356	22.8/26902	0.849	1.60
30-44	4.5/4060	1.109	50.5/34436	1.466	0.76
45+	12.2/4161	2.922	46.8/36177	1.295	2.26
Crude	40.8/26822	1.523	146.2/191257	0.764	1.99
MH-rate ratio					2.15

The observed (Tables 6.3.2a and b) and "corrected" (Tables 6.3.4a and b above) incidence rates are the same in the contact group but are different to each other in the non-contact group. The similarity in the contact group is expected as the number of cases and person-years at risk are corrected using the same sensitivity.

6.3.2.1 Variance of observed and corrected rate ratios

Table 6.3.5 shows confidence intervals of observed and corrected rate ratios for both males and females based on variances of "forward" sensitivity of contact status, observed and corrected rate ratios as explained in Section 5.4.1.

Table 6.3.5 Summary table of observed and corrected rate ratios (RR), with 95% confidence intervals, for risk of disease in household contacts compared to non-contacts.

Age group	MALE		FEMALE	
	Observed RR	Corrected RR	Observed RR	Corrected RR
0-9	1.84 (0.22, 15.26)	1.91 (0.19, 18.83)	7.08 (1.20, 25.10)	10.28 (1.48, 71.26)
10-19	4.20 (2.20, 7.99)	4.96 (2.23, 11.04)	3.05 (1.39, 6.72)	3.52 (1.34, 9.24)
20-29	2.02 (0.45, 9.13)	2.18 (0.38, 12.43)	1.55 (0.47, 5.12)	1.60 (0.43, 5.92)
30-44	0.87 (0.12, 6.47)	0.86 (0.11, 6.87)	0.76 (0.24, 2.44)	0.76 (0.23, 2.51)
45+	1.82 (0.72, 4.64)	1.87 (0.70, 5.04)	2.18 (1.07, 4.43)	2.26 (1.06, 4.82)
Crude	2.44	2.58 (1.74, 3.84)	1.94	1.99 (1.41, 2.82)
MH RR	2.57	2.74 (1.83, 4.10)	2.04	2.15 (1.51, 3.05)

The corrected crude RRs were based on aggregated (fractional) numbers of corrected cases and person years. The corrected MH RRs were based on (fractional numbers of) age-specific corrected cases and corresponding person-times using the formula directly. The corrected rate ratios for household contacts compared to non-contacts are higher than the observed rate ratios except for age group 30-44 (rate ratios less than 1) in both males and females. The correction moves the rate ratios from 1. The crude and MH corrected rate ratios are higher than those observed from the data for both males and females.

The confidence intervals for both rate ratios are wide, especially in age groups 0-9. The confidence intervals for the corrected rate ratios are in general much wider than the observed. This is partly because the variance of corrected rate ratios has an extra component based on variance of sensitivity of contact status (106).

Observed and corrected rate ratios are presented in Table 6.3.6, by age and sex. The table gives the lower and upper bounds for the corrected rate ratios based on the crude 95% confidence limits for sensitivity of contact status.

Table 6.3.6 Observed and corrected rate ratios (for household contacts compared to non-contacts) using crude 95% confidence limits of sensitivity, for males and females by age.

Age group (in years)	Males			Females		
	Observed rate ratio	Corrected rate ratios		Observed rate ratio	Corrected rate ratios	
		Lower	Upper		Lower	Upper
0-9	1.84	1.89	1.93	7.08	9.32	12.01
10-19	4.20	4.71	5.26	3.05	3.37	3.67
20-29	2.02	2.14	2.24	1.55	1.59	1.61
30-44	0.87	0.86	0.87	0.76	0.75	0.76
45+	1.82	1.86	1.89	2.18	2.23	2.29
Crude	2.44	2.54	2.63	1.94	1.96	2.02

Note: The "Lower" and "Upper" corrected rate ratios were obtained using upper and lower 95% confidence limits for sensitivity of contact status respectively

The confidence intervals based on variances of sensitivity are wider compared to those based on 95% crude confidence limits for sensitivity. The adjustments bias the estimates away from 1, but the differences between corrected and observed rate ratios are in general small.

6.3.3 Effect of varying sensitivity of initial contact status on corrected RR

This analysis assumed a fixed number of cases and their corresponding person-years of follow-up in the contact and non-contact groups in order to investigate the effect of varying sensitivity of initial contact status on the corrected rate ratios. Here, we use the aggregated observed number of male incident cases (22 and 106 cases in contact and non-contact groups) and person-years of follow-up (15043 and 177003 in contacts and non-contacts).

Figure 6.3.1 Variation in corrected rate ratios with sensitivity of initial contact status for (a) fixed number of cases and person years at risk for males only (Observed RR=2.44) and (b) varying observed RR.

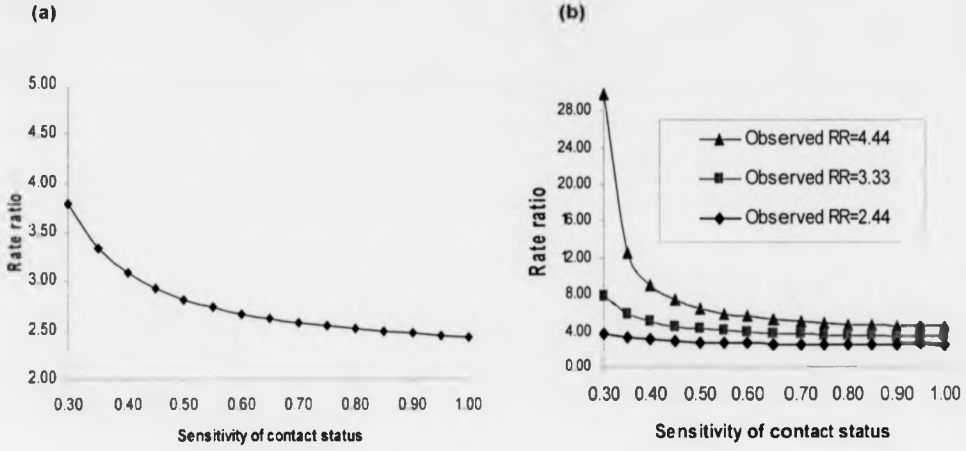
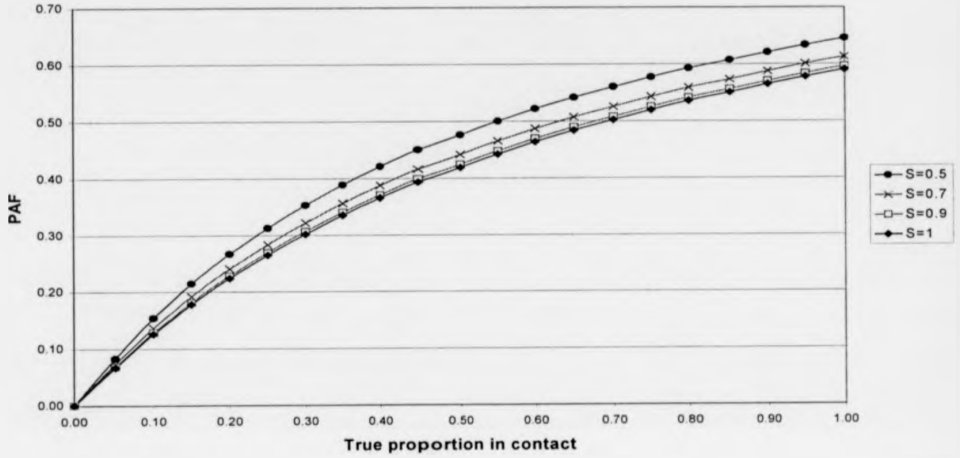


Figure 6.3.1(a) shows the trend in corrected rate ratio as a function of sensitivity. With a sensitivity of contact status of at least 0.6, the observed and corrected estimates are very similar. The proportion of the population in contact with index cases is important in determining the effect sensitivity of contact status on the rate ratios observed. Figure 6.3.1(b) shows that sensitivity becomes important with increasing relative risk of disease in the contact group. For the observed rate ratio of 2.44, the corrected and observed are similar even for low sensitivity. The larger the rate ratio, the further away the observed and corrected are, as sensitivity decreases.

Figure 6.3.2 shows the effect of varying the proportion of individuals in contact with index cases on the population attributable fraction (PAF), using the formula PAF=p(RR-1)/(p(RR-1)+1), based on the same data on males for different sensitivity values. In the formula, p is actual proportion of the population in contact with index cases and RR is the *observed* rate ratio.

Figure 6.3.2 Effect of varying proportion of population in contact with index cases on PAF (Observed RR=2.44) for different sensitivities of initial (observed) contact status.



Here we see that the proportion in contact has a greater influence on PAF than does the sensitivity. The observed proportion of the sample in initial contact was 10% and after simulations, we obtained an average of 15% (cumulative over 5 years) from 100 simulation runs. This translates to PAFs of between 12% and 18% respectively.

6.3.4 Rate ratios for contacts of MB and PB leprosy cases

The rate ratios corrected for contact status misclassification in the previous section were based on all index cases of leprosy combined. However, although MB cases were few in the population (296 out of 2176 index cases), they are known to be more infectious than PB cases. Thus, it is appropriate to look at rate ratios of disease for contacts of MB and PB index cases separately rather than aggregating cases, which led to a dilution of estimates.

Using observed person-years of follow-up and incident cases who were contacts of (MB and PB) index cases and non-contacts shown in Appendix Table G.3, rate ratios of disease for contacts of MB and PB cases were computed separately. Similar adjustments of observed estimates for contact status misclassification as in Section 6.3.2 were made. The same values of sensitivity of contact status were used in the two contact groups assuming non-differential contact status misclassification. Figures 6.3.3a and b show observed and corrected rate ratios, and their confidence intervals, for contacts of MB and PB cases relative to non-contacts by age, for males and females (see also Appendix Tables G.4a and b).

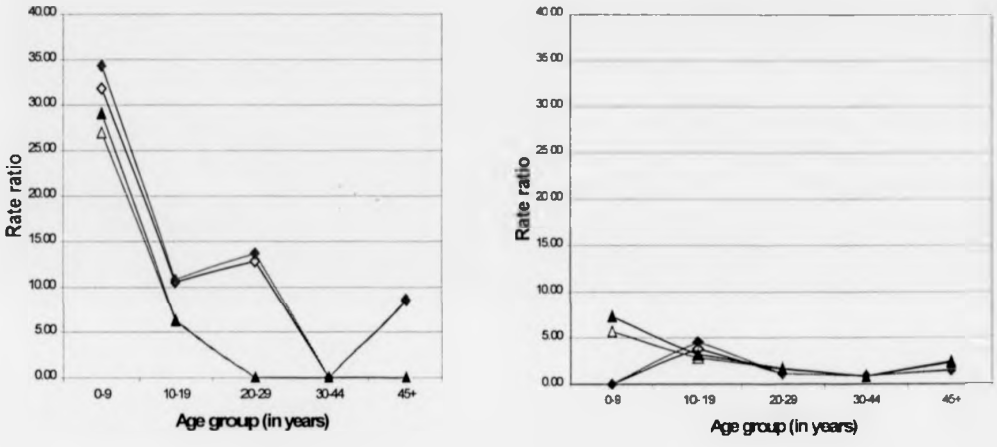
Due to further breakdown of the contact group into contacts of MB and PB index cases, some cells have very few (if not zero) incident cases, in particular among contacts of MB cases. Breaking these data further by BCG scar status led to even smaller numbers and larger confidence intervals. Previous published analyses had revealed no evidence for any confounding of BCG scar status on contact associated risks of leprosy (19).

The corrected rate ratios are similar or higher than the observed rate ratios in all age categories although the differences are small. Although some corrected rate ratios were higher than those observed, we note that apart from showing a stronger association of risk of disease associated with household contact, the significance of

results remains similar as in the observed estimates. This observation is also true for results shown in Sections 6.3.1 and 6.3.2.

Figure 6.3.3 Observed and corrected rate ratios for contacts of (a) MB and (b) PB cases of leprosy by age for females and males.

(a) Contacts of MB cases **(b) Contacts of PB cases**



Note: ◊, △ = Observed rate ratios for males and females respectively and ◆, ▲ = Corrected rate ratios for males and females respectively

The corrected MH estimates of rate ratios adjusted for age for male and female contacts of MB cases were 10.88 (5.16, 22.92) and 2.16 (0.68, 6.87) respectively and similarly 2.28 (1.47, 3.54) and 2.12 (1.48, 3.04) for male and female contacts of PB cases respectively.

The observed and corrected rate ratios of disease for contacts of MB cases were much higher than for contacts of PB cases. The observed Mantel-Haenszel estimate of rate ratio of disease for female and male contacts of MB cases adjusted for age were 1.14 (0.27, 4.78) and 4.84 (1.77, 13.21) times higher respectively than for female and male contacts of PB cases. The difference is statistically significant

for male, but not for female, contacts. The rate ratios for the 0-9 year old, with the exception of male contacts of PB cases, are high compared to those of other age groups.

Figures 6.3.3a and b show that the rate ratios for household contacts compared to non-contacts are highest for young children, especially 0-9 year olds, except for male contacts of PB cases. However, the rate ratios in the 0-9 year age group in Figure 6.3.3a were based on only 1 and 6 incident cases in the contact and non-contact groups respectively.

A high proportion of index cases of leprosy in households with young male and female contacts aged 0-9 years were the parents of the contacts, (38% and 36% respectively), as shown in Tables 6.3.7a and b. These proportions decrease with age to 3% and 2% for male and female contacts aged over 45 years. In addition, the proportion of index cases who were siblings of contacts is higher for young than old contacts. The reverse is true when index cases were themselves children. The proportion of contacts who were unrelated genetically to their index cases thus increased with age.

Table 6.3.7a Relative frequency table (%) for relationship between index cases and their contacts in households, by age of male contacts. Karonga District, northern Malawi 1979-85.

Age group (in years) of contacts	Index case in household					
	Parent	Sibling	Child	Unrelated	Other relations	Total
0-9	38%	11%	0%	29%	22%	1824
10-19	28	18	0	37	17	1338
20-29	22	15	0	50	13	620
30-44	11	9	7	67	6	488
45+	3	4	24	64	5	693
Total						4963

Table 6.3.7b Frequency table (%) for relationship between index cases and their contacts in households, by age of female contacts. Karonga District, northern Malawi 1979-85.

Age group (in years) of contacts	Index case in household					
	Parent	Sibling	Child	Unrelated	Other relations	Total
0-9	36%	11%	0%	30%	23%	1890
10-19	27	13	0	45	15	1258
20-29	9	7	1	79	4	694
30-44	4	2	11	81	2	792
45+	2	3	20	68	7	907
Total						5541

6.3.5 Adjusted estimates using "backward" sensitivity of initial contact status

It has been demonstrated (Section 6.2.3) that contact status misclassification can be more pronounced when we consider disease with long incubation periods, due to earlier unobserved contact. Although such was the case, the corrected rate ratios in Table 6.3.8 using "backward" sensitivity values in Table 6.2.4 show that even with a 9-year incubation period, the household dynamics described in this study indicate that the contact associated risks were only underestimated by a small amount (see also Appendix Tables G.5a and b).

Table 6.3.8 Summary table of observed and corrected rate ratios (RR), with 95% confidence intervals, for risk of disease in household contacts compared to non-contacts for long incubation periods.

Age group	MALE				FEMALE			
	Observed RR	Corrected RR			Observed RR	Corrected RR		
		Incubation period (in years)				Incubation period (in years)		
		5	7	9		5	7	9
0-9	1.84 (0.22, 15.26)	1.94	1.99	2.02	7.08 (1.20, 25.10)	11.95	16.87	23.91
10-19	4.20 (2.20, 7.99)	5.27	5.94	6.57	3.05 (1.39, 6.72)	3.62	4.03	4.29
20-29	2.02 (0.45, 9.13)	2.17	2.26	2.38	1.55 (0.47, 5.12)	1.59	1.61	1.62
30-44	0.87 (0.12, 6.47)	0.86	0.86	0.86	0.76 (0.24, 2.44)	0.76	0.76	0.76
45+	1.82 (0.72, 4.64)	1.88	1.91	1.92	2.18 (1.07, 4.43)	2.29	2.33	2.36
Crude	2.44	2.60	2.70	2.77	1.94	2.03	2.07	2.10
MH RR	2.57				2.04			

6.3.6 Effect of varying of specificity of initial contact status on corrected RRs

Misclassification of contact status can arise not only because of imperfect sensitivity (failure to recognise a true contact because an individual's experience of residing in the same house as a leprosy case was not observed) but also imperfect specificity. Imperfect specificity can occur if a person's observed residence in the same house as a leprosy case is in fact irrelevant, either because the apparent index case was non infectious, or because the minimum incubation period of disease is longer than the follow up period for the cohort study, in which case infection transmission at the start of a follow-up period could not manifest as disease during the study. Because of this misclassification, some individuals (apparent contacts) could be considered as having been "wrongly" classified as contacts, implying a specificity of contact status designation of less than 100%. In such situations, apparent index cases are not involved in the transmission of infection. We are thus dealing with "false positives" (false index cases or contacts).

The household dynamics model concentrated on investigation of the age/sex trend of sensitivity of contact status. Specificity deserves a more detailed similar investigation as for sensitivity than provided in this thesis. We here explore the effect of varying specificity of contact status as well as sensitivity, on the estimates of rate ratios of disease associated with contact.

Due to the thoroughness and completeness of the methods of diagnosis for leprosy used in the study (121) on which data this model was based, it is likely that very few cases were missed or misdiagnosed, and as such, high overall specificity was assumed.

It is well known that determination of bias in relative risk estimates in cohort studies is a function of sensitivity and more importantly specificity. Thus, it was important to investigate likely values of specificity in the Karonga population based on LEP data.

Among 112886 individuals in LEP-1, 2176 were diagnosed as having ever had leprosy. Table 6.3.9 below illustrates the computation of specificity (assuming 60% sensitivity and 2% leprosy prevalence). The true number of cases A+C is thus (0.02*112886=) 2258 and the number of cases correctly diagnosed as positive is (0.6*(A+C)=) 1355. Based on these figures, the specificity D/(B+D) is 99%.

Table 6.3.9 Frequency distribution of leprosy cases classified by their "true" and diagnosed status.

	"True" case		Total
Diagnosed case	Positive	Negative	
Positive	A=1355	B=821	A+B=**2176**
Negative	C=903	D=109807	C+D=110710
Total	A+C=2258	B+D=110628	A+B+C+D=**112886**

Thus, assuming an overall conservative estimate of sensitivity of 60%, the likely value for the minimum theoretical specificity was about 99% for leprosy prevalence of 1% and 2%; and 98% for a prevalence of 0.5%.

Table 6.3.10 shows observed and corrected estimates of rate ratios of disease associated with household contact, by age, for males and females separately. S_p in the table is specificity of contact status. Specificity values of 1, 0.98 and 0.95 were used, together with "forward" sensitivity of contact status values obtained earlier, in correcting estimates. Contact was considered for all (MB and PB) index cases of leprosy together.

We observe differences between the observed and corrected rate ratios after changing specificity from 1 to 0.98. For males aged 0-9 years, the observed rate ratio was 1.84. The corrected rate ratios in this age group with specificity values of 1, 0.98 and 0.95 were 1.91, 2.16 and 3.11 respectively. A similar pattern is repeated across all age groups for both males and females. The crude estimate increased from 2.44 (observed) to 5.06 for males and from 1.94 to 3.47 for females when specificity was 0.95.

Table 6.3.10 Effect of varying specificity as well as sensitivity of contact status on corrected rate ratios (for household contacts versus non-contacts)

Age group	MALES Sensitivity of contact status	RATE RATIOS Observed	Corrected		FEMALES Sensitivity of contact status	RATE RATIOS Observed	Corrected	
			S_p =0.98	S_p =0.95			S_p =0.98	S_p =0.95
0-9	0.69	1.84	2.16 (0.18, 26.58)	3.11 (0.15, 64.32)	0.66	7.08	12.73 (1.87, 86.45)	21.49 (3.25,142.09)
10-19	0.68	4.20	5.97 (2.66, 13.41)	9.46 (4.12, 21.68)	0.62	3.05	4.15 (1.54, 11.18)	6.32 (2.22, 17.97)
20-29	0.58	2.02	2.51 (0.39, 16.15)	3.74 (0.44, 32.09)	0.58	1.55	1.80 (0.40, 8.08)	2.78 (0.36, 21.68)
30-44	0.62	0.87	0.79 (0.03, 22.49)	0.0*	0.67	0.76	0.67 (0.11, 3.97)	0.22 (0.00, 48178.117)
45+	0.70	1.82	2.20 (0.71, 6.75)	3.95 (0.91, 17.24)	0.74	2.18	2.66 (1.16, 6.07)	4.40 (1.65, 11.71)
Crude		2.44	3.06	5.06		1.94	2.29	3.47

Note: *There was 1 observed case and negative (-0.09) adjusted number of cases in this age group. As such, the number of corrected cases was set to zero

6.3.7 Summary

Having estimated sensitivity of contact status from the model, the preceding sections showed results of adjusting observed RR for contact status misclassification to investigate the extent of bias from the "true" underlying (corrected) RR.

As expected, correction for sensitivity of observed contact status led to increase in estimates of contact associated rate ratios, but the effect was not great, given the numerical values involved. We noted that with a sensitivity value above 0.6 and specificity of 1, the observed and corrected rate ratios were very similar and that the extent of bias is directly related to the RR.

It is known that specificity is more important than sensitivity, in biasing assessment of relative risk estimates in cohort studies. False contacts (positives) might arise when apparent contact is with a (non-infectious) case. Such contact is irrelevant for infection transmission. Thus, it was important to get an estimate of specificity likely in such circumstances. The minimum value of specificity assuming a sensitivity of 60% and prevalence of disease of 0.5% was 98%. The lower the specificity of contact status for fixed sensitivity, the further away the corrected rate ratios were from the null. This is because we may be identifying apparent index cases who are not actually involved in the transmission of infection.

MB cases are known to be more infectious for longer than PB cases, and it has been argued that they may be responsible for most, if not all of the transmission of infection in a population. MB cases have longer incubation period than PB. The Mantel-Haenszel estimate of rate ratio of disease among contacts of MB cases adjusted for age were much higher than among contacts of PB cases, as expected.

The longer the incubation period, the further the corrected rate ratios are away from the observed. These corrected rate ratios are larger than those obtained using "forward" sensitivity of contact status showing the importance of prior unrecognised contact.

The rate ratios for household contacts relative to non-contacts are highest for young children (0-9 year olds) showing high susceptibility. This is consistent with the argument supporting age-related susceptibility. We suspect that exposure to environmental mycobacteria (which accumulates with age) provides some protection against *M. leprae*. High RRs in young children also can reflect social factors – i.e. young children have less extra-household exposure than do older individuals.

This high relative susceptibility in child contacts is consistent with (but does not confirm) genetic susceptibility. A high proportion of index cases of leprosy in households were parents of children (contacts) aged 0-9 years. As expected, this proportion decreases with age.

The proportion of individuals in contact with index cases of disease is important in determining the effect of sensitivity of contact status on observed population attributable fractions (PAFs). As sensitivity of contact status decreases, the greater is the overestimation of the PAF relative to the observed (although not much difference). With a proportion in contact of 15%, we observed a PAF of 18%, up from 12% in initial contacts.

CHAPTER 7 MODEL LIMITATIONS

This thesis describes a detailed stochastic simulation model of household dynamics. The primary purpose was to explore patterns of household change and, hence, contact status misclassification by age and sex through incorporation of selected demographic events. However, not even such a model can capture all the household changes of a rural society, and there will be many unrecognised intimate contacts in the real world, which may be better studied sociologically than mathematically.

This model is similar to other household models, in terms of basic assumptions controlling occurrence of demographic events. However, differences in assumptions exist on specific issues depending on the relevance to the society being modelled and on the intent of the underlying research. For example, household formation in other societies can occur through other means than marriage (in our model, it is mainly through marriage). The model's limitations are outlined as follows:

a) The model does not cover brief household changes or visiting e.g. for days or weeks. This could be included if data on such moves were available or could be estimated.

b) The parameters for household change were categorised by age and sex. However, we know from the LEP data that the propensity to change household also depends on membership status in household and relationship to head of household.

c) We also assumed independence in household change between individuals. However, in this as in most societies household change by parents may imply particular changes for their children. Groupings by family would have helped to model such events better. This was partly considered through the dissolution of a household when all remaining "members" under 18 years of

age were randomly allocated to other households if a head had changed household.

d) Generation of incident cases of leprosy in our model was based on age and sex patterns of leprosy incidence observed in the population. A more realistic approach would have been to generate incident cases based on infectiousness of source cases in households. This is because we expect that MB cases are more infectious than PB leprosy cases and, likewise, open pulmonary TB cases more infectious than other forms of TB. Young contacts of MB leprosy cases and sputum positive tuberculosis cases are particularly likely to become infected and hence to contract disease.

e) Although biased towards to neighbouring households, household change and allocation were random in that the model did not attempt to explore neighbour contact between households. However, in reality, people do not move randomly but are likely to preferentially visit relatives in neighbouring or distant households.

f) The extent of genuine "contacts" (population living in or close enough to source case households to become infected) is not clear. "Contact" in this study has been defined at household level. This definition is arbitrary and whether it should encompass contact at dwelling, family, casual or neighbourhood level for infection transmission is left unexplored.

g) When assigning a new head, only age and position in household were considered, and relationship to previous head and sex of an individual were not taken into account. However, this was implicitly considered in the model. When oldest "members" in households were considered for new headship, it was expected that the oldest remaining "member" would, in most cases, be the spouse of the previous head of household. In single parent households, the oldest remaining "member" would be the oldest child.

h) The LEP surveys did not record marriage partners and thus we did not model marriage explicitly. The marriage procedure used only assumed an average age difference of five years, with men older than women and the matching of partners was purely random.

i) This model did not consider transmission of infection through routes other than household contact. However, the household contact in this model encompasses shared environment, genetic susceptibility and intimacy. The influence of these various factors is difficult to disentangle. The key assumption in our model is that incidence of disease (leprosy or tuberculosis) is associated with contact.

j) In the simulation model, we have also assumed that incubation period is fixed. It is known that incubation period (e.g. of tuberculosis (134)) may depend on age at exposure and dose of the infectious agent. A more complex model would incorporate a variable incubation period. Because little is known on the distribution of the incubation period of leprosy, we assumed different fixed incubation periods.

k) Household change data used were collected during the 1980s and refer to an earlier less mobile period than the present. It would be of interest to know how these aspects of this society may have changed in recent years.

l) Based on LEP data, our model assumes no difference in household change for leprosy cases and non-cases in the Karonga population. This may not be true in other populations. Leprosy is associated with great social stigma in parts of Asia.

m) Just like many other household dynamics models, this model is limited in that it concentrated on investigating a particular issue – the age and sex trend of

contact status misclassification. Through simple modifications, application of this model could be extended to other issues in demographic research (household and population projections, and estimation of migration, mortality and fertility) in a developing country setting to determine future trends in the demographics of the population and assess their needs.

A model is by definition a simplification of reality and should not be over-interpreted. The model developed in this study has allowed us to explore how households change over time and the implications of these changes for household contact with infectious diseases.

CHAPTER 8 GENERAL DISCUSSION

This analysis of Karonga data provides a detailed description of a developing country population in Sub-Saharan Africa during the 1980s. Apart from providing input parameters for the micro-simulation model, it also provides baseline information for analyses of a variety of demographic, social, economic and epidemiological issues. Of particular importance today are questions relating to the demographic impact of HIV/AIDS which has grown into the dominant factor in this population in the years since collection of the data analysed here.

The detailed demography of Karonga District can be compared with other analyses of the Malawi population and of other Sub-Saharan countries in terms of fertility, mortality, age distribution and household mobility (45, 46, 49, 50, 79, 114). The LEP-based estimates of the age and sex distribution, mortality rates, mean household size (6.42 and 6.37 at LEP-1 and LEP-2 respectively) and heads of households (85%, 84% of heads at LEP-1 and LEP-2 respectively were males) are close to the 1987 census and 1992 MDHS results. The similarities between the results from the census, MDHS, LEP-1 and LEP-2 mutually support the reliability of each survey. Interestingly, the mean household size (4.9) for Karonga district during the 1998 census may indicate a recent decline in household size.

Computation of age as (interview/examination year - birth year) is a standard way used in many studies especially when reporting of birth dates is poor. As shown in this study, such computation leads to underestimation of infants (those under 1 year of age). To obtain precise estimates of birth rates, infant mortality and morbidity, one has to adjust for such underestimation.

Age heaping has often been attributed to digit preference at older ages in age reporting. In addition, when reporting is poor, local events calendars are often used to give approximate years of birth, and this may also result in age heaping. A

procedure for smoothing the age distribution by redistributing age of individuals with birth estimates based on those who gave precise years of birth was developed.

We note high child (0-4 year olds) mortality (259 deaths per 1000 live births in 1989 – based on indirect estimation technique) and adult (over 45 years of age) mortality. In adults, mortality rates were higher for males than females.

With the advent of HIV/AIDS, we expect the pattern of mortality by age to change with time in this population. HIV sero-prevalence during the 1990s reached more than 15% in women aged 15-45 and men 25-55 years of age (135). This is consistent with levels obtained in neighbouring countries (50). We expect mortality in young and old adults during the 1990s to be much higher than it was in the 1980s. However, this does not exclude other factors such as poverty and literacy level, which also affect mortality. HIV/AIDS is now the principal cause of adult mortality (and to a certain extent, child mortality through vertical transmission) in many African settings (46, 135-138).

We observe higher mortality in individuals with birth estimates compared to those with precise years of birth. This finding relates, in part, to social and economic status of individuals and level of education which in turn are related to poverty. Geographical patterns of mortality may reflect differential environmental factors. North Karonga is a hilly area with parts which are flooded during the rain season whereas south Karonga consists of the town centre for the district and a less hilly area with easier access to the lake and town (to supplement food and income), and to health centres. It is of interest to note that there was higher leprosy incidence in the north than south Karonga. Recently, it has been found that filariasis is also higher in the north. These geographical patterns deserve further investigations.

Not surprisingly, mortality was significantly higher in tuberculosis cases than non-cases. As more than 60% of tuberculosis cases are now HIV positive, the relative

mortality of tuberculosis cases has increased greatly above these estimates from the pre-HIV era (135).

As mortality attributable to HIV/AIDS is greatest in adults of reproductive age, this changes household structures and the age distribution of the population and leads to a rapid increase in orphanhood. The extent to which this will affect the Karonga population is yet unknown as the mortality estimates were based on pre-HIV era data.

Heads of households

The patrilineal nature of the northern region society was evident in this rural population. About 85% of heads of households at LEP-1 and LEP-2 were males, higher than the 1987 census and MDHS figure of 75%. Normally in this (rural) northern part of the country, upon marriage, a woman would leave her parental home to join her husband, as in many other Sub-Saharan African societies (46, 49, 50, 137).

One-person households

Most individuals in single person households were classified as heads, as expected. Their age distribution reveals a moderate proportion of heads under 20 years of age (16% of those aged 15-19). To what extent these figures represent orphanhood (young children staying on their own) and widowhood or separation (in old adults) is not known. However, studies within rural Sub-Saharan Africa (46, 49, 50, 137) have shown that these households are unstable (especially those which consist only of children) and more likely to be dissolved and absorbed into other households. The instability of one-person households is seen in these data (only 4.2% (40) of the single person LEP-1 households remained so at LEP-2). With the extended family system, such households especially with young children, are not really independent because they receive support from relatives in neighbouring households.

Such single individual households (although not shown by these data) could be on the increase due to high HIV/AIDS-attributable adult mortality. Although HIV/AIDS adult mortality is more pronounced in urban than rural areas of Malawi, affected families or households would normally move to their rural villages. These households become absorbed into existing households of the extended family group.

Household change and model results

Household dynamics can be analyzed using theoretical modelling techniques. Such modelling work benefits from reliable data on the long-term household dynamics situation in a given population, which are not available for many countries. Karonga was unusual in this respect, as data on households had been collected during two total population surveys.

Due to lack of data, there has been little progress in modelling household dynamics in developing country settings. The model developed in this thesis is a first attempt using LEP data, and provides relevant background for future improvements. Although this work was not solely to model household dynamics, it provides insights into such work and how the selected demographic events might contribute to infection transmission.

The data analysis and model results in this thesis make an important contribution towards the study of risk of disease among household contacts of source cases in a population. The work is particularly relevant to diseases with long incubation period for which incidence cases seen during follow-up can be attributable to unobserved contact which occurred prior to the onset of the study. Contact status misclassification can thus be high, leading to potential underestimation of RRs.

There are important age and sex patterns in household change in this rural population. Age and sex are thus important determinants of degree of contact status and, hence of infection transmission in households (40). Misclassification of contact in terms of sensitivity and specificity of observed contact status also vary as a function of age and sex. The distribution of sensitivity values with age is inversely related to the rate of household change, which is low in children but high in young adults. There was a relatively higher rate of household change in old females than males aged over 55 – probably a reflection of divorce and widowhood. Marriage and search for better employment opportunities are the most likely reasons for household change in young adults. Similar patterns have been shown in other studies in Sub-Saharan Africa (46, 49, 50, 137). Although both male and female members leave current households to get married, adult males leave at all ages to look for work and set up new households either within or outside the district.

The study has shown that the propensity of an individual to change household depends on his/her position in household. Although we found no differences in household change between leprosy cases and non-cases, if visitors or renters (a highly mobile group) are at an elevated risk of disease, this could have important implications on household contact-associated risk of disease. If (infectious) cases were common in visitors or renters (which might be so, as a result of lower socio-economic status), we expect higher frequency of household contacts, hence, higher rate of transmission and contact status misclassification.

In this population, we found that heads of households, who were less likely to change household, were twice as likely to be index cases of leprosy than "members" and non-members (includes visitors and renters) of household. Although this reflects age, as index cases were mostly adults, it also has implication on contact-associated risk of disease for individuals moving into such households.

Household change (which includes migration) also provides insights into STIs and HIV transmission. For STIs and HIV/AIDS, movements within rural areas or

between rural and urban areas may have different implications for transmission. Individuals moving within a study population, from an urban (with high HIV prevalence) to a rural area, might contribute to infection transmission in the community, especially in the sexually active age groups.

Change of household within a rural (study) area may not expose individuals to higher risks of HIV compared to out-migrating to urban areas. Female heads of households are more likely to have a partner who is deceased or at work in urban areas. Thus, when such males return (partly in-migration), their female partners are at an increased risk. Such patterns could be explored, extending the models developed here.

We expect adjustment of RRs for contact status misclassification to be particularly appropriate for young adults because of their active household change and migration. Younger people are more likely to migrate in and out of an area of exposure and, thus change risk associated with exposure. The type of area (for example, dwelling, household, village or district level) used in comparisons of rates strongly affects the extent of contact status misclassification and the amount of risk estimated (139). There was a consistently higher percentage of males out-migrating than females over the age of 15 years, in contrast to household change within the district where (young adult) females change households more frequent than males. Although there is active in-migration at older ages, the numbers are relatively small.

The longer the duration of follow-up in cohort studies investigating risk of disease associated with household contact, the greater the misclassification of contact status. Similarly, the longer the incubation period, the greater the likelihood for an actual contact to go unrecognised. Length of incubation period of disease under study affects the relative risk estimated because longer periods provide more opportunity of mixing between exposed and unexposed populations. Thus, disease seen during follow-up may be attributable to unobserved earlier contact, with source cases. In general, the low values for "backward" compared to "forward" sensitivity

are an indication of the importance of missed earlier contact. Using such a model as developed in this thesis, one can estimate the implications of movement of source cases and their contacts without having to conduct a classical cohort study over a long period of time.

Young children (0-9 year olds) in households with cases were at highest risk of leprosy, consistent with social factors or with genetic influence on susceptibility. For young children, the majority of index cases of disease in households are their parents, whereas in young adults, they are spouses (with whom they do not share any genes). For old adults, the index cases are likely to be siblings or other relatives of others in the household. Thus, index cases were mainly adults consistent with long exposure to infection and long incubation period.

In our simulation model, we assumed 100% specificity of contact status. This is because during the simulations our primary interest was to estimate the number of individuals who were ever in contact with index cases of disease. However, this assumption is likely to be invalid in certain scenarios. Misdiagnosis of cases leads to some individuals being wrongly classified as contacts. Although some individuals may have been in contact with a correctly diagnosed case at the start of follow-up, such contact may not lead to disease within the period of follow-up because of the long incubation periods. Thus, such individuals could as well be considered as having been wrongly classified as contacts. In general, the lower the specificity of contact status for fixed sensitivity, the further the corrected rate ratios were away from the null. Given that leprosy is so uncommon, it is better to work with false positive rates (% false among apparent positives) or specificity.

The discussion has touched upon a variety of issues that arose during the demographic analysis and modelling work based on pre-HIV era data, and has shown its wider implications. The value of this work is in the methods, concepts and broad results.

Patterns of mortality, migration and household change and structures presented in this thesis were pre-HIV and are likely to have changed appreciably as a consequence of high AIDS related mortality. The Karonga setting provides an opportunity to investigate selected demographic effects of HIV epidemic on rural households, which can be compared with demographic results in this thesis as baseline measure. An HIV cohort study (known as Family Health Study) following up HIV+ individuals, controls and their families is currently underway to assess the impact of HIV on the demography of the population (140). As individuals in the study were selected from the two LEP surveys and have been followed up for up to 18 years, the demographic effects of HIV may be compared to the baseline pre-HIV era data presented in this thesis.

CHAPTER 9 RECOMMENDATIONS FOR FUTURE WORK

This study began by investigating the implications of household dynamics for study of household contact associated risk of infection. It used leprosy in the context of a very large epidemiological study in Karonga District, Northern Malawi started in 1979. This investigation touched upon a variety of demographic, simulation modelling and epidemiological problems and techniques, and revealed several interesting issues which deserve further research.

There is little work on household dynamics modelling in developing countries. This work has important sociological and demographic impact on the population under study and adds to our understanding of the epidemiology of many diseases which are either transmitted in the context of households or else are responsible for household disruption. The most obvious application would be an investigation of the implication of HIV/AIDS, the most important public health problem in Africa.

KPS is ideally suited for further use in investigating the implication of HIV/AIDS on the population because of the long history of HIV data and family data. A Family Health Study following up HIV+ individuals, controls and their families is currently underway to assess the impact of HIV on the demography of the population.

Leprosy is declining in Malawi and many other countries. In Karonga it has been found that a high proportion of last cases have been from families known to include cases in the past. Studies of leprosy in households have shown increased risk of disease in contacts of known cases. Detailed studies of recent cases could be done in KPS to investigate whether this reflects genetics, environment or contact.

Studies of genetics, a major area of current interest in biomedicine, often start with family clustering although these may reflect either household contact, behaviour or

environment than genetics. The modelling work in this project might be expanded to include genetics in order to understand better the reasons for cluster of disease in families.

The importance of households for the study of tuberculosis is well recognized. However, this has not been analysed in this thesis. The KPS data provide an opportunity to conduct a detailed analysis of tuberculosis in households, as done for leprosy. A tuberculosis case-control study is being conducted investigating risk factors including genetics, contact with other tuberculosis patients, socio-economic variables, pregnancy and smoke exposure.

Such a study would provide parameters for development of a stochastic transmission model for tuberculosis, while incorporating knowledge of household dynamics, to study spread of infection in households of a developing country setting.

The analyses revealed differences in leprosy incidence, mortality and mobility by geographical zones. More studies need to be conducted on these geographic differences.

The problem of studying infectious diseases is particularly great when the incubation period is long and variable. The work presented in this thesis demonstrates some of the complicated social dynamic issues which affect efforts to study long incubation period diseases. Only one study (39) has attempted to describe the full incubation period of tuberculosis and this has never been attempted for leprosy.

There are several other implications that arise from this study, which are of relevance to leprosy, tuberculosis and other "chronic" diseases.

- It is obvious from this work that specificity of contact status plays a significant role in the estimation of household contact-associated risk of disease. In our study we assumed 100% specificity and were interested in identifying contacts of leprosy cases over the simulation period. However, some apparent contacts could be considered as having been wrongly classified as contacts, as discussed in the previous chapter, implying imperfect specificity. This issue needs further investigation similar to the one undertaken in this thesis for sensitivity of contact status.

- The movements of those people responsible for active transmission (infectious cases and their contacts) should be the focus of studies looking at household dynamics. Work in this study included estimation of the extent of household change and, hence, contact status misclassification. It would be of interest to also investigate the effect of movements between rural and urban areas (a change of disease environment) on transmission of infection or between households with an infectious case and those without. A move from one area (rural/urban) to another is often associated with a change in disease environment. Knowledge of population mobility can provide information on target groups for control and health education strategies. Many of these diseases are preventable and information that can assist in control should be recognized.

- Investigators should take steps to guard against the effects of household change and migration. High community infection transmission compared to within household infection transmission probably reflects a mobile population. There is need for a survey to look at national patterns of household change by age and sex. This can be incorporated in the National Census because such patterns do not change for long periods. Given household change patterns for several settings, it would be possible to adjust risk estimates for misclassification.

- The estimates of sensitivity of contact status obtained from the model should be accompanied by a range of scenarios in the face of uncertainty in our parameters. Variation of parameter estimates in our model was only done on household change, because we considered it the most important demographic event driving household contact misclassification.

- The effect of household change may be reduced by analysing data for only those areas where the change rates are low ensuring minimal contact status misclassfication. However, it should be noted that this can reduce the size of the population under study significantly.

- We observed low infant mortality in this population. Although not thoroughly investigated, it is well known that infant mortality in the first year of life is not uniform. It is quite high in the first few months of life than later in the year. There is need for a detailed investigation of the distribution of infant mortality in the first year of life.

APPENDICES

APPENDIX A

A.1 List of variables which were analysed in this research.

Data dictionary for variables used in this study, as extracted from LEP-1 and LEP-2 surveys.

VARIABLE	DESCRIPTION	CATEGORIES
IDENT	This is an identification number for an individual. It is a 6-digit unique identity number with a 7th as a check digit.	The format was XXXXXX-X
SEX	Whether an individual is male or female.	M for Male F for Female
BIRTHEST	Was an estimate of year of birth obtained using local events calendar in Malawi. These events were either in numerically coded (LEP-1 and LEP-2) or in letter-coded form (Health centre surveys). When local events were used, year of birth was the mid-point between years for adjacent events. Birth event used to code age, if appropriate.	1 = Before 1900 2 = Before battle of Karonga (World War I, 1914) 3 = Before major crops damage by locust in 1934 4 = Before passenger ship Viphya sank in 1946 5 = Before His Excellency Dr Hastings Kamuzu Banda returned to Malawi in 1958 6 = Before Malawi got independence in 1964 7 = At or after independence but 6 or more years ago 8 = Five or less years ago before interview.
BIRTH	Year in which an individual was born, if known or estimated from local events calendar.	
BIRTHMM	The month in which an individual interviewed was born.	1 = January 2 = February ... 12 = December
BIRTHIND	An index to indicate whether precise year of birth was given, or computed from numerically or letter coded events calendar.	1 = Precise year of birth given 2 = Birth year derived from numerically coded local events calendar 3 = Birth year derived from letter coded local events calendar
DEADLFTYR	The precise or estimated year when an individual died or left his/her household. It was left blank for those alive at the time of LEP-2 interview.	The format is the same as for BIRTH
ADL	An indicator variable that showed whether an individual was alive, dead or had left current household at time of household interview at	A = Alive and still in the household at time of interview whether seen or not D = Dead L = Left this household

	LEP-2	
AGEYRS	Age in years, computed by subtracting year of birth from year of interview. For those dead or who had left household, it was the age at either time of death or time of leaving household respectively. [Month of birth was not used because it was unreliable and very few individuals had given month of birth.]	
AGEGRP	This groups AGEYRS into 5-year age groups.	Either 0-4, 5-9, ..,15-19, ...70-74, 75+ or 0-14, 15-29, 30-39, 40-49, 50+
IDFATH	Identification number of the father of the individual interviewed.	See IDENT
IDMOTH	Identification number of the mother of the individual interviewed.	See IDENT
INTDD INTMM INTYY	Day, month and year on which the interview was actually conducted.	
HSEID	The identification number of the household in which an individual lives. A household was defined as a group of people living together and acknowledging a single individual as 'Head'.	A 5-digit household number. It was 90000 if outside LEP area.
IDHHSE	Identification number for the head of a household	See IDENT
HSEIDDL	The household for individuals who had left the household being interviewed at LEP-2. It is blank for all other individuals.	See HSEID
VIL	A 4-digit code for the village in which a household is located. These codes were the same as those used in the national census.	
POSN	A 1-digit code showing the relationship of an individual interviewed to the head of the household. For purposes of this analysis, these were grouped into 1=Head, 2=Member and 3=Other categories.	1 = Head of household 2 = member 3 = visitor less than 6 months 4 = visitor 6 months or more 5 = itinerant 6 = employed worker or servant 7 = renter 8 = patient 9 = relative of employed worker or servant
HSEYY	The year of joining current household. This may have been exact or an estimated year based on local events calendar.	See BIRTHEST
PREVHSE	The code of the previous household, where applicable, in which an individual had been living before moving into the current household.	See HSEID
PREVIL	The code of previous village, where applicable, in which the individual	See VIL

	lived before moving into current village.	
EDUCATION	Level of education (where applicable).	It was coded 1-8 representing classes at primary school, 0 if the individual is 5-18 years but not attending school, 9 for secondary/tertiary education. Otherwise it was left blank.
OCC1	Occupation of an individual. This was needed to extract schooling status.	Had codes for school going (primary and secondary) individuals. An individual could have been -a pre-school child, -at primary school, secondary school, -attending tertiary education or -not school-going though aged 5-18 years
OCCUPATION	This contains grouped occupation.	1= Fisherman and farmer 2= None and casual 3= Education 4= Salaried worker 5= Traditional and Trader
HSECONS	Housing standards of households based on construction material of dwellings.	1 = houses constructed with locally made burnt bricks, 2 = houses made with sun-dried bricks or pounded mud, 3 = houses constructed using wood or bamboo poles and interlaced twigs and rods for the walls which are later plastered with mud and 4 = temporary shelters made out of grass and other material.
VILEST	Estimated year since the household had been in the current village. It was based on local events calendar in Malawi. It is blank when VILYY is given.	See BIRTHEST
VILYY	Year since the household has been in the current village. It was missing when VILEST was given.	
LOC	This was a 10-digit number assigned from an aerial photograph of Karonga District to determine the location of a household. It was broken down into two location grid references, LOC1 and LOC2 which are 5 digits long each. The first 3 digits of LOC1 and LOC2 were used for grid references to give location to a square kilometre.	
OLD	Whether an individual was seen at LEP-1 or not. This variable was created at LEP-2.	
LEP	A variable indicating whether a record was from LEP-1 or LEP-2.	
ZONE	Karonga district was divided into 5 ecological zones on the basis of general ecological features.	A = The northern hills (hilly area with several streams) B = The northern lake shore (plains along the lake and lower Songwe river C = Southern hills (dry and sandy area) D = Semi-urban area (around the district capital) E = Southern lake shore (similar to C but hilly)
BCG SCAR	Evidence of BCG scar.	1 = Yes 2 = No 3 = Doubtful 4 = unknown

LEPROSY CASE	A variable created to indicate whether one was a case or not	1 = leprosy case 0 = non-case
LEPROSY TYPE		1 = Multibacillary 2 = Paucibacillary
DIAGNOSTIC CERTAINTY	Individuals with any evidence of leprosy were assigned any of the four categories.	1 = "Narrow" group (certain leprosy-- highest level of certainty) 2 = "Middle" group (probable leprosy) 3 = "Wide" group (possibly leprosy) 4 = "Out" group (leprosy diagnosis totally discarded)
INDEX CASE TYPE	All index cases of leprosy were also classified according to exposure-reference time (time elapsed between first LEP-1 household examination date and first registration date for leprosy).	NEW case - if the first registration date was within six months prior to LEP-1 date. OLD case - if the registration date was more than six months from LEP-1 date but earlier than 1/1/75. ANCIENT case - if the registration date was before 1/1/75.
HSIZE	Size of household based on number of individuals resident in the household at time of interview.	These were categorised into households of size 1, 2, 3-5, 6-10 and 11+.
IND	An indicator variable to show whether an individual has moved or not by LEP-2.	0= No change of household 1= Change of household
HDDEAD1	An indicator variable showing whether a head of household has either died or out-migrated or changed household in the preceding simulation year. Helps in determining need for assigning new head.	0=No such event for head 1= Head died or out-migrated or changed household

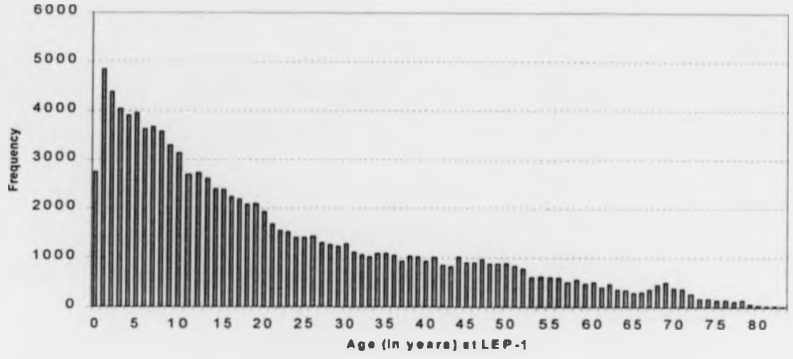
A.2 Age redistribution

The age distribution in Figure 4.1, showed apparent age heaping for older age groups, which was a reflection of birth estimates used. The age distribution of the population is skewed and, as such, the mean of an estimated period of birth may not be a good representation of the birth estimate of individuals born within that time period. This section explains the procedure for smoothing the age distribution by redistributing the ages of those with birth estimates proportionally according to those who gave exact years of birth within the same time period.

Individuals with birth year estimated according to the local events calendar were counted and their birth and age range calculated. Individuals who gave precise years of birth, and hence, had exact age were also counted. All the individuals with precise years of birth were counted according to the corresponding age ranges given in the birth estimates. This gave a total number and birth year distribution of individuals who were born within each birth estimate period. Finally, those with birth estimates were redistributed proportionally according to the age distribution of those with precise ages within the range of the birth estimate. For example, suppose we have 100 individuals who gave precise years of birth and, hence, had exact age and 30 individuals with birth estimate in the age group 50-54. The 30 individuals with birth estimates will be allocated ages based on the proportion of individuals with that exact age in the age group.

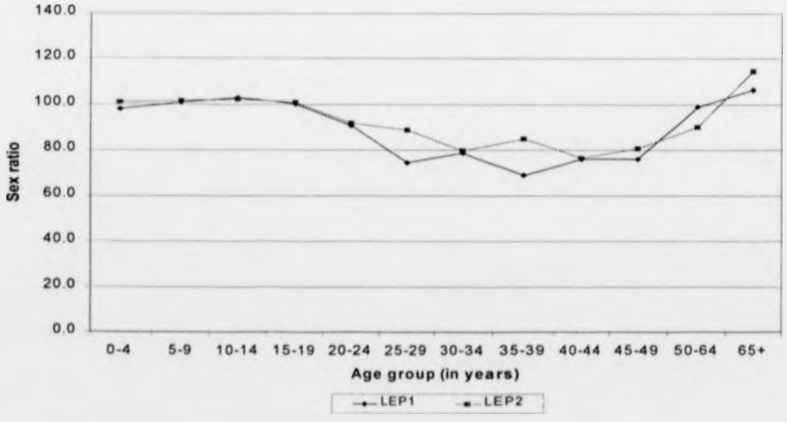
The age distribution in the figure below does not show any digit preference. There is a smooth decline with age except for a hump at about 70 years, which may reflect an artifact (i.e. old people claiming to be 70 years of age, and field staff then subtracting 70 from year of interview, and recording that figure as a "precise" birth year). The mean and median ages for the population as at LEP-1 were 21.92 and 16 years respectively.

Figure A.2.1 Age distribution after redistributing those with birth estimates proportional to those with precise years of birth in LEP-1 (1979-84) in Karonga, Malawi.



The sex ratios shown in the figure below are less erratic after age redistribution of those who gave birth estimates. The crude sex ratios for LEP-1 and LEP-2 studies after age smoothing were 93 and 95 respectively. The corresponding 1977 and 1987 census values for Karonga were 90 and 93 respectively.

Figure A.2.2 Age-specific sex ratios (male : female) during LEP-1 and LEP-2 after age redistribution of those with birth estimates. Karonga District, northern Malawi: 1979-89.



The sex ratios increase with age for older age groups. This increase may be due, in part, to errors in birth reporting by females and, hence selective misallocation of their ages. Females are less likely to give precise years and this increases with age. The extent of this misallocation depends on the proportion of individuals whose precise ages were within a birth estimate age range. The higher the proportion of individuals with precise age, the higher the number of individuals with birth estimates who would be allocated to that age hence the higher the chance of age category misallocation. This would be more marked if one were dealing with longer time periods between consecutive local events calendar used to approximate birth year.

In our age redistribution by proportional allocation we have assumed that individuals in a particular age range estimate are a random sample of that birth cohort. This, of course, is unlikely to be true – as even within an age range, the proportion with precise years of birth will probably decline with age. One could attempt to model and correct for this trend but we have considered this unnecessary in the context of this work.

These artificial ages were not used in this study, to avoid introducing age bias in the results. Bias may be introduced because individuals who had birth estimates may be allocated to one extreme of an age category whereas they actually belong to the other extreme.

APPENDIX B: Mortality during the first few years of life

During the first few years of life mortality can be represented by the function

$$l(x) = (1 + \alpha x)^{-\beta}$$

where $l(x)$ is the probability of surviving to age x, x can be age in months or years, α and β are constants which determine mortality (141). Our main interest from this analytic method is to illustrate how the values of α and β are obtained so as to extrapolate and determine level of mortality in first year of life and, hence, estimate the under-ascertainment of infants observed in Figure 4.1.

These values of α and β obtained from this estimation procedure are conservative estimates and should be treated with caution as infant mortality in the LEP study were greatly underestimated. The observed values of mortality from the LEP study are only used as an example to illustrate this method. The infant mortality rate was substituted with that from the 1992 MDHS.

The value of α is obtained from the ratio of the logarithms of the corresponding observed life tables selected as appropriate for the population under study at two ages, say 12 and 60 months, through iteration using the formula

$$Ratio = \frac{\ln l(12)}{\ln l(60)} = \frac{\ln(1 + 12\alpha)}{\ln(1 + 60\alpha)}$$

β is determined as $\frac{-\ln l(60)}{\ln(1 + 60\alpha)}$. Based on iterations shown in Figure B.1 and Table B.1, the value of α was 1.88, hence, β is 0.0455.

We then fitted the model to the LEP data as shown in Figure B.2. The figure shows the distributions of observed and fitted survival curves based on the computed values of α and β. The similarity of the distributions show that the values of α and β are reasonable.

Figure B.1 Plot of ratio of survival functions at age 12 and 60 months to obtain the value of alpha (α).

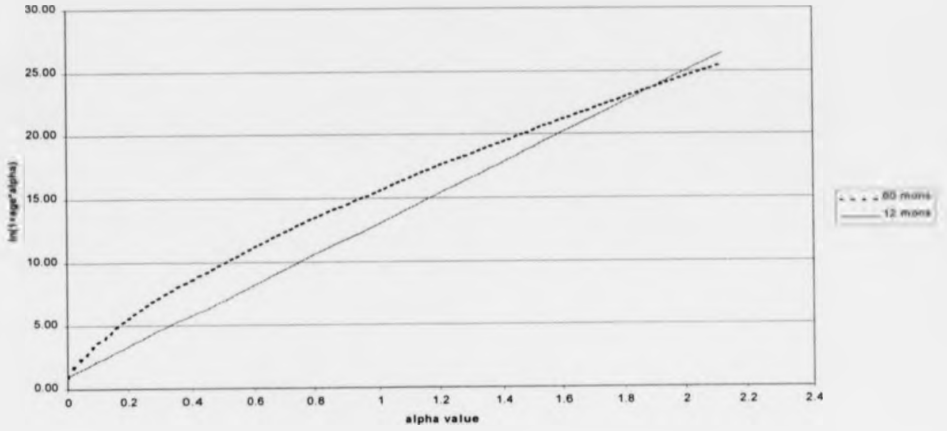


Figure B.2 Plot of observed and fitted survival curves for the first two years of life. Karonga District, northern Malawi. 1979-89.

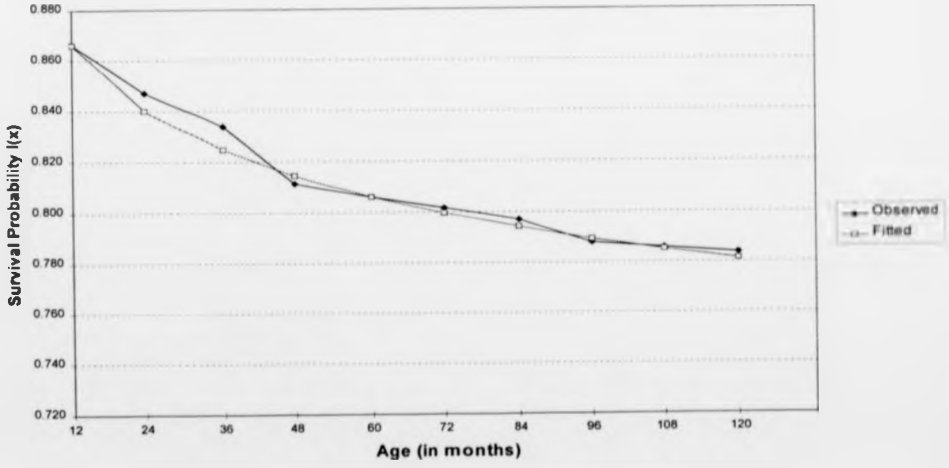


Table B.1 Computation of α and β based on mortality data for Karonga District, northern Malawi 1979-1989

Age (years)	General Mortality rate (/1000pyr)	Age (months)	Age-specific Survival	SURVIVAL Life table up to age x, l(x) (1+α*x)^(-β) Observed	Fitted
0	134	**12**	0.866	**0.866**	0.866
1	21.724	24	0.978	0.847	0.840
2	15.563	36	0.984	0.834	0.825
3	27.167	48	0.973	0.811	0.814
4	6.488	**60**	0.994	**0.806**	0.806
5	5.68	72	0.994	0.802	0.799
6	6.072	84	0.994	0.797	0.794
7	11.108	96	0.989	0.788	0.789
8	2.823	108	0.997	0.786	0.785
9	2.613	120	0.997	0.784	0.781
10	2.242				

Ln l(12) =	-0.1439	Ratio (R) =	0.6674
Ln l(60) =	-0.2156		
α=	**1.88**		
β=	**0.045533**		

Extrapolation of survival to the first year of life using the fitted model

α	(1+60*α)^R / 1+12*α 60 months	12 months	Age (in months)	(1+α*x)^(-β) Fitted l(x)	Number of children
0	1.00	1.00	0	1.000	1000
0.02	1.69	1.24	1	0.953	953
0.04	2.26	1.48	2	0.931	931
0.06	2.77	1.72	3	0.917	917
0.08	3.23	1.96	4	0.907	907
0.1	3.66	2.20	5	0.899	899
....			6	0.892	892
1.86	23.40	23.32	7	0.886	886
1.88	**23.56**	**23.56**	8	0.881	881
1.90	23.73	23.80	9	0.877	877
1.92	23.89	24.04	10	0.873	873
1.94	24.06	24.28	11	0.869	869
1.96	24.22	24.52	12	0.866	866
1.98	24.39	24.76			
2.0	24.55	25.00			

Appendix C

C.1 Under-ascertainment of infants

In order to check for under-ascertainment, it was first assumed that (a) each month of the year was equally likely to be a birth month, (b) people were interviewed at a constant rate throughout the year and were only asked their year of birth and (c) there was no mortality.

Let B_i and I_i be independent random variables representing month of birth and of interview respectively where $i = 1, 2, 3, \ldots, 12$ denoting January through December. The $Pr(B_i = i) = 1/12$ which is the same as the $Pr(I_i = i)$. If an infant was born in month B_i, then a correct age is possible if the month of interview is $I_i, I_{i+1}, I_{i+2}, \ldots, I_{12}$. Thus, the probability of a correct age for such an infant is

$$\begin{aligned} &Pr(B_i, I_i) + Pr(B_i, I_{i+1}) + Pr(B_i, I_{i+2}) + \cdots + Pr(B_i, I_{12}) \\ &= (12 - (i - 1)) \times Pr(B_i) \times Pr(I_i) \\ &= (13 - i) \times 1/12^2 \end{aligned}$$

The probability of being assigned a correct age throughout the year is the sum of these month-specific probabilities, i.e.

$$P = \sum_{i=1}^{12} (13 - i) \times (1/12^2)$$

The chance that an infant will be allocated a correct age based on month of birth and interview using the above formula is 0.542.

If we consider shorter time periods, say, days of birth and interview, the same argument holds. Births and interviews are uniformly distributed throughout the year. Thus $Pr(B_i = i) = Pr(I_i = i) = 1/365$ where $i = 1, 2, 3, \ldots, 365$.

The probability of an individual born on day i and being interviewed at least after day i is equal to

$$\int_{i=0}^{365} \int_{j=i}^{365} \frac{1}{365^2} dj\, di$$

$$= \frac{1}{365^2} \int_{i=0}^{365} (365 - i) di$$

$$= \frac{1}{2}$$

Therefore, the probability of an infant being classified in the correct age group, if there were no mortality, is 1/2.

C.2 Under-count of infants incorporating mortality

Infants are more likely to die in the first months of life than later in the year. During the first few years of life, survival can be represented by the function

$$l(x) = (1 + \alpha x)^{-\beta}$$

where $l(x)$ is the probability of surviving to age x, x can be age in months or years, α and β are constants which determine mortality (141), as given in Appendix B. Using this model one can extrapolate to estimate mortality in the first year of life.

Let i, j be the day of birth and day of interview respectively throughout the year where $0 \leq i, j \geq 365$. Age in days will be

$$x = \begin{cases} j - i & \text{if } j \geq i \\ 0 & \text{othewise} \end{cases}$$

Our function of proportion surviving to age x then becomes

$$l(x) = \begin{cases} (1 + \alpha(j - i))^{-\beta} & \text{if } j \geq i \\ 0 & \text{othewise} \end{cases}$$

Thus, taking mortality into account, the probability of an individual born on day i and interviewed at least after day i, at which time it was still alive, is then equal to

$$= \frac{1}{365^2} \int_{i=0}^{365} \int_{j=i}^{365} (1 + \alpha(j - i))^{-\beta} d_j d_i$$

$$= \frac{1}{365^2} \times \frac{1}{\alpha(1 - \beta)} \int_{i=0}^{365} (1 + \alpha(j - i))^{1-\beta} \,]_{j=i}^{365} \, d_i$$

$$= \frac{1}{365^2} \times \frac{1}{\alpha(1 - \beta)} \int_{i=0}^{365} [(1 + \alpha(365 - i))^{1-\beta} - 1] \, d_i$$

$$= \frac{1}{365^2} \times \frac{1}{\alpha(1 - \beta)} \times [\frac{1}{\alpha(2 - \beta)} \times (1 + \alpha(365 - i))^{2-\beta} \times (\frac{1}{-\alpha}) - i \,]_{i=0}^{365}$$

$$= \frac{364 + (1 + 365\alpha)^{2-\beta}}{365^2 \alpha^2 (1 - \beta)(2 - \beta)}$$

Given values of α and β, one can estimate the probability of correctly classifying age of an infant given survival till observation in the first year of life, hence, probability of their age being misclassified.

Appendix D

D.1 Adjusting for exposure misclassification in a cohort study

Let a, t_e and b, t_u be the number of observed incident cases and person times at risk in the contact (exposed) and non-contact (unexposed) groups respectively. The observed rate ratio is $RR^* = (a/t_e)/(b/t_u)$.

Let α, T_e and β, T_u be the number of true underlying incident cases and person times at risk in the contact (exposed) and non-contact groups respectively.

Let s and $s_p = 1$ be the sensitivity and specificity of contact status with which to adjust the observed incident cases and person times at risk in both groups. We assumed a specificity of 1 because our interest was in the number of individuals who were ever in contact with index cases. Due to the thoroughness and completeness of the methods of diagnosis used in the LEP study, very few cases were misdiagnosed, hence, high specificity assumed. Misdiagnosis would lead to recognising false (positive) contacts.

The observed incident cases, in terms of the true underlying incident cases, is given by $a = s \times \alpha$. Thus, the true underlying incident cases in the contact group is $\alpha = a/s$. Similarly, the true underlying person-times at risk in the contact group is $T_e = t_e/s$. If we know values of sensitivity, we can adjust our observed estimates of rate ratio for contact status misclassification.

Using the same notation, the observed number of incident cases in the non-contact group is $b = \beta + (1 - s) \times \alpha$ which implies that true underlying number of incident cases is $\beta = b - (1 - s) \times (a/s)$. Similarly, the true underlying person-times at risk in the non-contact group is $T_u = t_u - (1 - s) \times t_e/s$.

Thus, using the corrected number of incident cases and person-times at risk in both the contact and non-contact groups, the true underlying (corrected) rate ratio is

$$RR = (\alpha/T_e)/(\beta/T_u)$$

$$= \frac{\alpha T_u}{T_e \beta}$$

$$= \frac{(\frac{a}{s})(t_u - \frac{(1-s)}{s} \times t_e)}{(\frac{t_u}{s})(b - \frac{(1-s)}{s} \times a)}$$

$$= \frac{a(st_u - (1-s)t_e)}{t_e(sb - (1-s)a)}$$

If the sensitivity, $s = 1$ then the

$$RR = \frac{at_u}{t_e b} = RR^*$$

i.e. the observed rate ratio is equal to the true underlying rate ratio if there is no contact status misclassification.

D.2 Variance estimation of rate ratios

Greenland (1988) provides the same logic as above in both case-control and cohort studies but computes the variance of the rate ratios assuming non-differential misclassification of contact status.

He defines $e_1^* = \frac{a}{M}$ and $e_2^* = \frac{b}{N}$ as proportions of incident cases and person-time at risk classified in the exposed group respectively and $f_i^* = 1 - e_i^*$ as the correspondong proportion in the unexposed group where $i = 1, 2$. The true underlying proportions are e_i and f_i. $T = t_e + t_u$ and $M = a + b$ are the total person-times at risk and total number of cases respectively in both groups.

The natural logarithm of observed rate ratio is

$$L^* = ln(\frac{a \times t_u}{b \times t_e})$$

and its asymptotic variance is given by

$$V^* = \frac{1}{a} + \frac{1}{b}$$

Hence, the 95% confidence interval of the observed rate ratio is

$$exp(L^* + / - 1.96 \times sqrt(V^*))$$

The natural logarithm of corrected rate ratio will be

$$L = ln(\frac{\alpha \times T_u}{T_e \times \beta})$$

and its asymptotic variance under non-differential misclassification of exposure being

$$V = \frac{V_s(1/f_1 - 1/f_2)^2 + V_{s_p}(1/e_1 - 1/e_2)^2 + e_1^* f_1^* / (M e_1^2 f_1^2)}{s + s_p - 1}$$

where V_s and V_{s_p} are variances of s and s_p respectively. But $V_{s_p} = 0$ since $s_p = 1$. V_s is estimated from the simulation model runs based on the assumption of normality. The 95% confidence interval of the corrected rate ratio is

$$exp(L + / - 1.96 \times sqrt(V))$$

APPENDIX E: Selected schematic diagrams for procedures in the stochastic simulation model of household dynamics.

Figure E.1 Flow diagram for "Death" procedure.

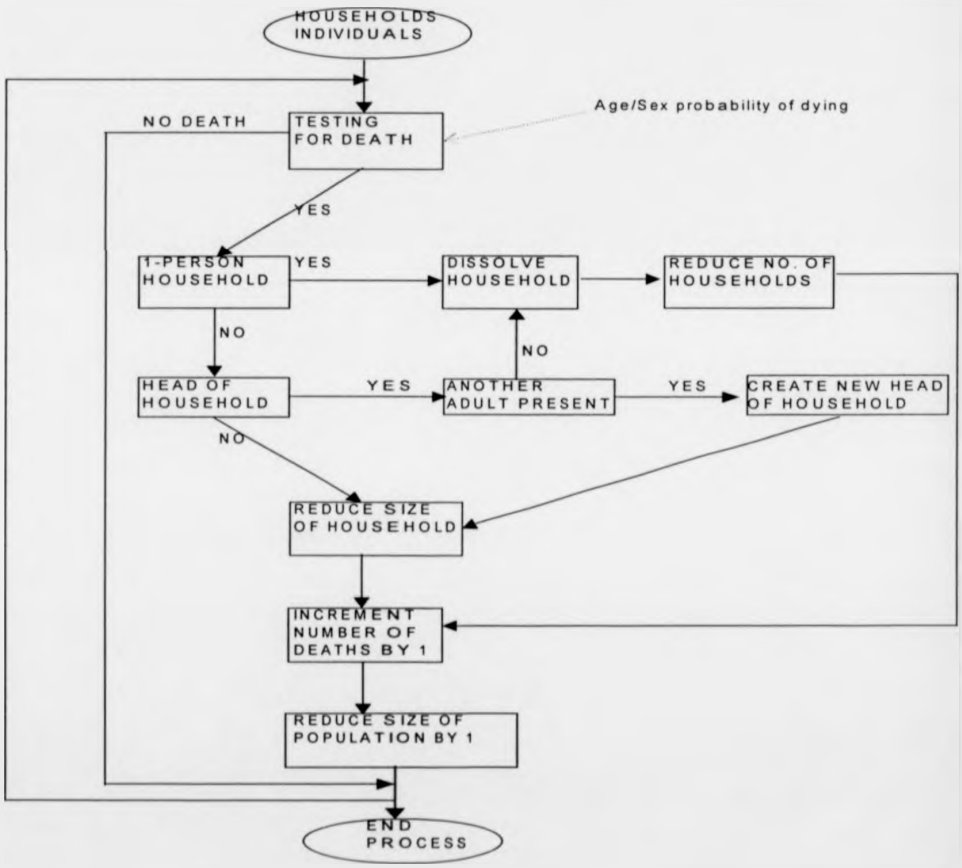


Figure E.2 Flow diagram for "Out-migration of individuals" procedure.

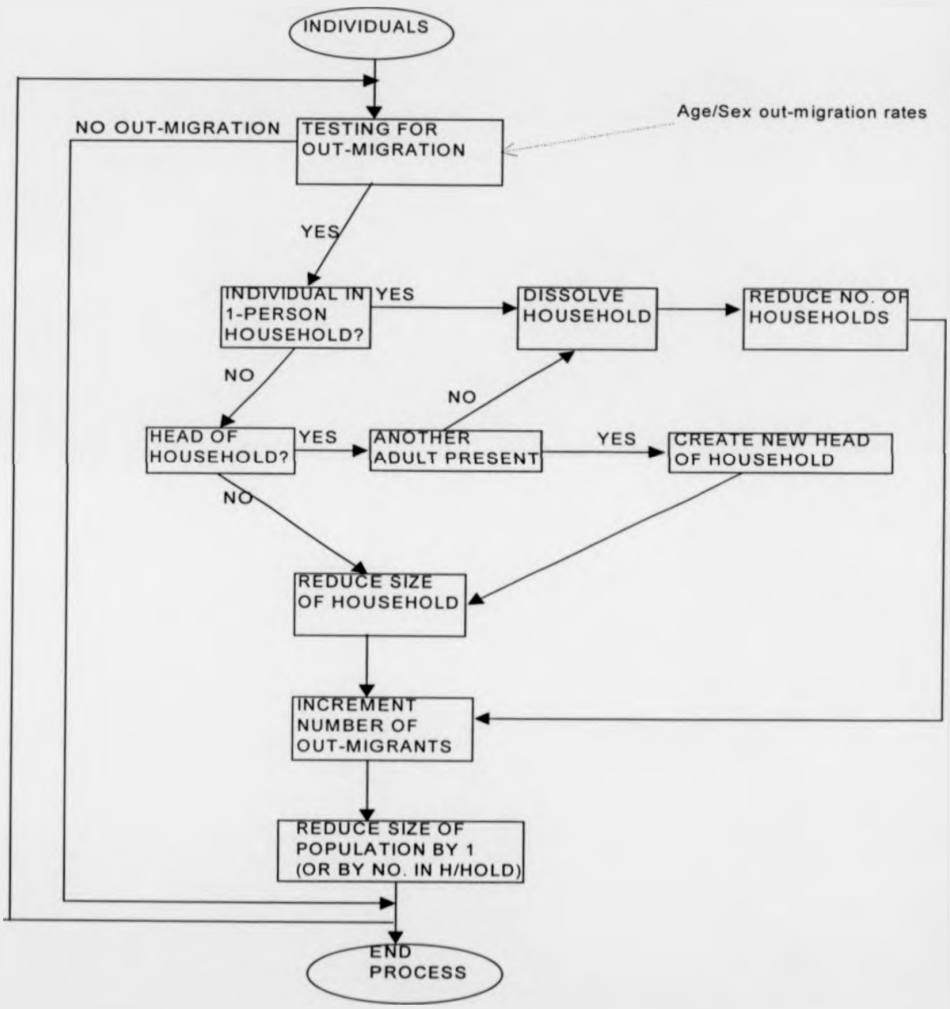


Figure E.3 Flow diagram for "Change of household" procedure.

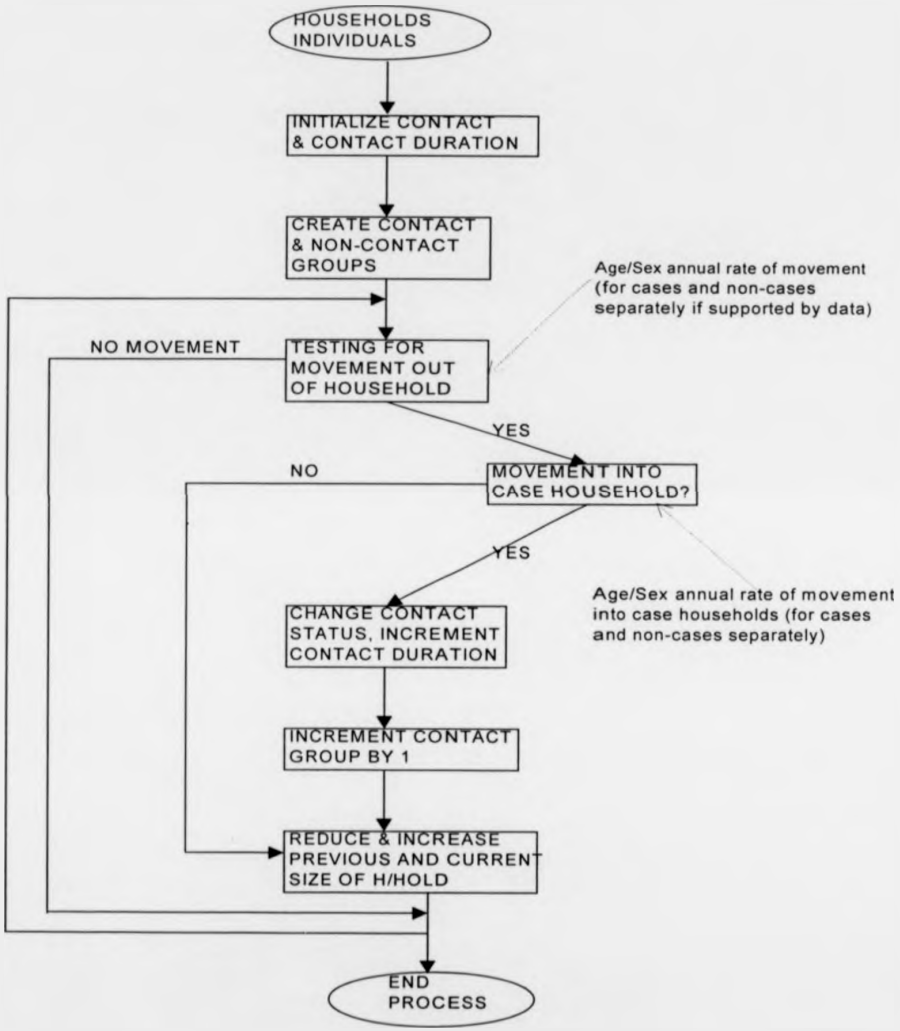
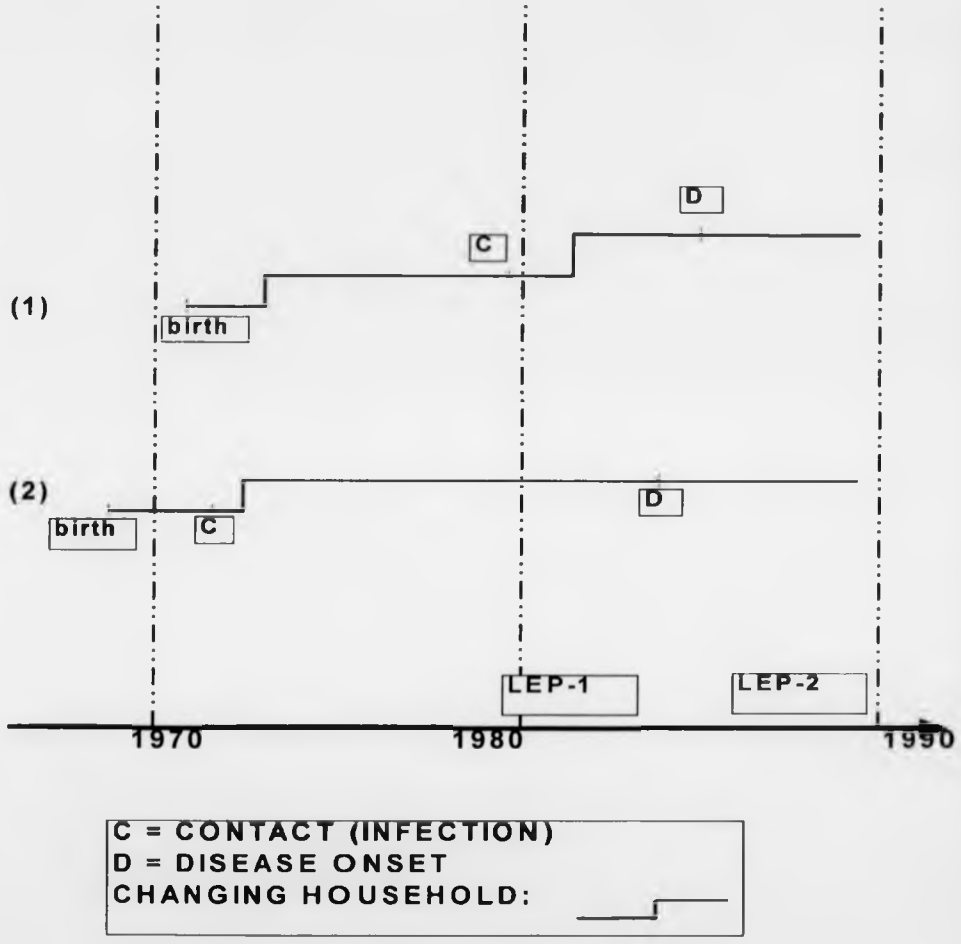


Figure E.4 Diagram illustrating implications of long incubation period and relevance of missed earlier contact with source cases of disease. For person (1) the household contact was observed in LEP-1, which led to disease onset observed in LEP-2 whereas for person (2) contact occurred earlier, but not observed in, LEP-1 but still led to disease onset during follow-up.



APPENDIX F: List of tables and figures from literature reviewed and from analyses of LEP data

Table F.1 Proportion of population under 15 years of age in some selected Sub-Saharan African countries

Country	Overall	Source
Malawi	46%; 47.4%;47.3%	Census 1987; MKAPH 1996; MDHS 1992
Uganda	49.8%;49%; 6.2%; 47.3%;47.3%;51%	Census 1980; UDHS 1988/9; Census 1990; Census 1991; ZDHS 1992; UDHS 1995
Zambia	49.8%;46.2%;47.3%; 46.8%	Census 1980;Census 1990;ZDHS 1992; ZDHS 1996
Kenya	52%; 52.5%; 49.1%	KCPS 1984; KDHS 1989; KDHS 1993
Zimbabwe	46%	ZDHS 1994
Tanzania	45.8%; 46.8%; 47.2%	Census 1988; TDHS 1991;TDHS 1996

Table F.2 Proportion of male heads of household by area

Country	Overall proportion	Rural	Urban	Source
Malawi	74.3%	73%	83.5%	MKAPH 1996
	75.4%	73.9%	87.4%	MHDS 1992
Uganda	74%	-	-	*Nakiyingi et al*
	72.2%	-	-	PHC 1991
	75.6%	76.1%	72.3%	UDHS 1995
Zambia	76.9%	75.2%	79.8%	ZDHS 1996
	83.8%	81.3%	86.9%	ZDHS 1992
Kenya	67.3%	65%	88%	KDHS 1993
Zimbabwe	67.3%	60.6%	81.4%	ZDHS 1994
South Africa		55%	69%	Based on SDHS 1996
Tanzania	70%	-	-	Census 1988
(mainland)	75.5%	83%	-	TDHS 1991
	78.2%	78.7%	76.7%	TDHS 1996

Table F.3 Mean household size by (rural or urban) area

Country	Overall	Rural	Urban	Source
Malawi	4.3	4.3	4.4	MKAPH 1996
	4.5	4.4	4.8	MDHS 1992
Uganda	4.8	-	-	Census 1991
	5.1	-	-	Nakiyingi et al
	4.8	4.8	4.2	UDHS 1995
Zambia	5.6	5.3	6.0	ZDHS 1992
	5.4	5.1	5.7	ZDHS 1996
Kenya	4.8	5.1	3.4	KDHS 1994
Zimbabwe	4.7	5.1	3.8	ZDHS 1994
Tanzania	4.2	-	-	Census 1988
	4.9	4.4	-	TDHS 1991
	4.9	5.1	4.3	TDHS 1996

Table F.4 Frequency distribution (%) of individuals changing households by age, sex and position in household. Karonga District, northern Malawi 1979-89

Age group	Head		Members		"Other" category		Total	
	Male	Female	Male	Female	Male	Female	Male	Female
0-9			20%	23%	81%	78%	15340	14839
10-14			21	44	80	89	5653	5021
15-19	22%	33%	35	62	80	89	4130	4352
20-24	7	44	54	41	84	85	2519	2617
25-29	6	31	61	27	88	85	2165	4048
30-44	4	20	51	18	87	85	5403	7742
45+	5	12	47	16	81	83	6747	7933
Total							41957	46552

Table F.5 Individual and household attributes of the starting population used in the stochastic micro-simulation of household dynamics

Attribute	Description
Individual Attributes	
Identification number	Unique to each individual
Age	Actual age in years
Sex	Male, Female
Position in household	Head, "Member" and Visitor/Renter
Leprosy case	1=yes;0=no
Observed contact status	1 if in contact with case; 0 if not. (Ultimately to be separated by MB and PB index cases of leprosy).
"True" contact status	Initialised to "observed" contact status and updated through simulation.
Household serial number	
Household Attributes	
Serial number	Unique to each household
Identification number of head of household	
Household size	1,2, 3-5, 6-10, 11+

Table F.6 Risk factors for household change between LEP-1 and LEP-2, Karonga District, northern Malawi 1979-89.

Risk Factor		Crude estimates			Estimates adjusted for age and sex only	Estimates adjusted for all risk factors in the table (including age and sex interaction)	
		Number of movers /Total	%age moving	Odds Ratio	Odds Ratio (95% CI)	Odds Ratio (95% CI)	P-value
Males (by age groups)	0-9	3275/15339	21.4	1	1	1	
	10-14	1223/5653	21.6	1.02	1.02 (0.94, 1.10)	1.05 (0.97, 1.13)	0.238
	15-19	1460/4128	35.4	2.02	2.02 (1.87, 2.17)	2.07 (1.92, 2.24)	<0.001
	20-24	1242/2518	49.3	3.59	3.59 (3.29, 3.91)	4.59 (4.17, 5.04)	<0.001
	25-29	733/2167	33.8	1.88	1.88 (1.71, 2.07)	5.67 (5.01, 6.43)	<0.001
	30-39	625/3556	17.6	0.79	0.79 (0.71, 0.86)	4.27 (3.74, 4.87)	<0.001
	40-49	333/3589	9.3	0.38	0.38 (0.33, 0.42)	3.99 (3.37, 4.72)	<0.001
	50+	391/5008	7.8	0.31	0.31 (0.28, 0.35)	3.79 (3.22, 4.47)	<0.001
Females (by age groups)	0-9	3613/14839	24.4	1	1	1	
	10-14	2256/5021	44.9	2.54	2.54 (2.37, 2.71)	2.59(2.42, 2.78)	<0.001
	15-19	2741/4353	63.0	5.28	5.28 (4.91, 5.68)	5.47 (5.08, 5.89)	<0.001
	20-24	1139/2618	43.5	2.39	2.39 (2.20, 2.61)	2.42 (2.21, 2.64)	<0.001
	25-29	1167/4046	28.8	1.26	1.26 (1.16,1.36)	1.31 (1.20, 1.42)	<0.001
	30-39	1039/4794	21.7	0.86	0.86 (0.80, 0.93)	0.84 (0.78, 0.91)	<0.001
	40-49	832/5368	15.5	0.57	0.57 (0.52, 0.62)	0.57 (0.52, 0.62)	<0.001
	50+	935/5514	17.0	0.63	0.63 (0.59, 0.69)	0.66 (0.60, 0.72)	<0.001
Leprosy case	Non-case	22591/86822	26.0	1	1	1	0.079
	Case (includes MB & PB cases)	413/1703	24.3	0.91	1.08 (0.96, 1.22)	1.12 (0.99, 1.28)	
Leprosy case type*	Non-case	22591/86822	26.0	1	1	1	
	New	123/562	21.9	0.80	0.95 (0.77, 1.18)	1.02 (0.81, 1.27)	0.889
	Old	139/494	28.1	1.11	0.99 (0.80, 1.22)	1.09 (0.88, 1.36)	0.433
	Ancient	125/595	21.0	0.76	1.18 (0.95, 1.45)	1.23 (0.98, 1.53)	0.077
Position in household	Head	869/14138	6.2	1	1	1	
	Member	20822/72571	28.4	6.06	8.33 (7.52, 9.23)	8.97 (8.03, 10.02)	<0.001
	Others	1513/1818	83.3	76.31	103.69 (88.57, 121.39)	89.31 (75.27, 105.98)	<0.001
Size of household	1	126/667	18.9	1	1	1	
	2	502/2221	22.6	1.25	0.66 (0.52, 0.83)	0.29 (0.23, 0.36)	<0.001
	3-5	4002/19560	20.5	1.10	0.59 (0.48, 0.73)	0.19 (0.16, 0.24)	<0.001
	6-10	8581/37083	23.1	1.29	0.72 (0.58, 0.88)	0.20 (0.16, 0.25)	<0.001
	11+	8615/26372	32.7	2.08	1.12 (0.91, 1.38)	0.29 (0.23, 0.35)	<0.001
Geographical zone	Northern Hills (zone A)	1299/5879	22.1	1	1	1	
	Northern lake shore (zone B)	8740/36036	24.3	1.13	1.14 (1.06, 1.22)	1.27 (1.18, 1.37)	<0.001
	Semi-urban area (zone D)	2315/8629	26.8	1.29	1.27 (1.17, 1.37)	1.31 (1.20, 1.43)	<0.001
	Southern Hills (zone C)	5786/21816	26.5	1.27	1.27 (1.19, 1.37)	1.37 (1.27, 1.48)	<0.001
	Southern lake shore (zone E)	4864/16163	30.1	1.52	1.50 (1.40, 1.62)	1.56 (1.45, 1.69)	<0.001
BCG Scar	Yes	10840/34507	31.4	1	1	1	
	No scar	9671/43775	22.1	1.82	1.02 (0.98, 1.06)	1.02 (0.98, 1.07)	0.257

Note: *Variable 'Leprosy case type' was considered in a separate full model of all risk factors from the one where 'leprosy case' was fitted. The estimates in the table above are adjusted for 'leprosy case' but not 'leprosy case type'.

Table F.7 Incidence rates of leprosy by age and sex for the 10 year period of simulation
Note: Values at year 10 were rates as observed at LEP-2. These rates were used as parameters for generating incident cases in the 10-year period simulation model.

50% incidence decline = 0.077016

Formula used = 2*(observed rate at year 10)*exponential(-(constant decline rate)*(year-1))

(a) Females

Year	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-44	45-54	55-64	65+
1	0.00009	0.00080	0.00238	0.00255	0.00270	0.00216	0.00423	0.00400	0.00446	0.00468	0.00367
2	0.00008	0.00074	0.00220	0.00236	0.00250	0.00200	0.00392	0.00370	0.00413	0.00433	0.00340
3	0.00008	0.00069	0.00204	0.00219	0.00232	0.00186	0.00363	0.00343	0.00382	0.00401	0.00315
4	0.00007	0.00063	0.00189	0.00203	0.00215	0.00172	0.00336	0.00317	0.00354	0.00371	0.00292
5	0.00007	0.00059	0.00175	0.00188	0.00199	0.00159	0.00311	0.00294	0.00327	0.00344	0.00270
6	0.00006	0.00054	0.00162	0.00174	0.00184	0.00147	0.00288	0.00272	0.00303	0.00318	0.00250
7	0.00006	0.00050	0.00150	0.00161	0.00170	0.00136	0.00266	0.00252	0.00281	0.00295	0.00231
8	0.00005	0.00047	0.00139	0.00149	0.00158	0.00126	0.00247	0.00233	0.00260	0.00273	0.00214
9	0.00005	0.00043	0.00128	0.00138	0.00146	0.00117	0.00228	0.00216	0.00241	0.00253	0.00198
10	0.00005	0.00040	0.00119	0.00128	0.00135	0.00108	0.00212	0.00200	0.00223	0.00234	0.00184

(b) Males

Year	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-44	45-54	55-64	65+
1	0.000000	0.000576	0.001844	0.003242	0.001466	0.002034	0.002576	0.002132	0.004070	0.002634	0.003654
2	0.000000	0.000533	0.001707	0.003002	0.001357	0.001883	0.002385	0.001974	0.003768	0.002439	0.003383
3	0.000000	0.000494	0.001581	0.002779	0.001257	0.001744	0.002208	0.001828	0.003489	0.002258	0.003132
4	0.000000	0.000457	0.001464	0.002573	0.001164	0.001614	0.002045	0.001692	0.003230	0.002091	0.002900
5	0.000000	0.000423	0.001355	0.002382	0.001077	0.001495	0.001893	0.001567	0.002991	0.001936	0.002685
6	0.000000	0.000392	0.001255	0.002206	0.000997	0.001384	0.001753	0.001451	0.002769	0.001792	0.002486
7	0.000000	0.000363	0.001162	0.002042	0.000924	0.001281	0.001623	0.001343	0.002564	0.001659	0.002302
8	0.000000	0.000336	0.001076	0.001891	0.000855	0.001186	0.001502	0.001244	0.002374	0.001536	0.002131
9	0.000000	0.000311	0.000996	0.001751	0.000792	0.001098	0.001391	0.001151	0.002198	0.001422	0.001973
10	0.000000	0.000288	0.000922	0.001621	0.000733	0.001017	0.001288	0.001066	0.002035	0.001317	0.001827

Table F.8 Age assignment using random uniform (0,1)

Outcome of random uniform (0,1) number	Age Groups				
	0	1-4	5-year age groups	30-44	45+
	Exact age				
< 0.3		1			
0.3-0.5		2			
0.5-0.7		3			
>0.7		4			
<0.2			Lower bound		
0.2-0.4			Lower bound+1		
0.4-0.6			Lower bound+2		
0.6-0.8			Lower bound+3		
>0.8			Lower bound+4		
<0.2				30	
0.2-0.3				35	
0.3-0.4				40	
>0.4				44	
<0.5					47
0.5-0.7					52
0.7-0.9					57
>0.9					60

Table F.10 Frequency distribution of individuals by age, sex and position in one-person households during LEP-1 and LEP-2 surveys.

Position in household	LEP-1 frequency (row %)				LEP-2 frequency (row %)			
	Female		Male		Female		Male	
	< 40 years	>=40 years	< 40 years	>=40 years	< 40 years	>=40 years	< 40 years	>=40 years
Head	36 (11)	292	179 (30)	417	43 (9)	413	317 (36)	558
Member	6 (30)	14	1 (33)	2	7 (29)	17	0	0
Servant	1 (100)	0	1 (100)	0	0	0	0	0
Total	43	306	181	419	50	430	317	558

Table F.11 Mean and median age for heads of households by sex during LEP-1 and LEP-2 surveys.

Head of Household	Mean Age (Std. Dev.)		Median Age (IQR)	
	LEP-1	LEP-2	LEP-1	LEP-2
Male	47 (14.6)	46 (15.1)	46 (35-58)	45 (34-58)
Female	55 (14.2)	57 (14.0)	57 (46-62)	61 (47-64)
Overall	48 (14.8)	48 (15.4)	47 (36-58)	47 (35-61)

Table F.12a Breakdown of all male "members" by relationship to head of household during LEP-1 survey. The table shows the number in each category and column percent.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Mother	Father	Unrelated	Other relations	Total
0-14	12065	54	515	1835	-	0	9459	535	24463
	75.2%	13.8%	55.8%	85.6%		0%	61.8%	58.1%	
15-29	3633	218	363	302	-	0	4801	313	9630
	22.7%	55.9%	39.3%	14.1%		0%	31.4%	34.0%	
30-44	301	69	41	6	-	0	721	37	1175
	1.9%	17.7%	4.4%	0.3%		0%	4.7%	4.0%	
45+	40	49	4	0	-	21	316	36	466
	0.2%	12.6%	0.4%	0%		100%	2.1%	3.9%	
Total	16039	390	923	2143	-	21	15297	921	35734

Table F.12b Breakdown of all female "members" by relationship to head of household during LEP-1 survey. The table shows the number in each category and column percent.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Mother	Father	Unrelated	Other relations	Total
0-14	11658	54	486	1824	0	-	9560	522	24104
	83.1%	13.3%	68.3%	89.3%	0%		27.6%	59.6%	
15-29	2091	80	189	210	0	-	10682	165	13417
	14.9%	19.7%	26.5%	10.3%	0%		30.9%	18.8%	
30-44	230	75	25	8	8	-	7988	62	8396
	1.6%	18.5%	3.5%	0.4%	1.4%		23.1%	7.1%	
45+	57	197	12	0	564	-	6352	127	7309
	0.4%	48.5%	1.7%	0%	98.6%		18.4%	14.5%	
Total	14036	406	712	2042	572	-	34582	876	53226

Table F.13a Percentage household change for male "members" broken down by relationship to head of household. The table shows the number (and percent) who changed household in each category and total.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Mother	Father	Unrelated	Other relations	Total
0-14	1087	17	178	545	-	-	2042	181	4050
	(10.7%)	(46.0%)	(42.4%)	(38.3%			(26.7%)	(43.2%)	(20.2%)
	10120	37	420	1424			7644	419	20064
15-29	948	93	149	117	-	-	1598	132	3037
	(34.0%)	(64.1%)	(58.4%)	(54.2%)			(48.7%)	(57.4%)	(43.9%)
	2791	145	255	216			3285	230	6922
30-44	119	35	24	2	-	-	266	15	461
	(46.9%)	(67.3%)	(70.6%)	(40.0%)			(53.3%)	(57.7%)	(53.0%)
	254	52	34	5			499	26	870
45+	9	17	3	-	-	3	115	18	165
	(25.0%)	(46.0%)	(75%)			(33.3%)	(55.6%)	(69.2%)	(51.7%)
	36	37	4			9	207	26	319

Table F.13b Percentage household change for female "members" broken down by relationship to head of household. The table shows the number (and percent) who changed household in each category and total.

Age group	Children	Sibling	Niece/ Nephew	Grand children	Mother	Father	Unrelated	Other relations	Total
0-14	1647	26	203	630	-	-	2675	195	5376
	(17.5%)	(68.4%)	(56.9%)	(45.3%)			(36.3%)	(50.1%)	(28.4%)
	9395	38	357	1391			7368	389	18938
15-29	988	38	106	115	-	-	3218	93	4556
	(65.0%)	(77.6%)	(80.9%)	(73.3%)			(38.6%)	(83.8%)	(44.2%)
	1520	49	131	157			8338	111	10306
30-44	78	31	6	2	0	-	1117	29	1263
	(44.6%)	(56.4%)	(35.3%)	(33.3%)	(0%)		(16.7%)	(64.4%)	(18.0%)
	175	55	17	6	5		6709	45	7012
45+	12	51	4	-	76	-	769	33	945
	(25.5%)	(37.2%)	(40.0%)		(22.0%)		(14.4%)	(47.8%)	(15.9%)
	47	137	10		345		5336	69	5944

Figure F.1 Percentage of individuals changing households between LEP-1 and LEP-2, by size of household, Karonga District, northern Malawi 1979-89.

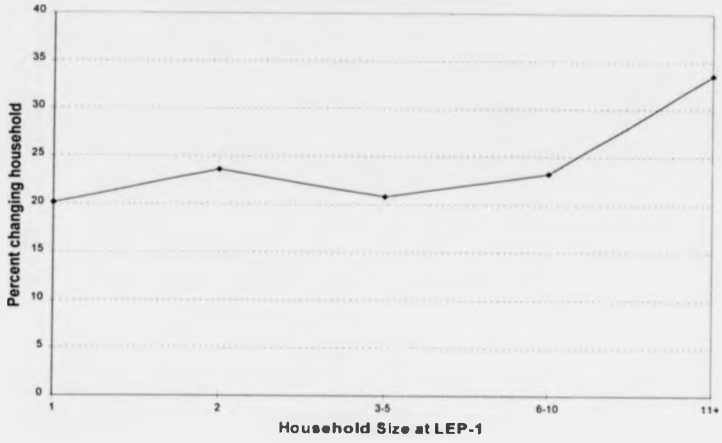


Figure F.2 Relative frequency distribution of household size for Karonga District, northern Malawi during LEP-2 survey (1986-89).

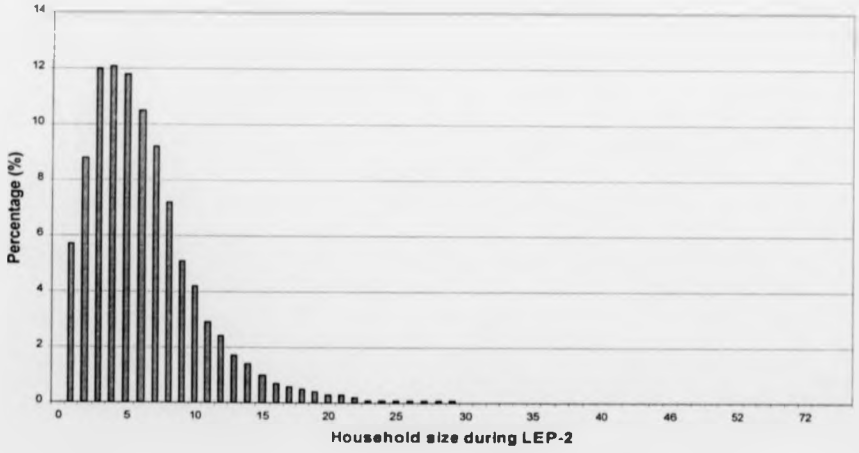
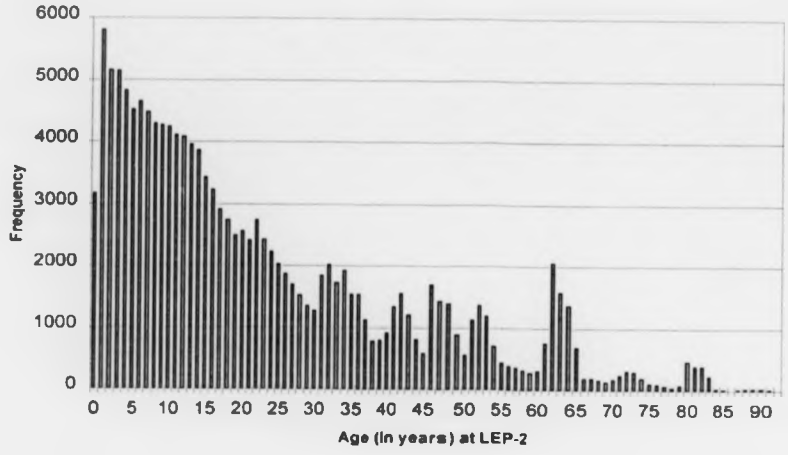


Figure F.3 Age distribution of population of Karonga District, northern Malawi during 1986-89 survey (LEP-2).



APPENDIX G: Selected list of tables and figures from simulation results.

Table G.1a Observed incidence rates of disease for male household contacts of index cases of leprosy, by age and BCG scar status.

Age group	BCG scar negative				BCG scar positive				Observed contact-associated RRs	
	Cases/Total person years		Incident rate (per 1000 pyar)		Cases /Total person years		Incident rate (per 1000 pyar)			
	Contacts	Non-contacts	Contacts	Non-contacts	Contacts	Non-contacts	Contacts	Non-contacts	BCG-	BCG+
0-9	1/2337	4/24928	0.428	0.160	0/2647	2/30025	0.0	0.067	2.67	0.0
10-19	6/1990	24/21224	3.015	1.131	7/2583	8/26006	2.710	0.308	2.67	8.81
20-29	2/564	17/6593	3.546	2.578	0/1252	7/13613	0.0	0.514	5.84	0.0
30-44	1/1139	17/20031	0.878	0.849	0/172	3/2740	0.0	1.095	1.03	0.0
45+	5/2171	37/29979	2.303	1.234	0/188	0/1864	0.0	0.0	1.87	-
Crude	15/8201	99/102755	1.829	0.963	7/6842	20/74248	1.023	0.269	1.90 (1.10, 3.26)	3.80
MH									1.94 (1.08, 3.41)	

Table G.1b Observed incidence rates of disease for female household contacts of index cases of leprosy, by age and BCG scar status.

Age group	BCG scar negative				BCG scar positive				Observed contact-associated RRs	
	Cases/Total person years		Incident rate (per 1000 pyar)		Cases /Total person years		Incident rate (per 1000 pyar)			
	Contacts	Non-contacts	Contacts	Non-contacts	Contacts	Non-contacts	Contacts	Non-contacts	BCG-	BCG+
0-9	4/2506	5/25613	1.596	0.195	0/2536	1/27957	0.0	0.036	8.18	0.0
10-19	7/2095	16/20821	3.341	0.768	1/2320	11/24688	0.431	0.446	4.35	0.97
20-29	2/1108	19/14207	1.805	1.337	1/1105	6/14289	0.905	0.420	1.35	2.16
30-44	3/2410	51/31572	1.245	1.615	0/295	1/4219	0.0	0.237	0.77	0.0
45+	9/2882	50/35022	3.123	1.428	0/198	0/2236	0.0	0.0	2.19	-
Crude	25/11001	141/127235	2.27	1.11	2/6454	19/73389	0.310	0.260	2.05 (1.34, 3.13)	1.20
MH									2.15 (1.41, 3.26)	

Table G.2 Frequency table for changes in position between LEP-1 and LEP-2 for surviving residents, Karonga District, northern Malawi 1979-89. The percentages of the total who stayed in the same household or changed are given in parenthesis.

		LEP-2 Household and position							
		Same Household				New Household			
		Head	Member	Others	Total	Head	Member	Others	Total
LEP-1 Position	Head	13084 (98.65%)	176 (1.33%)	2 (0.02%)	13262	272 (31.3%)	467 (53.74%)	130 (14.96%)	869
	Member	1178 (2.27%)	50603 (97.41%)	168 (0.32%)	51949	3055 (14.82%)	16308 (79.08%)	1258 (6.10%)	20619
	Others	6 (1.98%)	235 (77.56%)	62 (20.46%)	303	275 (18.18%)	1114 (73.63%)	124 (8.2%)	1513
TOTAL		14268	51014	232	65514	3602	17887	1512	23001

Table G.3 Observed incidence rates of disease for household contacts of MB and PB cases, by age.

Age group	Household contact with MB cases: Cases/Total person years		Incident rate (per 1000 pyr)		PB cases: Cases /Total person years		Incident rate (per 1000 pyr)		Non-contacts: Cases/Total person years		Incidence rate (per 1000 pyr)	
	M	F	M	F	M	F	M	F	M	F	M	F
0-9	1/288	1/331	3.47	3.02	0/4696	3/4711	0.00	0.64	6/54953	6/53570	0.11	0.11
10-19	2/280	1/268	7.14	3.73	11/4293	7/4147	2.56	1.69	32/47230	27/45509	0.68	0.59
20-29	1/143	0/112	6.99	0.00	1/1673	3/2101	0.60	1.43	11/20206	25/28496	0.54	0.88
30-44	0/62	0/204	0.00	0.00	1/1249	3/2501	0.80	1.20	20/22771	52/35791	0.88	1.45
45+	1/101	0/221	9.90	0.00	4/2258	9/2859	1.77	3.15	37/31843	50/37258	1.16	1.34
Crude	5/874	2/1136	5.72	1.76	17/14169	25/16319	1.20	1.53	106/177003	160/200624	0.60	0.80

Table G.4a Observed and corrected (for contact status misclassification) rate ratios, and their confidence intervals, for household contacts of multibacillary cases.

Age group	Observed rate ratio (95% c.i.) Males	Females	Corrected rate ratio (95% c.i.) Males	Females
0-9	31.80 (3.83, 264.16)	26.97 (3.25, 224.0)	34.32 (3.47, 339.15)	29.10 (2.95, 287.55)
10-19	10.54 (2.53, 43.99)	6.29 (0.85, 46.28)	10.83 (2.49, 47.22)	6.38 (0.84, 48.69)
20-29	12.85 (1.66, 99.5)	-	13.70 (1.53, 122.90)	-
30-44	-	-	-	-
45+	8.52 (1.17, 62.11)	-	8.61 (1.15, 64.18)	-
Crude	9.55 (3.90, 23.43)	2.21 (0.55, 8.90)	9.75 (4.66, 20.40)	2.15 (0.68, 6.82)
MH	10.62 (4.31, 26.16)	2.25 (0.56, 9.10)	10.88 (5.16, 22.92)	2.16 (0.68, 6.87)

Table G.4b Observed and corrected (for contact status misclassification) rate ratios, and their confidence intervals, for household contacts of paucibacillary cases.

Age group	Observed rate ratio (95% c.i.) Males	Females	Corrected rate ratio (95% c.i.) Males	Females
0-9	-	5.69 (1.42, 22.73)	-	7.32 (1.13, 47.56)
10-19	3.78 (1.91, 7.50)	2.85 (1.24, 6.53)	4.53 (1.90, 10.84)	3.20 (1.19, 8.61)
20-29	1.10 (0.14, 8.50)	1.63 (0.49, 5.39)	1.10 (0.12, 9.88)	1.69 (0.45, 6.26)
30-44	0.91 (0.12, 6.79)	0.83 (0.26, 2.64)	0.91 (0.12, 7.13)	0.82 (0.25, 2.72)
45+	1.52 (0.54, 4.28)	2.35 (1.15, 4.77)	1.55 (0.53, 4.52)	2.44 (1.14, 5.20)
Crude	2.00 (1.20, 3.34)	1.92 (1.26, 2.93)	2.12 (1.38, 3.26)	1.96 (1.37, 2.80)
MH	2.10 (1.25, 3.52)	2.03 (1.33, 3.09)	2.28 (1.47, 3.54)	2.12 (1.48, 3.04)

Note: Estimates of rate ratios (-) could not be made for 20-45 year old female and 30-44 year old male contacts of MB cases and 0-9 year old male contacts of PB cases because no incident cases were identified.

Table G.5a Observed and corrected (for contact status misclassification) rate ratios, and their confidence intervals, for household contacts of "old" cases.

Age group	Observed rate ratio (95% c.i.)		Corrected rate ratio (95% c.i.)					
	Males	Females	Males			Females		
			Incubation period (in years)			Incubation period (in years)		
			5	7	9	5	7	9
0-9	1.84 (0.22, 15.26)	7.08 (1.20, 25.10)	1.98	2.05	2.11	16.19	35.58	-
10-19	4.20 (2.20, 7.99)	3.05 (1.39, 6.72)	5.76	7.06	8.33	3.82	4.59	5.29
20-29	2.02 (0.45, 9.13)	1.55 (0.47, 5.12)	2.21	2.35	2.53	1.60	1.64	1.66
30-44	0.87 (0.12, 6.47)	0.76 (0.24, 2.44)	0.86	0.86	0.86	0.76	0.75	0.75
45+	1.82 (0.72, 4.64)	2.18 (1.07, 4.43)	1.90	1.94	1.95	2.33	2.39	2.42
Crude	2.44	1.94						
MH	2.57	2.04						

Table G.5b Observed and corrected (for contact status misclassification) rate ratios, and their confidence intervals, for household contacts of "new" cases.

Age group	Observed rate ratio (95% c.i.)		Corrected rate ratio (95% c.i.)					
	Males	Females	Males			Females		
			Incubation period (in years)			Incubation period (in years)		
			5	7	9	5	7	9
0-9	1.84 (0.22, 15.26)	7.08 (1.20, 25.10)	2.22	3.60	0.31	-	-	0.23
10-19	4.20 (2.20, 7.99)	3.05 (1.39, 6.72)	19.85	-	0.60	6.00	-	0.19
20-29	2.02 (0.45, 9.13)	1.55 (0.47, 5.12)	2.61	13.91	0.69	1.68	1.98	5.05
30-44	0.87 (0.12, 6.47)	0.76 (0.24, 2.44)	0.86	0.84	-	0.74	0.72	0.64
45+	1.82 (0.72, 4.64)	2.18 (1.07, 4.43)	2.11	2.93	0.64	2.77	5.90	0.58
Crude	2.44	1.94						
MH	2.57	2.04						

Table G.6 Forward sensitivity of contact status, and their 95% confidence intervals, for males and females by age

Age Group	Sensitivity of contact status: Mean M	Mean F	Median M	Median F	95% crude confidence limits: Males	Females	Relative precision (from +/- 20% mean annual household change): Males	Females
0	0.62	0.60	0.62	0.60	(0.45, 0.78)	(0.47, 0.74)	(0.56, 0.64)	(0.55, 0.65)
1-4	0.66	0.66	0.66	0.66	(0.54, 0.75)	(0.58, 0.73)	(0.62, 0.70)	(0.62, 0.70)
5-9	0.73	0.68	0.73	0.68	(0.66, 0.81)	(0.60, 0.76)	(0.70, 0.76)	(0.65, 0.74)
10-14	0.72	0.65	0.72	0.65	(0.65, 0.79)	(0.58, 0.72)	(0.69, 0.76)	(0.62, 0.69)
15-19	0.62	0.58	0.62	0.58	(0.53, 0.72)	(0.50, 0.67)	(0.59, 0.67)	(0.54, 0.63)
20-24	0.59	0.61	0.58	0.61	(0.52, 0.66)	(0.54, 0.68)	(0.57, 0.63)	(0.57, 0.67)
25-59	0.56	0.52	0.56	0.52	(0.45, 0.67)	(0.42, 0.64)	(0.53, 0.60)	(0.50, 0.58)
30-44	0.62	0.67	0.61	0.66	(0.54, 0.73)	(0.60, 0.76)	(0.59, 0.67)	(0.62, 0.70)
45+	0.70	0.74	0.70	0.74	(0.63, 0.77)	(0.68, 0.81)	(0.66, 0.74)	(0.71, 0.78)

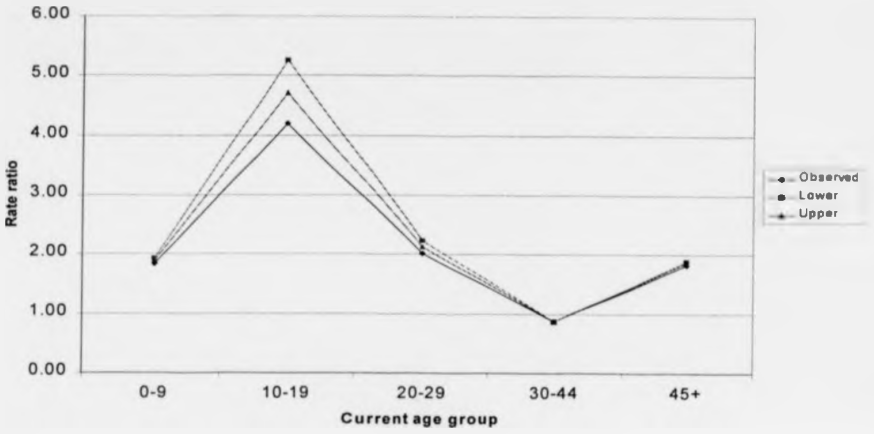
Note: Lower and upper (under relative precision) limits are average sensitivity of contact status obtained using upper and lower relative annual household change respectively.

Table G.7 Forward sensitivity of contact status, and their 95% confidence intervals, for males and females by age, broken down as in the contact analysis paper (19).

Age Group	Sensitivity of contact status: Mean M	Mean F	Median M	Median F	95% confidence intervals (assuming normality): Males	Females	95% crude confidence limits based on ordering: Males	Females
0-9	0.69	0.66	0.69	0.66	(0.68, 0.69)	(0.65, 0.66)	(0.62, 0.75)	(0.60, 0.71)
10-19	0.68	0.62	0.68	0.62	(0.67, 0.68)	(0.61, 0.62)	(0.62, 0.75)	(0.56, 0.69)
20-29	0.58	0.58	0.57	0.58	(0.57, 0.58)	(0.58, 0.59)	(0.51, 0.64)	(0.53, 0.63)
30-44	0.62	0.67	0.61	0.66	(0.61, 0.63)	(0.66, 0.67)	(0.54, 0.73)	(0.60, 0.76)
45+	0.70	0.74	0.70	0.74	(0.70, 0.71)	(0.73, 0.75)	(0.63, 0.77)	(0.68, 0.81)

Figure G.1 Corrected rate ratios of household contact for (a) males and (b) females obtained using crude 95% confidence limits compared to observed rate ratios, by age.

(a) Males



(b) Females

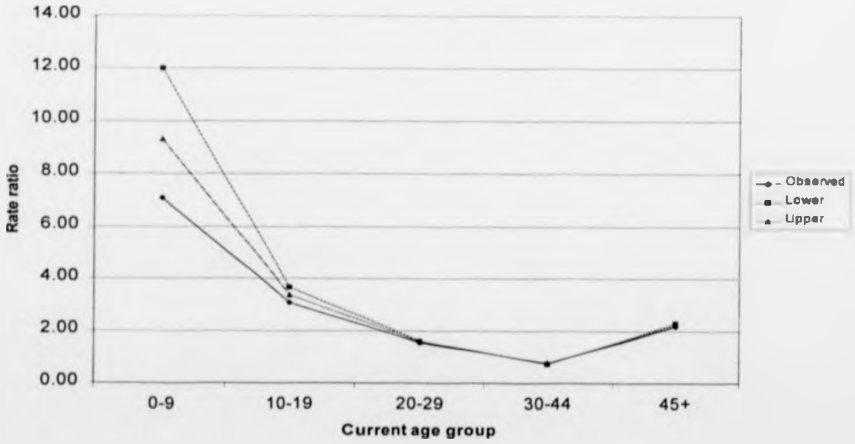


Figure G.2 Forward sensitivity of contact status for females by age and duration (in years) of follow-up.

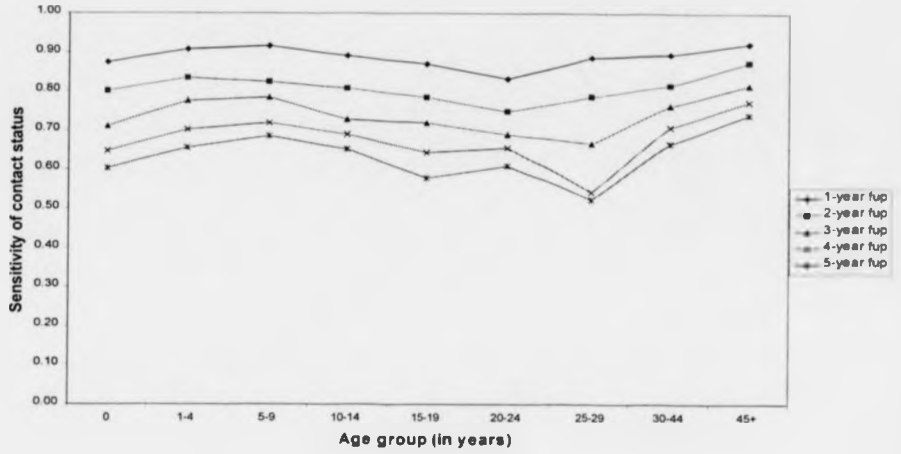
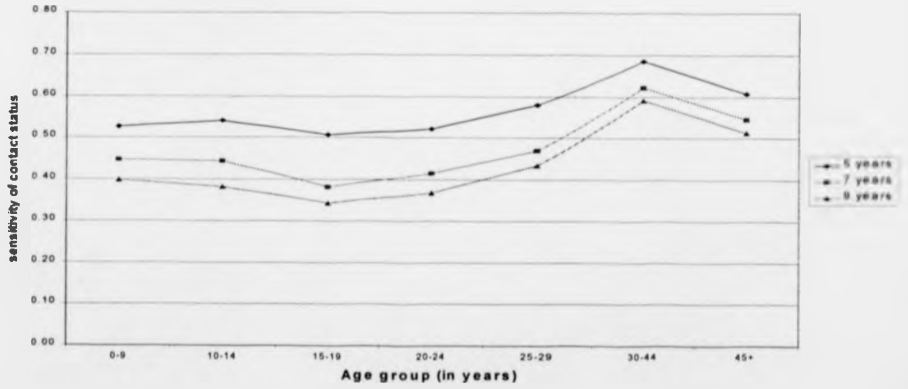


Figure G.3 Pattern of sensitivity of contact status for female contacts of old cases by length of incubation period (in years) and age.



REFERENCES

1. Ponninghaus JM, Fine PE, Bliss L, Sliney IJ, Bradley DJ, Rees RJ. The Lepra Evaluation Project (LEP), an epidemiological study of leprosy in Northern Malawi. I. Methods. Lepr Rev 1987;58:359-75.
2. Ponnighaus JM, Fine PE, Bliss L, Gruer PJ, Kapira Mwamondwe B, et al. The Karonga Prevention Trial: a leprosy and tuberculosis vaccine trial in northern Malawi. I. Methods of the vaccination phase. Lepr Rev 1993;64:338-56.
3. Frost WH. The familial Aggregation of Infectious Diseases. Am J Public Health 1938;28:7-13.
4. Hope Simpson RE. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). The Lancet 1952:549-554.
5. Noorden SK. The epidemiology of leprosy. In: Hastings RC, editor. Leprosy. 1st ed. New York: Churchill Livingstone; 1985. p. 15-30.
6. Frost WH. Risk of persons in Familial contact with Pulmonary Tuberculosis. Am J Public Health 1933;23:426-32.
7. Badillet M. Problems posed by measles, chickenpox and mumps in pregnancy. Sem-Hop 1983;59:314-8.
8. Holmes SJ. Review of recommendations of the Advisory Committee on Immunization Practices, Centers for Disease Control and Prevention, on varicella vaccine. J Infect Dis 1996;174:342-4.
9. Lee MS, King CC, Chen CJ, Yang SY, Ho MS. Epidemiology of measles in Taiwan: dynamics of transmission and timeliness of reporting during an epidemic in 1988-9. Epidemiol Infect 1995;114:345-59.
10. Fine PE, Jezek Z, Grab B, Dixon H. The transmission potential of monkeypox virus in human populations. Int J Epidemiol 1988;17:643-50.
11. Kemper JT. Error sources in the evaluation of secondary attack rate. Am J Epidemiol 1980;112:457-464.
12. Michael E, Grenfell BT, Isham VS, Denham DA, Bundy DA. Modelling variability in lymphatic filariasis: macrofilarial dynamics in the Brugia pahangi--cat model. Proc R Soc Lond B Biol Sci 1998;265:155-65.
13. Stoto MA. Models of the demographic effect of AIDS. In: Foote KA, Hill KH, Martin LG, editors. Demographic change in Sub Saharan Africa. Washington, D.C.: National Academy Press; 1993. p. 350-79.
14. Kaplan EH. Asymptotic worst-case mixing in simple demographic models of HIV/AIDS. Mathematical Biosciences 1992;108:141-56.

15. Anderson RM. The impact of the spread of HIV on population growth and age structure in developing countries. In: The global impact of AIDS. Proceedings of the First International Conference on the Global Impact of AIDS, co sponsored by the World Health Organization and the London School of Hygiene and Tropical Medicine, held in London, March 1988:95-106.

16. Fine PEM. Leprosy: the epidemiology of a slow bacterium. Epidemiol Rev 1982;4:161-88.

17. van Beers SM, Hatta M, Klatser PR. Patient contact is the major determinant in incident leprosy: implications for future control. Int J Lepr Other Mycobact Dis 1999;67:119-28.

18. Doull JA, Guinto RS, Rodriguez JN, Bancroft H. The Incidence of Leprosy in Cordova and Talisay, Cebu, P.I. Int J Leprosy 1942;10:107-131.

19. Fine PE, Sterne JA, Ponnighaus JM, Bliss L, Saul J, et al. Household and dwelling contact as risk factors for leprosy in northern Malawi. Am J Epidemiol 1997;146:91-102.

20. George K, John KR, Muliyil JP, Joseph A. The role of intrahousehold contact in the transmission of leprosy. Lepr Rev 1990;61:60-3.

21. Rao PS, Karat AB, Kaliaperumal VG, Karat S. Transmission of leprosy within households. Int J Lepr Other Mycobact Dis 1975;43:45-54.

22. Vijayakumaran P, Jesudasan K, Mozhi NM, Samuel JD. Does MDT arrest transmission of leprosy to household contacts? Int J Lepr Other Mycobact Dis 1998;66:125-30.

23. Jesudasan K, Bradley D, Smith PG, Christian M. Incidence rates of leprosy among household contacts of "primary cases". Indian J Lepr 1984;56:600-14.

24. White SJ, Stone MM, Howland C. Genetic factors in leprosy: a study of children in Uganda. J Hyg Lond 1978;80:205-16.

25. Hausfeld RG. An anthropological method for measuring exposure to leprosy in a leproxy-endemic population at Karimui, New Guinea. Bull World Health Organ 1970;43:863-77.

26. George R, Rao PS, Mathai R, Jacob M. Intrafamilial transmission of leprosy in Vellore Town, India. Int J Lepr Other Mycobact Dis 1993;61:550-5.

27. Rao PS, Karat AB, Karat S. Familial aggregation of leprosy. Leprosy in India 1969:261-5.

28. Noordeen SMA, Mohammed AP. A study of 579 families having multiple cases of Leprosy. First report. Leprosy in India 1964;36:176-81.

29. Selvasekar A, Geetha J, Nisha K, Manimozhi N, Jesudasan K, Rao PS. Childhood leprosy in an endemic area. Lepr Rev 1999;70:21-7.

30. Irgens LM, Melo Caeiro F, Lechat MF. Leprosy in Portugal 1946-80: epidemiologic patterns observed during declining incidence rates. Lepr Rev 1990;61:32-49.

31. Vidal R, Miravitlles M, Cayla JA, Torrella M, de Gracia J, et al. Increased risk of tuberculosis transmission in families with microepidemics. Eur Respir J 1997;10:1327-31.
32. Bogdanova EV. Impact of family contact on development of tuberculosis in infants and preschool children. Probl Tuberk 1997;4:9-11.
33. Kritski AL, Marques MJ, Rabahi MF, Vieira MA, Werneck BE, et al. Transmission of tuberculosis to close contacts of patients with multidrug-resistant tuberculosis. Am J Respir Crit Care Med 1996;153:331-5.
34. Ahiko T. Family contact examination for tuberculosis--the current situations and perspective. Kekkaku 1990;65:739-46.
35. Fine PEM. Immunities in and to Tuberculosis: implications for pathogenesis and vaccination. In: Porter J, McAdam KP, editors. Tuberculosis: Back to the future: John Wiley & Sons; 1994. p. 53-78.
36. Vynnycky E. An investigation of the transmission dynamics of M. tuberculosis. London: London School of Hygiene and Tropical Medicine, University of London; 1996.
37. Medical Research Council M. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. Bull World Health Organisation 1972;46:371-385.
38. ten Asbroek AH, Borgdorff MW, Nagelkerke NJ, Sebek MM, Deville W, et al. Estimation of serial interval and incubation period of tuberculosis using DNA fingerprinting. Int J Tuberc Lung Dis 1999;3:414-20.
39. Vynnycky E, Fine PEM. Lifetime risks, incubation period, and serial interval of tuberculosis. Am J Epidemiol 2000;152:247-63.
40. Madico G, Gilman RH, Checkley W, Cabrera L, Kohlstadt I, et al. Community infection ratio as an indicator for tuberculosis control [see comments]. Lancet 1995;345:416-9.
41. Nations DoEaSA-U. Manual VII. Methods of Projecting Households and Families. In: Manuals on methods of estimating population. New York; 1973.
42. Burch TK. Household and family demography: a bibliographic essay. Population Index 1979;45:173-195.
43. Willekens F. A life course perspective on household dynamics. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling household formation and dissolution. Oxford, England: Clarendon Press; 1988. p. 87-107.
44. National Statistical Office Malawi. Malawi Family Formation Survey, 1984. Zomba, Malawi. National Statistical Office; 1986.
45. National Statistical Office Malawi. Malawi Demographic and Health Survey, 1992. Calverton, Maryland, Macro International; 1994.

46. Foster G, Makufa C, Drew R, Kralovec E. Factors leading to the establishment of child-headed households: the case of Zimbabwe. Health Transition Review 1997;2:155-68.

47. Amin S. Family structure and change in rural Bangladesh. Population Studies 1998;52:201-13.

48. Kuijsten A, Vossen A. The relevance of household models. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling Household Formation and Dissolution. Oxford, England: Clarendon Press; 1988. p. 3-7.

49. Zulu E, Sibanda A. Racial Differences in Household Structure in South Africa. In: Population Association of America Annual Meeting; 2000; Los Angeles, California; 2000.

50. Nakiyingi JS, Malamba SS, Kamali A, Carpenter LM, Nunn AJ, Whitworth JAG. Household composition and the HIV-1 epidemic in a rural Ugandan population: Medical Research Council, Ugandan Virus Research Institute.

51. Van de Walle E. A first look at household data in a Census: Botswana 1981. Philadelphia: Population Studies Centre, University of Pennsylvania.; 1999.

52. Keilman N, Keyfitz N. Recurrent issues in dynamic household modelling. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling household formation and dissolution. Oxford, England: Clarendon Press; 1988. p. 254-85.

53. Keilman N. Emerging issues in demographic methodology. In: European population, 1993. p. 483-508.

54. Brown SP. Analysis of a hypothetical stationary population by family units. Population Studies 1951;IV:380-394.

55. Illing W. Population, Family, Household and Labour Force Growth to 1980. Economic Council of Canada, Staff Study 1967;19:49-69.

56. George MV, Perreault J. Projecting Households and Families in Canada by the Headship Rate Method. In: Woods R, Rees P, editors. In: Population structures and models: developments in spatial demography. Boston, Massachusetts/London, England: George Allen and Unwin; 1986. p. 385-402.

57. Harsman B, Snickars F. A method for disaggregate household forecasts. Tijdschrift Voor Economische En Sociale Geografie/Journal Of Economic And Social Geography 1983;74:282-90.

58. Galler H. Microsimulation of household formation and dissolution. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling Household Formation and Dissolution. Oxford, England: Clarendon Press; 1988. p. 139-157.

59. Keilman N. Dynamic household models. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling household formation and dissolution. Oxford, England: Clarendon Press; 1988. p. 123-38.

60. Clarke M. Demographic processes and household dynamics: a microsimulation approach. In: Woods R, Rees P, editors. Population structures and models: developments in spatial demography. Boston, Massachusetts/London, England: George Allen and Unwin; 1986. p. 245-72.

61. Orcutt GH, Greenburger M, Rivlin A, Korbel J. Micro-Analysis of socio-economic systems: a simulation study. New York: Harper & Row. 1962.

62. Orcutt G, Caldwell S, Wertheimer R. Policy exploration through microanalytic simulation (Urban Institute, Washington DC). 1976.

63. Wilson A, Pownall C. A new representation of the urban system for modelling and for the study of micro-level interdependence. Area 1976;8:246-54.

64. Bonsall P. Micro-simulation of mode choice: a model of organized car sharing. In: Planning and Transportation Research; 1979; University of Warwick; 1979.

65. Kreibich V. Modelling car availability, modal split and trip distribution by Monte Carlo simulation: a short way to integrated models. Transportation 1979;8:153-66.

66. Wegener M. A simulation study of movement in the Dortmund housing market. Tijdschrift voor Economische en Sociale Geografie 1983;74:267-81.

67. Nakamura M, Nakamura AO. On microanalytic simulation and its application in population projection. Journal of Operational Society 1978;29:349-60.

68. Keilman N, van Dam J. A dynamic household projection model: an application of multidimensional demography to lifestyles in the Netherlands. The Hague, The Netherlands, Netherlands Interuniversity Demographic Institute 1987:72.

69. Nelissen JH, Vossen AP. Projecting household dynamics: a scenario-based microsimulation approach. European Journal Of Population 1989;5:253-79.

70. Willekens F, Baydar N. Forecasting place-to-place migration with generalized linear models. In: Woods R, Rees P, editors. Population structures and models: developments in spatial demography. Boston, Massachusetts/London, England: George Allen and Unwin; 1986. p. 203-44.

71. Menken J, Trussell J, Stempel D, Babakol O. Proportional hazards life table models: an illustrative analysis of socio-demographic influences on marriage dissolution in the United States. Demography 1981;18:181-200.

72. Teachman JD. Modeling family formation and dissolution. In: Readings in population research methodology 1993;4:15-64.

73. MacDonald MM, Rindfuss RR. Earnings, relative income, and family formation. Demography 1981;18:123-36.

74. Kotowska IE. Population dynamics in Poland, 1950-2050: internal migration and marital status changes. Laxenburg, Austria, International Institute for Applied Systems Analysis [IIASA] 1994:94-74.

75. van Imhoff E. Modelling the impact of changing household structure on social security in the Netherlands. In: Gonnot JP, Keilman N, Prinz C, editors. Social security, household, and family dynamics in ageing societies. Dordrecht, Netherlands: Kluwer Academic; 1995. p. 181-208.

76. Keilman N, Prinz C. Modelling the dynamics of living arrangements. In: Gonnot JP, Keilman N, Prinz C, editors. Social security, household, and family dynamics in ageing societies. Dordrecht, Netherlands: Kluwer Academic; 1995. p. 21-45.

77. Menken J. Simulation studies. In: Chandrasekaran C, Hermalin AI, editors. Measuring the effect of family planning programs on fertility. Dolhain, Belgium: Ordina; 1975. p. 351-379.

78. Murphy M, Nelissen JH, Vossen AP, Wang J, Crommentuijn L. The dynamic household as a logical concept and its use in demography. European Journal Of Population/Revue Europeenne De Demographie 1996;12:363-81.

79. Timaeus I. Mortality in Sub-Saharan Africa. In: Symposium on Health and Mortality; 1997 19-22 November, 1997; Brussels, Belgium; 1997.

80. Courgeau D, Lelievre E. Estimation of transition rates in dynamic household models. In: Keilman N, Kuijsten A, Vossen A, editors. Modelling household formation and dissolution. Oxford, England: Clarendon Press; 1988. p. 160-76.

81. De Francisco A, Fauveau V, Sarder AM, Chowdhury HR, Chakraborty J, Yunus MD. Measles in rural Bangladesh: issues of validation and age distribution. Int J Epidemiol 1994;23:393-9.

82. Mostafa G, Wojtyniak B, Fauveau V, Bhuiyan A. The relationship between sociodemographic variables and pregnancy loss in a rural area of Bangladesh. J Biosoc Sci 1991;23:55-63.

83. Phillips JF, Macleod BB, Pence B. The household Registration System: Computer Software for the rapid dissemination of Demographic Surveillance Systems. Demographic Research 2000;2.

84. Hashimoto S, Nakai S, Tsubono Y, Nishino Y, Fukao A, Hisamichi S. Effects of misclassification and temporal change of response in food frequency on risk ratio. J Epidemiol 1997;7:153-9.

85. Goldberg DJ. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. J. Am Stat Assn 1975;70:561-7.

86. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 1977;105:488-95.

87. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. Am J Epidemiol 1986;123:736-51.

88. Wells AJ, English PB, Posner SF, Wagenknecht LE, Perez Stable EJ. Misclassification rates for current smokers misclassified as nonsmokers. Am J Public Health 1998;88:1503-9.

89. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 1990;132:1176-84.

90. Spiegelman D, Valanis B. Correcting for bias in relative risk estimates due to exposure measurement error: a case study of occupational exposure to antineoplastics in pharmacists. Am J Public Health 1998;88:406-12.

91. Chen TT. A review of methods for misclassified categorical data in epidemiology [see comments]. Stat Med 1989;8:1095-106.

92. Bross I. Misclassification in 2x2 Tables. Biometrics 1954;10:478-486.

93. Diamond EL, Lilienfeld AM. Misclassification errors in 2x2 tables with one margin fixed. Some further comments. Am J Public Health 1962;52:2106-2110.

94. Diamond EL, Lilienfeld AM. Effects of errors in classification and diagnosis in various types of epidemiological studies. Am J Public Health 1962;52:1137-1144.

95. Newell DJ. Errors in the Interpretation of Errors in Epidemiology. Am J Public Health 1962;52:1925-8.

96. Harper D. Misclassification in epidemiological surveys. Am J Public Health 1964;54:1882-1886.

97. Gullen WH, Bearman JE, Johnson EA. Effects of misclassification in epidemiologic studies. Public Health Rep. 1968;83:914-8.

98. Holcroft CA, Spiegelman D. Design of validation studies for estimating the Odds Ratio of exposure-disease relationships when exposure is misclassified. Biometrics 1999;55:1193-1201.

99. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. J Clin Epidemiol 1988;41:1167-74.

100. Wynder EL, Cornfield J, Shroff PD, Doraiswami KR. A study of Environmental Factors in Carcinoma of the Cervix. Am. J. Obst. & Gynec 1954;68:1016-1052.

101. Barron BA. The effects of misclassification on the estimation of relative risk. Biometrics 1977;33:414-8.

102. Farrington CP. Quantifying misclassification bias in cohort studies of vaccine efficacy. Stat Med 1990;9:1327-37.

103. Reade Christopher SJ, Kupper LL. Effects of exposure misclassification on regression analyses of epidemiologic follow-up study data. Biometrics 1991;47:535-48.

104. Sosenko JM, Gardner LB. Attribute Frequency and Misclassification bias. J Chron Dis 1987;40:203-7.

105. Espeland MA, Hui SL. A general approach to analyzing epidemiologic data that contain misclassification errors. Biometrics 1987;43:1001-1012.

106. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat Med 1988;7:745-57.

107. Selen J. Adjusting for errors in classification and measurement in the analysis of partly and purely categorical data. American Statistical Association 1986;81:75-81.

108. Greenland S. The effect of misclassification in the presence of covariates. Am-J-Epidemiol 1980;112:564-9.

109. Hackshaw AK. Lung cancer and passive smoking. Stat Methods Med Res 1998;7:119-36.

110. Brown J, Kreiger N, Darlington GA, Sloan M. Misclassification of exposure: Coffee as a surrogate for Caffeine Intake. Am J Epidemiol 2001;153:815-20.

111. Schulman J, Selvin S, Shaw GM, Malcoe LH. Exposure misclassification due to residential mobility during pregnancy in epidemiologic investigations of congenital malformations. Arch Environ Health 1993;48:114-9.

112. Shaw GM, Malcoe LH. Residential mobility during pregnancy for mothers of infants with or without congenital cardiac anomalies: a reprint [corrected and republished article originally printed in Arch Environ Health 1991 Sep-Oct;46(5):310-2]. Arch Environ Health 1992;47:236-8.

113. Murphy M, Sullivan O, Brown A. Sources of data for modelling household change with special reference to the OPCS 1% longitudinal study. In: Keilman N, Kuijsten A, Vossen A, editors. In: Modelling household formation and dissolution. Oxford, England: Clarendon Press; 1988. p. 56-66.

114. National Statistical Office Malawi. Malawi Population and Housing Census: Population Characteristics; 1987.

115. Timaeus I, Graham W. Measuring adult mortality in developing countries: a review and assessment of methods. London School of Hygiene and Tropical Medicine, Centre for Population Studies 1988;55.

116. Timaeus IM. Adult mortality. In: Foote KA, Hill KH, Martin LG, editors. Demographic Change in Sub-Saharan Africa. Washington, D.C.: National Academy Press; 1993. p. 218-255.

117. National Statistical Office Malawi. Population Characteristics. Zomba, Malawi: Government Printer; 1984.

118. National Statistical Office Malawi. Fertility, Mortality and Population projections. Zomba, Malawi: Government Printer; 1984.

119. Ponnighaus JM, Fine PEM, Sterne JA, Malema SS, Bliss L, Wilson RJ. Extended schooling and good housing conditions are associated with reduced risk of leprosy in rural Malawi. IntJ Lepr Other Mycobact Dis 1994;62:345-52.

120. Appleyard R, Wilson A. Migration and HIV/AIDS. Geneva, Switzerland, International Organization for Migration [IOM] 1998;36.4.

121. Ponnighaus JM, Fine PEM, Maine N, Bliss L, Kalambo M, Ponnighaus I. The Lepra Evaluation Project (LEP), an epidemiological study of leprosy in northern Malawi. II: Prevalence rates. Lepr Rev 1988;59:97-112.

122. Sterne JA, Ponnighaus JM, Fine PEM, Malema SS. Geographic determinants of leprosy in Karonga District, Northern Malawi. Int J Epidemiol 1995;24:1211-22.

123. Brass W. Methods for estimating fertility and mortality from limited and defective data, based on seminars held 16-24 September 1971 at the Centro Latinoamerico de Demografia (CELADE) San Jose, Costa Rica. International Program of Laboratories for Population Statistics 1975;159.

124 David PH, Bisharat L, Hill AG, Bennett S. Measuring childhood mortality: a guide for simple surveys. Amman, Jordan; 1990.

125. Feeney G. Estimating infant mortality trends from child survivorship data. Population Studies 1980;34:109-128.

126. Karonga Prevention Trial Group. Randomised controlled trial of single BCG, repeated BCG, or combined BCG and killed Mycobacterium leprae vaccine for prevention of leprosy and tuberculosis in Malawi. Lancet 1996;348:17-24.

127. Kristensen I, Aaby P, Jensen H. Routine vaccinations and child survival: follow up study in Guinea-Bissau, West Africa. BMJ 2000;321:1435-8.

128. Fine P. Commentary: an unexpected finding that needs confirmation or rejection. BMJ 2000;321:1439.

129. Cree IA, Smith WC. Leprosy transmission and mucosal immunity: towards eradication? Lepr Rev 1998;69:112-21.

130. Irgens LM, Skjaerven R. Secular trends in age at onset, sex ratio, and type index in leprosy observed during declining incidence rates. Am J Epidemiol 1985;122:695-705.

131. Fine PEM, Ponnighaus JM, Maine N, Clarkson JA, Bliss L. Protective efficacy of BCG against leprosy in Northern Malawi. Lancet 1986;2:499-502.

132. Ponnighaus JM, Fine PEM, Sterne JA, Bliss L, Wilson RJ, et al. Incidence rates of leprosy in Karonga District, northern Malawi: patterns by age, sex, BCG status and classification. Int J Lepr Other Mycobact Dis 1994;62:10-23.

133. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med 2000;19:1141-64.

134. Vynnycky E, Fine PEM. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect 1997;119:183-201.

135. Glynn JR, Warndorff DK, Fine PEM, Msiska GK, Munthali MM, Ponnighaus JM. The impact of HIV on morbidity and mortality from tuberculosis in sub-Saharan Africa: a study in rural Malawi and review of the literature. Health Transition Review 1997;7:75-87.

136. Gregson S, Garnett GP, Anderson RM. Assessing the potential impact of the HIV-1 epidemic on orphanhood and the demographic structure of populations in Sub-Saharan Africa. Population Studies 1994;48:435-58.

137. Urassa M, Ng'weshemi JZ, Isingo R, Kumogola Y, Boerma JT. Orphanhood, child fostering and the AIDS epidemic in rural Tanzania. In: The socio demographic impact of AIDS in Africa. Based on the conference organized by the Committee on AIDS of the International Union for the Scientific Study of Population 1997.

138. Konde Lule JK, Wawer MJ, Nalugoda F, Gray RH, Menon R. HIV infection in rural households, Rakai district, Uganda. In: The socio demographic impact of AIDS in Africa. Based on the conference organized by the Committee on AIDS of the International Union for the Scientific Study of Population 1997.

139. Polissar L. The effect of migration on comparison of disease rates in geographic studies in the United States. Am J Epidemiol 1980;111:175-82.

140. Crampin AC, Floyd S, Glynn JR, Sibande F, Maluwa D, et al. Long term follow-up of HIV positive and negative natural history cohorts in rural Malawi. Unpbublished 2001.

141. Brass W, Blacker J. The estimation of infant mortality from proportions dying among recent births.: Centre for Population Studies, London School of Hyigiene & Tropical Medicine.