The Social Epidemiology of Tuberculosis:
A Study in Zambia

Delia Boccia

January 2010

Department of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine

Thesis submitted to the London School of Hygiene and Tropical Medicine for the degree of Doctor of Philosophy

Dedicated to my mother. Because I promised this to her.

Acknowledgements

One thing they never tell you when you start a PhD is that life generally doesn't care about public health. You're there busy with statistics, numbers and theories that you believe can change the world when suddenly, and quite disrespectfully, life happens, diverting your attention and taking you through unpredictable storms. Fortunately, I am a blessed person and over these past 4 years (and 47 days!), a long list of people have contributed to keep me strong and making sure I reached the end of this journey.

Probably James Hargreaves in his next life will be a rock star, but I am glad that in this one he happened to be an epidemiologist and most importantly my supervisor. To him goes my deepest debt: his intelligence and competence have shaped my thoughts and turned them into meaningful epidemiological concepts. His analytic rigor and honesty have strengthened every single page of this thesis. His lucid passion has taught me that to be useful social epidemiology has to be grounded on solid scientific theories and methods. Otherwise it is just bad science and ideology, both things that TB patients do not need. Meeting him was a gift and I will never thank him enough.

I'm most grateful to the rest of my advisory panel: Bianca De Stavola and Katherine Fielding. I will never forget their precious patience, their help, and their extraordinary human and teaching skills. Thanks also to John Porter and Alex Coldham for their intellectual and practical contributions to this thesis, but most of all for the rock cakes, the bad coffees and the wonderful time we shared at the school.

I'm proud to say that this study was part of a much bigger 'dream' led by Helen Ayles and Peter Godfrey-Faussett. Without their scientific, logistic, and economic investment this study would have never been possible. Thanks for all that they taught me about TB and for introducing me to a research area that has become a great passion of mine: they knew long before than me what I was meant to do in life. Thank you. Special thanks go to all the people working in the Zamstar project, in particular Ab, Inonge and Arrival. Their friendship and commitment during my fieldwork gave value and quality to this study and to my time in Zambia.

My experience in Zambia was the people I've met. In particular, Jean Mulenga and her kids. Thanks to the love and the joy that they brought into my life and to the things that they have taken from me I became for the first time the person I've always wanted to be. To them and for their lesson of life I remain eternally indebted.

I would have never finished this PhD without my friends. Thanks to Mara, Claudia, Marta, Silvia, Costanza, and Bianca for their unconditional trust in me, my mysterious life choices and my hopeless dreams. Without them I'd be nothing. Thanks to Roberto, for being the guardian of my nicest memories; and to my 'cat' Claudio, for seeing every part of me, all my light and all my dark, and - despite this - being still here.

Immense love and gratitude to my dad and my sister. They kept me sane and have stopped me from giving up during the hardest moments our family went through. This PhD and anything good I will ever do in life is for them and for all of my mum that still lives through them.

This PhD is about TB, but also about injustice, unhappiness and things that must be changed. In this world there is someone who has chosen to dedicate every day of his life to these challenges with inspirational creativity, unreserved devotion, and a considerable degree of madness (!). I'm lucky enough to say that this person is not just one of the greatest TB scientists I know, but also the love my life. Thank you always, Carlton.

I, Delia Boccia, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Delia Boccia

Abstract

Study rationale

While household socioeconomic position (SEP) may be clearly a root cause of tuberculosis (TB), it is more challenging to understand *how* household SEP affects the risk of TB, *how* this effect is mediated by risk factors that are on the causal pathway, *how* this effect differs at different level of analysis and different pathogenic stages of TB, and *how* this evidence can inform control strategies.

Objectives

The aim of this study was to: 1) explore the ecological correlation between community living conditions and TB prevalence rates; 2) quantify the association between household SEP and TB disease and TB infection and, 3) explore the causal mechanism underlying these associations.

Methods

This research project was nested within a population-based HIV-TB prevalence survey conducted in 2005-2006 in two Zambian communities, one rural and one urban. Each community is divided into smaller tracts called Census Advisory Areas (CSA). Prevalent TB was diagnosed through culture and molecular testing of sputum samples collected from the prevalence survey participants. The correlation between community living conditions and TB prevalence was assessed through an ecological analysis conducted at CSA level. The association between household SEP and TB disease and infection was assessed respectively through a case-control and a cross-sectional study. CSA living conditions were explored through poverty mapping. Variables accounting for four different domains of household SEP were recorded (human resources; food availability; housing quality; and access to services) and combined into a composite index using principal component analysis. In addition four more SEP indices were developed: one for each household SEP dimension considered. The analysis of the mediation pathway between household SEP and TB (disease and infection) was driven by a pre-defined conceptual framework, including household SEP and individual-level risk factors. Adjusted Population Attributable Fractions (PAF) were estimated.

Main findings

Ecological analysis - The overall cluster-adjusted TB prevalence rates was 870/100,000 (95%CI: 570-1160/100,000) Urban CSA had both significantly higher TB prevalence rates and wealthier living conditions compared to rural CSA. Although not significantly, TB prevalence rates tended to increase with the worsening of urban CSA living conditions.

Case control study - At *household level*, prevalent TB was significantly associated with lower household SEP [aOR = 6.2, 95%CI: 2.0-19.2 and aOR = 3.4, 95%CI: 1.8-7.6 respectively for low and medium household SEP compared to the baseline]. The food availability domain was more strongly associated with prevalent TB than the other domains considered [aOR = 4.2, 95%CI: 1.8-9.2]. At *individual level*, TB prevalent cases were

significantly more likely to have a diet poor in proteins [aOR= 3.1, 95%CI: 1.1-8.7], to be not BCG vaccinated [aOR = 7.7, 95%CI: 2.8-20.8], to be HIV positive [aOR= 3.1, 95%CI: 1.7-5.8], and to have migrated [aOR = 5.2, 95%CI: 2.7-10.2] than controls. These associations all persisted after controlling for household SEP. The association between household SEP and TB appeared to be mainly mediated by inadequate nutrition. Before mediation, PAF for household SEP was equal to 30%. The adjusted PAF for inadequate nutrition and HIV infection was equal to 42% and 36%, respectively.

Cross sectional study – TB infection was associated with higher, rather than lower, household SEP [aOR = 0.4, 95%CI: 0.2-0.9 and aOR= 0.4, 95%CI: 0.2-0.8 for high and medium household SEP compared to low SEP]. This association was driven by the household SEP domain on access to community services [aOR = 2.7, 95% CI: 1.0-7.1]. None of the investigated risk factors appear to mediate the association between household SEP and TB infection.

Conclusions

In this setting, urban communities were wealthier but also had greater prevalence of TB. The low power of the ecological analysis does not allow any conclusions; however – as for other health indicators, even for TB it seems that the apparent advantage of living in an urban setting can mask severe TB inequalities across smaller urban communities.

This study revealed an unexpected and counterintuitive result: low household SEP was significantly associated with a higher risk of TB disease, but with a lower risk TB infection. Higher household SEP may be associated with lifestyles increasing social mixing and thus the risk of infection. Among the infected people, poorer individuals may be more likely to progress to TB disease and, therefore, are more likely to be identified in prevalence surveys. In the association with TB disease, household SEP seems to operate mainly through inadequate nutrition. While both HIV and inadequate nutrition were important social determinants of prevalent TB, these findings suggest that interventions addressing food insecurity may prevent more TB cases in this setting.

Because of low study power and possible selection biases results this interpretation will have to be confirmed by more powered and better designed studies. In the meantime, this study suggests that interventions addressing household SEP, through food and financial support, may effectively reduce TB prevalence and complement currently control strategies mainly targeted to HIV positive people.

To evaluate the impact and cost-effectiveness of this approach is a priority.

Table of contents

List of tables

List of figures

List of Appendices

Glossary of terms and abbreviations

ABSM	Area Based Socioeconomic Measures
AF	Attributable Fraction
ARV	Antiretroviral therapy
BCG	Bacillus Calmette-Guérin
BMI	Body Mass Index
CI	Confidence Intervals
CSA	Census Supervisory Area
CSO	Central Statistical Office
CSDH	Commission on Social Determinants of Health
DOTS	Directly Observed Therapy Short Course
ELISA	Enzyme-linked immunosorbent assay
HIV	Human Immonodeficiency Virus
LRT	Likelihood Ratio Test
LSMS	Living Conditions Monitoring Survey
LSMSIII	Zambia Living Conditions Monitoring Survey 2002-2003
NGO	Non-governmental Organization
OR	Odds Ratio
PAF	Population Attributable Fraction
PCA	Principal Component Analysis
PLWHA	People living with HIV/AIDS
QFT	QuantiFERON®-TB Gold In Tube
SEA	Standard Enumeration Area
SEP	Socioeconomic position
SES	Socioeconomic status
TB	Tuberculosis
TST	Tuberculin Skin Test
WHO	World Health Organization

My strength lies solely in my tenacity

Pasteur L[1].

1. Introduction

One learns, I would hope, to discover what is right, what needs to be righted – through work, through action.

Berrigan D[2], 1971

1.1 Tuberculosis and poverty: do we need more research?

Since Robert Koch discovered *Mycobacterium tuberculosis* in 1882, spectacular scientific advances have been in made in our understanding of tuberculosis (TB) and ability to treat patients. The famous Madras studies carried out by the Madras Tuberculosis Chemotherapy Centre and the National Tuberculosis Centre in Bangalore in the 1950s demonstrated that treatment of TB was possible and efficient even in very poor population [1-2]. This was interpreted as the proof that poverty was no longer an obstacle to successful TB control. The antibiotic revolution of 1940-50 led to such a prevailing optimism that in the 1960s international organisations and experts declared that TB was to be eradicated as a public health problem by the year 2000 [3].

Unfortunately this aspiration has not been achieved. In 2006 the World Health Organization (WHO) documented an estimated 9.2 million new cases of TB together with an estimated 14.4 million prevalent cases. In the same year approximately 2 million deaths were reported, of which more than 2/3 were among HIV negative people [4].

These estimates are shocking not just because TB is a preventable and treatable disease, but also because of the significant inequalities in the distribution of these cases over the world: 95% of the TB deaths occur in developing countries [5]. Over 44% of TB deaths occur in the poorest 20% countries, while only 2% occur in the wealthiest 20% of countries [6].

We certainly know already a lot about socioeconomic determinants of TB, but we might not know enough yet, or at least not enough to translate this evidence into coherent public health policies to control TB. While poverty may clearly be a root cause of TB and poverty eradication might be an important goal in its own right, it is more challenging to understand

how poverty affects the risk TB, *how* this effect differs in different contexts and *how* this evidence can inform pragmatic control strategies.

Despite the extensive literature available, there is still need of more research on the link between poverty and TB:

- Although at societal level, the association of TB with poverty has been widely proved, most of these studies are ecological and by definition may overestimate the strength of the association or assign erroneously group-based measures of disease-exposure association to persons within the group. On the other hand, analytical studies did not consistently find an association between poverty and TB at individual level [7-8].

- The relative importance of socioeconomic conditions in the different pathogenic stages of TB (exposure, infection, disease, outcome) remains poorly understood.

- Despite the acknowledgement that TB has strong socioeconomic determinants, there has been little epidemiological research to identify key mediators on the causal pathway of this association.

- It remains unclear what is the place of poverty in the epidemiology of TB in the HIV-era.

- Poverty has often been conceptualised in terms of deprivation of income or basic needs. Today it is widely acknowledged that poverty is multidimensional, involving more than the narrow lack of material consumption or resources, to encompass notions of material well-being, an absence of infrastructure, a lack of power and voice, and unravelling of social structures [9]. Describing the impact of socioeconomic determinants on TB requires clear definitions and appropriate measurement tools. However, the more sophisticated our conceptualisation of poverty becomes, the more difficult is to develop practical tools to measure it; the broader and more intangible our definition of poverty is, the more difficult it is to identify entry points for TB control.

To address properly these issues with rigorous and transparent methods, transcending mere opinions and ideology, is a core public health function of epidemiology [10] and the scope of this thesis.

The aim of this doctoral research is to explore the social epidemiology of TB, that is to move beyond the focus on individual risk factors for TB to look at those processes and mechanisms that generate the differential distribution of these risk factors across the

socioeconomic groups and ultimately to the observed TB inequalities. Specifically, this thesis poses the following research questions:

1. Is low household socioeconomic position (SEP) associated with TB?

2. What is the possible causal pathway through which low household SEP increases the risk of TB?

3. What is the epidemiological importance of HIV and other more proximal, individual, risk factors once household SEP has been taken into account?

4. Is household SEP equally important in affecting the risk of TB infection and TB disease? And if not, what differential mechanisms can be identified?

The research project was undertaken in two communities in the southern African country of Zambia, a poor country with a high TB and HIV burden. To answer these questions I have employed three different strategies: 1) the ecological analysis of the results of the latest population-based HIV-TB prevalence survey conducted in Zambia to assess the geographical correlation between community living conditions and TB prevalence rates; 2) a case control study to assess the role of household SEP on the risk of TB disease; 3) a cross sectional investigation to explore the association between household SEP and the risk of TB infection.

1.2 Structure of the thesis

This thesis has been organised into the following chapters:

Chapters 2 and 3 provide the background and the literature review. Chapter **2** describes the living conditions and the burden of TB in Zambia, introduces the concept of social epidemiology and why the use of social epidemiology concepts and methods might be useful to a better understanding of the current TB epidemic. The study of the social epidemiology of TB requires the preliminary definition of socioeconomic determinants and the mechanism responsible of the observed inequalities. For this reason, chapter 2 describes how socioeconomic position has been conceptualised and measured over time and discusses some of the key methodological challenges in this field. It finally illustrates the most relevant aetiological models used for the interpretation of health inequalities. In order to better understand the association between household SEP and TB. Chapter **3** will provide a review of the most relevant analytical studies exploring this association. The limitations and the knowledge gaps identified in the literature will inform the rational of this thesis.

Chapter 4 presents the study methods. It includes the study objectives and the conceptual framework employed to generate hypotheses about how socioeconomic factors may affect the risk of TB. The chapter provides details on the three study components of this research project and describes how the study framework was operationalised and key constructs measured.

Chapter 5, 6 and 7 contain the results of the research. Chapter **5** describes in detail the living conditions of the communities enrolled in the latest population-based TB prevalence survey conducted in Zambia and how a selected list of area-based measures of SEP correlate with the TB prevalence rates detected in these communities. Chapter **6** describes the findings of the case-control study on the association between socioeconomic position and prevalent TB, whereas Chapter **7** describes the socioeconomic distribution of TB infection among the disease-free controls enrolled in the case-control study.

Chapter 8 shows the sensitivity of the main results of this thesis to the choice of different methods and approaches for the measurement of socioeconomic position.

Chapter 9 summarise the main study results and describes strengths and limitations of the research. Findings are evaluated in the light of the preceding empirical and theoretical literature. The chapter concludes with the discussion of the programmatic implications of the study including recommendations to improve TB control in these two communities and directions for future research. Final remarks are presented briefly in **Chapter 10**.

2. Background

Our scientific view of the universe should be as simple as possible, but not more simple than that.

Einstein A[3], 1933

Introduction

The aim of this chapter is to describe:

1. The study setting, including a summary of the living conditions in Zambia and the burden of TB in this country.

2. The basis for a social epidemiology of TB. After introducing some key elements of TB epidemiology, I will illustrate some recent evidence suggesting the importance of considering driving forces other than HIV in the current TB epidemic, especially when interpreting the impact of the current TB control strategies. I will show how this evidence calls for a more systematic approach to the study of TB epidemiology, integrating biological, behavioural and socioeconomic factors in a multilevel framework.

3. The essential theories and concepts necessary to apply a social epidemiological approach to the study of TB, such as the definition and measurement of socioeconomic position (SEP) and the most influential aetiological models developed for the study of health inequalities, including TB.

2.1 The context

Zambia is a southern Africa country with a population of approximately 11 million people. It is predominantly a rural country with much of the population involved in agriculture; however, approximately 70% of Zambia's population are concentrated in the urban provinces of Copperbelt and Lusaka.

Like much of Sub-Saharan Africa, Zambia experienced a long period of European colonial rule (by the United Kingdom, 1923-1964). Zambia achieved independence peacefully and has avoided violent conflicts since becoming an independent nation state unlike some of its immediate neighbours, such as Angola, Mozambique, The Democratic Republic of Congo, and Zimbabwe. Zambia's peaceful environment and mineral wealth enabled it to fuel large public sector investment in the 1960s and 1970s making it one of the most prosperous countries of sub-Saharan Africa. Unfortunately, rising oil prices in the mid 1970s coincided with a major drop in the world copper price and led to significant economic decline in Zambia where copper was the mainstay of the economy. The droughts that were experienced in the early 1990s, combined with increasing cattle morbidity and mortality rates, contributed to a decline in agricultural production [11]. Liberalisation and structural adjustment of the economy have, at least in the short term, denied the farming community access to markets, both for agricultural inputs and for the sale of products, and this has in many areas reinforced the tendency toward declining production. Furthermore, adjustment has led to increased unemployment and livelihood insecurity due to retrenchment in public services and mining [11].

In recent years, Zambia has experienced its longest period of sustained growth since independence, averaging 2% annual GDP per capita growth (2003)[11-12]; however, given the low pattern of growth during the 1990s, most of the welfare indicators show only small changes and Zambia is still among the poorest countries of Africa [13].

2.1.1 The burden of poverty in Zambia

The Central Statistical Office of Zambia defines absolute poverty based on two thresholds relating to extreme and moderate poverty lines. The extreme poverty line is set at the monthly cost of a basic-needs food basket, while the moderate line relates to the monthly cost of all basic needs including non-food items. Accordingly, households whose monthly expenditure is below the cost of the food basket (extreme line) are classified as extremely poor, while those whose expenditure falls short of the moderate poverty line are classified as moderately poor. The food basket represents the food–expenditure needed to meet the daily caloric requirements of 12,564 for an average family of 6, or 2,094 calories per person per day.

The cost of the food basket depends on the type of food included in it. In order to reflect the consumption pattern of the poor, the food basket should obviously contain food items normally consumed by the poor. Technically this can be achieved by including in the food

basket the food expenditure values of the households whose food expenditure per adult equivalent is 20% of the national median food expenditure.

According to the Zambian Living Conditions Monitoring Survey for year 2002 and 2003 (LCMSIII) data [14], in 2002 the extreme poverty line coincided with US$ 15 per adult per month (approximately US$ 0.5 per day), whereas the moderate line this was equal to approximately $US 22 per adult per month.

The poverty analysis of 2002-03 showed that 67% of the population fell below the poverty line: 46% were classified as extremely poor and 21% as moderately poor. The overall poverty prevalence varied from 57% in Lusaka province to 81% in Northern Province. While the proportion of population living in moderate poverty did not vary much between provinces, there were considerable variations in terms of the proportion of the population living in extreme poverty across provinces (**Figure 2.1**). The prevalence of extreme poverty ranged from 36% in Lusaka province to the 63% in the Northern Province.

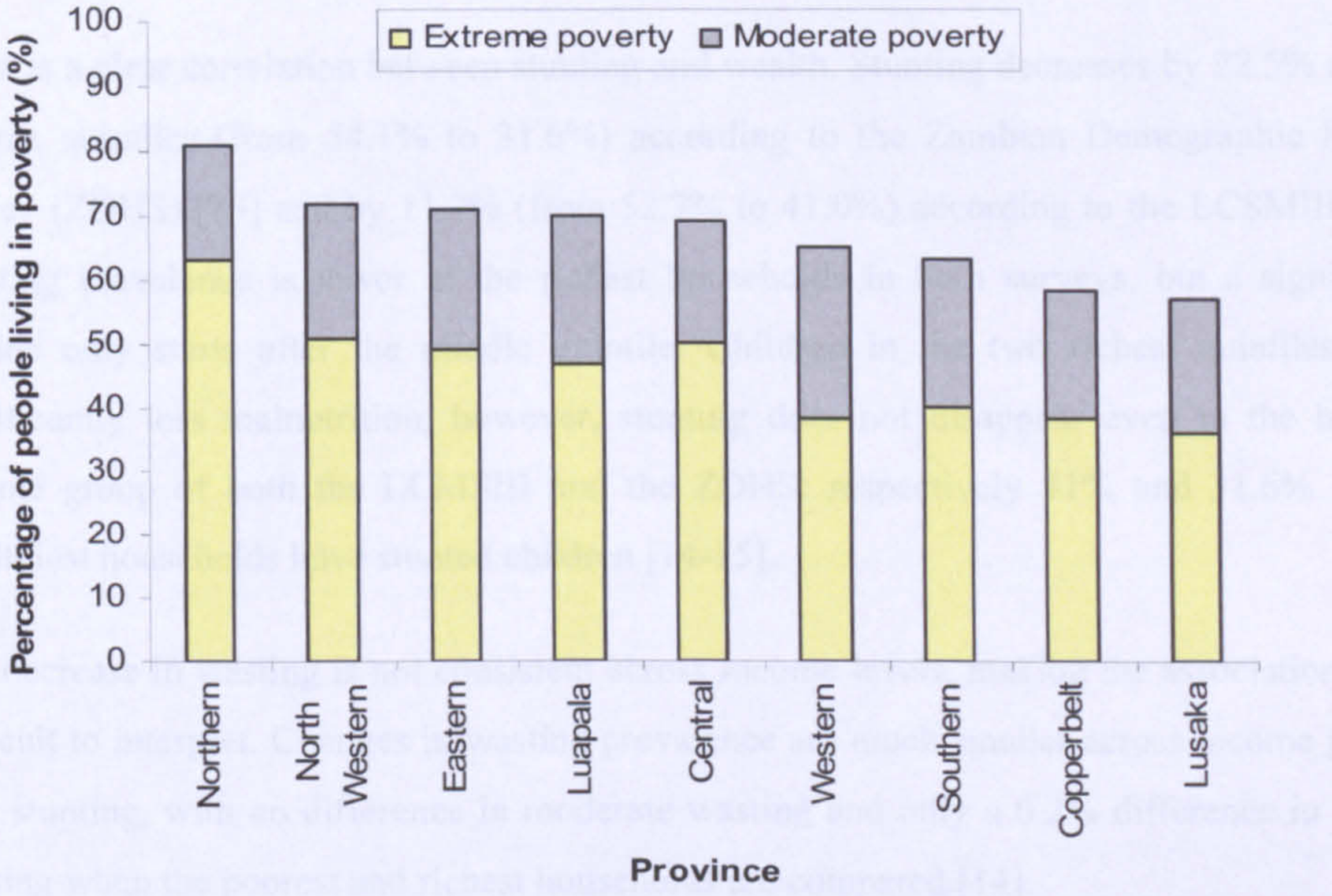
The observed high level of extreme poverty in Northern, North Western, Central and Eastern Provinces has an important effect on the food security situations in these regions: in these areas the majority of the households (51%) can only afford two meals in a typical day, 38% of the households are able to have at least three meals and over in a day and a remaining 11% can only afford 1 meal per day. The proportion of households that could not manage at least three meals in a day was higher in rural than urban areas.

(a) Malnutrition and poverty

Malnutrition is a serious problem in Zambia. The nutritional status of the entire population progressively deteriorates between November and March, coinciding with the pre-harvest season and the increased transmission of diseases during the rainy season. Seasonality is particularly important in the rural areas: available surveys [11] show that household supply of maize, the staple of the local farming system, are typically very low between November and February and often completely run out in March and April when people rely on other non-staple food.

Next I will present some data particularly relevant for this thesis as they informed the choice of variables selected for the creation of the SEP indicators employed in this research project.

Figure 2.1: Prevalence of poverty by Province, Zambia, 2002-2003



The prevalence of stunting (a marker of chronic malnutrition) in children under-five years of age increased from 40% in 1992 to 48% in 2002/3 [12]. By contrast, the prevalence of wasting (a marker of acute malnutrition) remained constant at 5% over the same period.

There is a clear correlation between stunting and wealth. Stunting decreases by 22.5% across income quintiles (from 54.1% to 31.6%) according to the Zambian Demographic Health Survey (ZDHS) [15] and by 11.7% (from 52.7% to 41.0%) according to the LCSMIII [14]. Stunting prevalence is lower in the richest households in both surveys, but a significant decline only starts after the middle quintile. Children in the two richest quintiles have significantly less malnutrition, however, stunting does not disappear even in the highest income group of both the LCMSIII and the ZDHS: respectively 41% and 31.6% of the wealthiest households have stunted children [14-15].

The decrease in wasting is not consistent across income levels, making the association more difficult to interpret. Changes in wasting prevalence are much smaller across income groups than stunting, with no difference in moderate wasting and only a 0.2% difference in severe wasting when the poorest and richest households are compared [14].

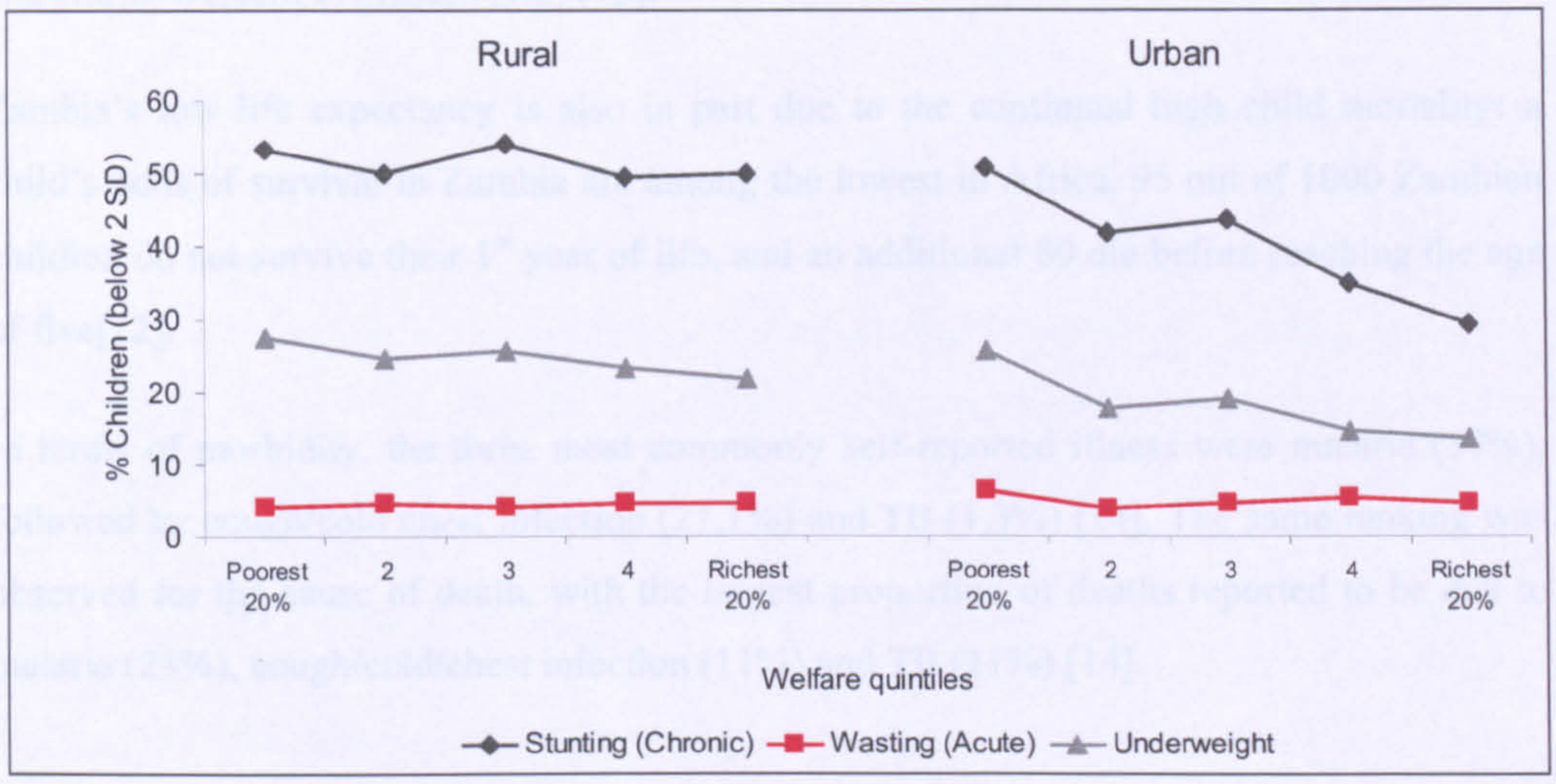
The overall trend is therefore that wasting is similar across income groups whereas stunting decreases as income increases.

Urban areas are characterised by a declining trend across income quintiles of both overall stunting (43% decrease) and severe stunting (36% decline). In the rural areas the difference between the richest and the poorest are less evident with only a 6% decline in stunting and 20% decline in severe stunting from the poorest to the richest quintile. Despite this lesser inequality in the rural areas, the prevalence of stunting is higher than in urban areas, overall and by each quintile considered (**Figure 2.2**) [12].

(b) Additional indicators of well-being

Table 2.1 provides a summary of additional indicators related to education, HIV/AIDS, childhood mortality, and child malnutrition that are discussed below.

Figure 2.2: Nutritional status in children under 5 years by urban and rural area and by income quintiles



Modified from [12]

Education – Among sub-Saharan African countries, only South Africa and Botswana have populations with higher average years of schooling than Zambia. In 2002/03, the proportion of the population attending school in the age groups 5-6, 7-13, 14-18 and 19-22 years were respectively 13, 75, 64 and 21%. This is largely the legacy of the post-colonial period, when the government invested income from copper wealth into education. Nonetheless, because of the poor quality of the education services provided, Zambia's adult literacy rate is no higher than that of nearby countries like Rwanda and Uganda that have a much lower schooling rate [12].

Life expectancy – Life expectancy at birth in Zambia is the fifth lowest in the World, according to UN estimates [12]. Life expectancy for those born 2000-2005 was just 37.4 years, a drop of 14 years from 25 years earlier. The decline in life expectancy at birth has largely been driven by HIV/AIDS: Zambia is one of the ten countries worldwide with double digit HIV prevalence rates [16]. 16% of all adults are infected with HIV and infection rates are particularly high in Zambia's cities and towns, however HIV/AIDS in Zambia is not uniformly distributed across geographic and demographic categories [12]. The evidence suggest that, while males were disproportionately affected during the early phase of the epidemic, the preponderance of new infections now occur among Zambian women, with 40% of the women in their 30s infected with HIV. The rates of infection in urban areas are double than those of rural areas (23.1% and 10.8% respectively) and women seem to be the most affected (26.3%) [15]. The highest rates are observed in the urbanised provinces of

Copperbelt (20%) and Lusaka (22%), whereas rates are relatively low in the Northern (8.3%) and North-Western Provinces (9.2) [15].

Zambia's low life expectancy is also in part due to the continued high child mortality: a child's odds of survival in Zambia are among the lowest in Africa. 95 out of 1000 Zambian children do not survive their 1st year of life, and an additional 80 die before reaching the age of five[12].

In terms of morbidity, the three most commonly self-reported illness were malaria (37%), followed by cough/cold chest infection (21.1%) and TB (1.3%) [14]. The same ranking was observed for the cause of death, with the largest proportion of deaths reported to be due to malaria (23%), cough/cold/chest infection (11%) and TB (11%) [14].

Table 2.1: Measures of well-being in Zambia, by sex

Indicator	Overall	Male	Female
Education			
Adult literacy rate (%), ages 15 and above	67	77	58
Mean years of schooling, ages 15 and above	5.6	6.4	4.9
School attendance rate (%), ages 7-13	75	75	76
School attendance rate (%), ages 14-18	64	71	56
HIV/AIDS			
Prevalence rate (%), Urban and Rural, ages 15-49	16	13	18
Prevalence rate (%), Urban, ages 30-34	38	34	43
Life expectancy and Childhood mortality			
Life expectancy at birth	37.4	37.9	36.9
Infant mortality (deaths/1000 before age 1)	95	95	93
Under-5 mortality (deaths/1000 births before age 5)	168	178	160
Child malnutrition			
Stunted (%), ages 5 and under	47	48	46
Wasted (%), ages 5 and under	5	6	4

(c) The main causes of poverty in Zambia

The World Bank has identified three main causes for poverty in Zambia [12]:

1) The lack of assets coupled with low and uncertain returns from these assets. The labour of poor men and women is concentrated in subsistence agriculture or low-paid casual work. Only a small proportion of available land is irrigated. Labour is a binding constraint for many of the rural poor and most of them cannot afford to purchase seeds, fertilizer and other

productivity-enhancing agriculture inputs. Thus, poverty in Zambia is caused by scarce and uncertain levels of private assets.

2) The scarcity of public assets and services. Public services such as health and education frequently by-pass the poor and the services that are available are often under-funded, poorly managed and capacity constrained. Infrastructure development, including provision of roads, communication, electricity, drinking water and irrigation, is low in rural Zambia, and access is particularly limited in more rural regions of the country, where many of the poor reside. Infrastructure is more developed in urban areas but it suffers from inadequate capacity and years of poor maintenance.

3) The poverty of social relations. Zambia society is comprised of a system of overlapping kinship networks whose members are obliged to offer each other assistance and support. However, the deteriorating economy has put traditional coping mechanisms under considerable strain.

2.1.2 The burden of tuberculosis in Zambia

In 2007 Zambia ranked as 9th in the list of 15 countries with the highest estimated TB incidence rates per capita. The annual TB incidence rate during this period was 506/100.000corresponding to a total of 60,337 incident cases notified. [17] Of these, 41,954 people were co-infected with HIV (reflecting a TB incidence rate of 352/100.000 among HIV positive patients). Zambia had among the highest number of prevalent cases of TB: in 2007 approximately 47,000 cases of prevalent TB were estimated in the general population, 20,997 of these among people living with HIV/AIDS (PLWHA), corresponding to a prevalence of 387/100.000 and 176/100.000, respectively. In 2007, the HIV prevalence in incident cases of TB was 70%. In the same year, 13,661 people died of TB, 10,624 of which PLWHA. The proportion of Multidrug Resistant TB (MDR-TB) is still below 2% among the new TB cases and 2.3% among previously treated cases [17].

Data on case detection rate and percentage of cases successfully treated are available only for year 2006: according to WHO estimates these were respectively equal to 58% and 85%, suggesting that Zambia has achieved the WHO target on successful treatment rate, but is still far from the case detection rate WHO target (equal to 70%).

The TB epidemic of Zambia has been called an epidemic of reactivating TB in those who are HIV infected [11]: retrospective data from the Zambian Ministry of Health show that over the period 1964-2000 there was a 12-fold increase in TB case notification rates. This

increase only started by the mid 1980's and coincided with the onset of the HIV epidemic in the country [18]. On the other hand, as mentioned earlier, these years also coincide with a more evident contraction of the national GDP of Zambia mainly due to drought and the end to most government subsidies for agriculture [12].

The occurrence of this double epidemic led Zambia to be one of the first countries promoting an integrated approach to TB and HIV with the ProTEST initiative. This program aims to facilitate the collaboration between TB and HIV services by using voluntary counselling and testing (VCT) as entry points for TB screening, preventive therapy and home based care. Such integration is becoming increasingly efficient in identifying TB patients among HIV positive people, especially since 2004 when antiretroviral therapy for HIV has become more widely available [17]. The overlap between TB and HIV services has gone further since mid 2007, with the introduction of a diagnostic and testing policy for TB patients, according to which patients are told they will be tested for HIV unless they wish to opt out [19].

In Zambia government-run TB services are free of charge: suspect TB patients are normally seen by a clinician or a nurse in an outpatient department, and asked to submit three sputum samples. Confirmed TB cases are, instead, seen in specially designated areas called "the TB corner". TB treatment in Zambia is a 6 months chemotherapy regimen, during which patients living close by the TB services undergo DOTS (Directly Observed Therapy Short Course). Patients living far from the clinic or who are too sick to come there are supposed to be assisted at home by home-based care givers or community volunteers (treatment supporters), but these services are erratic and often do not operate efficiently [19].

Despite being free of charge, several studies in Zambia have documented severe economic barriers to accessing timely and quality diagnosis and treatment. Before being correctly diagnosed and started on treatment, patients often go through an expensive and tortuous path involving private clinics, governmental health centers, traditional and spiritual healers. The costs of this protracted diagnostic search can often reduce productivity by of adults and affected family.

In a 1998 study on health seeking behaviour TB patients reported a mean of 63 days of symptoms duration before accessing the TB services. Among those indicating a reason for delay, 38% blamed lack of money as the primary cause of delay. According to the study, in seeking diagnosis patients incurred in costs up to 137% of their mean monthly income corresponding to approximately 59 US$ (mainly for transportation, costs for special food and inadequate alternative diagnosis and treatment). Patients also lost an average of 18 working days before being diagnosed with TB [20]. These data were largely confirmed in a more

recent study (2008). This study also showed that the total direct costs as a proportion of income were significantly higher for women than men and three times higher for patients on the clinic-based DOTS, compared to patients on self-administrated treatment strategy [21].

After TB diagnosis, it typically takes several months before a TB patient can return to work. A recent qualitative study has documented how the farming economy of rural Zambia is especially vulnerable if a household member becomes critically sick during the farming season. In this study, half of the TB-affected Zambian households were left severely short of food in the 2006/7 season and such food insecurity was also proven by the anthropometric measurements of children under five [19]. This study found the impoverishment of TB in Zambia being exacerbated by stigma and the lack of any welfare support or food aid by the government or any NGO operating in the study area.

2.2 Basis for a social epidemiology of tuberculosis

2.2.1 The epidemiology of tuberculosis

(a) The prevalent cases of Tuberculosis

TB is a multistage disease in which a susceptible person exposed to an infectious case, may first become infected and then, after a variable interval of time, may or may not develop the disease depending on a variety of circumstances (**Figure 2.3**). Thus, at one given time it is possible to identify in a community:

1. Those who are TB-free;

2. Those who have asymptomatic TB infection;

3. Those who have TB and have been already diagnosed through passive case finding from the health care system and possibly put on treatment. These cases are those notified by the TB services.

4. Those who have TB, but are still undiagnosed and not on treatment and therefore still potentially infectious. Some of these individuals may be unlikely to access the health service for a long time, if at all, and are more likely to be detected through an active case finding approach, such as a prevalence survey. These cases can be also called *prevalent cases* of TB. Because of their prolonged infectiousness, prevalent cases of TB are very important for TB transmission and, thus, particularly relevant for TB epidemiology. TB

prevalent cases should not be confused with incident cases of TB: whereas the first group represents the number of new TB cases detected cross-sectionally 'at one given time', incident cases represent the number of cases of TB observed longitudinally over a time period (usually a year for TB).

Much of this research project is focused only on prevalent cases of TB and not incident or notified cases of TB. It is important to stress this distinction not just because for the need of working with clearly defined epidemiological groups, but at least for two other reasons:

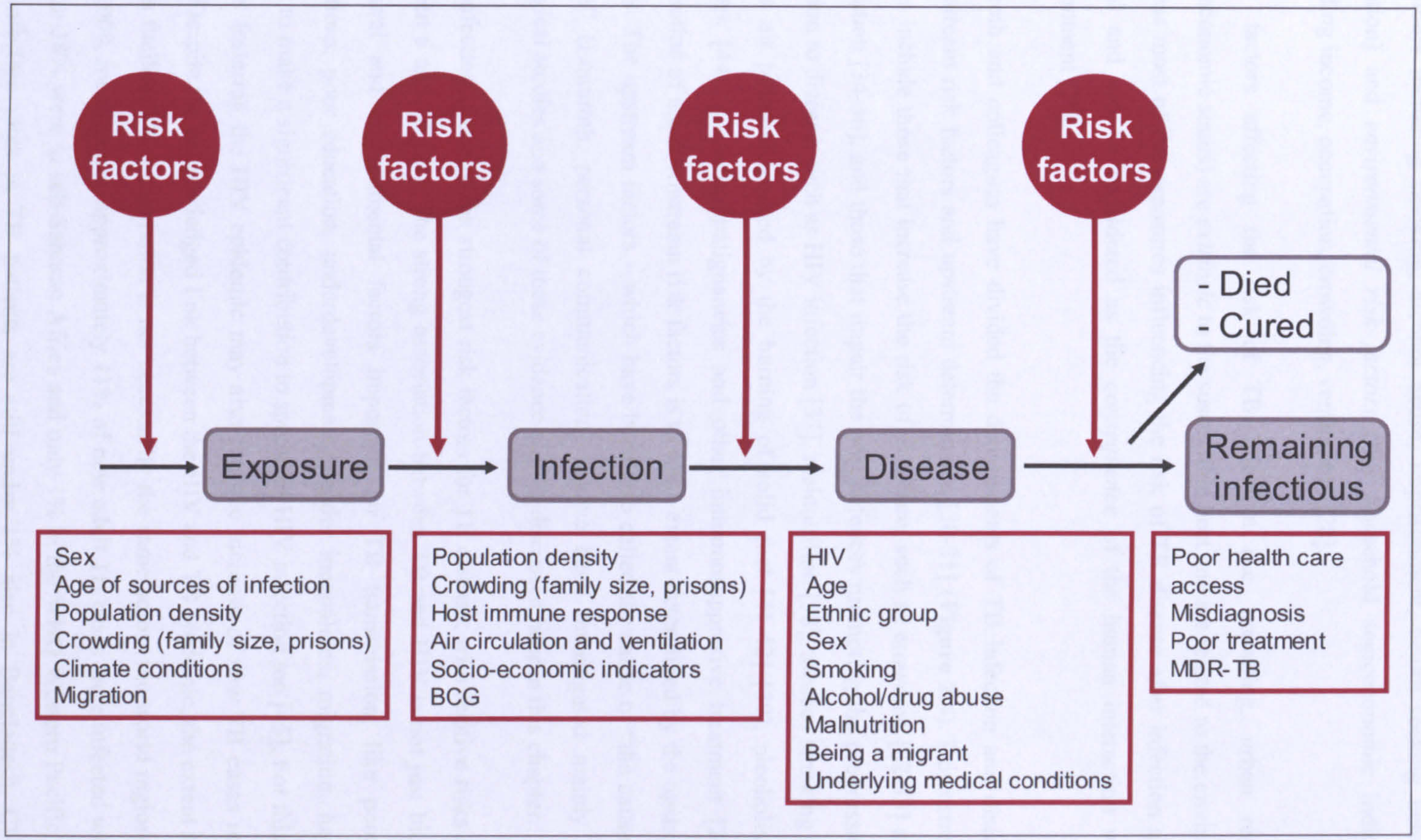
1. Because they spontaneously sought medical attention overcoming all the economic barriers to TB care, notified cases are likely to represent a selected, probably wealthier, population group. As a result, studies on these cases may often result in the paradoxical impression that TB is more common among the wealthiest group of the population [7] or more common in wealthier area [22]. The collection of SEP data from prevalent cases makes possible to reduce this detection bias [23-24].

2. Prevalent and incident TB cases are likely to give different contributions to TB transmission and may be also characterised by different risk factors. HIV is an example of that: a study conducted among gold-miners found that HIV was much more strongly associated with incident cases of TB rather than prevalent cases and therefore that HIV had a much less pronounced effect on TB transmission than expected [25]. A possible explanation is that TB-HIV co-infected patients tend to have a shorter infectious period and therefore are less likely to be identified in prevalence surveys. The shorter infectiousness is partly due to a more rapid progression to extra-pulmonary TB in HIV patients and partly due to the more frequent detection of TB among HIV patients (because of TB screening, for example).

(b) Risk factors for tuberculosis: from exposure to disease

As mentioned above, being a TB case is the result of being exposed to a succession of various factors including: 1) the risk of exposure; 2) the risk of infection; 3) the risk of developing the disease [26] (**Figure 2.3**). The risk factors for infection might be quite different from those for the development of the disease after infection and from those associated with remaining sick after the disease develops. This has important implications for TB prevention and control [27].

Figure 2.3: A model for tuberculosis epidemiology



Modified from[26]

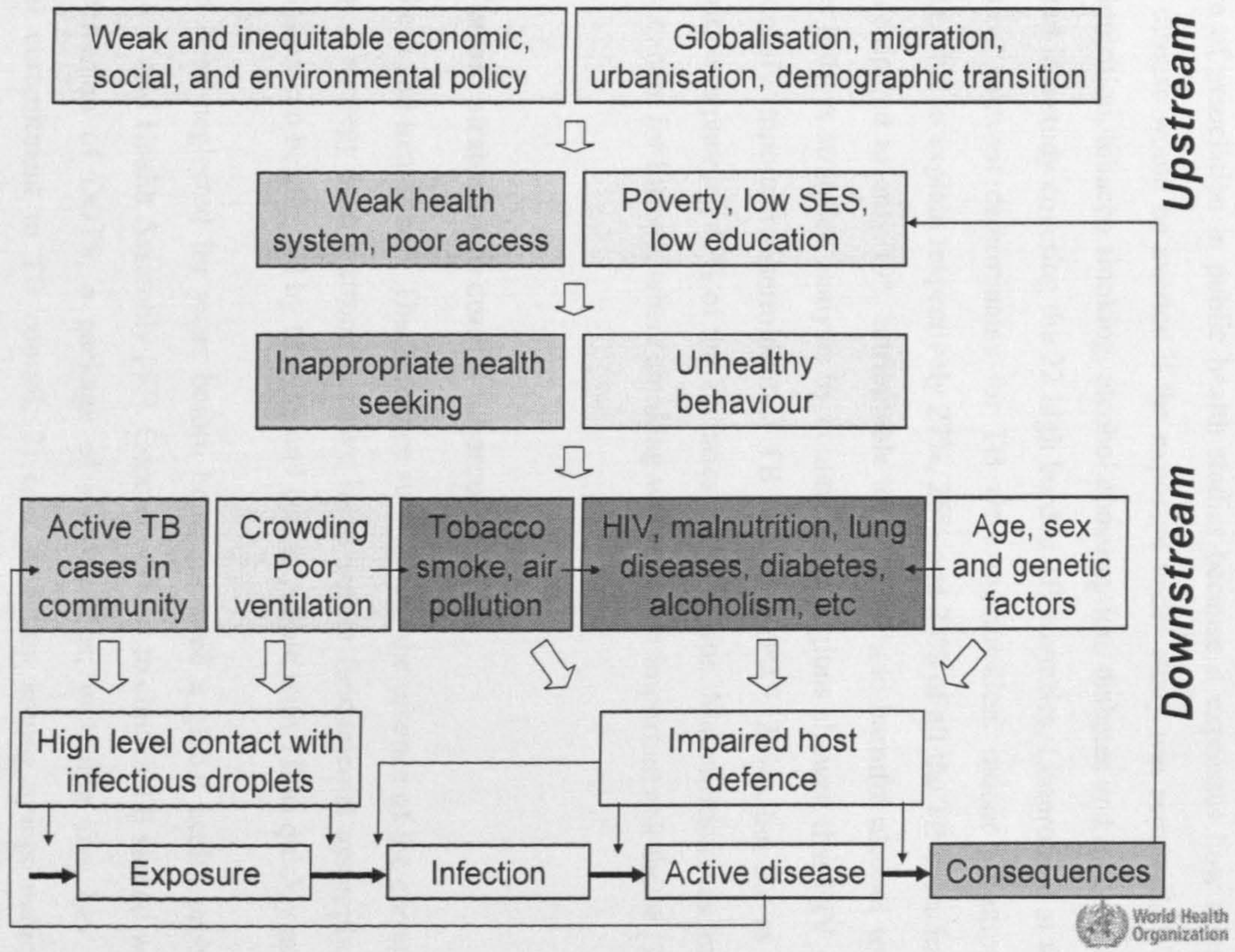
Lienhardt divides the risk factors for TB infection and disease into *host-related factors* (like age, sex, race/ethnicity, immune response, HIV and other underlying illnesses, and hazard behaviours including smoking, alcohol abuse, imprisonment, health seeking behaviours, migration) and *environmental risk factors* (like household socioeconomic indicators - including income, occupation, crowding, ventilation) [28].

Many factors affecting the risk of TB infection (i.e. crowding, urban residence, socioeconomic status) are *extrinsic* to the susceptible host and are related to the environment, whereas most of the exposures influencing the risk of TB disease after infection are host-related and can be considered as the *consequence* of the human interaction with the environment [29].

Lonnroth and colleagues have divided the determinants of TB infection and disease into downstream risk factors and upstream determinants [30-31] (**Figure 2.4**). Downstream risk factors include those that increase the risk of exposure, such as crowding [32-33] and poor ventilation [34-36], and those that impair the host defences promoting the progression from infection to disease, such as HIV infection [37], malnutrition [38], tobacco smoking [39-40], indoor air pollution caused by the burning of solid fuel [41-42] [39], alcoholism [43], diabetes [44], silicosis, malignancies and other immunosuppresive treatment [26]. The distribution of the downstream risk factors is to some extent determined by the upstream risk factors. The upstream factors – which have been also called the cause of "the causes of the causes" (Lonnroth, personal communication) - have been investigated mainly through ecological studies and some of these evidence will be discussed later in this chapter.

HIV infection is one of the strongest risk factors for TB disease, with relative risks reported between 6 and 10 [31]. The strong association between TB and HIV is not just biological: structural and environmental factors important for TB transmission, like poor living conditions, poor education, underdevelopment, gender inequalities, migration, have been found to make a significant contribution to spread of HIV infection too [45]. For this reason, factors fostering the HIV epidemic may also fuel the increase of new TB cases related to HIV. Despite the acknowledged link between the HIV and TB epidemic, the extent to which HIV is fuelling TB transmission is not necessarily the same across the world regions: in the year 2000, for example, approximately 11% of new adult TB cases were infected with HIV, of them 38% were in sub-Saharan Africa and only 1% in the WHO Western Pacific Region. HIV infection rates in TB patients are still under 1% also in Bangladesh, China and Indonesia, suggesting that HIV infection is not the most important driving force of the TB epidemic in every region of the world [46].

Figure 2.4: Downstream risk factors and upstream determinants of tuberculosis and related entry points for interventions



*From Lonnroth et al [30]. The authors have indicated in green the areas were the current global TB strategy has currently his main focus (i.e. case finding, improving access and health system strengthening). Red boxes indicate where National TB control programs could intervene jointly with other disease control programs within the general health care system. Yellow boxes indicate entry points for interventions outside and beyond the health system.

Malnutrition, tobacco smoking, alcohol consumption, diabetes and indoor pollution are generally found to be weaker risk factors for TB than HIV infection, with relative risks ranging between 2 and 4 [31].

However, the public health relevance of a risk factor is not only due to the strength of its association with the health outcome, but also to its prevalence in the population. These two aspects are combined within the Population Attributable Fraction (PAF), a very popular measure of association in public health studies because it expresses how many cases of a certain disease would be averted if the exposure under study was removed. When the PAF for malnutrition, tobacco smoking, alcohol consumption, diabetes and indoor pollution were computed in a study covering the 22 high burden TB countries, Lonnroth et al found that the three most relevant determinants for TB were malnutrition, indoor pollution and active smoking, able to explain respectively 27%, 26% and 23% of all the TB cases included in the study, compared to only 19% attributable to HIV, 13% to harmful alcohol use and 6% to diabetes [30]. A stratified analysis by countries and regions showed that HIV was clearly a tremendously important determinant of TB in Africa (PAF=28%), but not as important as malnutrition explaining 47% of the TB cases in this region. Malnutrition was important in all regions, except for Europe, where smoking was the most important risk factor [31].

2.2.2 Current strategies to control tuberculosis

'Find them and treat them'. This sentence summarises the essence of the current TB control strategy, a strategy with a strong curative focus whose fundamental assumption is that TB transmission can be stopped by treating and curing people with TB as quickly as possible.

TB had been neglected for years before being declared a global health problem in 1991 during a World Health Assembly [47]. Concrete efforts to control TB started with the WHO implementation of DOTS, a package of interventions, including five key elements: 1) political commitment to TB control; 2) case detection among symptomatic patients; 3) standardised short course chemotherapy to at least all confirmed smear sputum positive cases, provided under direct observation (from which the program takes its name); 4) a system of regular drug supply; and 5) a monitoring system for program supervision and evaluation.

DOTS is now widely accepted as one of the most successful public health program: by 2006 over 32 million TB patients were treated under the DOTS algorithm over the world and that over 25 million of them were cured [31]. World Bank analysis also suggests that DOTS is

also one the most cost-effective public health program ever implemented in terms of cost per adjusted quality of life [31].

The DOTS strategy has been recently updated by the WHO [48] in response to a number of challenges that were not included in the original DOTS strategy, including the limits imposed by weak health systems, the inadequate diagnostic and therapeutic tools, and the emergence of multi drug resistant strains of TB (MDR-TB) and the impact of HIV. Especially these last two element led experts in the field to speak about a 'new tuberculosis'[49].

WHO and the STOP TB Partnership have set two epidemiological impact targets for TB: to halve the TB prevalence and deaths rate by 2015 as compared with 1990 levels and – in the long term – to eliminate TB as a public health concern by reducing TB incidence to 1 or less per million by 2050 [31]. Mathematical modelling suggested that these two objectives can be achieved through an annual TB incidence decline of approximately 5-10%, a rate attainable only by ensuring at least the detection of 70% of the incident cases of highly infectious TB and the successful treatment of at least 85% of them (the so called '70/85' algorithm). For these reasons, since 1990, these two outcomes have become the common parameters against which to monitor the achievement of the elimination targets and the performance of the current Global Plan to Stop TB (2000-2015).

Despite the intensified efforts, the recent estimates included in the 2008 Global Tuberculosis Control Report from WHO (WHO) (1990-2006) [4], show that TB incidence, prevalence and mortality are not declining as fast as predicted: globally TB incidence seems to have peaked in 2003 and is now slowly declining by 0.5% per year, a rate less than expected and less than required to achieve the TB elimination targets. The prevalence and death rates registered in 2006 are also far from the TB reduction targets. The TB elimination goal by 2050 seems to be even more unlikely: recent modelling has suggested that even if TB incidence declined by 5-10% as initially predicted, by year 2050 it would still be about 100/100.000 (i.e. approximately 100 times higher than the desired).

Lonnroth et al have warned against excessive trust in the algorithm '70/85' for the detection and treatment of TB cases, seeing this as one possible explanation for the unmet targets [31]. They consider the model to be 'too simplistic' as it takes little account of treatment delays, relapse rates, changes in the virulence of *M. tuberculosis* strains and the role played by risk factors for TB other than HIV. Through mathematical modelling Dye showed that TB incidence falls can be enhanced only by using a combination of control methods, including

drug treatment for active cases of TB, prevention of infection (through vaccination) and the prevention of progression from TB infection to TB disease [50].

Both Lonnroth and Dye do not dismiss the importance of early diagnosis and treatment, but claim that to accelerate the decline in TB incidence enough to get close to elimination by year 2050, it is essential to conjugate the current curative strategy with an intensified effort in primary prevention interventions able to reduce the exposure and vulnerability to TB.

Similar conclusions were drawn by Dubos more than 50 years ago, when he wrote [51]:

> *'In the final analysis, the fight against Tuberculosis can be carried along two independent approaches, by preventing the spread of the bacilli through procedures of public health, and by increasing the resistance of man through a proper way of life'.*

2.2.3 The scope for a social epidemiology of tuberculosis

The acknowledgement of the importance of prevention is a fundamental step forwards the expansion of the current TB control paradigm; however, it still does not answer the more general questions: *why do people have TB today?*

Historical data from Europe and North America suggests that TB was in widespread decline from the middle decades of the nineteenth century onwards, before the introduction of any systematic attempts to prevent its spread [3]. The significant time lapse between the recorded decline of TB and coordinated control efforts forms the basis of the historian Thomas McKewon's theory, according to which TB and other infectious diseases declined as a consequence of wide, social and economic advances rather than because of any specific medical intervention [52]. Today it has been convincingly demonstrated that McKeown was only partially correct: although medical care per se had a marginal role in the incidence or mortality decline for TB before World War II, economic growth alone was not the full explanation. Specific public policies such as patient isolation and the elimination of bovine TB were also important [53-55].

Davies and colleagues have further argued that the relationship between living conditions and historical TB trends requires the careful consideration of environmental (i.e. bovine TB and environmental mycobacteria) and genetic factors (i.e. native immunity), which may have contributed to protection against the more serious and life-threatening pulmonary TB in

some regions of the world [56]. Davies has also argued that while TB rates fell by about 1.7% from the mid-1800s to the I World War as a result of improved living conditions, the TB decline rate approached 11% after the introduction of treatment in the 1950 together with BCG vaccination [57].

The picture has been further complicated by the emergence of HIV and multi-drug resistant TB (MDR-TB) and some authors have questioned whether it is appropriate to extrapolate the evidence for Europe and North America to modern developing countries [58]. In particular, is it still correct to use historical data on TB and socioeconomic development to predict changes in TB trends? Even if the present world order would permit a rapid socioeconomic improvement in developing countries would this be enough to revert the current TB epidemic [58]?

These questions are at least partly addressed in a recent study showing that indicators of human development (such as the Human Development Index, Gross Domestic Product per capita, under 5 mortality, access to clean water and adequate sanitation, and health expenditure per capita) account for the recent TB incidence decline observed in 134 countries over the period 1997-2006 more than DOTS [30]. The notion that DOTS programs have successfully reduced TB morbidity and mortality, but have had far less impact on TB control is well documented in another recent article from Obermeryer and colleagues [59] and from a series of case studies at country level. For example, many countries (i.e. Indonesia, part of Latin America and the Easter Mediterrenean region) have experienced significant reduction in TB incidence before the implementation of DOTS [60-61]. By contrast, in countries like Vietnam [62] and Morocco [63] the decline in the incidence has been less than anticipated despite the successful implementation of the DOTS programs.

In conclusion, DOTS programs have not yet become the main determinants of trends in TB incidence in any region of the world. Rather, broader socioeconomic development and access and quality of health services seem to be of more importance.

If it is a 'new tuberculosis' that we are facing [49], resulting from the complex interaction of biological and socioeconomic factors, then 'a new approach' to the study of TB epidemiology is also needed, able to investigate the risk factors for TB and, at the same time, "those economic and social forces that produce differential exposures that often yield health disparity". This is ultimately the scope of social epidemiology [64].

Grange and colleagues have called this new approach a 'biosocial model', a paradigm for the study of TB, according to which TB epidemiology cannot be explained by the simple recourse to the 'nature' versus the 'nurture' argument, but only through the integration of

sociological and economic, behavioral factors as well as the possibility of variations in the native immunity across populations and the change in the virulence or resistance of the tubercle bacillus [58].

This biosocial model reflects the concept of 'embodiment' formulated by Nancy Krieger [65], according to whom:

> *'We, as humans, are simultaneously social beings and biological organisms. [...]. No aspect of our biology can be understood in the absence of knowledge of the history and the individual and societal ways of living'*

In a biosocial or embodied view, none of the exposures discussed in this chapter intrinsically 'cause' TB infection or disease, but rather acts as mediators in the causal pathway between socioeconomic factors and TB. As argued by Grange, although complex, such a model is needed if we are to give differential weighting to the risk factors and the process affecting the distribution of the risk for TB infection and TB disease.

A social epidemiological approach to TB requires firstly the definition of those socioeconomic factors that shape people's chance to be healthy [66] and the development of hypotheses about the causal pathway linking these factors to TB [58]. In the following sections I will introduce how socioeconomic exposures have been conceptualised and measured in time and some of the most influential models of aetiological pathway applied to the interpretation of health inequalities.

2.3 Socioeconomic position: theories and methods

2.3.1 Definitions and potential sources of ambiguity

Despite the increasing importance of health inequalities in the public health agenda, studies on the socioeconomic determinants of health still suffer from at least two limitations when it comes to the definition of the exposure of interest: 1) *linguistic*: the terms poverty, socioeconomic status, socioeconomic position, social determinants and social class are often considered interchangeably despite their different theoretical bases; 2) *conceptual*: ambiguity in the terminology adopted also generates confusion concerning which aspect of the measure employed represents a risk for health (and why) and therefore what intervention may be more appropriate to reduce the health disparities observed [67].

One of the terms greatly contributing to the generation of the above ambiguities is 'poverty'. The 1995 World Summit for Social Development in Copenhagen [68] and the Millennium Development Goals (http://www.un.org/millenniumgoals/) both define poverty as a multidimensional phenomenon that goes well beyond a narrow lack of material consumption or resources, to encompass notions of material well-being, an absence of infrastructure, a lack of power and voice, and unravelling of social structures. Similarly, the World Bank defines poverty as an unacceptable deprivation in well-being, in terms of material deprivation, human deprivation, vulnerability, destitution and social isolation [69]. A joint paper between WHO and the Organisation for Economic-Co-operation and Development on poverty and health has provided an operational definition of poverty including five core dimensions of deprivation in human capabilities: economic (income, livelihoods, decent work), human (health, education), political (empowerment, rights, voice), socio-cultural (status, dignity) and protective (insecurity, risk, vulnerability) [70].

Currently no epidemiological quantitative method adequately measures all the different components of poverty. In order to achieve a better understanding of the mechanism underlying the association between poverty and health, our definition and conceptualisation of poverty should be as comprehensive as possible. On the other hand, the use of a too articulate or complex poverty construct in the studies of health inequalities may make the interpretation of the results cumbersome and – more importantly – make difficult the translation of these observations into policy actions.

Another limitation is that poverty identifies a complex construct that can be defined in relation to: "need", "standard of living", "limited resources", "lack of basic security", "lack of entitlement", "deprivation", "exclusion", "inequality", "class", "dependency", and "unacceptable" hardship [71]. Because the term of 'poverty' implicitly assumes a low socioeconomic status, its use in epidemiological studies may prematurely assign a direction in the association between living conditions and the health outcome under study.

To avoid some of these definitional issues, 'SEP' has become the favoured umbrella term to describe ranked socioeconomic measures in studies of health inequalities, and I will adopt it throughout this thesis. In the following paragraphs I will define SEP, illustrate how it differs from the definitions of socioeconomic status, social class or socioeconomic determinants, and discuss how the use of these other terms may be less appropriate than SEP.

(a) Socioeconomic position (SEP)

According to the Commission on Social Determinants of Health (CSDH), people's socioeconomic position (SEP) is the result of the socioeconomic stratification operated by a broad spectrum of structural, cultural, and functional factors in a given societal system [72]. These attributes create social stratification and hierarchies mainly through power, prestige and access to resources, which ultimately configure the differential health opportunities of people in a society [72]. In more health terms, SEP has been also defined as: "the social and economic factors that influence what position(s) individuals and groups hold within the structure of society, i.e., what social and economic factors are the best indicators of location in the social structure that may have influences on health" [73]. Another definition describes SEP as reflecting social hierarchies in which persons or groups can be arranged along a ranked order of some attribute such as income or educational level [74]. These definitions both acknowledge two different aspects of SEP: resource-based and prestige-based measures [71]. Resource-based measures refer to the material and social resources and assets, including income, wealth, and educational credentials; while prestige-based measures refer to individuals' rank or status in a social hierarchy, typically evaluated with reference to people's access to and consumption of goods, services, and knowledge, as linked to their occupational prestige and educational level [71].

Because SEP is understood to be multi-dimensional authors should always clarify what they are actually attempting to measure in their work and how they have operationalised this choice [71]. Additionally, the proposed causal pathway to health should be specified if authors aim to suggest meaningful interventions able to reduce the observed health inequalities [75].

The term SEP is often erroneously used in place of socioeconomic status (SES). Krieger and colleagues have argued that differently from SEP, SES represents a narrower concept, privileging the 'status' of an individual over material resources [71]. In this respect, SES is better identifiable with the prestige-based component qualifying someone's SEP [71]. It follows that SES and SEP should never be used as synonymous since two people may share the same 'status' but occupy different 'position'.

The concept of social stratification is helpful to clarify the distinction between SEP and SES. The term stratification is used in sociology to identify social hierarchies in which individuals or groups can be arranged along a rank order of some attribute [72]. The variables most commonly identified as social stratifiers are income, education and occupation; it is also

increasingly recognised that gender, race/ethnicity and sexuality are becoming important forms of social stratification mainly through discrimination and social exclusion [76].

Both in industrialised and developing countries, SEP and SES are obviously two forms of stratification of a population, but differ because in the case of SEP people are stratified by social *position* and economic resources, whereas with SES, the economic resources are not relevant for the stratification process. The term social position is sometimes confused with social class although it is widely accepted that they represent two different concepts.

(b) Social class and social position

The definition of social class involves the idea of hierarchy or ranking 'from top to bottom' of a society and incorporates concepts like 'prestige' and 'social honour'[77]. According to Krieger, social class is a construct created by societies and arising logically and materially prior to its expression in the distribution of occupation, income, wealth, education and social position. In this view, social class is the upstream stratifier distributing those opportunities that ultimately contribute to someone's SEP [78].

Most measures of SEP are indicators of *position* in the social hierarchies in accordance with education, income or prestige, whereas social class is more appropriate to describe the position of the individual within a web of social relationships based on economic, political and cultural power [74]. However, since it is very difficult to measure dimensions such as power, control or prestige, researchers have increasingly combined occupation and education as the sole measure of class.

Two figures have mainly contributed to the conceptualisation of social class: Karl Marx and Max Weber. In a Marxist view of the world, societies are stratified into social classes that are characterised by the inherent conflicts between exploited workers and the exploiting capitalists who control the means of production. According to Marxism, class is not an inherent property of individual human beings, but arises from the organisation of capitalist societies [79].

The Weberian tradition views society as stratified in multiple ways, including class, status and power. Class implies ownership and control of resources and is essentially an economic measure; status is considered to be prestige and honour in the community while power is related to the political context [72]. Differently from Marx, Weber recognised the limits of the capitalist system not because it exploited and alienated workers, but because it generated

different set of skills, knowledge and assets contributing to what Weber called 'life chances' or more simply individuals "opportunities". For Weber, the real issue in the capitalist system was not exploitation, but the unequal distribution of "opportunities" produced by this system [79].

Most epidemiological studies of health inequalities are based on a Weberian view in the sense that more than the skills, knowledge, or resources held by an individual per se, what is really relevant to health are those structural factors driving the distribution of these 'life chances' or 'opportunities' [79]. This last point raises another source of ambiguity concerning the term of SEP.

(c) Social determinants of health and social determinants of inequalities

Graham has argued that in the study of health inequalities there is the tendency to conflate the concept of social determinants of health and social determinants of inequalities [80]. In particular, SEP (at household and community level) is often erroneously used to indicate those social determinants promoting or undermining the health of individuals and populations, when, in fact, it represents the social process underlying the unequal distribution of the social determinants of health between groups occupying unequal position in the society. By contrast, the social determinants of health instead refer to all those material, psychological and behavioural circumstances linked to health and generically indicated as 'risk factors' in the conventional epidemiological language [80].

This distinction is more than just conceptual because it can feed the policy assumptions that health inequalities can be diminished by policies that focus only on the social determinants of health [72]. The history of the last 30 years suggests that even if the significant improvement in health determinants (i.e. rising living standards, declining smoking) has caused a parallel improvement of people's health, it has not broken the link between SEP and health. Thus, individuals' and populations' health improved, but inequalities persisted [80].

As argued by Graham, the distinction between the social causes of health and the factors determining their distribution in the population should be emphasised because "tackling the determinants of health inequalities is about tackling the unequal distribution of health determinants" [80].

Recently, the Commission for Social Determinants of Health (CSDH) has tried to overcome this linguistic ambiguity by introducing a new terminology: they adopted the term *structural determinants* to refer to the social determinants of health inequalities used by Graham. The structural determinants operate through a series of *intermediate factors or mediators*, corresponding to the social determinants of health [72].

(d) Socioeconomic position at area level

SEP can be defined at different points of the lifespan (e.g. infancy, childhood, adolescence, adulthood), whereby the relevant time period depends on the presumed exposure and the postulated causal pathway [78]. SEP can be also measured at three complementary levels: individual, household and neighbourhood. Each level may independently contribute to the distribution of exposures and outcomes.

To effectively examine the effect of community SEP it is necessary to be clear on how community is defined and how area-based socioeconomic measures (ABSM) are operationalised. In health research, the terms neighbourhood and community have often been used mutually to refer to 'a person's immediate resident environment, which is hypothesised to have both material and social characteristics potentially related to health' [81]; Despite this formal definition, a recent review on the 'neighbourhood effect' [82] revealed little consistency in the way neighbourhood is defined in the literature. Such variability is probably due to the fact that the size and the definition of the relevant geographic area may vary according to the process through which the area effect is hypothesised to operate and the outcome being studied [81].

Moreover, often researchers conducting large quantitative studies on the community characteristics effect have to rely on existing administrative definitions for which standard data are available (i.e. census data). This approach may be the most practical alternative, however very rarely the political, administrative, geographical boundaries of an area correspond to the actual social pattern of the individuals and to the real spatial distribution of the causal factors linking the context to health.

2.3.2 Measurement strategies

(a) At household and individual level

Because measures of SEP are context and time specific no universally accepted measure of SEP has emerged. As a general rule, the choice of measure of SEP should depend on how SEP has been conceptualised and how SEP is thought to be linked to health. To review all the different strategies implemented for the measurement of SEP goes beyond the purpose of this thesis, I will therefore concentrate only on the two approaches most commonly used:

Income and expenditure data: Initially economists and analysts focused on these two measures, based on the assumption that a person's material standard of living largely determines his/her well-being. Income or expenditure data have been called 'the best measure of the economic component of living standards'[83]. Because income is closely linked to material living conditions, its use is particularly appropriate when the lack of tangible, material, resources is considered to be the explanation of the uneven distribution of a health outcome across the socioeconomic gradient.

It is not money itself that can have a link with health, but the way this money is converted into health enhancing commodities and services via expenditure that may be more relevant in the interpretation of the relationship between income and health. As a result, expenditure – the total sum of the household expenditures over a wide range of items and time - is often considered a better and more direct measure for the analysis of health inequalities. Ignoring home-produced food would greatly understate the expenditure levels of rural households. For this reason, researchers prefer to speak in terms of consumption expenditure: this measure takes into account the household own production (not captured by expenditure only), and in this respect represents an even better measure of welfare than simple expenditure alone.

The choice of consumption expenditure over income is also dictated by a variety of difficulties involved in measuring income in developing countries, including the seasonal variability in such earnings, and the large share of income in developing countries that are from self-employment both in and outside of agriculture [69].

The use of expenditure data presents limitations too. Firstly, unlike industrialised countries, consumption and expenditure surveys are intermittent at best and may present some reliability issues: in developing countries, in fact, consumption expenditure data are collected with at least a recall period of 14 days. These recall data are prone to measurement error, not all of which is random.

Secondly, when computing the aggregate measure, it is necessary to make a number of significant assumptions, including: 1) the choice of price deflators: price deflators are statistical tools used to convert current prices into inflation-adjusted prices and, thus, used to

make prices comparisons over time. In low income countries the regional and seasonal variability of prices can be dramatic, making the choice of deflators quite problematic; 2) the choice of an equivalence scale. These scales allow the estimates of household expenditures adjusted by different household size and composition. Even if robust measures of material resources can be obtained, disagreement remains whether they should be adjusted by the number of adults per household or number of household members; and, 3) the type of expenditure that should be taken into account.

Finally, the collection of expenditure data generally requires lengthy, time-consuming, longitudinal surveys over a long follow up period, because expenditures – as much as income - are unlikely to be stable in time.

The assets based index: because of the limitations of income and expenditure data, experts have been increasingly committed to developing an alternative measure of SEP. Ideally this alternative measure should: 1) be robust, but less data-intensive and subject to smaller measurement error than income/expenditure data; 2) be feasible and compatible with the financial means of and technical capabilities of government statistical offices; 3) be able to identify and characterise the profile of the poor, but also provide sufficient information on the SEP spectrum of the population under study, across which to read and interpret health inequalities; finally 4) measure a meaningful, concrete, aspect of SEP that can be addressed by social policy interventions [69].

The most widely adopted method to meet the above requirements is the asset index approach. Often in literature, wealth index, asset index and socioeconomic index are used interchangeably. In this thesis I will use the term asset index as it seems to more accurately reflect the nature of the index.

An asset-based index is a composite measure describing the household living conditions in terms of durable goods ownership, housing characteristics and access to community services. By pooling together different aspects of a household's material condition, the asset-based index captures at least one dimension of SEP and it considered to be a rational, simple, reliable alternative to consumption expenditure [84-85].

The use of asset-based indices has become common in the analysis of Demographic Health Surveys (DHS): these surveys do not contain income and expenditure data, but collect information on ownership of a range of durable assets (e.g. radio, TV, fridge), housing characteristics (e.g roofing, flooring material, toilet facilities) and access to basic services (e.g. primary schools, clinics, markets, electricity, water supply). Initially all these indicators were observed individually in respect to their effect on health, subsequently researchers from

the World Bank, United Nations Development Programme and Macro International, attempted to try to combine them within composite indices using different methodologies [69, 85-86].

A key issue for building an asset index is how to assign a weight to each indicator, so as to establish the relative contribute of each indicator to the overall index. A number of different techniques have been used in the literature, ranging from very sophisticated weighting strategies to the aggregation of assets by simply creating a count variable based on the number of assets owned [87].

Among the weighting strategies, the simplest approach is to assign equal weights to the ownership of each asset or presence of each household dwelling characteristic. Such an approach assumes that the value of the ownership of a radio is the same as having access to a flush toilet, which in turn is the same as having safe drinking water. Clearly this is not the case because assets are unlikely to have equal meaning in terms of SEP. This approach is not only too simplistic and arbitrary, but there is also evidence that equal weights wealth indices have poor agreement with consumption data [88].

Statistical weighting procedures include Principal Component Analysis (PCA), which is one of the most frequently employed in the literature on socioeconomic determinants of health [89], factor analysis (a similar alternative to PCA) [90] and data regression [91]. PCA and data regression are the two measurement strategies adopted in this thesis: the rational of this choice as well as the methodological aspects will be described in **Chapter 4.**

A limitation of asset-based indices is that they measure relative rather than absolute SEP. Consequently an asset-based index cannot be used to construct poverty lines like income and expenditure can. However, this is often not necessary for health studies because we are interested in exploring the socioeconomic gradient in a health outcome distribution (i.e. health inequalities), rather than knowing the prevalence of a certain disease in those falling below a certain poverty line. These indices also generally measure household SEP making the index of limited use when we are more interested in individual-level SEP and when we assume that individual and household SEP could affect health through different pathways [92-93]. Finally, an asset index does not incorporate the value nor the age and the quality of the assets [69].

Despite these limitations, the advantages of using an asset index in place of income/expenditure data are compelling [90]: assets are fewer and easier to measure; they are considered less subject to reporting bias as they are based on simple question or direct observation by the interviewer; the standardisation of the questionnaire is less of a problem

and in many cases when data on physical assets such as durables, human capital and housing characteristics are collected there is no need to consider currency deflation. Finally, because they are likely to be based at least partially on economic wealth and because households assets are unlikely to change in response to short economic shocks, assets-based measurements are considered to be able to represent the long-term SEP of a given household.

(b) At community level

As for the household SEP indices, area-based socioeconomic measures (ABSM) of SEP should be derived from valid, reliable, and accessible data using appropriate statistical techniques and should be created so that they represent constructs suitable for policy interventions [94].

A recent review of the types of SEP constructs used in studies addressing the effect of area-level socioeconomic indicators, revealed that a wide range of indices is normally taken into account, in contrast to a limited diversity of approaches used to measure these constructs [95]. Usually, these studies either consider separately one or more indicators of deprivation or combine them into composite indices using data reduction strategies[96]. The advantage of using separately a number of socioeconomic indicators or domains is that we can attempt to explore the multidimensional nature of the context and assess the relative contribution of each specific dimension; on the other side, the use of composite indices has the advantage to provide simpler understanding of the results. For these reasons, authors are often advised to use a combination of single indicators and composite indices so to increase their chance to assess and understand the effect of SEP at area level on a health outcome [82].

Whatever the approach adopted, the measures adopted in the analysis should reflect the theory postulated from the investigator on how the context affect health, in other words the measures have to be conceptually based. All too often the selection of community indicators is driven by the data availability in population surveys and census rather than theories. As a result, one of the main problems in the study of the association between the ABSM and health is the lack of theories and conceptual frameworks about plausible, social, physical and psychological links between specific features of the context and the specific heath outcome [97].

After this overview of the main definitions and methods used for the measurement of SEP at household and community level, I will now move into the description of the most relevant

theories developed to explain health inequalities. Some of these theories are suitable for the interpretation of TB inequalities and have inspired the hypotheses formulated in this thesis.

2.4 Theories of diseases distribution

2.4.1 Psychosocial and material/neo-material explanations for health inequalities

According to Mel Bartley [77], the many aetiological models developed over the past years to explain health inequalities are essentially a combination of four theories: 1) the behavioural and cultural; 2) the psychosocial, 3) the material/neo material one; and 4) the life-course approach one.

Nancy Krieger [53] has adopted a slightly different classification and identified three main frameworks currently invoked by social epidemiologists. These theories are not mutually exclusive and can be designated as follows: 1) psychosocial approaches; 2) social production of disease/political economy of health (corresponding to the material/neo material mentioned above); and 3) ecosocial theory and related multilevel frameworks. All these approaches represent what Krieger has defined as 'theories of disease distribution'. They differ in their respective emphasis on different aspects of social and biological conditions in shaping population health, in how they integrate the biological and social explanations, and, thus, in their recommendations for action.

In this thesis I have decided to focus my attention only on the material/neo-material and the psychosocial model not just because in the literature they have been often seen as opposite and competing views in the explanation of health inequalities [98], but also because they are the theories that drove the formulation of my hypotheses in the study of the TB inequalities in Zambia.

(a) The material model

According to the material model, health inequalities arise as a result of differential exposure to material disadvantage that accumulates over the life course.

A simplistic interpretation of this model is that the health status of an individual can be affected by: better homes (i.e. warmer, cleaner, better ventilated) and workplaces; better occupation (either as type of job, higher income or less hours of work); less exposure to unhealthy places, and better diet (in terms of quantity and quality) [99].

In a more complex definition from Blane and colleagues, the materialist model refers to all those experiences arising as a consequence of social structure and organisation, over which the individual has no control [100]. This model, also identifiable with the 'political economy of health' theory described by Lesley Doyale and Jamie Breilh in 1979, explicitly addresses economic and social privilege and inequality as the root or 'fundamental' causes of inequalities in health [53].

More recently, a *neo-materialist* explanation has been proposed according to which the different SEP of individuals in a community is the result of historical, cultural and political-economic processes. These processes influence the private resources available to the individuals and shape the nature of public infrastructures – education, health services, transportation, environmental controls, availability of food, quality of housing – that form the neo-material matrix of life [101]. The central element of the neo-materialistic explanation is that the observed health inequalities across countries have to do with the differential provision of public services [102]: under this interpretation, the effect of income inequality on health reflects both the lack of resources held by individuals and the systematic under-investment in a wide range of community infrastructures [103]. The added value of the neo-material model is the introduction of the context as a fundamental dimension to the analysis of socioeconomic determinants of heath inequalities.

(b) The psychosocial model

Over the last two decades an increasing number of evidence, mainly from industrialised countries, is suggesting that material deprivation at individual or community level are unlikely to fully explain health inequalities and that we should move beyond the simplistic view of the material model: firstly, strong evidence suggest that even in rich countries there is a clear socioeconomic gradient in health and that to low socioeconomic position always corresponds lower morbidity or mortality despite the fact that people in the lowest socioeconomic group still live well above the poverty line threshold below which the causes of material deprivation operate [104]. Secondly, among industrialised countries it is not the richest ones that have the best health, but those that have the smallest income inequalities between the rich and the poor groups of the society [105]. Finally, there is consistently little relationship between average income (gross domestic product per capita) and life expectancy in rich countries, suggesting that absolute material standards are not, in themselves, the key [106].

These observations all form the theoretical background for the formulation of the psychosocial model according to which – in addition to direct effect of absolute material circumstances – psychosocial factors such lack of social support, insecurity, social isolation, socially hazardous environment, poor control and autonomy in the work place as well as adequate reward and depression contribute to health inequalities in a population. The psychosocial model for health inequalities is one of the most widely researched theories of health inequalities nowadays because it moves from the simple analysis of the physical or material threats for the health to the study of how income inequalities make people feel [77].The most convincing evidence supporting this model come from neuroscience studies showing how the effect of psychosocial stressors on the neuroendocrine system results into a variety of biological processes that overall determine the health status of an individual [107]. Most of these effects are known under the "fight or flight" response, according to which when the body is under external threats it responds by activating one or two of two responses circuits: the sumpathetic-adrenomedullary and the pituitary-adrenocortical. The effect of these two responses have been mainly put in relation of work-related stressors and cardiovascular diseases: several studies have proved not only that low control in the work place predict heart diseases independently from social status [108], but also that low control in the work place accounted for about half of the social gradient in cardiovascular diseases [109]. More recently, studies on happiness have shown that beyond some point average happiness within a country is almost completely unaffected by increases in its average income level [106]. Probably the most convincing argument on the importance of psychosocial factors is the observation that not always to poor material conditions correspond poor health and, as opposite, relatively wealthier living conditions do not necessarily translate in better health: for example, in 1996 Black American men had a median income of $26,552 and a life expectancy of only 66.1 years. Men in Costa Rica had a mean income of only $6,410 and still their life expectancy was 75 years. The fact that a four time higher mean income still results in a life expectancy nine years shorter can be only explained by assuming that the health of Black people in the Unites States has more to do with psychosocial effects of relative deprivation, such as education discrimination, racism, violence, social and family disruption and gender discrimination than material conditions [106].

The most relevant aspect stressed by the psychosocial explanation is the importance of relative deprivation, over absolute deprivation, as a predictor of the observed health inequality. If these observations are correct, health is determined not just by what someone has, but what someone has relative to what other people have [102]. Another important element introduced by the psychosocial model in the discussion concerning health

inequalities is that in a society the uneven distribution of income affects the health of everyone, not just the most deprived individuals. This accounts for the fact that in a society we often observe a health gradient across the socioeconomic groups of the population, not just between the poorest and the richest.

The great emphasis given to psychosocial factors and their physiological effect is the real innovative element of this model, but, at the same time, it makes the model not only difficult to test, but also somewhat controversial. As argued by Bartley, the idea that the perception of relative disadvantage (either material or social) may represent a threat to our health can be challenging to our image of human beings. In other words, can we really die of envy? [102]. Other authors have argued that not only do structural, political-economic processes that generate socioeconomic differentials exist before their effects are experienced at individual level, but also that the psychosocial model encourages understanding of the effects on health in a vacuum [101]. The authors do not deny the negative psychosocial consequences of income inequality, however, they argue that the interpretation of links between income inequality and health must begin with the structural causes of inequalities and not just focus on the perception of them [110].

The material/neomaterial and psychosocial explanations of health inequalities are not mutually exclusive. Rather they should be considered as complementary and both contributed to the conceptual frameworks that have been developed over the past 15 several to explain health inequalities, including the present study.

A full understanding of health inequalities may require the combination of these theories into a single, more sophisticated, explanation. For Nancy Krieger, for example, both the theories above are somewhat inadequate: in the psychosocial model the material circumstances creating the conditions for the psychosocial insults to health are not properly acknowledged, at the same time in the material/neomaterial model is the role of the physiological effect of the material circumstances remains 'opaque' [53]. Moving forward requires multilevel frameworks integrating the social, biological, historical and ecological perspectives to develop new insights into the determinants of population distribution of diseases. According to Krieger, 'more than adding the 'biology' to 'social' analysis, or 'social factors' to 'biological' analysis, the ecosocial framework begins to envision a more integrated approach capable of generating new hypotheses, rather than just simply reinterpreting factors identified by one approach (e.g. the biological) in terms of another (e.g. the social) [53]. Again, central in the development of this theory is the concept of embodiment 'referring to how we literally incorporate biologically influences from the material and social world in which we live, from conception to death' [111].

2.4.2 Theoretical models for the effect of community living conditions on health

As anticipated in this Chapter, few studies addressing the context effect provide an explicit description of the conceptual framework or theories underlying the link between the context level variables explored and the health outcome under study. The poor conceptualisation of this association has been mainly attributed to the lack of specific theories in the field of sociology, social geography, and community psychology applicable to health outcomes. Most of these disciplines, in fact, have formulated theories relevant for deviant behaviours, crimes, and discrimination, which may not adequate for public health interpretations [94, 96]. Social epidemiologists have been therefore invited to start developing frameworks and testable hypotheses or to adapt existing theories in order to identify plausible and testable processes to interpret the community effect [96].

One theory formulated by Sampson is that neighbourhoods living conditions may be responsible for a negative effect on individuals' health by undermining social cohesion or collective efficacy, which the author defined as 'the linkage of mutual trust and shared willingness to intervene for the public good' [82]. Sampson has identified at least two other potential neighbourhood mechanisms, based on social capital and institutional resources (e.g. the quality and quantity, and diversity of institutions in the community that addressed the needs of the population resident in a certain area). Sampson, however, notes that the social and institutional processes in many cases do not explain all – and not even most – of the associations observed [82]. More often factors such as concentrated disadvantage, community affluence and stability remain the most important predictors of the health outcomes of interest. In conclusion, social capital and collective efficacy do not occur in a vacuum, but appear to emerge mainly in environments with sufficient socioeconomic resources and stability [82].

In another model proposed by Robert, the community socioeconomic context affects health through two major pathways: a) by shaping the socioeconomic position of the individuals, and b) by directly affecting the social, service and physical environments of communities shared by the residents, which then affect the individual characteristics, conditions, and experiences of individual that more directly impact health [112]. These two pathways are likely to affect each other, but they can be also seen as independent.

The first pathway suggests that communities with different socioeconomic profiles essentially represent those opportunities and constraints that ultimately determine people education's attainment, job prospects, and income level achievement. The social environment reflects the web of social relationships, social organisation of the individuals in a society. These social ties can influence individuals' behaviours. So for example, living in

communities with low socioeconomic profiles may negatively affect a person's health promoting attitudes and behaviours by exposing this person to neighbours that are less likely to practice health promoting behaviours compared to neighbours with higher socioeconomic position. The social links among individuals are also the social circuits along which information may flow about how to recognise and respond to health threats [113]. The community environment can also negatively impact people's health by affecting the access to high quality services for all the residents or by increasing the exposure of people to pollution (i.e. lead, asbestos, pest infestation), poor housing quality and unhealthy working conditions [112].

Several authors [112, 114] have hypothesised that even if the community socioeconomic context may affect the health of all or most residents, the nature and the extent of this affect may be somewhat different depending on the individual socioeconomic position. Thus, that community and individual socioeconomic characteristics may interact with each other.

For what concerns specifically communicable diseases, Montgomery and colleagues have introduced the concept of 'health externalities' to refer to the spatial proximity of diverse urban populations. These externalities are ultimately responsible of the contamination and transmission patterns that may make the life in a urban area more unhealthy compared to a rural one, despite the generally easier access of the urban residents to transport and better health services [113].

2.4.3 The social epidemiology of tuberculosis: historical and modern studies

(a) Historical studies

To the best of my knowledge theories of disease distribution have not been formally employed to help explain the large TB socioeconomic inequalities observed between countries, within countries, within community and over time. However, most of the available literature on the distribution of TB has focused on the importance of material living conditions such as nutrition, working conditions, housing quality and crowding and all the other physical and social circumstances able to affect the risk of TB above and beyond the characteristics and behaviour of the individual.

This *material/neomaterial* approach to the study of TB started in mid 1960s with the influential historical and demographic studies of Thomas McKeown. As described earlier, according to McKeown the population growth observed in England and Wales in the 19th

century was essentially due to a drastic reduction in the bulk of decline in airborne diseases mortality, including TB. This reduction was mainly due the incremental secular changes in population standards of living, and particularly nutrition, rather than to medical advances. The overall conclusion of McKeown's work has been reassumed in a more recent book on the rise and decline of TB in Japan [115]:

> *Nutrition is the one of the most powerful of all socially and environmentally determined influences on the development and retardation of active Tuberculosis [...]. Neither medicine nor public health measures had a significant impact on the mortality from the disease until after World War II.*

McKeown's conclusion was essentially the result of an exclusion process, in which possible explanations of the population growth, including the changing virulence of microorganisms, clinical interventions and better living conditions, were ruled out by logical reasoning rather than through empirical testing. It is this lack of quantitative data derived, for example, from animal models, the production of food in England and Wales or the more recent experience from developing countries [116], that contributed to generate the vivid debate between the 'nutritionists' (those adhering to McKeown thesis and emphasised the importance of nutrition to enhance resistance to TB) and the 'contagionists' (those that emphasised the importance of sanatoria and the isolation of patients to reduce the source of infection) [3].

Historian Leonard Wilson, for example, argued that McKeown did not review the extensive earlier work of Arthur Newsholme (1908), according to whom TB had declined in England as an indirect consequences of the 1834 Poor Law infirmaries and workhouses [117], a law requiring that poor people affected by TB were sent to these institutions rather than the new and relatively scarce sanatoria until they were cured. Newsholme reviewed data from nutrition, living standards and segregation of consumptive patients from different countries and found that segregation, rather than the improvement in nutrition or standards of living, was the only factor consistently correlated to the declining deaths rates from consumption observed in England and Wales, Scotland, Prussia, Minnesota and New York city [117].

Wilson used Newsholme's evidence to completely dismiss McKeown's theory. However, like the nutritionists he did not provide a body of quantitative data to support his critics: the mortality data for 19th century are, in fact, incomplete and neither Wilson or Newsholme provided a direct measure of the rate of patients segregation or controlled simultaneously for potentially confounding variables, such as education, urbanisation, poverty, etc. Finally

Wilson simplified Newsholme's work to a monocausal explanation of TB declined, whereas Newsholme never undermined the importance of nutrition and living conditions. He simply argued that among the factors considered, segregation probably had a more predominant effect and varied more consistently with the variation of TB deaths rate.

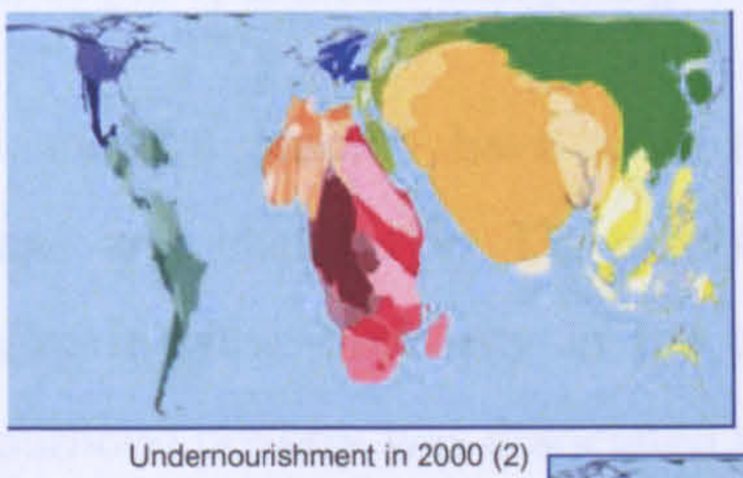
Fairchild and Oppenheimer [116] have summarised the strengths and limitations of both nutritionist and contagionist arguments and argued that neither is entirely correct. Most importantly, they argued that over time the interpretation of the historical decline of TB mortality has been severely biased by the political sensibilities and values of the historians called to examine it. The authors conclude that the complexity of the historical data for TB works against any generalised model of explanations and advocate for a multifactorial model based on local data in which nutrition, housing, ventilation, and segregation data are tested for their separate effect while controlling for the others. The existence of a multifactorial explanation for the declining trend in TB mortality has also been suggested by Sretzer, according to whom the 'pre-bacteriology' public health campaign against TB was 'a complicated history of struggle and pressure for relevant clauses in Factories and Workshop acts, Housing and Crowding Acts, and the enforcements of building regulations and by-laws' [55].

Given the limits of each model proposed, authors have called for more studies on the TB trends able to provide a broad spectrum of data in different times, places and populations [116]. In the following paragraphs I will summarise some of the most recent research that has been conducted in this field: typically the association between TB indicators and living conditions has been explored along two dimensions: in space (i.e. geographically) and in time (i.e. to identify determinants of changing trends in TB).

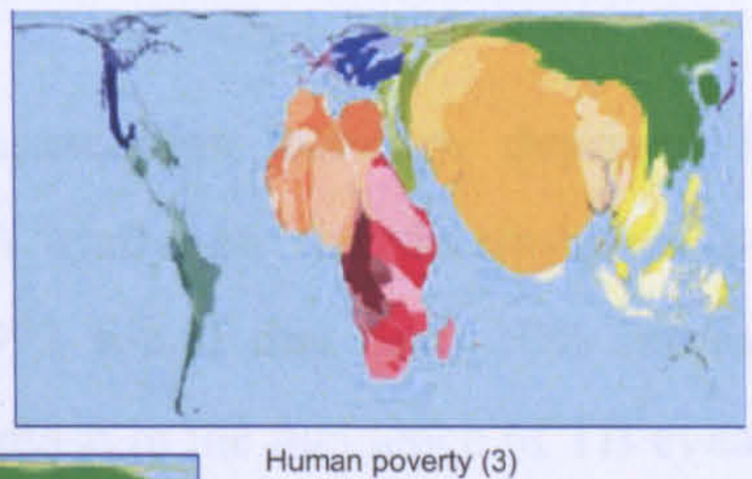
(b) Modern studies

In 2006, 9 million TB cases occurred globally [4]. Of these, only 1% occurred in the European Union and in North America, while Africa and South East Asia contributed more than 65%. Among the 22 highest burden countries, 17 were in Africa, 16 of which were in the lowest quartile of per capita domestic product (GDP) [118]. Over 44% of global TB deaths occur in the poorest 20% countries compared to only 2% observed in the wealthiest 20% of countries [6]. **Figure 2.5** shows the striking geographical overlap between TB prevalent cases as reported in 2003 and some indicators of living conditions extracted form different sources.

Figure 2.5: Geographical overlap between some indicators of living conditions and the proportion of TB cases reported in 2003 worldwide

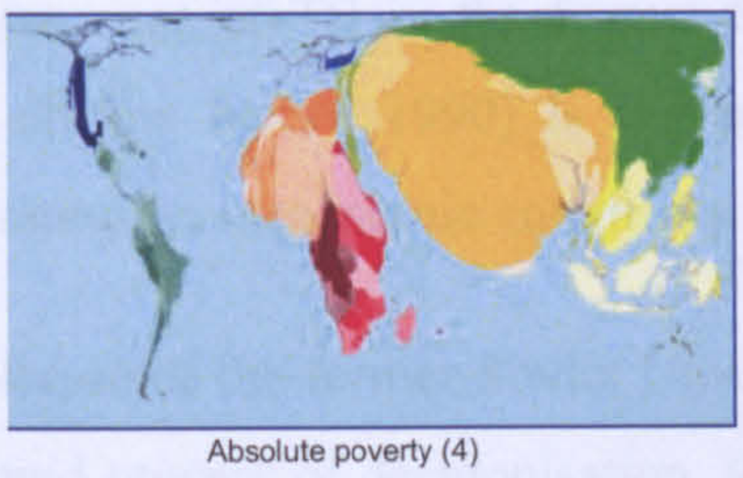


Undernourishment in 2000 (2)

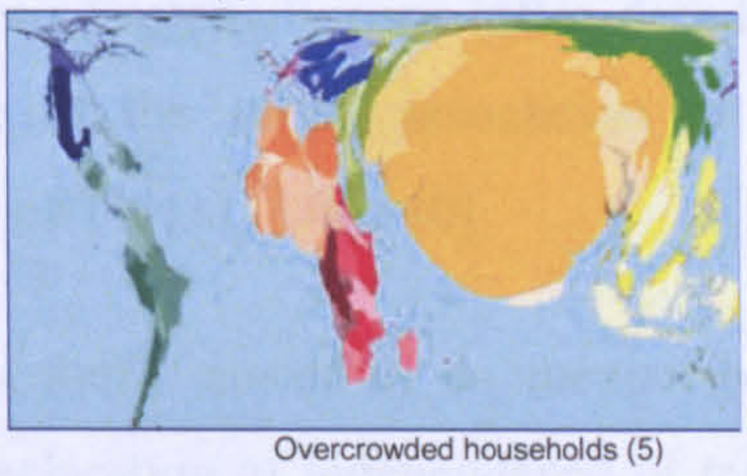


Human poverty (3)

Worldwide distribution of prevalent cases of TB in 2003 (1)



Absolute poverty (4)



Overcrowded households (5)

Reproduced from Worldmapper (http:www.worldmapper.org). Worldmapper is a collection of world maps. The maps presented on this website are equal area cartograms, otherwise known as density-equalising maps. The cartogram re-sizes each territory according to the variable being mapped. The source of data used to create the maps is reported in the website.

(1) Territory sizes show the worldwide proportion of prevalent cases reported in 2003.

(2) Territory sizes show the worldwide proportion of people that between 1999 and 2000 lived on an inadequate amount of food.

(3) Territory sizes show the worldwide proportion of people living in poverty in 2000. The human poverty index uses indicators that capture non-financial elements of poverty, such as life expectancy, adult literacy, water quality, and children that are underweight.

(4) Territory sizes show the worldwide proportion of people living on less than or equal 2 USD.

(5) Territory size show the worldwide proportion of people living in overcrowded households living there. Overcrowding is defined here as when there are more than two people for each room in the house.

The latest article published on the association between socioeconomic indicators and TB is one exploring the association between the incidence of TB and GDP indices using data from the World Bank and WHO. Linear regression analysis showed a 38.5% decrease in the incidence of TB for each doubling of GDP [119].

The existence of a strong spatial overlap between indicators of poor living conditions and TB has been consistently reported also in studies conducted in the Unites States [105-108], Japan [109, 110], United Kingdom [120], Germany [121], Turkey [122], Scotland [123], and France [124]. The same observations are available from studies conducted in middle and low

income countries like India [125], Brazil [126-130], Peru [131], South Africa [132-133], Philippines [134] and Pakistan [135].

Several modern studies have also reported the association between deprived living conditions and changes in TB trends over time. A study on the determinants of TB resurgence in New York city in the period 1984-1992 found that each 10% increase in neighbourhood poverty was associated with a 33% increase in the incidence of TB even after adjusting for AIDS, proportion of foreign born, and race/ethnicity [136]. Similarly, a study in London found that migration and overcrowding were strongly associated with the average TB notifications rates, but only the increase of unemployment from 1981 to 1991 were significantly associated with the rate of change in notification over the same period [137]. The resurgence of TB in the former Soviet Union and Eastern Europe (after decades of steady decline before 1990) is the perfect example of the strong association between deteriorating socioeconomic conditions and resurgence of TB [31, 138-140].

The collapse of the former Soviet Union has been essentially caused by an unexpected and quite rapid process of decolonisation, following the declaration of independence of most of the Baltic States forming the Soviet Union [141]. The consequent economic crisis, the conflicts, the inefficiencies in the distribution of the resources, and the unpreparedness to self-governance are the key elements of a political disaster coinciding with a dramatic worsening of several health indicators, including TB incidence. In just one decade (1990-2000) TB incidence rate almost trebled in Russia [139]. 120,000 new cases of TB and 30,000 deaths have been reported in 1999, making TB the leading cause of death for infectious disease in the same year [142]. Two possible causes have been identified: the weakening of the health system resulting in the collapse of the TB control services and the increased susceptibility of the population caused by social and economic factors [142]. According to government statistics more than 50 million people were living with less than 15 USD per day [143]. Living conditions were characterised by unemployment, inadequate nutrition and crowding [138] and further exacerbated by the deterioration of social capital and equality [144], war at the geographic margins of Russia, rise of alcoholism, accidents and violence [145], migration and the mixing of prisoners with civilian population [142]. Recent data suggest that since 2003, TB incidence in Belarus, Estonia, Latvia, Lithuania and Russia has been falling [139]. It remains unclear whether this reversed trend reflects the effect of TB control programs, the general recovery in population health lowering susceptibility to TB, or because of the exhaustion of new people susceptible to infection. However, the effect of wealth appears to be relevant because the fall in incidence is far more conspicuous in the richer states of Soviet Europe than in the poorer countries of central Asia [139].

2.4.4 Tuberculosis inequalities: limitations of ecological analyses

The studies described so far are all ecological: descriptive epidemiologic studies in which both independent and dependent variables are measured for groups, and variability in outcomes across groups are examined as a function of group-level variables [146].

Ecological studies are considered by many epidemiologists to be 'inherently inferior' [147] as they are seen as 'crude attempts to ascertain individual level correlations' with limited usefulness for anything other than hypothesis generation. This is because of their principal limitation: the ecological fallacy [148], a bias that occurs when 'inferences regarding the association between an individual-level exposure and an individual-level outcome are drawn on the basis of group-level associations between the corresponding aggregate (or derived) group level exposure and disease rate or the mean outcome for the members of the group' [149]. Numerically, the existence of ecological fallacy is proven when there is a lack of correspondence between the individual-level association (e.g. TB and income at the individual level) and the group level association (e.g. TB and mean of income in a community). Conceptually, the ecological fallacy arises from assuming that because there is certain size and direction of association between two variables measured at aggregate level, the same relationship will be observed at individual level too.

Another limitation of ecological analyses, including the ones presented for TB, is that they are unable to assess whether the area-level variables may be related to TB because of the characteristics of the individuals who live in an area, more than the characteristics of the area itself. For example, it could be that the highest prevalence of TB observed in a determined area is due to the age and gender distribution of the people living in the area or the HIV prevalence or number of smokers in the community, rather than to some physical or resource related factor defined at community level (like crowding, air pollution or mean income).

The terms contextual and compositional effect have recently come into use to describe this phenomenon. The remaining effect of 'group-level' variables on health after controlling for individual level characteristics has been called *contextual effect* [150]. However, the same terminology is often used to indicate generically the effect of area-level variables. Alternatively there may be circumstances in which the association observed is not explained by the characteristics of the context, but by the characteristics of residents of an area that make these individual more or less vulnerable to the outcome of interest. This phenomenon is called the *compositional effect* and has been defined more formally as that situation occurring 'when inter-group (or inter-context) differences in an outcome (for example diseases rate) are attributable to differences in group composition (that is the characteristics of the individual from which the groups are comprised) [150].

In conclusion, although numerous, ecological studies have neither quantified nor explained the association between SEP and TB.

The only studies likely to provide evidence on the causal pathway linking SEP and TB are analytical studies. This dissertation thesis aims to explore more in detail the processes and mechanisms involved in the pathway between SEP and TB through a proper analytical design. For this purpose, it was thought useful to review first the results and methodological features of existing similar studies.

Summary

Zambia suffers from a severe burden of poverty and TB and a vast HIV epidemic. Consequently it provides an appropriate setting in which to explore the association between SEP and TB in the HIV-era.

This research dissertation coincides with a renewed interest in the social determinants of health with increasing evidence that socioeconomic factors remain key in the epidemiology of TB. This evidence can at least partially explain why the current TB control strategy – based on case finding and treatment – is efficiently reducing the prevalence and mortality rates of TB in several regions of the world, but is not succeeding in reducing TB transmission. It also supports the potential value of a more integrated approach to the study of TB epidemiology, combining biological, behavioural and socioeconomic factors.

In this chapter I have illustrated how SEP has been conceptualised and operationalised so far at household and community level. I have introduced the concepts of structural determinants of health inequalities and social determinants of health and I have presented the most relevant theories developed to explain the causal pathway underlying the association between socioeconomic position and health. Although TB has been historically associated with poor living conditions, so far theories of diseases distribution have not been formally applied to the study of TB inequalities. Current thinking on TB inequalities is heavily influenced by McKeown's theory, which has inspired a wide number of ecological studies exploring the time and place distribution of TB according to socioeconomic indicators defined at the group level. The chapter concludes by providing an overview of these studies and discussing why these studies suffer from limitations hampering our understanding of the association between SEP and TB. In the next chapter I will explore in more detail this association by reviewing the most relevant studies that have addressed this link through an analytical approach.

3. Literature review

Never ignore, never refuse to see, what may be thought against your thought.

Nietzsche F[4].

Introduction

The epidemic of TB is considered to consist of two phases: the first one, in industrialised countries, peaked in 1900 and has been largely attributed to social, economic, and environmental factors including the increased population density, urbanisation, and poor nutrition. The second phase, mainly involving developing countries commenced in 1950 and for the moment shows slow signs of diminishing [1]. This emerging pandemic is more complex and therefore more difficult to contain than the first because it results from the complex interplay of old and new factors, like HIV, the emergence of antimicrobial drug resistance and more virulent strains [30, 58].

The modern study of the SEP-TB interaction requires the careful consideration of the elements charactering the "new tuberculosis" pandemic and raises questions that can be summarised as follows: is it correct to consider SEP as important as in the past for the epidemiology of TB in the HIV era and especially in developing countries? In other words, has the HIV epidemic obscured the epidemiological importance of living conditions? Not only has the biology of TB become more complicated, but also the way we conceptualise and measure SEP. Have recent studies on SEP and TB incorporated these concepts and these measurement tools? And if so, how have these new concepts and methods changed what we know about the association between SEP and TB? What are the aetiological models hypothesised and is the analysis driven by an a priori formulated conceptual framework?

To answer these questions I have reviewed the main analytical studies on the association between SEP and TB. The aim of this chapter is to show how the findings and the knowledge gaps identified through this review have been used to inform the objectives and the methods of this thesis.

3.1 Objectives

1. To collate the most relevant analytical studies on the association between household and/or individual measures of SEP and TB disease and TB infection.

2. To critically review the above evidence with reference to the following criteria:

   - The strength and the direction of the association

   - The definition and measurement of the socioeconomic risk factors explored

   - The confounding and mediating factors accounted for the analysis and whether or not they have been incorporated into a conceptual framework

3.2 Methods: search strategy, inclusion and exclusion criteria

Available references were retrieved through PubMed at the beginning of the project and updated regularly until January 2009.

The review aimed to include all studies in which the association between socioeconomic factors and TB was the main research question by searching publications including in their title any of the MeSH term indicated in the following query:

> Tuberculosis[Title] **AND** (socioeconomic[Title] **OR** poverty[Title] **OR** deprivation[Title] **OR** expenditure[Title] **OR** income[Title] **OR** inequalities[Title] **OR** social[Title])

This query led to the identification of 504 articles. Eligible papers identified through this query were identified as 'core studies'.

To increase the sensitivity of the search strategy, the following query was also run:

> Tuberculosis[Title] **AND** (case control[Title] **OR** cross sectional[Title] **OR** Cohort[Title] **OR** risk factors[Title])

Aim of this query was to identify broad risk factors studies that although not focusing on SEP, may still contain relevant data. This query identified 284 references including duplicates of articles retrieved in the first query. This search aimed to identify studies whose main objectives was generically to identify risk factors for TB. The articles identified through these queries were referred to as 'secondary studies'. Papers were selected if they

reported socioeconomic risk factors among their most important results, regardless the direction of the association.

Initially, the terms 'disease' and 'infection' were excluded from any query in PubMed, because it was assumed that the term 'tuberculosis' was broad enough to encompass both stages of TB. The analysis of the references retrieved, however, revealed that only a very small number of papers on TB infection were identified. For this reason, in a second stage, I run an additional search including terms more relevant to the stage of infection:

> **Tuberculosis[Title] OR TB[Title] AND (infection[Title] OR Transmission[Title] OR TST[Title]) AND (risk factors OR socioeconomic OR poverty OR deprivation OR income OR inequalities OR social OR poor).**

This query initially identified 713 articles that were referred to as 'Infection studies'

After running the three queries above, I applied the following selection criteria:

Linguistic - Only papers in English were considered eligible because resources for translation were not available and because the assessment of the study methodology only from the abstract in English was considered inadequate. The same applies to articles whose abstract was available, but that could not be retrieved even through inter-library loan.

Outcome - Only papers focusing on either TB infection or pulmonary TB were included in the review process. Outcomes excluded were extrapulmonary TB, TB mortality, relapse, treatment failure, treatment compliance, TB patient management, drug resistant TB, delay to TB care, social and economic impact of TB. In the case of TB infection, all the studies aiming to assess sensitivity and specificity of new tests for TB infection were excluded.

Selected population groups - Papers specifically targeting homeless, migrants, refugees, drug users, special professional categories (i.e. health care staff, gold miners), and prisoners were excluded. This was done because it would have been difficult to extrapolate the evidence gathered to the general population and because the association between low SEP and TB in such vulnerable groups is somehow expected, regardless of how SEP is measured or conceptualised. Articles focusing on HIV-coinfected people and people suffering from any underlying disease (e.g. diabetes, cancer, silicosis) were also excluded because it would have been difficult to disentangle the effect of these conditions from the effect of SEP on the risk of TB.

Type of articles - Only original research articles were included and among them only those examining socioeconomic factors. Systematic review and meta-analysis were also initially collected, but just as a potential source of additional references not identified through PubMed. Although often containing a detailed socioeconomic analysis, articles addressing the effect of TB on SEP (i.e. the reverse association) were not included in the review because they were considered not to be relevant for this thesis.

Once eligible articles have been selected, their reference list was searched to ensure the inclusion of any relevant studies not detected through the queries above.

The selection process applied to the articles retrieved is summarised in the flowchart in **Figure 3.1**. Only 234 of the 504 articles retrieved for the 'core studies' category were in English and, of them, only 9 met the eligibility criteria. 234 articles in English were also found for the 'secondary studies': of them, 208 were discarded and 26 considered potentially eligible. Of these 26, 15 were excluded because they did not report sufficient information on socioeconomic risk factors. The majority of articles concerning TB infection (680 out of 713) focused on the diagnostic aspects and TB immunology. Among the 33 potentially eligible articles, 32 were in English and 22 of them were considered for inclusion in the review. Of them, only 11 reported socioeconomic risk factors among their main results and were therefore included.

At the end of this screening process 31 articles were included in this review overall: 20 on the association between SEP and TB disease and 11 on the association between SEP and TB infection.

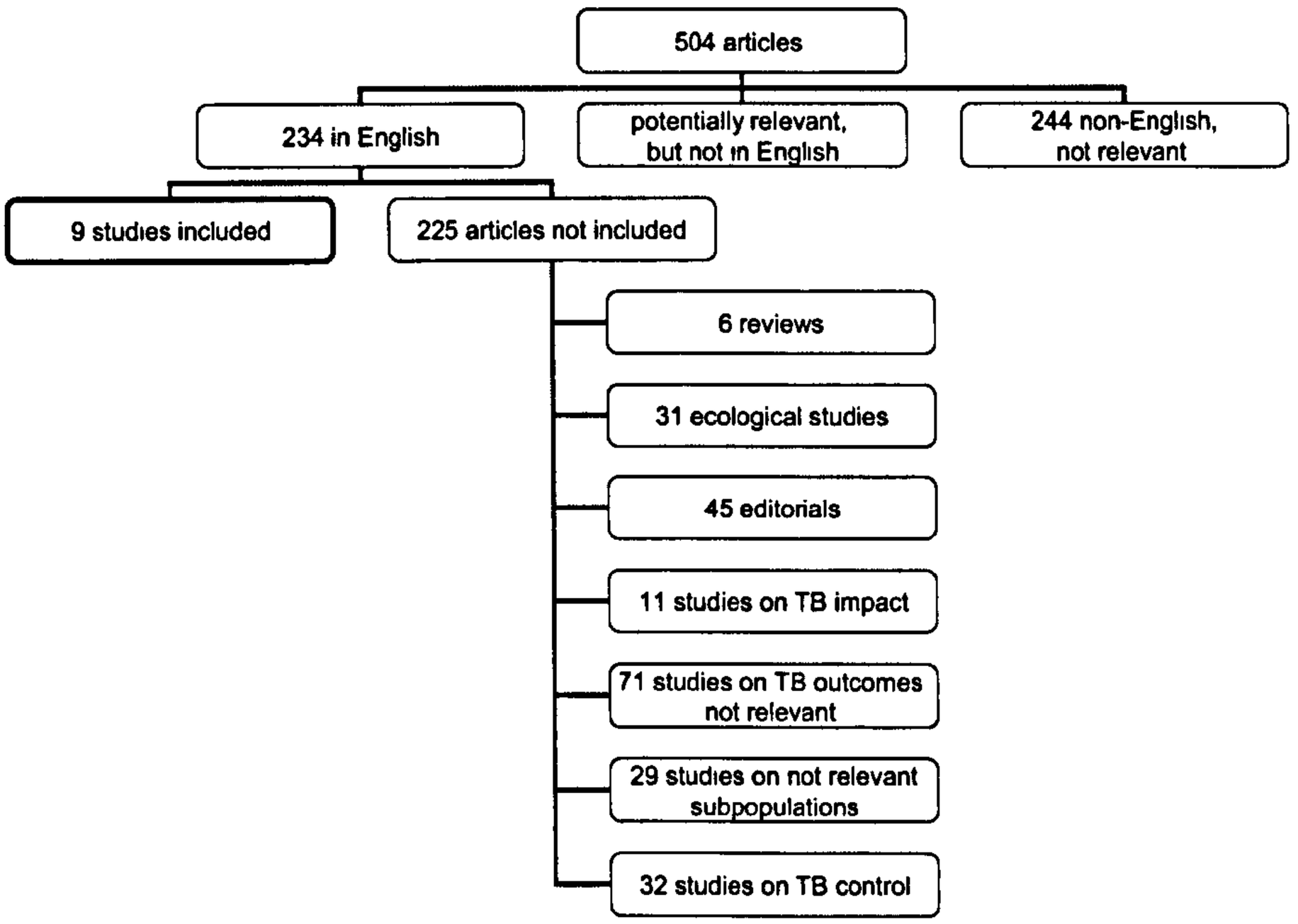
3.3 Results

Table 3.1 provides a summary of the articles included in the present review divided by world region where the studies were undertaken.

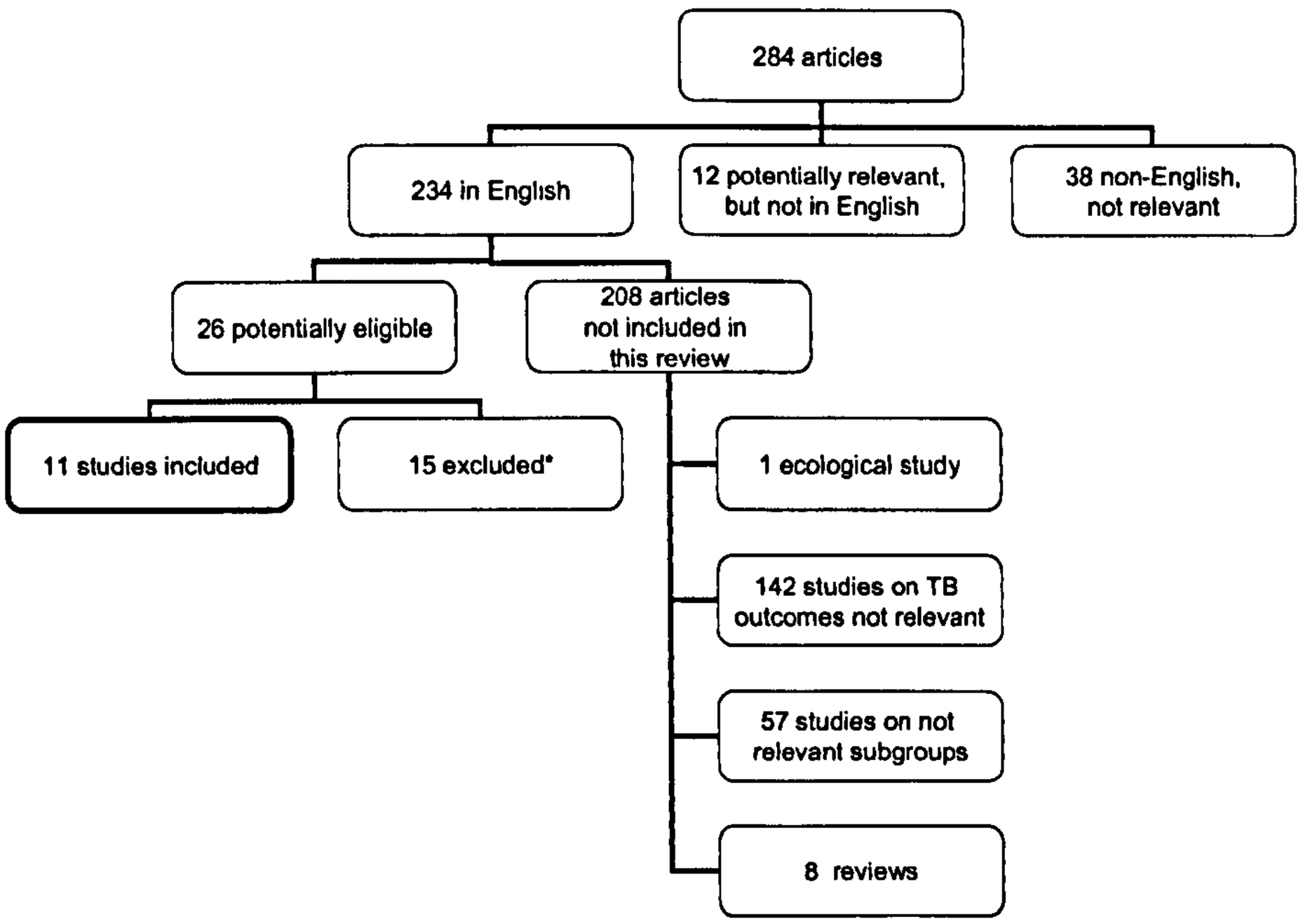
Overall, eligible studies originated from India [151-154], Pakistan [155], Indonesia [156], Thailand [157], China [158], Peru [159-160], Mexico [161-162], US [163-164], England [165-166], Italy [167], Estonia [168] and Russia [169].

Figure 3.1: Flowchart of the article selection by study category

Core studies



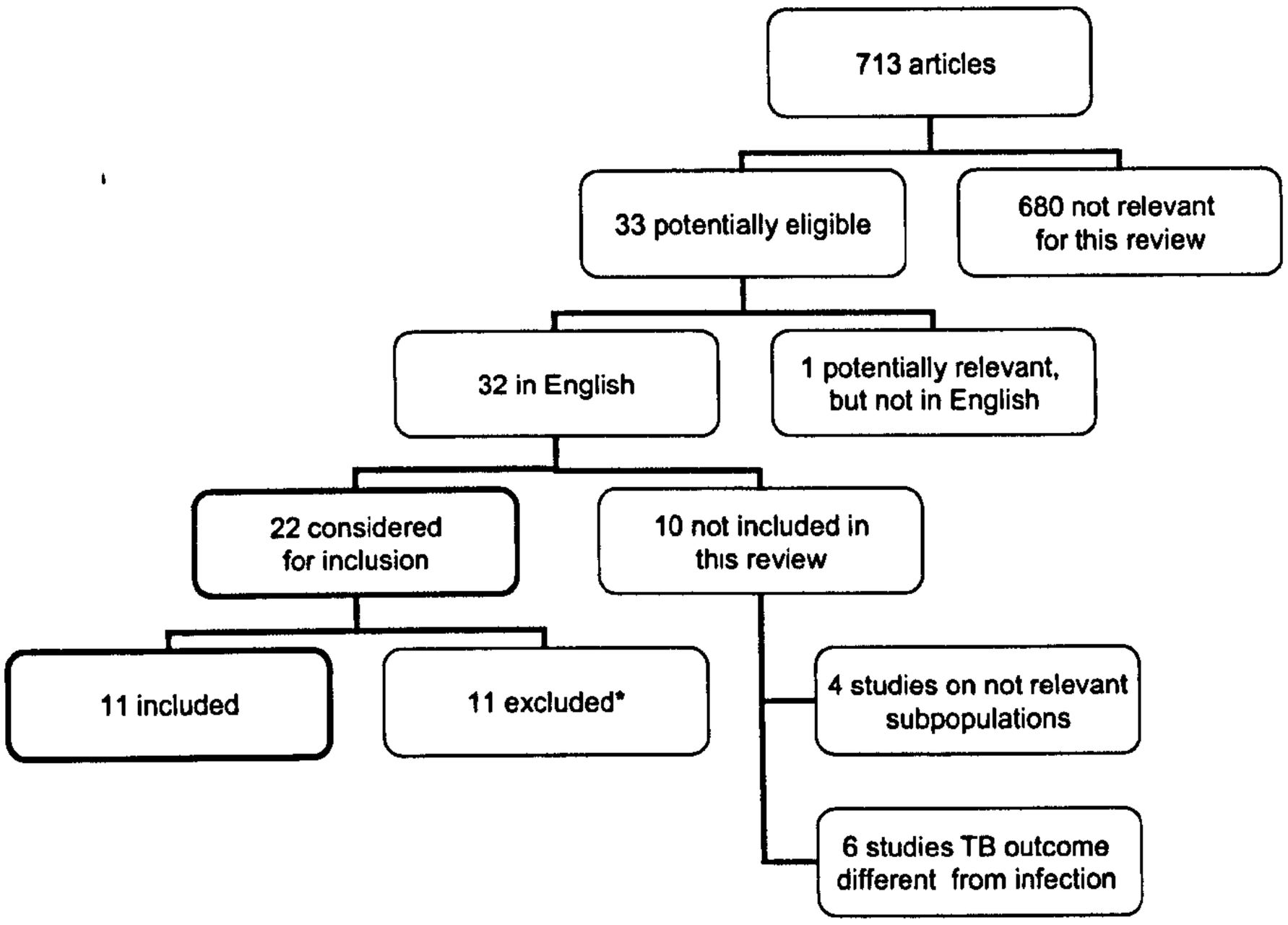
Secondary studies



*Reporting data on few socioeconomic indicators

Figure 3.1: Continued

Infection studies



*Reporting data on few socioeconomic indicators

Table 3.1: Synopsis of the articles included in the review

Setting	Study Category*	Study Period	Study Design†	TB outcome‡	Study population	Study size
African region						
Malawi [7]	C	1981-1996	R-COH	D-P/I	TB cases from prevalence survey and TB clinic attendees	69 TB cases, 11172 controls
Malawi [22]	C	2001	CSS	D-I	TB clinic attendees on treatment	179 cases of TB
Malawi [170]	S	1996-2001	FM-CC	D-P/I	Enhanced passive TB case finding	
Malawi [171]	I	2003-2005	CSS	I-RT	Children	195 household contacts of TB cases
Guinea-Bissau [172]	S	1996-1998	TB-SP	D-P/I	TB cases active and passive case finding	247 TB cases
Guinea Bissau, Bissau, Gambia [173]	S	1999-2001	IM-CC	D-I	All new TB cases diagnosed at the urban health centre	846 TB cases, 702 household controls, 828 community controls
Guinea Bissau [174]	I	1999-2000	CC	LTBI	Adults and children	1059 household contact of TB cases, 921 household contact of community controls
Gambia [175]	S	2002-2004	IM-CC	D-I	All new TB cases diagnosed in the outpatient centre	100 TB cases, 200 controls
Gambia [176]	I	1999-2000	CC	LTBI	Adults and children	305 household contacts of TB cases, 315 household contacts of community controls
Gambia [177]	I	1999-2001	CSS	I-RT	Only children	384 household contacts of TB cases
South Africa [8]	C	NA^	IM-CC	D-I	Clinic register of TB patients on treatment	84 TB cases, 84 controls
South Africa [178]	C	1996-1998	PS	D-P	Self-reported TB cases*	69 recent TB, 369 TB ever told they had TB
South East Asia Region						
Indonesia [156]	C	1997-1998	RS	D-I	Untreated TB outpatients	90 TB cases
Thailand [157]	I	2002-2003	CSS	I-RT	Only children	500 household contacts of TB cases
Pakistan [155]	I	1999	CSS	LTBI	Adults and children	385 household contacts of TB cases
India [154]	C	1998-1999	PS	D-P	Self-reported TB cases**	3001 TB cases, 93486 Households
India [153]	S	2001-2003	IM-CC	D-I	All new TB cases diagnosed in a clinic in Bangalore	189 cases, 189 controls
South India [151]	C	2004-2006	PS	D-P	All TB cases detected in the prevalence survey	32780 Households, 151 TB cases
India [179]	I	NA	CSS	I-RT	Children	140 household contacts of smear positive TB cases, 141 household contacts of smear negative TB cases
China [158]	C	2002-2004	IM-CC	D-I	All new TB cases registered at the county TB clinics	160 TB Cases, 320 Controls
American Region						
United States [163]	S	1988-1990	CC	D-I	TB patients clinic attendees	151 TB cases, 545 controls
United States [164]	I	1996-1998	IM-CC	LTBI	Only children	96 cases TB infected, 192 controls
Peru [160]	I	1997-2000	CSS	LTBI	Attendees of local health post	2410 people tested for TST
Peru [159]	I	NA	CSS	LTBI	Adults	212 adults
Mexico [162]	S	1998-1999	CC	D-I	TB patients clinic attendees	288 TB cases, 545 controls
Mexico [161]	S	1998	PS	D-P	All residents in 32 areas of Chiapas	17 TB cases

Table 3.1: Continued

Setting	Study category*	Study period	Study Design†	TB outcome‡	Study population	Study size
European Region						
England [166]	C	1982-1990	RS	D-I	Notified cases of childhood TB	63 TB cases
England [165]	S	1995-1997	IM-CC	D-I	Notified TB cases	430 TB cases
Estonia [168]	S	1999-2000	IM-CC	D-I	TB patients admitted to hospital during the study period	248 TB cases, 248 controls
Russia [169]	S	2003	IM-CC	D-I	All new TB cases diagnosed at any TB clinic and on treatment	334 cases, 334 controls
Italy [167]	I	1997-1998	CSS	LTBI	Adults and children	346 household contacts of TB cases

*Study category: Indicates the type of study included in the review as from the search criteria. C= Core study; S= Secondary study; I= Infection study
†Study design: CC = Case Control study; PS= Prevalence survey; FM-CC = Frequency matched Case Control study; TB-SP= TB Surveillance Program; CSS = Cross Sectional Survey IM-CC= Individual Matching Case Control Study; RS= Retrospective Survey; R-COH= Retrospective Cohort Study;
‡TB outcome: D-P= TB disease, prevalent cases; D-I= TB disease, incident cases; D-P/I= TB disease, both prevalent and incident cases; LTBI = Latent Tuberculosis Infection; I-RT= Infection, Recent Transmission;
*From the 1998 South Africa Demographic Health Survey
** From the 1998 National Family Health Survey
^NA= Not available

12 studies were conducted in the African region accounting for just five countries from Sub-Saharan Africa, including Malawi [7, 22, 170-171], Gambia [173, 175-177], Guinee [173], Guinea Bissau [172, 174] and South Africa [8, 178]. In fact these studies represent a very small proportion of the population of Africa: two of the four studies in Malawi [170] [7], and all the studies from Gambia, Guinea Bissau and Guinee come from the same research groups and therefore probably overlapped in terms of study population and methods.

Studies included in the review covered a period of time ranging from 1981 [7] to 2006 [151].

Among the studies on TB disease, one implemented a retrospective cohort study [7], one used data coming from a TB surveillance program [172], four used data from prevalence surveys [151, 154, 161, 178], three conducted a cross-sectional survey [22, 156, 166]; eleven adopted a case control study design, of which two were unmatched [162-163], one frequency matched [170] and eight individually matched [8, 153, 158, 165, 168-169, 173, 175]. Of the four studies reporting data on prevalent cases of TB, only two included cases detected within an actual TB prevalence survey [151, 161]. The remaining two studies referred to self-reported TB cases detected within demographic health surveys which had not been clinically nor microbiologically confirmed [154, 178].

The studies on TB infection can be divided into those focusing on latent TB infection (LTBI) (mainly conducted among adults) and those assessing recent TB transmission (mainly conducted among subjects under the age of 15). Most of these studies (N=6) were cross sectional surveys in which household contacts of known pulmonary TB cases were tested for TB infected with the Tuberculin Skin Test (TST) [155, 157, 159, 167, 171, 177]. A variation to this design was the study from Singh and colleagues that compared household contacts of smear positive and smears negative TB patients [152]. The remaining five studies included a cross-sectional survey of households contacts of TB cases and household contacts of non-TB cases identified in the community [174, 176], one case control study among children under 5 years of age that were age and clinic matched to healthy controls [164], two TST surveys conducted among a random samples of adults identified in different geographic zones [159-160].

3.3.1 Measurement of socioeconomic position

For clarity, the description of the methods used will be provided by main study category.

The core studies - Six of the 'core studies' reported the use of some kind of SEP composite index alone or in combination with single socioeconomic indicators [8, 22, 151, 154, 158, 178], whereas 3 studies employed only single socioeconomic indicators [7, 156, 165].

Among the nine 'core studies', only Shoeman and Simwaka explicitly mentioned the attempt to develop a locally appropriate socioeconomic index as one of the main objective of their investigation [8, 22] (**Table 3.2**).

Of the six 'core studies' reporting the use of SEP composite indices only two made an explicit mention about the weighting strategy adopted [22, 178]. In the study from South Africa [178], authors used factor analysis to weight a list of 9 assets including, main source of drinking water; type of toilet facility; type of fuel used; main materials of floor and walls; food affordability and the ownership of assets. This study was also the only one to introduce two further measurement of SEP: at individual level (through the number of years of education and unemployment) and at community level (through two measures of absolute poverty, headcount poverty rate and rate of unemployment, and two measures of income inequalities).

The cross sectional study in Malawi [22] used regression analysis of six different assets, including source of cooking energy, electricity, having a car, household size, gender of the head of the household and education of the head of the household, to predict per capita consumption expenditure.

The two studies in India employed the so called Standard Living Index (SLI), a summary household index computed by assigning scores to information on type of house, toilet facility, source of lighting, main fuel used for cooking, source of drinking water, separate kitchen, ownership of house, agricultural and irrigated land and livestock and durable goods [151, 154]. In one case [151], the author mentioned that this index was developed from the World Bank [151], suggesting that the SLI may be the equivalent of the wealth index included in Demographic Health Surveys and therefore is presumably based on principal component analysis or factor analysis.

In terms of type of socioeconomic indicators collected, the nine 'core studies' showed a wide variability: the most frequently variables collected fell into the human resources (i.e. education and occupation), followed by housing quality and assets ownership category, but only one study mentioned what the indices aimed to measure and how they were intended to be linked to TB [178](**Table 3.2**). The other most frequently collected variables were BMI and or food availability [8, 156, 178], crowding [7-8, 156], income [151, 156, 158], and access to community services [8, 22, 166].

Table 3.2: The association between socioeconomic factors and TB disease (1991-2008): results from the 'core studies'

Setting	TB outcome‡	Socioeconomic indicators§								Weighting strategy*	Conceptual framework^	Adjusted OR** (95%CI)	Controlling for:	
		HR	HQ	AS	CR	FA/ BMI	CO	IN	SP				Confounding	Mediation
South Africa [8]	D-I	■	■	■		■	■			None	None	Not significant	None	None
England [166]	D-I						•			None	None	Living in a poor area: 4.0 (2.4-6.7)	Ethnicity	None
Malawi [7]	D-P/I	•	•							NA	None	Farmers: Reference Salaried workers: 3.7 (1.5-9.5) Casual workers: 3.1 (1.0-9.1) High quality dwelling: References Low quality dwelling: 0.3 (0.09-0.8)	Age, sex, HIV Housing Schooling Occupation	None
Indonesia [156]	D-I					•		•		NA	None	No OR(a)	None	None
India [154]	D-P	■	■	■						None	None	High SLI(b): Reference Medium SLI: 1.5(c) Low SLI: 1.9(c)	Sex Area of residence Ethnicity Education Type cooking fuel Type of housing	None
China [158]	D-I			■					■	None	None	Lowest quarter assets index: Reference II: 0.74 (0.41-1.32) III: 0.52 (0.28-0.94) IV: 0.48 (0.25-0.92) Lower third self-poverty index: Reference Middle/upper third: 0.2 (0.1-0.4)	Age, sex Area of residence Smoking Marital status Migration	None
South India [151]	D-P	■	■	■				•		None	None	High SLI: Reference Medium SLI: 1.8(c) Low SLI: 3.7(c) Living below the poverty line: 1.6(c)	None	None

Table 3.2: Continued

Setting	TB outcome‡	Socioeconomic indicators§								Weighting strategy*	Conceptual framework^	Adjusted OR** (95%CI)	Controlling for:	
		HR	HQ	AS	CR	FA/ BMI	CO	IN	SP				Confounding	Mediation
Malawi [22]	D-I	■		■			■			Regression	None	No OR[(d)]	None	None
South Africa [178]	D-P	■	■	■		•				Factor analysis	Not tested	I Poorest quintile: 1.6 (1.0-2.3) II: 1.0 (0.7-1.4) III: Reference IV: 0.9 (0.6-1.6) V: 0.8 (0.5-1.3) Ever missed a meal: 1.7 (0.9-3.3) BMI < 18.5: 2.9 (1.9-4.2)	Community-level	None

‡ TB Outcome: D-P = TB disease, prevalent cases; D-I = TB disease, incident cases; D-P/I= TB disease, TB incident and prevalent cases.

*Weighting strategy adopted if using a composite indicator of SEP. None = Not Mentioned. NA = Not applicable as not using composite indices in the study.

§Socioeconomic indicators grouped into most relevant socioeconomic categories: HR= Human Resources; HQ= Housing Quality; AS= Assets Ownership; CR= Crowding; FA/BMI=Food availability and/or BMI; CO = Community services (i.e. water supply, electricity) or residence in a community considered to be poor; IN= Income; SP= Self-reported poverty status. A yellow box indicates that one or more indicator from the category were employed in the study. An orange box indicates that the indicator was used in a composite index covering more than one SEP aspect. The • in a yellow box indicates that the indicator from the category was significantly associated with TB after controlling for confounding and/or mediating factors. The ■ in an orange box indicate that the composite index was significantly associated with TB after controlling for confounding and/or mediating confounding factors.

^None indicates that the use of conceptual framework is not mentioned in the article. Not tested means that the conceptual framework is mentioned, but it was not formally tested through the analysis of the available data.

**Adjusted for any of the confounding and mediating factor reported in the next columns. Reference group is underscored.

(a)No Odd Ratios Reported. They just reported that 71% of their cases belonged to households with a median monthly income below the poverty line for the urban areas of Indonesia. In this study TB patients belonging to families with members currently or previously suffering from TB had a significant lower BMII than those who had no such contact (P=0.04)

(b)SLI= Standard of living index.

(c)No 95% confidence intervals reported.

(d)No Odd Ratios reported. They just report the overall poverty headcount of 78% among TB patients compared to the 30% of the general population of Lilongwe, Malawi.

No papers provided a definition of SEP used in the study. Consequently, it was difficult to judge what the authors intended to measure and if the socioeconomic variables were appropriately chosen for this purpose.

Secondary studies - Among the 'secondary studies', six studies reported an attempt to combine the socioeconomic indicators into a composite index [162-163, 170, 172, 175]: apparently, in all these studies but one [162] the generation of the composite index was not driven by any statistical analysis, but based on the personal judgment of the author (e.g housing quality score) or the aggregation of the socioeconomic data into a count variable, resulting from the sum of the different socioeconomic factors.

The only study that mentioned the construction of a weighted composite index did not mention any weighting strategy, but only that income per dependent was the indicator receiving the heaviest weight [162].

In order to account for the large variety of information collected, data were grouped into 13 principal socioeconomic indicators (**Table 3.3**).

Household income data were overall reported in two studies [153, 165], whereas in one study from Estonia, authors collected information on individual income [168]. The collection of income data was always in addition to, rather than in place of, assets data, but authors never mentioned whether the econometric index and the assets data were actually measuring different aspects of SEP. The most frequently collected information include education-related data (collected in every study but the one from Gustavson and the one from Coker [169, 172]), occupation [153, 161, 165, 168, 170, 173, 175] [169], housing quality (collected from all the studies but the one in Russia) [169], crowding [153, 161-162, 168-170, 172-173, 175], followed by and assets ownership [153, 161, 165, 173, 175, 180]. Three studies reported the collection of BMI data [153, 163, 168]. In these studies, BMI was used to ascertain malnutrition and was not discussed as a socioeconomic proxy. However, in order to maximise the information available, I have decided to include in this review all the variables reflecting somehow SEP rather than just relying only on authors' classification. In the specific case of BMI, although the association between malnutrition and SEP is likely to be less direct than the association between food insecurity and SEP, it is also well accepted that in extremely poor settings like Zambia malnourished people are far more prevalent among the lower socioeconomic strata of the society. For this reason I considered as appropriate to

use BMI as a proxy of SEP and – as a result of that - increase the amount of information available for the literature.

Finally, the study in Estonia collected data on self-perception of poverty [168].

Infection studies - No study employed a SEP composite index (**Table 3.4**). As for the studies on TB disease, even in the case of TB infection, investigations showed a large variety of data collected. The variables most consistently collected were those about housing quality [157, 159-160, 164, 167, 172, 176-177] crowding [157, 159-160, 164, 171, 174, 176-177] and BMI [152, 159-160, 171, 177].

Table 3.3: The association between socioeconomic factors and TB disease: results from the secondary studies

Setting	TB outcome‡	Socioeconomic indicators§											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for†		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			D	B	C
United States [163]	D-I	■		■	■						●		(a)	0 criteria: Ref. 1: 3.0 (1.4-6.8) 2: 3.4 (1.6-7.8) 3: 3.7 (1.5-5.9) Highest BMI quartile: Ref III quartile: 1.6 (0.9-3.1) II quartile 2.6 (1.5-4.9) I quartile: 2.2 (1.2-4.0)	+		
England [165]	D-I				●				●				NA	Having more than 1 bathroom: Ref Having 1 bathroom: 1: 4.0 (1.3-12.4) Having diary products weekly: Ref Having diary products less than weekly: 3.8 (1.4-10.3)	+		
Mexico [162]	D-I	●		■	■								(b)	No odds ratio reported#			
Mexico [161]	D-P												NA	Not significant			
Estonia [168]	D-I	●	●	●	●				●		●	●	NA	University education.: Ref Secondary: 3.2 (1.6-6.5) Basic: 13.2 (5.3-32.6) Primary: 15.5 (4.1-58.4) Average income: Ref Up to average income: 2.8 (0.9-9.8) Up to minimal income: 3.0 (0.7-12.2) Minimal income: 9.9 (2.2-43.9) No fixed income: 12.3 (3.2-47.3)	+		

Table 3.3: Continued

Setting	TB outcome‡	Socioeconomic indicators§											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for†		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			D	B	C
Estonia [168]	D-I	●	●	●	●				●		●	●	NA	Type of residence: Own house/flat = Ref Room in hostel: 8.9 (2.6-31.2) Nursing home: 11.3 (1.0-123.9) Not Uneployed: Ref. Currenlty unemployed: 5.1 (2.7-9.7) Previusly unemployed: 1.0 (0.6-1.8) Occupational group: Group 2= Ref[#¤] Group 3: 2.2 (1.0-4.7) Stable economic situation: Ref Improved compared to last year: 0.3 (0.2-0.7) Enough food: Ref Sometimes shortage of food: 1.4 (0.7-2.6) Often shortage of food: 5.2 (1.1-24.2) Normal weight: Ref. Overweight: 0.2 (0.06-0.4) Obesity: 0.3 (0.08-1.2) Underweight: 2.0 (0.8-4.9)			
Malawi [170]	D-P/I				■	●							(c)	Household possession score 0= Ref 1-3: 1.2 (0.8-1.9(4-7: 0.7 (0.5-1.1) 8+ 0.7 (0.4-1.1) ≥ 3.5 people sharing sleeping room = Ref 3: 0.6 (0.4-0.9) 2: 0.6 (0.4-1.0) 1: 0.8 (0.5-1.2)		+	++

Table 3.3: Continued

Setting	TB outcome[†]	Socioeconomic indicators[§]											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for[†]		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			D	B	C
Guinea-Bissau [172]	D-P/I				■	●							NM	House foundation: Bricks/cement: Ref. Mud/earth: 1.5 (1.0-2.1) 1-2 adults/household: Ref. 3-4: 1.7 (1.1-2.4) 5-8: 1.7 (1.0-2.6) 9+: 1.9 (1.0-3.3)	+	++	
Multicenter study [173]	D-I					●	●						NA	1-5 adults in the household: Ref. 6-10: 1.4 (1.0-1.8) > 10: 2.8 (1.7-4.6) Ownership of the house: Ref No: 1.4 (1.1-1.9)	+	++	
India [153]	D-I	●			●						●		NA	No education: Ref. 3-10 yrs schooling : 0.6 (0.2-1.5) > 10 yrs schooling: 0.3 (0.1-0.8) Household income < 1000 Rupees (Rs): Ref 1000-5000 Rs: 0.9 (0.5-1.5) > 5000 Rs: 0.5 (0.2-1.0) Having a separate kitchen: Ref. Not having a separate kitchen: 3.3 (1.2-8.5) BMI ≥ 18.5: Ref < 18.5: 11.1 (5.7-21.9) MUAC[‡‡] ≥ 24 cm: Ref < 24 cm: 6.8 (3.9-11.9)	+	+	
Gambia [175]	D-I				●	■							NM	Crowding index 1: Ref 2: 2.6 (0.7-9.7) 3: 5.1 (1.7-14.6) Having ceiling: Ref Not having ceiling: 2.1 (0.9-4.9)	+	+	+

Table 3.3: Continued

Setting	TB outcome‡	Socioeconomic indicators§											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for†		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			D	B	C
Russia [169]	D-I		•			•	■		•				NM	Most assets: Ref	+	+	
														2: 1.1 (0.6-2.1)			
														3: 3.6 (1.8-7.1)			
														4: 3.6 (1.8-7.0)			
														Least: 8.8 (4.0-19.3)			
														No shortage of food: Ref.			
														Shortage of food: .2.1 (1.2-3.8)			
														Being employed: Ref.			
														Being unemployed: 5.8 (3.8-9.0)			
														Most living space/person: Ref.			
														2: 1.7 (0.9-3.2)			
														3: 1.9 (1.0-3.5)			
														Least: 3.8 (2.1-9.0)			

‡TB Outcome: D-P = TB disease, prevalent cases; D-I = TB disease, incident cases; D-P/I= TB disease, TB incident and prevalent cases.

*Weighting strategy adopted if using a composite indicator of SEP. NM = Not Mentioned. NA = Not applicable as not using composite indices in the study. (a) Socioeconomic status was based on family income, housing condition, and years of education, each dichotomised as either high or low. Low SES was defined as being in the lower category of at least two of those factors; (b) authors used the socioeconomic classification of the National Institute of Respiratory Diseases in Mexico City. This index includes 7 categories and is a weighted composite of income per dependent (heaviest weight), household characteristics, owning or renting the house, and the presence of sick relatives; (c) The construction of walls, floor, roof and windows were scored according to building materials and the component scores were summed. Assets owned were scored according to their relative values and the score for individual items were then summed.

§Socioeconomic indicators grouped into most relevant socioeconomic categories: ED= Education; OC = Occupation; IN = Income; HQ= Housing Quality; AS= Assets Ownership; CR= Crowding; CO = Community services (i.e. water supply, electricity) or residence in a poor community; FA=Food availability; VU= Vulnerability; BMI= Body Mass Index; SP= Self-reported poverty status. Yellow box indicates that one or more indicator from the category were employed in the study. Orange box indicates that the indicator was used in a composite index covering more than one SEP aspect. The • in a yellow box indicates that the indicator from the category was significantly associated with TB after controlling for confounding and/or mediating factors. The ■ in am orange box indicate that the composite index was significantly associated with TB after controlling for confounding and/or mediating confounding factors.

**Adjusted for any of the confounding and mediating factor reported in the next columns. Reference group is underscored. #Authors reported that TB cases had lower income and education than controls. ## Reference group = technicians, clerks, service workers; Group 2= skilled agricultural and fishery workers, craft workers, plant and machine operators.

†Controlling for: D = Demographic variables (including age, sex, marital status, education, occupation, ethnicity, area of residence); B= Biological and behavioural risk factors (including HIV, BMI, chronic underlying illnesses, alcohol consumption, smoking, drugs abuse, imprisonment, migration); C= Contact with a TB case (including crowding, exposures to TB cases and attending public venues). + indicates that the association between SEP and TB has been adjusted for one or more variable in the category; ++ indicates that the association between SEP and TB has been adjusted for HIV.

††MUAC = Mid upper-arm circumference.

Table 3.4: The association between socioeconomic factors and TB infection: summary of results

Setting	TB outcome‡	Socioeconomic indicators§											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for†		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			**D**	**B**	**C**
Italy [167]	LTBI		■		■								NA	Not significant	+	+	+
United States [164]	LTBI	■		■	■	■				■			NA	Not significant	+	+	+
Gambia[176]	LTBI		■		■	■	■	■					NA	Not significant	+	+	+
Pakistan [155]	LTBI	■											NA	Not significant			
Gambia [177]	RT		■	■	■	•	■				■		NA	<u>> 10 people in the house: Ref.</u> 6-10: 3.9 (1.2-13.2) 1-5: 1.3 (0.6-3.2)	+	+	+
Peru [160]	LTBI	■			■	■					■		NA	Not significant	+		
Thailand [157]	RT	■	■	■	■	•								<u>Crowding, ≤ 2 persons/room = Ref.</u> 2.1-3: 1.3 (0.6-2.9) > 3: 2.6 (1.2-5.8)	+		+
India [152]	RT										•		NA	<u>Malnutrition[(a)], Grade 1-2: Ref.</u> Grade 3-4: 3.9 (2.2-7.2)	+	+	+
Peru [159]	LTBI			■	•	■			■		•		NA	Poor ventilation[(b)]: 3.6 (1.4-9.0) CAMA[(c)]: 1.05 (1.01-1.10)			

Table 3.4: Continued

Setting	TB outcome‡	Socioeconomic indicators§											Weighting Strategy*	Adjusted OR** (95% CI)	Controlling for†		
		ED	OC	IN	HQ	CR	AS	CO	FA	VU	BMI	SP			D	B	C
Malawi [171]	RT												NA	Not significant			
Guinea-Bissau [174]	LTBI												NA	Not significant			

‡TB Outcome: LTBI= Latent Tuberculosis Infection; RT= Recent transmission.

*Weighting strategy adopted if using a composite indicator of SEP. NM = Not Mentioned. NA = Not applicable as not using composite indices in the study.

§Socioeconomic indicators grouped into most relevant socioeconomic categories: ED= Education; OC = Occupation; IN = Income; HQ= Housing Quality; AS= Assets Ownership; CR= Crowding; CO = Community services (i.e. water supply, electricity) or residence in a poor community; FA=Food availability; VU= Vulnerability; BMI= Body Mass Index; SP= Self-reported poverty status. A yellow box indicates that one or more indicator from the category were employed in the study. An orange box indicates that the indicator was used in a composite index covering more than one SEP aspect. The • in a yellow box indicates that the indicator from the category was significantly associated with TB infection after controlling for confounding and/or mediating factors. The ■ in an orange box indicate that the composite index was significantly associated with TB infection after controlling for confounding and/or mediating confounding factors.

**Adjusted for any of the confounding and mediating factor reported in the next columns. Reference group is underscored.

†Controlling for: D = Demographic variables (including age, sex, marital status, education, occupation, ethnicity, area of residence); B= Biological and behavioural risk factors (including HIV, BMI, chronic underlying illnesses, alcohol consumption, smoking, drugs abuse, imprisonment, migration); C= Contact with a TB case (including crowding, exposures to TB cases and attending public venues). + indicates that the association between SEP and TB infection has been adjusted for one or more variable in the category.

(a)Malnutrition was classified according to the Indian Academy of Pediatrics. Grade 1= weight 71-80% of expected; Grade 2 = weight 61-70% of expected; Grade 3 = weight 51-60% of expected; Grade 4 = weight less than 50% of expected. Grade 1 and 2 were classified as mild malnutrition and Grade 3-4 as severe malnutrition.

(b) Household ventilation was assessed as poor if construction techniques prevented the free flow of air through the building.

(c) CAMA= Corrected (bone free) Arm Muscle Area. It is an indicator of body protein content and it is associated with chronic malnutrition.

3.3.2 The association between socioeconomic position and tuberculosis

For clarity, results on the association between SEP and TB disease and TB infection will be presented separately. When possible the results hereby presented refer to the ones obtained after adjustment (if any). The type and effect of confounding and mediating variables will be discussed in the next paragraph.

(a) TB disease

The core studies - Of the nine core studies, seven reported evidence of a positive association between low SEP and TB disease, one study in Malawi reported a reverse association [7] and one in South Africa no association at all [8] (**Table 3.2**). This latter study[8] did not find any association between TB and any of the three SEP indices adopted in the study, apart from lower consumption of meat and chicken observed in TB patients (for which no OR was reported). With the exception of the study from Schoeman and colleagues [8], SEP was always associated with TB when using a SEP composite index [22, 151, 154, 158, 178].

Three studies reported a significant association between a measure of absolute poverty and TB [22, 151, 156]: in Indonesia 71% of the TB patients were found to live in households with a median monthly income per capita below the poverty line [156]. The odds of having TB among people living below the poverty line were almost doubled in South India [151]. In Malawi, 78% of the TB cases recruited in the prevalence survey lived below the poverty line compared to the 38% of the general population [22]. However, of the 139 patients classified as poor, 60% were from high density permanent and traditional planned areas, and 15% only from squatter settlements, suggesting that many cases coming from the poorest areas were missed because of the inclusion in the study of TB cases attending the clinic and in treatment. In one study in China [158], 29% of the TB households were living below the poverty line compared to 23% the control households (P = 0.05), however the average annual income of the TB cases enrolled in this study was not significantly lower than the one reported by controls (P = 0.2). In this study, TB appeared significantly associated with the asset based measure of relative SEP and the self-reported level of poverty[158].

The only core study reporting data on prevalent cases of TB was the one in South India documenting a clear socioeconomic gradient in the TB cases distribution: the rates of prevalent TB were equal to 343, 169 and 92/100.000 in the population groups respectively classified as low, medium and high according to the standard of living index (SLI) [151]. People in the low SLI group were almost four times more likely to be a prevalent case of TB

compared to the reference group (i.e. high SLI). This result contrasts with what reported in another prevalence survey conducted in Chiapas, Mexico [161]. In this survey, pulmonary TB was found to be not statistically associated with any demographic or socioeconomic variable. According to the authors an explanation for these results is that the survey was conducted in 32 districts of Chiapas characterised by extreme poverty, so that the study may have had not enough power to detect any socioeconomic significant difference among the participants.

The other two investigations that enrolled prevalent cases of TB actually referred to self-reported cases of TB. Both studies consistently reported a strong association with all the socioeconomic indicators used: in South Africa, people belonging to the poorest quintile of the asset base index were approximately as twice as more likely than the people in the middle quintile to report an episode of TB in their life [178]. Also in India, the association between low SEP and self-reported TB showed an OR of almost 2, even after controlling for area of residence, education, ethnicity, housing quality, cooking type of fuel [154].

Two of the core studies reported a strong association between BMI and TB [156, 178]: in the study from South Africa TB patients were almost three times more likely to have a BMI below 18.5 compared to the non-cases [178]. TB patients were also two times more likely to have ever missed a meal due to the lack of funds.

The secondary studies - The large variability of data collected and the different measurement adopted made difficult to compare and comment meaningfully the results coming from the secondary studies. **Table 3.3** summarises the main evidence gathered.

In general, as for the core studies, articles were more likely to find an association between SEP and TB when using composite indices of socioeconomic position. When looking at the single socioeconomic indicators, lower education was found to be associated with TB in four [153, 162-163, 168] of the nine studies collecting this information [161, 165, 170, 173, 175]. Occupational related variables were collected in eight studies [143, 153, 161, 165, 168, 170, 173, 175], but were found to be associated with TB only in the two studies from Easter Europe [143, 168]: in both studies unemployed people were five times more at risk of pulmonary TB compared to employed people.

Income was recorded in five studies [153, 162-163, 165, 168] and in four cases was associated with TB: in two studies income was part of a composite SEP index [162-163] so it is difficult to quantify the specific association of this variable with TB; in the remaining two

studies income was analysed as an independent indicator [153, 168], but only in the study from Estonia [168] there was strong evidence of association with TB (with the highest risk among those with no fixed income, OR= 12.3, 95%CI = 3.2-47.3). In the study from India there was only a modest association [153].

The comparison between evidence was further complicated when looking at variables such as crowding and housing quality because of the heterogeneity of definitions and measurements adopted: in the case of crowding, for example, only five [169-170, 173-175] of the nine studies collecting this information reported an association after adjustment. Of the ten studies reporting housing quality data, four studies included this information into a composite index [162-163, 170, 174] that was found to be associated with TB so it is not possible to estimate the relative contribute of housing quality to the risk of TB. In the remaining six studies [153, 161, 165, 168, 173, 175], housing characteristics were explored as single indicators and were found to be associated with TB in four of them [153, 165, 168, 175]

Inadequate food availability was also more frequent among TB patients in another study from South Africa, in which people who ever missed a meal due to lack of funds were twice as likely to be TB patients compared to those who did not. Three studies reported data on BMI [153, 163, 168], of which only two found a significant association with TB [153, 163]. In the study from India [153], TB patients were 11 times more likely to have a low BMI (BMI <18.5 kg/m^2), whereas in the study from the United States people in the lowest BMI quartile were more than twice as likely to have TB [163]. In the study where no association was found, weight and height were self-reported by the study participants and thus likely to be inaccurate (as people are generally prone to overestimate their height and underestimate their weight) [168]. Two studies reported a significantly less adequate food intake among TB patients: in England TB patients were almost four times more likely to report a less frequent consumption of diary products [165], whereas in Estonia they were 5 times more likely to have often experienced shortage of food [168].

(b) TB infection

The search criteria applied in this review did not identify studies having as their main objective the assessment of the role of socioeconomic factors in the risk of TB infection.

The only identified studies reporting some kind of socioeconomic data reported modest associations between SEP and TB infection. Only four studies documented the association

with potential proxies of SEP, such as crowding [157, 177] and malnutrition [159] [152]; however, in all these studies these variables were investigated principally as potential route of transmission (i.e. crowding) or factors interfering with the accuracy of the TB infection diagnosis (i.e. malnutrition), rather than as measures of deprivation. In one study, for example, the number of people in the household was expressed as the household size more than as a crowding index [177]. Another study in Peru [159], for example, showed that for each cm^2 decrease of CAMA, an anthropometric indicator of chronic malnutrition, there was a 5% increase in the odds of being TST positive. This finding was imputed to an antigen-specific anergy to tuberculin in those with lower body protein reserves, rather than a biologically plausible protective effect of chronic malnutrition to TB exposure and infection. In contrast, a study from India showed that even severely malnourished children could respond to tuberculin test being almost 4 times more likely to be TST positive compared to mildly malnourished children [152].

Finally, Pelly et al found a strong association between poor house ventilation and TB infection, but it is not entirely explicit whether this was more common among low SEP households [159].

3.3.3 The assessment of confounding and mediation and aetiological pathways

In epidemiology, a confounding factor is a variable able to distort the association under study because of its relationship with both the exposure and the disease of interest. Unadjusted results for known and unknown confounding factors can be misleading and lead to wrong conclusions. A mediator is also a variable related both to the exposure and the disease under study, but differently from a confounding factor a mediator is in the causal pathway linking the exposure and the outcome. A mediator does not bias an association, rather it can help to explain it.

Likewise for confounding factors, after controlling for a mediating factor the association of interest appears attenuated [181]. However, in this case the reduction in the size of the measure of association is not indicating a bias, but rather how much the effect of the exposure of interest is captured or explained by the mediator. Whereas it is always recommended to control for confounding factors, controlling for mediation can result in an over-adjustment and an underestimate of the strength of the association under study [182].

Whether a variable is a confounder or a mediator is an *a priori* decision largely driven by our knowledge of the effect of that variable and our hypotheses on how the variable works. More details on these concepts will be given in the next chapter.

For the purpose of this review, I have investigated whether in the 'core studies' authors have adjusted the association between SEP and TB for at least another variable, whether this third variable was treated as a confounding or mediating factors, and whether this decision was part of a priori defined conceptual framework.

Four of the nine 'core studies' did not report any evidence of adjustment [8, 22, 151, 156]. Among the remaining five 'core studies', only the one from Glynn and colleagues reported results adjusted by HIV [7]: as reported earlier, in this study the adjustment for HIV did not affect the significant association between proxies of higher SEP and TB. The association between socioeconomic factors and TB was adjusted for HIV only in three of the 11 'secondary studies' [170, 172-173]. In two of these studies the inclusion of HIV in the multivariable analysis did not significantly change the size of the OR of any of the socioeconomic factor investigated and the authors concluded that socioeconomic factors were associated with the risk of TB independently from HIV [172-173]. In the study from Crampin and colleagues, after adjusting for sex, age and HIV, TB was still more common among those with fewer possessions; however, there was no longer an association between TB and quality of housing and schooling level [170]. The association between TB and occupation was attenuated after adjustment for HIV, but this was due more to restricting the analysis to those with HIV rather than to confounding [170]. In a fourth study, because of the small number of people tested for HIV, authors could only repeat the analysis after excluding the HIV positive subjects. They found no difference in the results from the analysis including also HIV cases, but this it is not the same as adjusting for HIV [175].

In the lack of a clear assumption about confounding and mediating factor it is difficult to judge whether the adjustment or non-adjustment was correct. The study in South Africa is the only study trying to separate the concept of confounding and mediation [178]: in this study smoking, alcohol consumption and BMI appeared to be strong predictors of TB in the univariable analysis and the multivariable analysis revealed that this association was not confounded by the household SEP index. However, the size of the odds ratio for the household SEP index resulted 'attenuated' after the inclusion of individual risk factors, suggesting a potential mediation effect of these more proximal factors. They also found that individual education achievement, recent unemployment and low household SEP, were independently associated with the increased risk of TB suggesting that in this study setting the measures of SEP at individual and household level act through separate causal pathways.

Among all the 20 references collected, only two studies mentioned theories of diseases distribution and invoked the psychosocial model to explain the association between socioeconomic factors and TB. The study from South Africa [178] proposed a conceptual framework to describe the causal pathway linking income inequality at community level to the different stages of TB, including infection, disease progression, health care access and finally appropriate disease diagnosis. In their model, they suggested that income inequality affects the structure of the community in which TB is prevalent, leading to more psychosocial stress that would be present in more egalitarian settings. This increased stress may reduce the effectiveness of an individual's immune response to infection and thus raise the risk of developing TB. Unfortunately this conceptual framework was a *post-hoc* model developed to explain the results observed and was not empirically tested in this study. A possible psychosocial explanation was also suggested in another study from Tocque [165]; however, this hypothesis was also not formally tested in this thesis.

3.4 Conclusions

The aim of this review was to answer a list of open questions on the association between SEP and TB. In this section I will first summarise the key evidence gathered and then discuss the evidence gaps identified.

3.4.1 Evidence summary

(a) Is TB still a plague of poverty in the HIV era?

Two studies conducted in Sub-Saharan countries did not find an association between their measurement of SEP and TB. In one case, authors did find a inverse association [7-8]. In the first case, the authors suggested that the choice of socioeconomic indicators might have not been sensitive enough to detect difference in living conditions in the study population. Beyond this, it seems also plausible that the lack of any evidence in the association between SEP and TB could be attributed to the enrolment in the study of notified cases of TB.

Detection bias was also the explanation given from Glynn and colleagues to explain the reverse association observed in Malawi [7]. In this study TB was more common among those with a better occupation and living in better quality houses after adjusting for age, sex, and HIV status. TB appeared also to be more common among those with higher education, but this association did not persist after the adjustment for housing and occupation. At the time

of this study in Malawi, HIV was known to be more common among wealthy groups of the population and one can argue that the observed association between markers of higher SEP and TB was reflecting the link between SEP and HIV. However, the inverse association between SEP and TB persisted after adjusting for HIV and other factors, suggesting that, even if some residual confounding effect from HIV is possible, this is not the full explanation. The authors concluded that these results are more likely to be explained by the different chances of being diagnosed (i.e. more likely among wealthier people) because 48 out of the 69 cases included in the study were passively found by the health services [7].

The selection bias can explain the lack of association between socioeconomic indicator and incident TB, but even when dealing with prevalent cases results are inconsistent: the prevalent survey from from India [151] showed a strong association between living conditions and prevalent TB, whereas the one from Chapas (Mexico) did not [161].

The inclusion in this review of the secondary studies further offered the opportunity to verify whether the HIV epidemic in developing countries could have introduced some discontinuity in the epidemiology of TB, with crowding and socioeconomic factors becoming less important in TB epidemiology [58]. The evidence is too scattered to draw conclusions: even if results from single studies give the impression that risk factors for TB in developing countries are today the same as were identified more than 100 years ago in industrialised countries (i.e. crowding, poor housing), when the evidence are aggregated this conclusion is no longer so convincing. For example, almost half of the studies included in this review and reporting data on crowding found no association between TB. Similarly, only one third of the studies having data on housing quality were able to show an association between this exposure and TB. The same observations are valid when looking at the risk factors for TB infection.

Probably the most relevant methodological limitation of all these studies lies on their cross-sectional nature. No cross sectional study can accurately measure the effect of these exposures *before* the onset of TB infection and disease and it is almost impossible to conclude whether affected cases have TB because they are poor or they are poor because they have TB [5, 183]. Because of the vicious cycle between low SEP and TB [184], it is possible that authors are observing the impoverishment of TB rather than the effect of socioeconomic factors on the risk of TB. The main consequence of this phenomenon, also called *reverse causality* or *endogeneity* – is that the lack of a clear chronological sequence between the exposure and the outcome hampers the inference of a ia causal role of socioeconomic factors.

A second major methodological limitation is the frequent lack of information on the HIV status of the study participants and thus the inability to account for its confounding or mediation effect. This is particularly relevant for studies from Sub Saharan Africa, where the HIV epidemic may have become more important than socioeconomic factors in explaining cases. Unfortunately, the evidence from this region are still too few and too little diversified to allow conclusions.

A third methodological limitation is that most of the studies present crude or minimally adjusted results. Among those reporting adjusted results, few make explicit assumptions about what they consider to be the confounding or mediating factors of the association between the SEP and TB. The lack of this information makes it difficult to interpret the results implications.

(b) Is SEP equally important in the aetiology of TB disease and TB infection?

Only four of the eleven 'infection studies' included in this review provided evidence of the association between SEP and TB infection; however, in most of these cases the identified risk factors were treated more as route of transmission, rather than socioeconomic factors. These results contrast with what observed in three studies not included in this review because not containing enough methodological information in North America between 1963 and 1978: in a study from Maryland in 1963, authors found that high school students belonging to households characterised by higher educational level of the head, lack of crowding and high quality housing had lower rates of TST reaction [185]. Similar results were found in a small survey in the United States, that found that TST positivity rates were lower among people who were better educated, had higher income and better jobs [186]. Finally, in a study from New York city, tuberculin reactors were more common among persons of lower SES (22.4%) than among those of higher SES (5.5) after adjustment for age, sex and ethnic group [187]. The lack of more methodological details makes it difficult to make proper comparisons with the evidence observed in this review.

The scarce evidence on the association between SEP and TB infection can be due to the study design of these investigations. Most of the studies focused on household contacts of TB cases. It was, thus, possible that the enrolled households were sharing the same impoverishment effect from having a TB case and that they were indirectly overmatched by socioeconomic status. If this is the case, it is possible that these studies were not enough powered to detect a difference in terms of household socioeconomic characteristics. The 'overmatching' explanation is also plausible in two other studies [155, 164]: in the study

from Saiman and colleagues children were clinic-matched, and therefore potentially matched by area of residence and health seeking behaviours which are both associated with SEP [164]. In the study from Rathi, authors compared household contacts of TB smear negative and smear positive patients: as before, these contacts may have shared the same impoverishment effect of TB, regardless the clinical presentation of TB [155].

If the overmatching explanation is plausible remains difficult to explain why studies using as reference group households not affected by TB (as reflection of the general population) did not report a positive association between SEP and TB. It could simply be, as concluded by most of the references, that the strongest risk factor for TB infection is being in contact with TB cases and that this risk factor plays a much more relevant role than SEP. However, as well as for TB disease, even for TB infection the number of evidence investigating socioeconomic factors is too small to draw definitive conclusions.

(c) How SEP is measured? What causal pathway is hypothesised?

A full understanding of the real effect of SEP on the risk of TB is further hampered by the large variability of methods used for the measurement of SEP. The use of composite SEP indices was reported in a minority of studies and even in these cases only two papers make explicit the weighting strategies adopted [22, 178]. In the remaining cases either the weighting strategy was not mentioned or data were combined into a composite index according the judgment of the authors

Similar observations can be made for studies assessing the risk factors for TB.

The absence of a conceptual framework also impairs interpretation. Two studies suggested a psychosocial model to explain the association between SEP and TB. However, these hypothesis were not formulated *a priori* and were not empirically tested [165, 178]. It is difficult to say whether in the remaining studies these issues were not taken into account or they were simply not reported in the publications.

3.4.2 Evidence gaps

(a) The association between SEP and TB

- The evidence from analytical studies is less conclusive compared to ecological studies. Someone may argue that observed inconsistencies should not undermine the general

conclusion emerging from these studies largely suggesting the persistence of the association between socioeconomic factors and TB even in the HIV era. This might be fair, but are the references presented here methodologically strong enough to support this conclusion?

- TB is widely associated with poor living conditions. However, the evidence reviewed above suggest that this belief is often based on a surprisingly small number of analytical studies, that provide inconsistent results and that have at least two main methodological limitations: reverse causation and inadequate adjustment for confounding factors.

- Very few studies report data on HIV status. Lack of data on individual HIV status could be of less concern for interpretation of the association between socioeconomic risk factors and TB if HIV was considered to be a mediator (and not a confounder) in the pathway between the exposure and the outcome. In such circumstances, the inclusion of HIV would be considered inappropriate as leading to an overadjustment in the analysis. No studies, however, make an explicit assumption on whether HIV should be considered as a mediator or a confounder and therefore doubts on the inaccuracy of the results remain.

- The articles reviewed in this chapter showed modest evidence of association between SEP and TB infection. These results can be interpreted as follows: 1) they simply reflect the revision of an inadequate number or type of references; 2) they are due to study design limitations; 3) factors other than SEP are more important determinants of TB infection. The limited number of studies hampers the possibility to identify the most likely cause.

(b) Heterogeneous measurement of SEP

- The limited use of composite indices, often based on different construction strategies, hampers the comparisons of results from different studies.

- None of the studies included in this review provided a clear definition of SEP nor described how this definition was operationalised, and how this exposure was meant to be linked to TB. It could well be that this rational exists but it was simply not reported in these articles and it is also true that that logistic and practical constrains may dictate the quantity and type of data collected. In both cases, in the lack of better explanations the

selection of socioeconomic variables and the measurement of SEP appear to be incompletely explained.

(c) The use of conceptual frameworks

- Only two studies mentioned a possible pathway underlying the association they detected. However, these hypotheses were not formally tested. Thus, the vast majority of studies are still focused on *how much* TB is due to socioeconomic factors, rather than *how* TB is due to socioeconomic factors.

- Because of the lack of a clear conceptual framework it is difficult to fully understand the role of the risk factors hypothesised to be involved in the different stages of TB (**Figure 2.3**) and whether or not they would be still as important after controlling for household SEP.

- The main disadvantage of not formally conceptualising a pathway and testing the hypotheses originating from the conceptual framework is that it is difficult to identify possible entry points for interventions or to imagine to what extent the interventions proposed are likely to have an impact in TB control.

3.5 The study rationale

The results gathered from the review presented in this chapter were used to identify the objectives of this study and to propose strategies able to improve the methodological rigor of studies on SEP and TB.

The review highlighted the need for a study providing evidence on the separate effect of SEP on the risk of TB disease and infection and elucidating *how* household SEP can influence someone's likelihood of being a TB case or being TB infected.

For such a study, two aspects appeared methodologically critical:

1. The generation of a composite tool for the measurement of SEP based on the explicit definition of SEP, how the index is intended to measure this concept of SEP and the statistical method used to weight the items included in the index.

2. The formulation of study hypotheses based on a priori defined conceptual framework. The use of a conceptual framework requires the preliminary definition of what variables

are relevant in the association between SEP and TB, how they are hierarchically linked, and whether they should be considered confounding or mediating factors. For this reason, the definition of a conceptual framework can provide guidance for the variable collection, the analysis of the data and even the interpretation of the results. Such a conceptual framework can also help to identify interventions at different level that may be useful to reduce TB inequalities.

In the next chapter I will describe more in details how these methodological challenges have been addressed and the study objectives.

4. Methods

The main function of epidemiology is to discover groups in the population with high rates of disease, and with low, so that causes of disease and of freedom from disease can be postulated. [...] The biggest promise of this method lies in relating diseases to the ways of living of different groups, and by doing so to unravel 'causes' of disease about which it is possible to do something.

Morris JN[5], 1955

Introduction

The aim of this chapter is to describe the study objectives, provide a justification and a detailed description of the study methods and describe the conceptual framework from which the study hypotheses originate.

This dissertation consists of three main components, all nested within a large scale population based TB-HIV prevalence survey conducted in two communities in Zambia during 2006: 1) an ecological study; 2) a case control study; 3) a cross sectional study. After introducing the study objectives and the study hypotheses, these three components will be presented separately to facilitate the understanding of the specific methodologies implemented.

4.1 Study objectives

The objectives of this study are:

1. To describe the community and household living conditions of the people involved in a population-based TB and HIV prevalence survey conducted in two Zambian communities (**Chapter 5** and **Chapter 6**).

2. To describe the ecological association between community living conditions and TB prevalence rates detected in the prevalence survey (**Chapter 5**)

3. To explore in detail the association between household SEP and TB in two communities of Zambia by (**Chapter 6 and 7**):

- Estimating the size and directions of the association between household SEP and cases of TB disease detected within the prevalence survey.

- Quantifying the association between household SEP and TB infection among a disease-free population using a more sensitive test for TB infection (QFT® - TB Gold (In Tube) [QFT]).

    - Assessing which aspects or domains of household SEP are the most important driving force accounting for the associations under study.

    - Investigating the mediation effect of individual risk factors included in the hypothesised causal pathway between household SEP and TB (both infection and disease).

4. To verify to which extent the study findings are robust to the choice of the household SEP indicators through a sensitivity analysis of the main results (**Chapter 8**).

5. To discuss how the study findings can be translated into concrete policy actions and to indicate directions for future research in the study of TB inequalities (**Chapter 9**).

4.2 Conceptual framework

In developing the conceptual framework for this study I have tried to achieve the following: a) make visible the concepts discussed in **Chapter 2** (i.e. the theories of disease distribution, SEP conceptualisation and measurement); b) identify different levels of causation (i.e. structural vs mediators); c) describe a plausible causal pathway by which SEP at household level might generate the unequal distribution of TB risk factors at individual level.

This conceptual framework is inevitably a simplification of the reality, representing just one of the many possible mechanisms underlying the association between household SEP and TB.

Throughout the thesis I will adopt the same nomenclature suggested by the Commission on Social Determinants of Health [72]: in particular, I will refer to community and household socioeconomic conditions as the *structural determinants* of TB inequalities, whereas the individual biological, behavioural and material risk factors will be identified as the *social determinants* of TB.

In the study setting of this thesis, TB inequalities can be explained by a psychosocial interpretation: even in a context of extreme poverty, in fact, the poorest of the poor can perceive the extent of inequalities surrounding them and translate this frustration into some physical condition increasing the vulnerability to TB. However, because of the severity of deprivation experienced from the most marginalised individuals in these communities, I have hypothesised that stress generated by the perception of income inequalities, lack of social support, poor employment conditions (in terms of poor control, safety and reward in the work place), anxiety, racial or gender discrimination, depression, violence , would be likely to have a less prominent role compared to the material living conditions, like inadequate nutrition, bad air ventilation or bad housing quality.

Following to this, in the development of the conceptual framework I have preferred to adopt a material/neomaterial approach, trying to combine the materialistic view of the first one and the 'ecological' perspective of the second one. This decision reflects my hypotheses of how household SEP may be associated to TB, but it also results from pragmatic reasons: the adoption of the psychosocial model would have required the time consuming and expensive collection of biological samples, making the study not only far more complicated, but also unfeasible given the resources available.

The conceptual framework includes also some elements of the behavioural model: I have, in fact, hypothesised that people belonging to low household SEP are more likely to engage in unhealthy behaviours, such as attending smoking, abusing alcohol etc., which all increase their risk of TB (disease or infection). However, this model explains only one possible pathway through which household SEP generates TB inequalities. Moreover, this model seems to me somewhat embedded into the material/neomaterial model: as argued Bartley, people's unhealthy behaviours are rarely the result of their personal or psychological characteristics, but they are more often the reflection of their low income and their disadvantaged position in the society [188]. More specifically, in this thesis the difference in the distribution of TB-relevant behaviours have been hypothesised to be a consequence of people's household SEP rather than an individual choice. This – although indirectly – leads to a material explanation of TB inequalities.

4.2.1 The pathway

In this framework, I have hypothesised that TB (either infection or disease) might be the result of interactions between different levels of causal conditions, embedded with each other

in a hierarchical fashion: (A) the community; (B) the household; and (C) the individual level **(Figure 4.1)**.

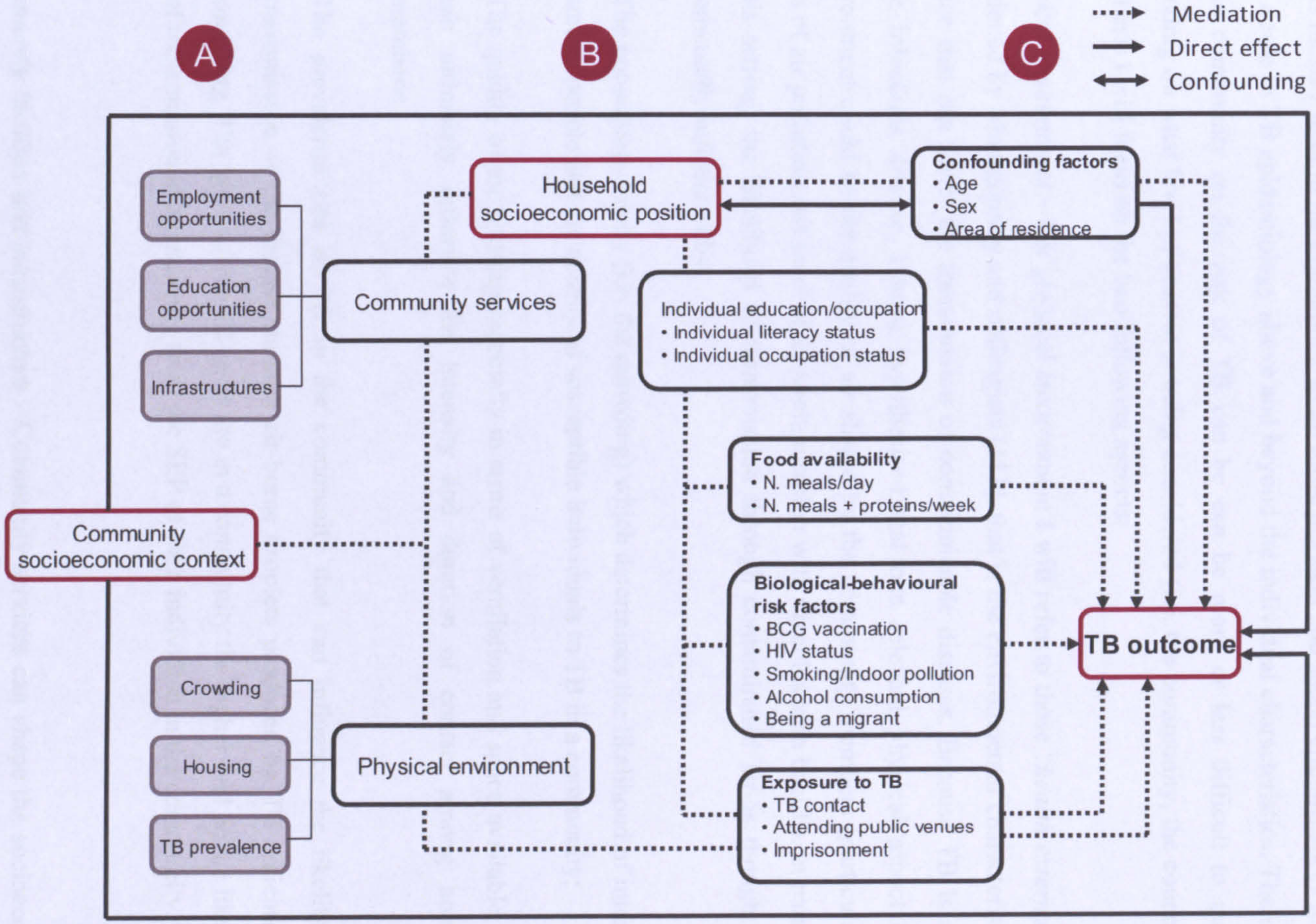
The hypothesised causal pathway can be summarised as follows: the *community*, environment through infrastructures and utilities, can contribute to generate the socioeconomic stratification at *household* level which, in turn, could be responsible for the differential distribution of risk factors for TB among the *individuals*. In other words, the hypothesis is that the standards of living achieved in a community can shape the household SEP, which is itself ultimately responsible for the individual's *opportunities* (in terms of education, working conditions, nutrition, housing quality, and social interactions) and his/her *health related behaviours* (e.g. smoking, alcohol consumption, HIV, BCG vaccination)

The *community* can affect the risk of TB also through its physical characteristics which can contribute to create an unhealthy environment increasing vulnerability to TB and influencing the transmission patterns.

The community and household socioeconomic characteristics together constitute the *structural determinants* (or the *'social determinants of health inequalities'*), that is the social process underlying the unequal distribution of the TB risk factors between groups occupying unequal position in the society [80]. These risk factors are seen as the potential *mediators (or the social determinants of health)* of the association between SEP defined at household level and TB measured at individual level, both as disease and infection. In other words, the risk factors traditionally investigated in the epidemiology of TB infection and TB disease do not confound the association with household SEP. On the contrary they are on the causal pathway that underlies this association.

As suggested by Lonnroth [30], these mediators have been divided into factors increasing the risk of exposure and infection and those promoting the progression from infection to disease through immune system impairment. In this framework I have hypothesised three different mediating pathways, through: 1) malnutrition; 2) health risky behaviours; and 3) increased exposure to TB. They will be discussed more in detail later.

Figure 4.1: The conceptual framework



A: Community level; B: Household level: C: Individual level

4.2.2 The structural determinants

(a) Community socioeconomic position

In this thesis I refer to the community as all those physical and social circumstances that can have a role in TB epidemiology above and beyond the individual characteristics. The impact of the community on the risk of TB can be can be more or less difficult to quantify depending on what level of context is being considered (i.e. the community, the country). In this thesis I will focus on the two following aspects:

Physical environment – For physical environment I will refer to those "*health externalities*" mentioned by Montgomery and colleagues [113], that is the environmental characteristics of a place that can foster the transmission of communicable diseases. Because TB is an air-borne infectious disease, I have hypothesised that one relevant physical aspect of the environment could be the quality of air shared by the community members (particularly in terms of air pollution and level of air contamination with mycobacteria bacilli). In particular, in this setting, the likelihood of transmission through contaminated air is thought to be fundamentally influenced by:

1. The population density (i.e. the crowding) which determines the likelihood of interaction among people and the density of susceptible individuals to TB in a community;

2. The quality of the housing especially in terms of ventilation and space available, which can ultimately influence the intensity and duration of contact among household members;

3. The prevalence rate of TB in the community that can influence the likelihood of transmission of the disease through air-borne droplets produced by TB patients when coughing. The more active TB cases are in a community the higher will be the likelihood of transmission independently from the SEP of each individual in the community.

Community facilities and infrastructure - Community services can shape the socioeconomic profile of a household and therefore, indirectly, the risk of TB at individual level. The determinants of household SEP are difficult to establish and likely to be complicated: in a simplified model it can be hypothesised that the community characteristics might influence the likelihood of access to public services such as health services, transport, markets and, most importantly, schooling and occupation opportunities of the individuals. The availability

of these services might ultimately determine the occupational status of the household and its level of education. Occupation and education themselves influence the household earning capacity and, therefore, the food availability and the material living conditions shared by the household members.

(b) Household socioeconomic position

Consistent with the definition provided by Lynch [79], in this thesis I use the term SEP to reflect the social and economic factors that influences what position(s) individuals and groups hold within the structure of a society. Household SEP has been conceptualised as a multidimensional phenomenon comprising four domains: 1) human resources; 2) food availability and vulnerability; 3) housing quality and assets ownership; and 4) access to community services.

These dimensions are hypothesised to be associated with prevalent TB disease or infection in the following way:

Human resources - The human capital of a household reflects its capacity to generate income, to consume and to have assets, which are all key economic aspects of SEP. Key indicators include the education and occupational status of household members. In a Weberian prospective, occupation can be seen as related to health through prestige, power and social networks or more directly through income and access to material resources [189]. Education might influence earning potential, but also reflects access to knowledge and information on health and through this individual health behaviour [189].

In respect to the postulated conceptual framework, it is assumed that this domain is likely to affect the work and educational opportunity of an individual and consequently is behaviours and the access to material resources, especially nutrition, and the likelihood of contact with TB affected people.

Food availability and vulnerability - Food security is "[...] a situation that exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life"[190].

The attainment of food security is influenced by its availability, access and utilization. Food availability is described as when sufficient quantities of appropriate, necessary types of food from domestic production, commercial imports or donors, are available to the individuals or are within reasonable proximity to them or are within their reach [191].

Food availability reflects household SEP in several ways. Poorer household may consume less food on a less regular basis than wealthier households and/or may consume less costly food. In many low income countries food may make a large share of consumption expenditure. It is therefore possible that the inclusion of simple measure of food consumption can increase the agreement of the wealth index with consumption expenditure [192].

Vulnerability has been defined as the probability or risk today of being in poverty or of falling into deeper poverty in the future and reflects the capacity of an household to react to different shocks at household level (i.e. illness or health), at community level (i.e. pollution, riots) or at national or international level (i.e. national calamities or macroeconomic shocks) [193]. It is a key dimension of welfare, because a risk of large changes in living conditions may constrain households to lower investments in productive assets, to diversify their income sources, and may influence household behaviour and the coping strategies implemented by the household.

Vulnerability has been also defined as the people's propensity to fall, or stay, below a food security threshold [194]. In other words, inadequate food availability can be considered as a marker of vulnerability and vulnerable households are more likely to be food insecure. Because these two aspects can mirror each other, I have decided to include them in the same SEP domain.

Food availability at household level might influence the nutritional status of individual household members and therefore their vulnerability to TB. Vulnerability can influence the risk of TB indirectly because of its relation to food insecurity and by changing the health seeking behaviour of the household or by inducing coping strategies that increase the risk of TB exposure (for example, migration or moving into smaller, poorer quality houses).

Housing quality and assets ownership - It may reflect the material living conditions of the household because they reflect the accumulation of permanent assets over time. Housing quality is reflected not just by the size, but also by the materials used for its construction and the type of facilities available to and used by the household. Relevant assets include

productive assets (e.g. land, animal ownership) and durable assets (e.g. TV, radio, bicycle or car). Asset ownership is generally thought to reflect a long term aspect of SEP because it less prone to fluctuation in time than income.

This dimension describes all those physical circumstances, mostly ventilation, indoor pollution, physical and chemical hazards and crowding, that are linked to health.

Access to community services – This is a broad concept including access to public goods and utilities as power, piped water supply, sanitation, and sewerage, solid waste collection and disposal and to public facilities as roads, transport, schools, public buildings and health care posts. This dimension reflects not only the community features of the place where an individual lives, but also his/her behaviour in terms of social mixing and contact pattern and health seeking behaviour.

4.2.3 Potential confounding factors

As anticipated in **Chapter 3**, confounding factors are variables that produce a mixing of the effect of the exposure under study on the disease with that of a third factor [195]. For a variable to be a confounder it must be associated with the exposure and, independently from that exposure, be a risk factor for the disease. A potential confounder should not be an intermediate link in the causal chain between the exposure and the disease under study.

The main confounding factors considered in this study were age, area of residence and gender. The rationale of this choice is discussed below.

(a) Area of residence

TB tends to be more common in urban settings for reasons probably related to population density, crowding, environmental pollution and housing characteristics influencing ventilation issues [196]. Living conditions are also known to be quite different between urban and rural area, and are generally lower in rural areas [197].

(b) Age

Large differences in TB incidence are often observed by age, with higher incidence of disease with increasing age [26], while individual age can affect the household SEP by its association with the level of economic dependency ratio in the household. The economic dependency ratio measures the extent to which the economically inactive population is dependent on the economically active (the labour force) [14]. The distinction between economically active and inactive depend also on age as some people may not be able to work because too young (i.e. full-time students) or too old (i.e. retired).

(c) Gender

TB notification rates in males are globally higher than those of females for all ages over 15 [198]. It is not clear yet whether this is an epidemiological feature of TB or whether female cases of TB in developing regions are under-reported because of stigma and barriers to access that are gender-specific. This latter hypothesis is supported by studies comparing active with passive case finding, which found that women with TB were consistently under-notified to public health authorities when relying on passive case-finding. A review suggested that not only TB among women is under-reported because of cultural and socioeconomic factors, but also these factors can actually increase the chances of women to be exposed to TB bacilli and to progress from TB infection to TB disease (through for example malnutrition and HIV co-infection) [199].

Poor living conditions are also experienced differently by men and women. Female poverty is more prevalent and typically more severe than male poverty [70]. Gender inequality concerns all dimensions of poverty: women and girls in poor households may receive less than their 'fair share' of private consumption and public services; they typically have much less access to job opportunities; females owning enterprises often cannot access credit and when they work for wages they generally earn less than men; even when they have the same education of men and work experience. Finally, they are more likely to be illiterate as well as politically and socially excluded in their communities. Women in general have less access than men to assets that provide security and opportunity. Hence, women's abilities to overcome poverty are generally different from those of men.

4.2.4 Mediating factors

Mediation has been defined as "the totality of process that explains an observed relationship between exposure and disease" [200]. In epidemiological studies, a mediation model is introduced when the aim of the analysis is to explain and quantify the mechanism that underlies the relationship between an independent variable and a dependent variable by the inclusion of a third explanatory variable, hypothesised to be in the causal pathway and known as a mediator variable [181].

In my conceptual framework I have hypothesised that the effect of household SEP on TB might be mediated through at least the three different mechanisms that are described below. A fundamental assumption in this hypothesis is that household SEP is determined temporally earlier than the occurrence of the mediating factors that increase vulnerability to TB.

(a) Inadequate nutrition

Malnutrition has been notoriously associated with TB: inadequate food intake enhances the development of active TB and active TB makes malnutrition worse. Biologically, it has been suggested that generalised malnutrition, by reducing the expression of gamma interferon and other mycobactericidal substances – may selectively compromise portions of cell mediated response that are important for preventing TB infection from progression [201]. However, very few epidemiological studies have clearly proved whether malnutrition leads to development of active TB or active TB leads to malnutrition [202].

In this thesis I hypothesise that food unavailability might increase the likelihood of TB disease progression and expose people to a higher risk of TB infection when exposed. As discussed previously, time series data show the temporal relationship between trends of TB and food security in recent history: TB mortality has rapidly increased during the First World War, declined subsequently, and increased again during the period of monetary inflation with serious food shortages in the years 1922-1923. In the Netherlands, TB mortality increased with the beginning of the Second World War, but the increase was accentuated when a German food embargo led to a sharp drop in available calories [26].

(b) Biological and behavioural risk factors

This group includes HIV infection, lack of BCG vaccination, alcohol abuse, smoking (including tobacco and indoor pollution), and being a migrant. These variables are the risk

factors traditionally explored in TB epidemiology: for most of them it is possible to identify an extensive literature showing a variable degree of association with TB infection and disease. The effect of BCG vaccination among adults is controversial [203]; however - because there are studies documenting some protective effective [204] - it was felt appropriate to take BCG vaccination into account. In this thesis, migration has been investigated more in terms of individual's mobility history increasing his/her chance to get TB, rather than a socio-demographic phenomenon with a structural role on TB epidemiology. As for food availability, more than to investigate their association with TB, this doctoral research will try to find out whether the association between SEP and TB is due to a differential distribution of these risk factors across different SEP groups.

(c) Contact with other tuberculosis cases

The environmental factors associated with lower socioeconomic position that can increase the likelihood of TB contacts are:

- an increased likelihood of crowded living conditions;
- an increased likelihood of exposure to places characterised by an increased likelihood of TB transmission;
- an increased likelihood of exposure to poor people who are not only more likely to be affected by TB, but also more contagious because they are less likely to be under proper treatment.

In this thesis the concept of crowding has been explored in relationship with SEP more than as a route of transmission. For these reason, detailed data were not collected on the type and duration of contacts in the households, ventilation characteristics of dwellings, nor on the size of these households. Consistent with other references, crowding has been defined in terms of number of people sharing the same bedroom [35].

As for crowding, in this thesis, attendance of public venues and imprisonment has been considered more as a proxy of life style rather than as a route of transmission. I have only taken into account the frequency of attendance of these places (frequency of visits and duration of the visits) and no data has been collected on the physical characteristics of these places. Recent studies of TB outbreaks [205] have shown how complex social networks involving public places can be critically important in TB transmission patterns. The importance of social mixing and places, including bars [206-207], churches [208-209],

prisons [210] in maintaining TB transmission has been documented in a number of studies. There are a number of reasons that can explain these findings, including the level of crowding of these places, their construction characteristics and the ventilation features as well as the different life style of the people attending regularly these venues [206-207, 209], all of which put them at higher risk for TB infection and disease.

4.3 The study hypotheses

Following from the conceptual framework, three main study hypotheses have been formulated:

1. There is a socioeconomic gradient in the distribution of TB prevalence across the Zambian communities under investigation and this can be detected at ecological level by area-based socioeconomic measures (ABSM) derived from the census data.

2. Low household SEP is associated with both an increased risk of prevalent TB and TB infection.

3. In this context SEP might influence TB risk through at least four different mediating pathways at individual level, including: 1) education and occupation; 2) food availability; 3) health behaviours; and 4) the likelihood of exposure to TB.

4.4 Study design

4.4.1 The source population: the latest TB-HIV prevalence survey in Zambia

The present research project was *nested* into a population-based TB-HIV prevalence survey conducted over the period June 2005 to March 2006 in two communities of Zambia, one rural and one urban. This survey was part of the ZAMSTAR project, a community randomised intervention aimed to reduce the burden of TB in Zambia and South Africa by reducing TB transmission (through enhanced case finding and early treatment), reducing the reactivation of latent TB (through TB preventive chemotherapy) and reducing HIV prevalence. The total population included in this study is 1.2 million people, distributed across 24 communities in the two countries. The purpose of the prevalence survey was to estimate the burden of TB in areas representative of the ZAMSTAR interventions Zambian sites, to be used as baseline estimates to measure the impact of the trial [211]. The people recruited in this doctoral research originate from this survey.

Ethical approval for this study was obtained from the Ethics Committee of the London School of Hygiene and Tropical Medicine and the University of Lusaka, Zambia (see **Appendix A** and **B)**

4.4.2 The study setting

Zambia is divided into geographical and political regions, including province (N=9), districts (N=72), constituencies (N=155) and wards. Each ward is further divided into smaller administrative tracts called Census Supervisory Areas (CSA), which are themselves divided into Standard Enumeration Areas (SEA). As stated above, the prevalence survey has been conducted in two areas of Lusaka province: one mostly rural and one mostly urban.

Consistent with the definition provided by the Central Statistical Office of Lusaka, in this thesis a rural area is defined as an area with a predominant agrarian economy; an urban community is an area with a population of 5000 and above. Ideally, an urban centre should be provided with a minimum of social facilities and services such as piped water, electricity, banking facilities etc. For analysis purposes, sometimes it is justifiable to define as 'urban' areas with even less than 5000 inhabitants, but presenting with a minimum of facilities in which the dominant economic activity is not agriculture.

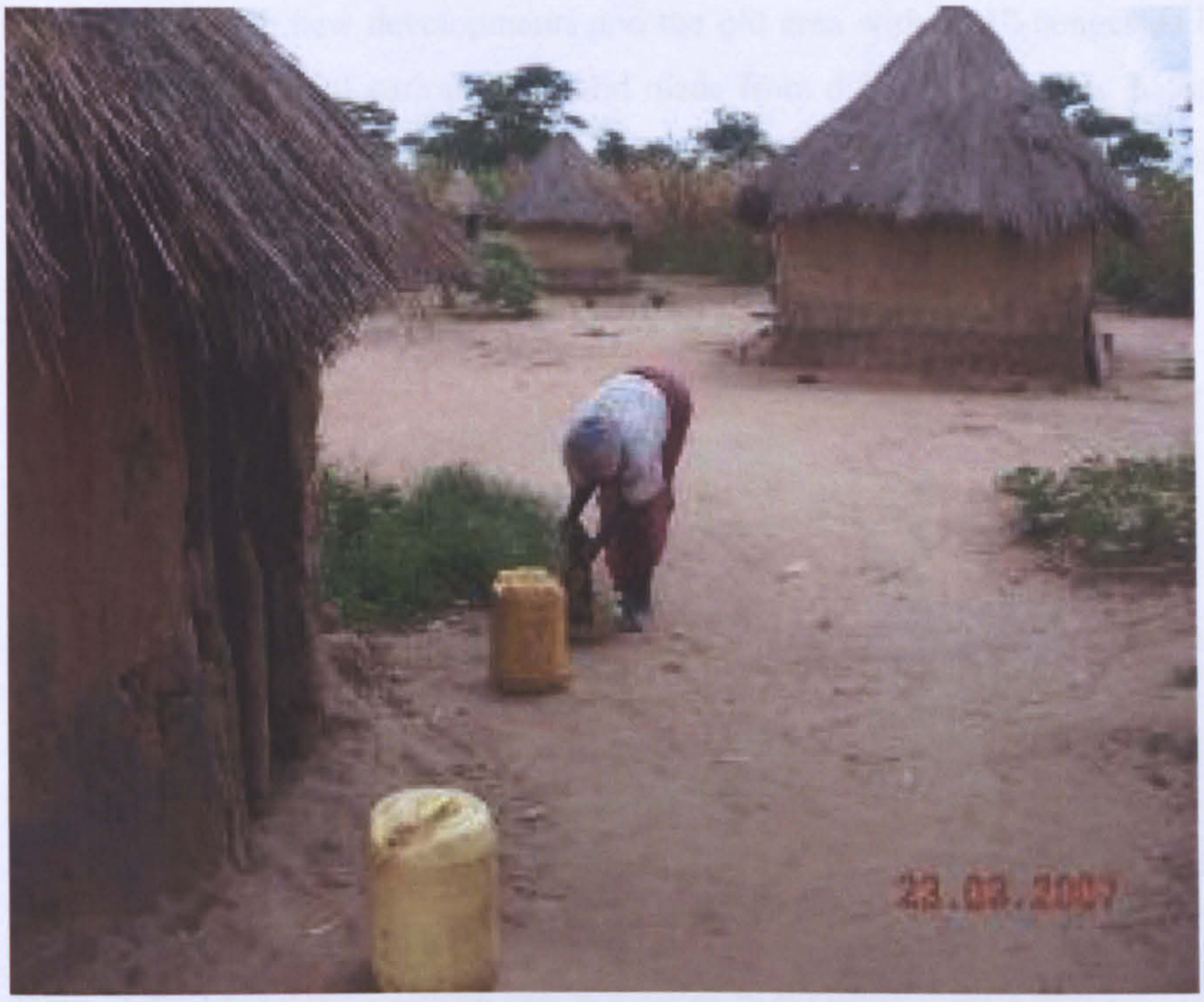
(a) The rural community

The rural community involved in the prevalence survey lies 31 km to the east of Lusaka city and it is home to approximately 17,000 people [212]. Despite mostly rural, this site has also peri-urban settlements along the main road running through it and surrounded structures such as housing, government institutions, council offices, cooperatives, private businesses, shops, markets, guesthouses etc. The surrounding areas are quite rural with dispersed villages, farms, government settlements, and council plots.

Houses in the villages are made of mud bricks and thatched roofs, the farms have a mixture of houses; some built with mud bricks and thatched roofs and some built with concrete blocks and iron roofing **(Figure 4.2)**. The town centre has mostly houses built with concrete blocks, iron roofing, with or without windows. Most of the houses have pit latrines, only few of the houses in the town centre have flushable toilets.

In order to account for the semi-urban nature of the rural site, cases and controls coming from the urban CSA of the rural site where re-classified as urban and the analysis was repeated. It can be anticipated that this has not changed the results presented in this thesis.

Figure 4.2: Typical house in the rural area



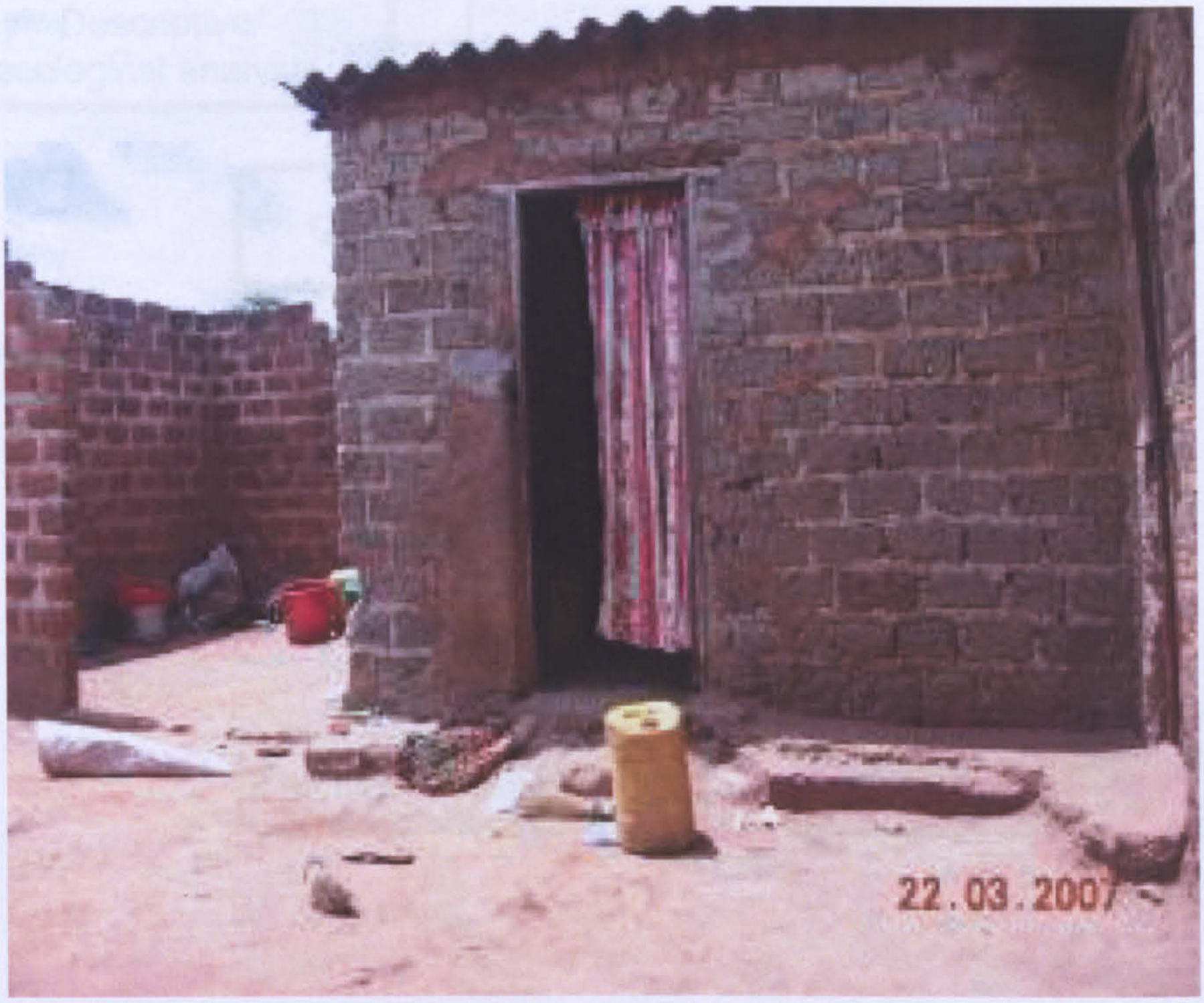
A small proportion of people in the rural community are in formal employment. The area is rather a small farming community, where peasant farmers grow cassava, rape and other vegetables, which are later sold in Lusaka at Soweto market. Other livelihood options are mainly petty vending. People survive doing different kinds of activities: rearing chickens, owning small workshops (for battery charging, welding and bicycle repairing, drawing water), working on farms for piecework, or working as casual workers in the new mines that have opened up around the area. Around town, the men buy fuel from truckers and resell it, whereas women and children sell fritters, groundnuts, maize etc in the market and open areas. In 2004 the TB notification rates for the rural area was 275/100,000.

(b) The urban community

The urban community selected for the prevalence survey is representative of the urban settlements in the major towns of Zambia, but is not as crowded as the high-density areas in the capital, Lusaka. From the data reported in the 2000 Population and Household Census for the province of Lusaka it has been estimated that approximately 11,000 people were living in the urban community involved in the prevalence survey [212].

It is a small shanty compound generally surrounded by farms that can be divided into two areas: one new area with new developments and the old area with small-congested houses. This urban area has houses of various sizes and made from different materials. Most of the houses are made from burnt bricks while others from concrete blocks. For the roofing, some houses are thatched, other have asbestos and others have iron roofing (**Figure 4.3**).

Figure 4.3: Typical house in the urban area

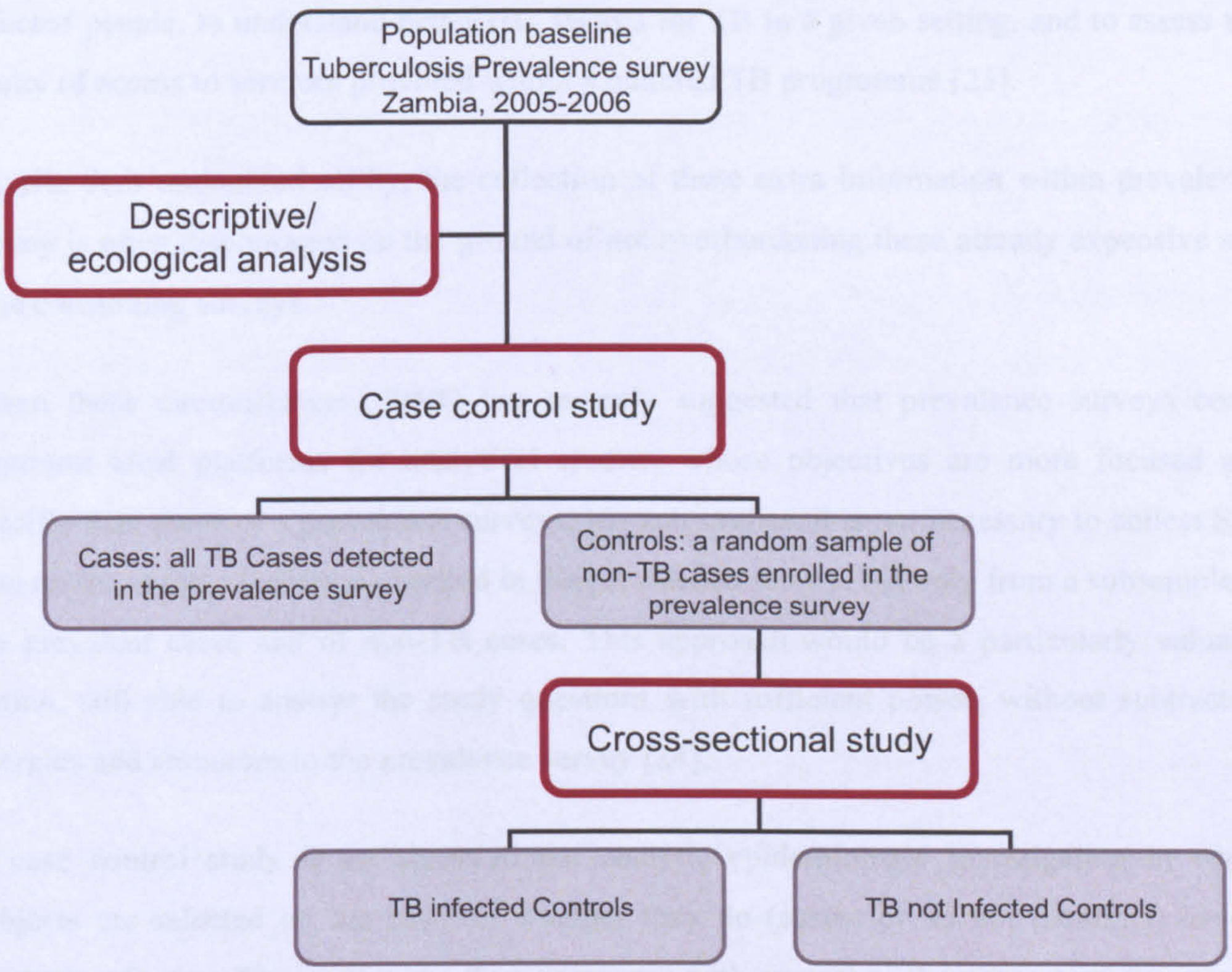


This urban settlement is generally poor with very few livelihood options. There are a number of retirees who have settled in this area. People generally find piecework in the surrounding farms, others also work in small gardens, where they grow vegetables. People spend more time daily looking for ways and means to survive because of the poor status of the compound. In 2004 the TB notification rate for the urban area was 438/100,000.

4.4.3 Study components

In order to meet the research objectives, three different studies have been implemented: 1) an ecological analysis of data from the prevalence survey; 2) a frequency matched case control study; 3) a cross-sectional study. In this section I will provide a brief description of these components and a justification for the selected methodology. More details will be provided later in this Chapter. The overall study design is outlined in **Figure 4.4**.

Figure 4.4: Overall study design



(a) The ecological study

The aim of the ecological study was to provide a socioeconomic profile of the communities where the TB-HIV prevalence survey was conducted. This profile is simply a comprehensive socioeconomic comparison to assess how living conditions vary across CSA under study. This analysis also aimed to identify any correlation between a selected list of area-based socioeconomic measures (ABSM) and the risk of prevalent TB at population level in these areas.

(b) The case control study

The aim of the case control study was to assess the association between household SEP and TB disease. For this purpose, household SEP was compared between a subsample of TB cases detected in the prevalence survey and a random sample of the people classified as not TB cases.

The nested design within the prevalence survey follows the recommendation included in a recently published document from WHO [24]: the collection of socioeconomic data together with TB risk factors and health seeking behaviours within prevalence surveys represents an invaluable opportunity to provide an unbiased description of the SEP profile of the TB affected people, to understand better risk factors for TB in a given setting, and to assess the equity of access to services provided within a national TB programme [23].

Despite their undoubted utility, the collection of these extra information within prevalence survey is often discouraged on the ground of not overburdening these already expensive and time consuming surveys.

Given these circumstances, WHO has recently suggested that prevalence surveys could represent ideal platforms for analytical studies, whose objectives are more focused and specific than those of a prevalence surveys. In such studies, it is not necessary to collect SEP data on the entire population enrolled in the prevalence survey, but only from a subsample of the prevalent cases and of non-TB cases. This approach would be a particularly valuable option, still able to answer the study questions with sufficient power, without subtracting energies and resources to the prevalence survey [24].

A case control study is an observational analytic epidemiologic investigation in which subjects are selected on the basis of whether they do (cases) or do not (controls) have a particular disease. The groups are then compared with respect to the proportion reporting a history of exposure or characteristic of interest.

The case control design was chosen because of the following advantages: 1) case control studies are well suited for the evaluation of uncommon diseases (as in this study project), which would otherwise require the follow up of a large number of individuals in order to observe a sufficient number who develop the outcome; 2) case-control studies are an efficient way of studying diseases with a long latency period (as in the case of TB), because investigators can ask affected and unaffected individuals to describe prior exposures without having to wait years for the disease to develop.

(c) The cross-sectional study

The aim of the cross-sectional study was to assess the association between household SEP and TB infection. For this purpose, I have conducted a cross-sectional investigation restricted only to the disease-free controls enrolled in the case control study.

Cross sectional studies are descriptive epidemiologic investigations in which the status of an individual with respect to the presence or absence of both exposure and disease is assessed at the same point time.

In this study, each individual without active TB was classified as TB infected or non-TB infected on the basis of further laboratory tests and these two groups were compared in terms of their socioeconomic characteristics. No additional participants were enrolled in the study and no extra information was collected from the controls enrolled. The *nested* nature of this study seemed a particularly convenient option because it made possible answering another important research question with very limited extra resources investment. Moreover, because the information analysed are the same for the case control study and because the method of measurement of SEP is the same adopted for the case control study, this design allows the assessment of the differential effect of SEP on TB disease and TB infection.

Another added value of this investigation is the adoption of QuantiFERON® - TB Gold (In Tube) [QFT], a relatively new *in-vitro* interferon-gamma release assay for the diagnosis of TB infection. As better discussed later in **section 4.7.2**, this test presents features that are likely to enhance the diagnosis of TB and minimise the chance of biases associated with the tuberculin skin test (TST), the test traditionally employed for the diagnosis of TB infection.

After this overview I will now move into the detail of each single study.

4.5 The ecological study

4.5.1 Sample size and sampling strategy of the prevalence survey

According to Ayles and colleagues [211] a sample of approximately 5000 individuals in each study site was necessary to detect the expected population prevalence of TB of approximately 1000/100.000 with a 95% certainty.

Figure 4.5 shows the maps and a flowchart of the areas of Lusaka province that were involved in the TB-HIV prevalence survey: in the rural community, the area sampled for the

prevalence survey was located in the ward of Chongwe (which is itself included in the constituency and district of Chongwe). All the CSA and SEA of this ward were sampled.

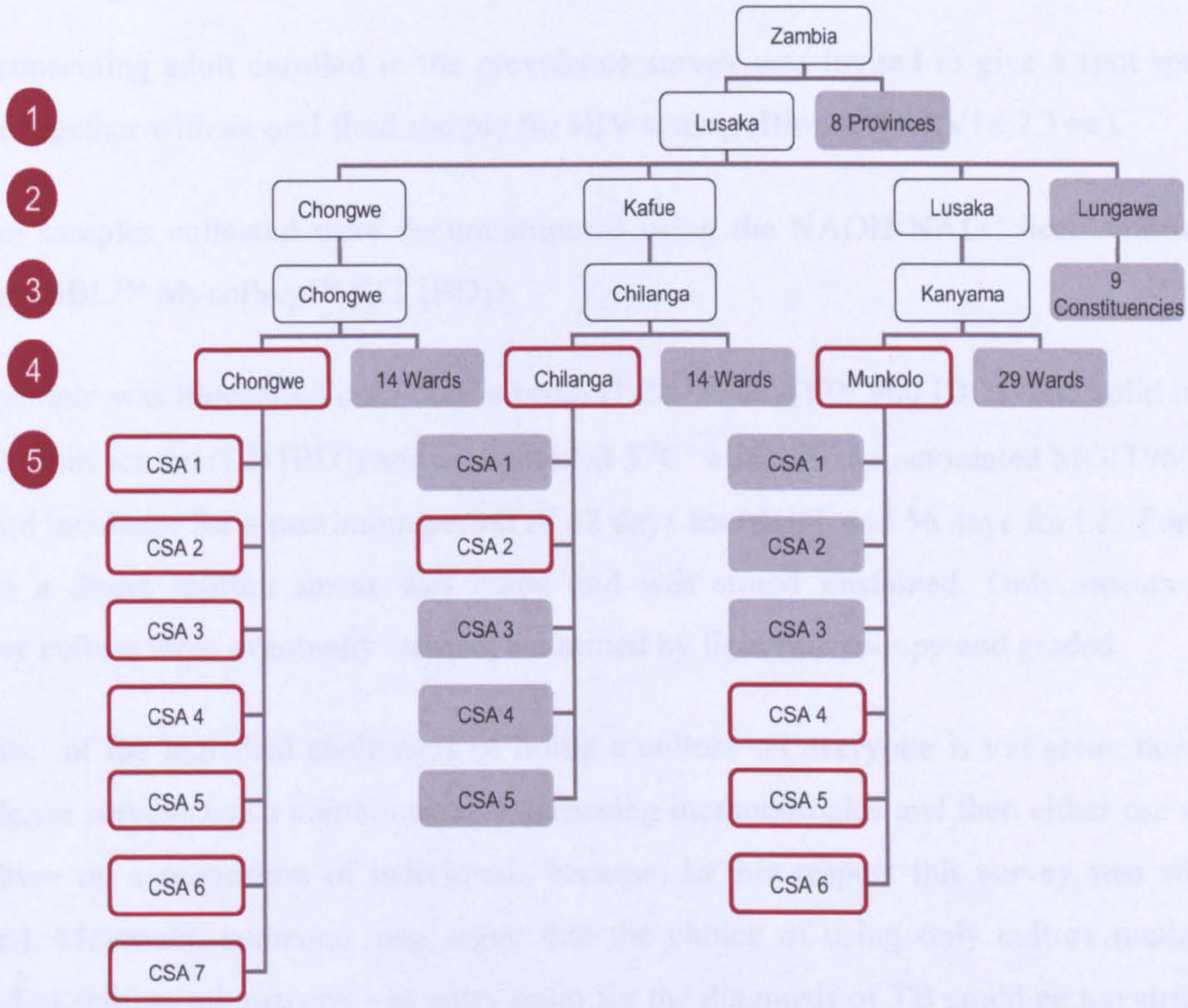
In the urban site, only parts of two wards have been sampled for the prevalence survey. The first one was the Chilanga ward, which is included in Chilanga constituency and Kafue district. Only CSA 2 has been sampled of this ward. The second one is the Munkolo ward, which is included in the Kanyama constituency and Lusaka district. In this ward only CSA 4, CSA 5 and CSA 6 have been sampled.

For the purpose of this thesis, the CSA have been renamed as follows:

Area	Ward	CSA	Nomenclature for the thesis
Rural	1	1	R-11
	1	2	R-12
	1	3	R-13
	1	4	R-14
	1	5	R-15
	1	6	R-16
	1	7	R-17
Urban	2	2	U-22
	3	4	U-34
	3	5	U-35
	3	6	U-36

Each SEA in the selected CSA was assigned a randomised order using a simple randomisation and all households in each SEA were visited.

Figure 4.5: Areas involved in the population-based TB-HIV prevalence survey



Numbers in red dots indicate the level of aggregation: 1) Provinces; 2) Districts; 3) Constituencies; 4) Wards; 5) CSA. Grey boxes indicate the areas that have not been involved in the population-based TB-HIV prevalence survey (for example: of the 4 districts of Lusaka province, Lungawa was the only one not included in the sampling)

4.5.2 The diagnosis of prevalent tuberculosis

Each consenting adult enrolled in the prevalence survey was invited to give a spot sputum sample together with an oral fluid sample for HIV testing (Bionor™ HIV1&2 Test).

Sputum samples collected were decontaminated using the NAOH-NALC decontamination method (BBL™ MycoPrep™ KIT [BD]).

Each sample was inoculated onto both a broth (BBL™ MGIT™ 960 [BD]) and solid media (Löwenstein Jensen (LJ) [BD]) and incubated at 37C° either in the automated MGIT960 or a standard incubator for a maximum period of 42 days for MGIT and 56 days for LJ. For each sample a direct sputum smear was made and was stored unstained. Only smears from positive culture were eventually stained, examined by light microscopy and graded.

Because of the logistical challenges of doing a culture on everyone is too great, normally prevalence surveys use a combination of screening methodologies and then either use smear or culture on a proportion of individuals because. In this respect this survey was slightly unusual. Moreover, someone may argue that the choice of using only culture methods – instead of sputum microscopy - as entry point for the diagnosis of TB could be too strict and potentially able to cause a loss of cases; The prevalence team has justified this choice as follows:

- Culture is the gold standard for TB Diagnosis;

- Symptoms or CXR screening is subjective and misses some cases.

- According to WHO figures, In Zambia the prevalence of HIV in incident cases of TB is 70%. In the prevalence survey from which this study derived, 30% of TB cases were HIV positives. Because of such high co-infection proportion, a lot of the TB cases in Zambia are smear negative. This implies that the use of smear microscopy alone instead of culture for the diagnosis of TB would have resulted in the loss of a considerable number of TB cases.

- Compared to broth culture, the smear microscopy technique has notoriously poor sensitivity (30-60%, depending on the laboratory techniques and patient population). Further, the technique is strongly dependent on the experience, skills and especially the time invested by the lab staff and the number of samples to process.

- In order to minimize the loss of TB positive subjects two culture methods were used: a liquid and a solid culture. The combination of liquid and solid media for the culture of TB is known to be able to increase the diagnosis sensitivity and the use of broth culture generally reduces the time of diagnosis.

Ayles and colleagues have acknowledged that because of poor growth and contamination problems the prevalence of TB they detected is likely to be an underestimate [211]; however this is unlikely to have major implications for the present study: the number of missed cases is unlikely to be big so the sample size would be hardly bigger than this. Finally, for this to be a bias we should assume that the likelihood to be diagnosed by culture is linked to the household SEP of the individuals tested which is unlikely.

At the time of the start of this study research, the microbiological algorithm used for the identification of positive *M. tuberculosis* cultures was based on the response to the niacin accumulation test and spoligotyping.

Niacin functions as a precursor of nicotinic acid [213]. Although all mycobacteria produce nicotinic acid, comparative studies have shown that because of the blocked metabolic pathway for conversion of free niacin to nicotinic acid mononucleotide, *M. tuberculosis* accumulates niacin and excretes it into the culture media from which it can be extracted. Although niacin-negative strains of *M. tuberculosis* isolates are extremely rare, the niacin test should not be used alone to identify *M. tuberculosis* [213].

Spoligotyping allows the simultaneous detection and typing of *M. tuberculosis* in clinical specimens and reduces the time between suspicion of the disease and typing from 1 or several months to 1 or 3 days. The method is based on polymorphism of the chromosomal DR locus. The method is referred to as spacer oligotyping or "spoligotyping" because it is based on strain-dependent hybridization patterns of in vitro-amplified DNA with multiple spacer oligonucleotides. Most of the clinical isolates tested showed unique hybridization patterns, whereas outbreak strains shared the same spoligotype [214]. In order to limit any risk of lab cross contamination – a concrete possibility when using culture methods, all strains that were processed on the same day were spolygotyped and fingerprinted.

All cultures growing organisms positive either to the niacin accumulation test and/or identified by Spoligotyping as *M. tuberculosis* strains were considered positive for *M. tuberculosis*. Discordant results were also double checked using polymerase chain reaction (PCR).

The enrolment of cases and controls during the fieldwork was based on this microbiological definition.

Several months after the completion of the data collection for this thesis, all the positive cultures detected in the prevalence survey were retested using a different, and more specific, technology, the Genotype Mycobacteria CM assay (HAIN, Life Science), and re-classified as positive for *M. tuberculosis* or non-tuberculous mycobacteria. The change in the microbiological definition made necessary the exclusion from the case control study of cases that did not longer meet the new microbiological case definition. This had implications for the present study that will be discussed later in this Chapter.

The final results of the prevalence survey in this thesis refer to this latter HAIN methodology.

All individuals positive for *M. tuberculosis* were traced back and two further samples (spot and early morning) were collected for smear and culture. As part of the follow up procedure, these individuals were further interviewed about their symptoms, examined by trained physicians and taken to the tertiary hospital for chest X-ray examination. These extra information gathered allowed the classifications of individuals into suspect or confirmed cases of TB depending on the clinical algorithm shown in **Figure 4.6**.

A prevalent case of TB was defined as any individual enrolled in the prevalence survey, from whom *M. tuberculosis* was cultured from the original sputum sample. The prevalence of *M. tuberculosis* was defined as the proportion of individuals with a positive sputum sample for *M. tuberculosis* among the total number of people from whom an evaluable sample was available [211].

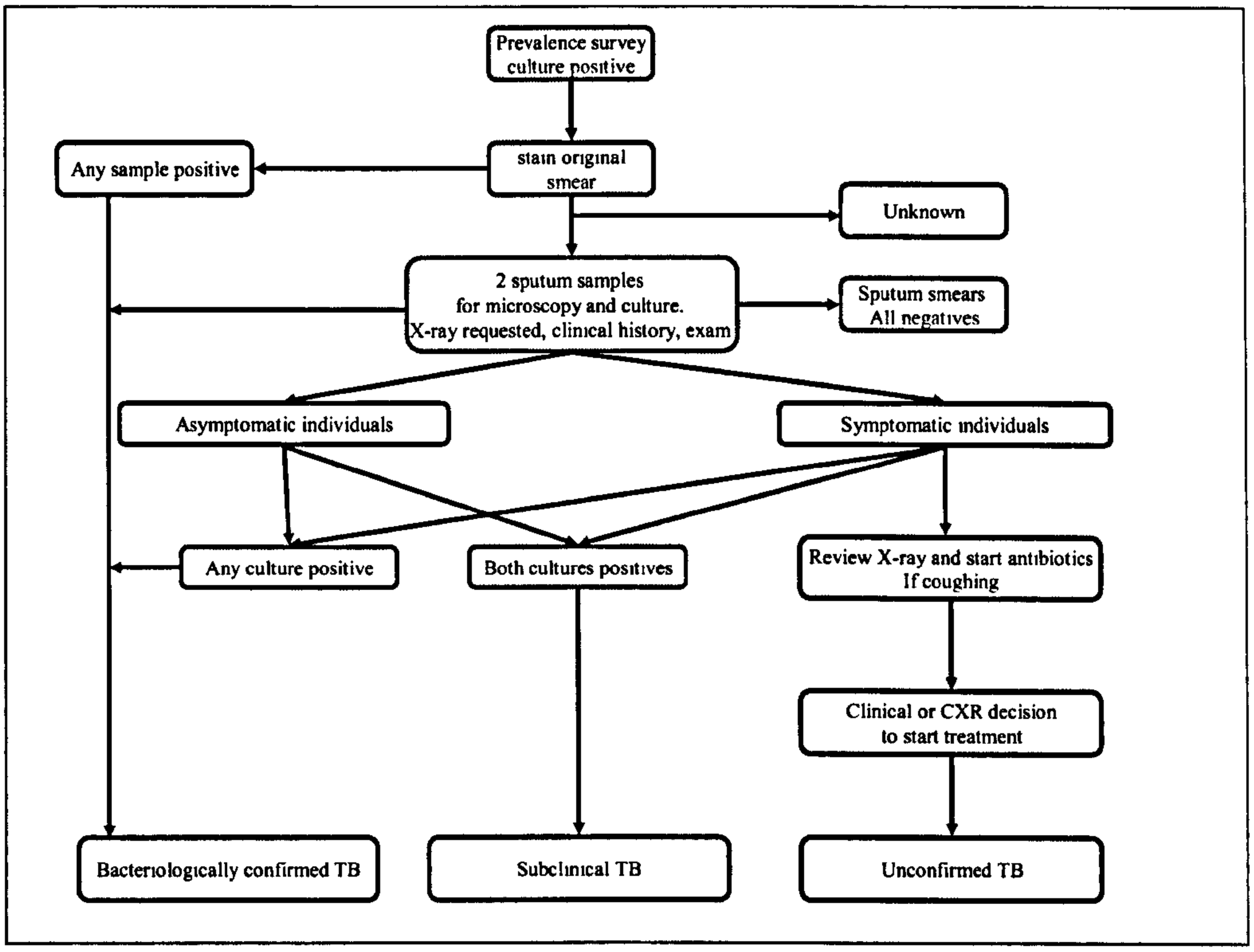
According to the clinical algorithm, it was possible to distinguish three different types of prevalent cases:

- **Bacteriologically confirmed tuberculosis disease:** Any individual with at least two positive results either culture or smear.

- **Bacteriologically unconfirmed TB:** Any individual with a single *M. tuberculosis* positive initial culture who was started on TB treatment due to symptoms or X-ray findings, but with no further microbiological proof of TB, either by smear or culture.

- **Subclinical TB:** Any individual with an initial single *M. tuberculosis* positive culture, but with no further positive smear or culture and no symptoms or significant X-ray abnormalities.

Figure 4.6: Clinical algorithm for the diagnosis of clinical tuberculosis



Reproduced from Ayles et al. [211]

4.5.3 Data sources

For the ecological analysis, the prevalence survey data were supplemented by publicly available datasets in the form of the Zambian Living Conditions Monitoring Standard dataset, year 2002-2003 and the Zambian Census dataset, year 2000, for Lusaka Province. These datasets were considered secondary data not originally collected to meet the specific bjectives of this work.

(a) The Living Standard Monitoring Survey, year 2002-2003 (LSMSIII)

Living Standard Monitoring Surveys (LSMS) were introduced in 1980 by the World Bank in order to improve the quality and type of household socioeconomic data available in developing countries. LSMS are typically multi-topic surveys designed for the collection of data on household livelihood, including consumption, income, saving, employment, health, education, fertility, housing, nutrition and migration. The aim of these surveys is to assess

the household living conditions and behaviour and to evaluate the effect of policies on the socioeconomic conditions of a population [215].

The LSMS survey employed in this thesis was the LSMSIII: a national multi-topic survey conducted in 2002-2003 in Zambia offering a rich database upon which to base a detailed analysis of poverty in this country. The survey was designed to cover 520 Standard Enumeration Areas (SEAs) or approximately 10,000 non-institutionalised private households residing in both the rural and urban areas of all nine provinces in Zambia. The LCMSIII survey employed a two-stage stratified cluster sample design: during the first stage, 520 SEAs were selected with probability proportional to estimated size. During the second stage, households were systematically selected from an enumeration area listing [14].

This survey includes three main sections:

I. A set of core living condition indicators monitored over a period of twelve months and covering the households' demographic characteristics, migration pattern, health and education status of the household, the ownership of amenities, assets and the housing quality, the access to community facilities, the agricultural production and finally the self assessed poverty status of the household together with the coping strategies implemented to respond to times of hardship;

II. The household budget of the household over a period of twelve months; the expenditure data were collected using a diary method and covered the consumption of several types of goods which were aggregated into: 1) food-items, 2) Housing related costs (i.e. rent); and 3) non-food items, including durable goods and also the expenses for health, education, transport, clothing, house wares, electricity etc etc. Excluded items included funeral expenses and costs of marriages. The three sub-aggregates were then added together to obtain a measure of nominal total monthly household consumption expenditure. Because SEP analysis is much more meaningful when it is done per capita items, the total monthly consumption expenditure was also divided by household size to obtain estimates in per-capita consumption terms [14].

III. The third part is the *price questionnaire*, which collects information about a large number of commodity prices in each community where the survey is undertaken. This is useful because it allows analysts to correct for differences in price levels by region, and over time.

Although LSMS provide comprehensive information on a broad range of dimensions of living standards and their determinants and correlates, like income and consumption, these

surveys cover only a relatively small subset of households or individuals, and estimates are only representative of a random sample of a population at relatively high aggregated levels – such as provinces. LSMS surveys are, thus, generally considered not ideal to conduct poverty analysis in smaller geographical areas. For this purpose, the 2000 Census of Population and Housing for the Lusaka Province in Zambia is a more suitable tool.

(b) The Census dataset for the Lusaka Province, year 2000

A census is usually an extremely large and expensive survey in which each member of a population is enumerated and enrolled in the data collection. Census datasets are considered better tools when conducting analysis of community effect: census data can be compiled for small administrative areas, communities, villages or town. They provide information on all individuals and households, thereby allowing for the finest geographic disaggregation. However, censuses are not carried out very frequently (usually once a decade) and they only collect information on a limited set of indicators; in particular, income or consumption expenditure data are typically not available.

It is interesting that the utility of census surveys in the interpretation of epidemiologic data and public health planning was initially recognised in studies on TB conducted in the 1920-1930, after the establishment of the first census tract in New York City: these studies were sponsored by the National Tuberculosis Association and aimed to assess the risk to TB in relationship of the socioeconomic conditions of the census tracts, that were initially called 'sanitary areas' [216].

In this thesis I have used only data relative to the Lusaka province that included data at individual and household data, including: demographic structure of the resident population, education and economic activity of each household member, household housing quality (including building materials and number of bedrooms), assets ownership (including durable goods and productive goods like land and animals), type of water sanitation and main type of water supply, source of energy (for lighting, cooking and heating) and type of waste disposal [212].

4.5.4 The spatial unit of analysis

In the ecological analysis the CSA was the spatial unit of analysis. This level of analysis has been chosen since it is the smallest level of aggregation ensuring a relatively good

approximation of the actual socioeconomic characteristics of the context where the prevalence survey was conducted.

A CSA is a grouping of SEA and is meant to be assigned to one supervisor during the census data collection. It is designed to comprise from two to five SEA. A SEA is a politically demarcated area with a population ranging between 300-500 people and 600-800 people respectively in the rural and urban area.

According to the latest census dataset for the province of Lusaka, in year 2000 each CSA in the rural area had on average 2500 residents and 540 households. In the same year, the mean number of residents and households per CSA in the urban area was 2800 and 660 respectively.

Overall, 11 CSA have been included in the prevalence survey (seven from the rural area and four from the urban one).

4.5.5 Measurement of community living conditions

Two criteria guided the selection of the area-based socioeconomic measures (ABSM): 1) they had to be theoretically-justified, that is they had to measure those socioeconomic constructs that were hypothesised in the a priori conceptual framework [217]; 2) they had to be indicators suitable for policy interventions [94].

For the measurement of ABSM I employed a list of 23 census variables (**Table 4.1**) combined into:

- Two area-based composite measures of SEP: a composite measure of relative SEP, describing the proportion of households classified as low SEP in each CSA; and a composite measure of absolute SEP, describing the proportion of households living below the poverty line in each CSA. These two indices have been built by including a selected subgroup of the variables reported in **Table 4.1** into two composite indices using two different aggregation strategies: the first index is an assets-based index, the second one is based on the estimates of household per capita consumption expenditure.

The use of two different area-based measures of SEP was considered appropriate because: 1) I have hypothesised that the detection of a TB inequalities across CSA can be sensitive to the use of different ABSM; and 2) studies have shown that the use of assets-based indices and indices based on consumption expenditure data can give markedly different community SEP ranking when comparing communities [91].

- Five different area-based socioeconomic domains, including: the household structure (4 variables), crowding (1 variable), the education and occupation opportunities (respectively 5 and 2 variables), the access to community services (4 variables) and the housing characteristics (7 variables) (**Table 4.1**).

Within each domain, variables were collapsed into a composite index resulting in the creation of as many indices as domains. All the area-based socioeconomic domains derive from households level attributes; however, by summarising into composite indices they can reflect the key aspects of community that I have hypothesised to be relevant in explaining the socioeconomic gradient of TB prevalence rates across different CSA.

As argued by Messer and colleagues [218] the inclusion of more than one variable in each domain – instead of using a single indicator - can provide a richer picture of the specific domain considered. For example, it can be said that the proportion of households with illiterate adults and the proportion of households with no woman with primary education have different meanings, but both can inform us on what level of access to education these households have in the community (i.e. the *education opportunities*), either in terms of infrastructures and rights to education. Moreover, the use of composite indices for the SEP domain avoids the arbitrary selection of just one variable to represent the domain of interest. This does not apply to the 'overcrowding' domain, which was defined by the number of people sharing the same sleeping room and that could be therefore described only by one variable.

The advantage of using a broad array of indicators is that I can assess the overall effect of the community socioeconomic conditions and, at the same time, attempt to examine their multidimensional nature and explore the relative contribution of each specific domain considered [218-219]. Furthermore, the analysis of different domains can help to understand which interventions are more likely to have an impact in improving public health problems.

In the next two sections I will describe the methods adopted for the creation of these ABSM.

Table 4.1: The measurement of the area-based socioeconomic measures: domains and variables

Domains and variables	Operational definition
Household structure	
Household size	% Households with more than 5 members
Age dependency ratio (ADR)[(a)]	% Households with an average age dependency ratio above the mean ADR for Lusaka province
Female headed households	% Households with a female head
Female widows household heads	% Households with a female and widow head
Crowding	
N. people sharing the same sleeping room	% Households with more than 3.6 people per bedroom
Education opportunities	
Households with no literate adults	% Households with no adult (> 15 years old) able to write and read
Households with illiterate children	% Households with at least one child (≤ 15 years) having not completed the primary school
Households with illiterate women	% Households with no woman completing primary school
Households reporting no school attendance	% Households with no adult who has ever gone to school
Household highest school attainment	% Households with primary school as the reported highest education level achieved
Employment opportunities	
Households with no member involved in any formal job	% Households with no member employed in the formal sector (i.e. mine, farm, construction)
Households with no member involved in any economic activity	% Households with no member either employer, employee or self-employed
Community services	
Electricity	% Households with no electricity
Water supply	% Households with no water piped directly into the house
Water sanitation	% Households with no protected water (i.e. water from bore hole, unprotected well, river)
Waste disposal	% Households with no collection system of disposal (i.e. dumped on the road)

Table 4.1: Continued

Domains and variables	Operational definition
Housing quality and assets ownership	
Assets ownership[(b)]	% Households reporting no assets ownership
Roof quality	% Households with poor roof quality (i.e. grass, iron)
Wall quality	% Households with poor walls quality (i.e. mud bricks, iron, board, pole, grass, other)
Floor quality	% Households with poor floor quality (i.e. mud, other)
Lighting energy	% Households using poor sources of lighting (i.e. fire, candles, other)
Cooking energy	% Households using poor sources of energy for cooking (i.e. wood, coal, other)
Toilet facilities	% Households with no toilet facilities

(a) According to the Census definition the age dependency ratio refers to the 'joint account of variations in the proportion of children, aged persons, and persons of working age'. It is therefore the ratio of children aged 0-14 years and persons aged 65 years and older per 100 persons in the working age group 15-64 years old [212].

(b)The list of assets included radio, TV, fridge, phone, bike, motorbike, vehicle, plough, canoe, scotch kart.

4.5.6 The area-based measure of absolute socioeconomic position

The proportion of households living below the poverty line in each CSA was estimated through *poverty mapping* [220]. This technique has been developed to address the measurement of poverty and inequalities in small geographic areas in order to capture the heterogeneity of living conditions and better detect pockets of poverty in a region [91]. One of the most useful practical applications of this methodology is probably in making comparisons between local patterns of socioeconomic indicators and health indicators in order to investigate if and how they are related. From a policy perspective, to map poverty at very small level of aggregation can be useful because it can help to better target interventions.

As anticipated, both LSMS and census datasets are inadequate tools for the analysis of the poverty profile of small communities: LSMS surveys include income and expenditure data, but they are usually representative for broad regions of a country, whereas censuses can provide reliable estimates at highly disaggregated levels, but do not contain the income and expenditure data.

Poverty mapping allows overcoming these limitations by combining information from censuses and LSMS. Briefly, the methodology involves imputing into the census data a measure of household per capita consumption expenditure starting from the regression analysis of data included in the LSMS. Poverty mapping is therefore a method allowing the prediction of the per capita consumption expenditure of each household living in a community by statistical inference. Because the poverty line is based on a measure of per capita consumption expenditure, poverty mapping allows the estimate of households living below the poverty line.

The poverty mapping required the following steps:

1. *To build a model of per capita consumption expenditure* - I have first estimated a model of consumption using LSMSIII data. The creation of the model that best predict household per capita consumption expenditure can be done only through linear regression of those explanatory variables that are common to both the LSMSIII and the census. The model is specified as follows:

$$\ln y_i = x'_i\beta + \varepsilon_i$$

where the dependent variable (y_i) is the logarithm of per capita consumption expenditure of the households, x_i represent the independent variables, ε_i is the random error term and β is the coefficient that we want to estimate.

All the variables included in **Table 4.1** were considered for inclusion in the linear regression. Prior to the analysis, categorical variables were converted into binary variables based on my judgment of the most appropriate dichomisation. The entries coded as 'other' were examined in relation to the other variables in order to decide the most appropriate classification.

In order to select the variables to include in the final model I have first run a set of submodels, one for each socioeconomic domain specified. Only variables showing a P value ≤ 0.01 have been finally combined into an overall model, from which I have screened out variables not reaching statistical significance ($P \leq 0.01$). This cut-off was arbitrarily chosen to reduce as much as the number of variables included in the model and therefore increase its stability.

Before applying the final model to the census, the actual and the predicted measure of household per capita consumption expenditure were studied both as continuous and categorical variables. For this latter purpose, variables were grouped into quintiles. The household SEP ranking agreement between the two measures were assessed by looking at the household misclassification among quintiles and tested with Kappa statistic, a test providing the expected probability that two observations are classified in the same group by chance. A kappa value of less than 0.5 is generally considered an indication of poor agreement [221].

2. *The prediction of per capita consumption expenditure* – Once specified the model, I applied the regression coefficients to the correspondent variables in the census dataset of the Lusaka province. The imputation exercise consists of multiplying the estimated coefficients by the census variables included in the regression model to obtain a predicted value of per capita consumption expenditure for each household in the census dataset.

For the purpose of the ecological analysis, the prediction of per capita consumption expenditure was explored both as continuous and categorical variable. In this latter case a cut-off of 33% was used resulting in the creation of three groups identified as 'Low', 'Medium' and 'High' per capita consumption expenditure.

The estimate of per capita consumption expenditure was also used to estimate the proportion of households living below the poverty line. For this purpose, I have first created the empirical cumulative distribution function of the imputed per capita consumption expenditure variable through the STATA command "`cumul`". The use of this command allows the creation of a new variable cumulating and normalising the estimated consumption expenditure values so that their maximum is 1. Once this cumulative variable is created, the proportion of people living below the poverty line can be easily computed by counting the number of households below the two thresholds of extreme and moderate poverty set for Zambia (i.e. 64,530 and 92,185 Kwacha, respectively) [14].

4.5.7 The area-based measure of relative socioeconomic position

This measure was obtained by combining a subset of the variables included in **Table 4.1** using principal component analysis (PCA).

(a) The data reduction method

PCA is essentially a data reduction strategy used to identify, or extract, within a group of correlated variables, the underlying components that are common to these variables and which cannot be measured directly. PCA has several different applications, from education science, psychometrics to economic science. In this latter field, PCA it is often used to effectively combine socioeconomic information from various indicators into a single measure of SEP. PCA was firstly implemented by Filmer and Pritchett [85] and today is one of the most widely used strategies by big governmental agencies such as the World Bank and Macro International Inc [86, 222].

PCA identifies as many components, called principal components, as the number of variables included in the analysis. These components are assumed to capture a unique, unobserved attribute shared by households.

These principal components represent the linear combination of all the original variables, in which each variable receives a weight depending on the amount of information shared with the remaining variables and the unobserved attribute. Because the first component explains the largest proportion of the total variance, it is generally used to represent this unobserved attribute. The variables weights are derived from the correlation matrix of the data. The

weights of each variable from the first principal component are then used to generate a household score representing the households SEP in relation to all other households in the sample.

This scoring system allows the households to be ranked according to a measure of relative SEP.

In this thesis the application of PCA follows from guidelines available in Vyas et al [223] and Henry [224].

(b) The area-based SEP index construction

Only socioeconomic variables presenting a certain degree of variability across a group can contribute to differentiate the SEP of the group members: if, for example, all the households, or the vast majority of them, own a certain asset, it is impossible to establish any socioeconomic difference among these households based on that asset. For this reason, in PCA it is good practice to perform a prior descriptive analysis of the frequency distribution of the available variables in order to decide which are suitable for the inclusion in the PCA. Variables with low variability would exhibit low weight in the PCA and would be eventually discarded from the analysis in any case [225].

The descriptive analysis of the original 23 variables considered for PCA inclusion resulted in the elimination of 4 variables, including – female household head, head of the household female and widow, having no waste disposal system and having no toilet facility because their mean frequency distribution across sites was either lower than 20% or bigger than 80%. These two thresholds were chosen arbitrarily as a measure of too low or too high frequency of the variables in the sample.

The remaining 19 variables were included in the PCA. Prior to performing PCA, all the CSA were merged so to have one common area-based SEP index applied to all the CSA. The analysis was performed in STATA 9.0.

There is no statistical test or any formal way to establish which combination performs better in predicting the underlying index. In this case, I have used the method suggested by Messer and colleagues [194] based on the evaluation of the weights assigned to each binary indicator entered in the PCA model. In PCA the size and the sign of each indicator's weight are often used to screen out those variables that have least contribute to the explanatory power of that PCA model.

More specifically, I adopted the following steps:

1. I have run PCA using all the 19 binary variables to obtain an overall area-based SEP index.
2. I have re-run the same PCA stratified by CSA to obtain 11 different CSA-specific area based SEP indices.
3. I have then compared the weights assigned to each socioeconomic indicator included in the of the overall area-based SEP index with the CSA-specific ones.
4. Since the aim of this analysis was to develop an overall area-based SEP index allowing the comparison of relative SEP across different CSA, variables appearing to have low weight (less than 0.1) in any of the 11 CSA were screened out.

Following this screening approach, only 7 of the 19 variables originally considered were retained in the final PCA model. PCA was then re-run with only these 7 variables to obtain the final overall area-based SEP index.

The performance of this index was assessed by looking at the following outputs provided from STATA 9.0:

1. Eigen value –The size of the Eigen value of each principal component indicates the amount of the variance that was explained by this component. The larger the Eigen value for a component, the more the underlying SEP construct is explained by the linear combination of the indicators (i.e. the bigger is the amount of information they share and the bigger is their explanatory power). As stated before, only the first principal component is usually used because it represents the principal component with the largest Eigen value.
2. The percent share of common variance – This indicates the contribution of each principal component to the total variance.

(c) The creation of the area based SEP score

The weights of the first principal component were used to compute an area-based SEP score that was assigned to each household living in the 11 CSA. This area based score was explored both as linear and as categorical variable: in the first case, CSA were ranked according their median area-based SEP score; in the second case terciles of this score were created indicating 'Low', 'Medium' and 'High' area-based relative SEP. This analysis was

repeated stratifying by CSA so to compute the proportion of households following in the 'Low' category in each area. Terciles, rather than the conventional quintiles, were used in order to avoid the problem of small cell-sizes for CSA specific analysis and to simplify the interpretation of the SEP categories.

4.5.8 The area-based socioeconomic domains

As specified earlier, all the domains described in **Table 4.1**, except household structure, were used as ABSM. Apart from the overcrowding domain, for which only one variable was employed, these indices were all created with PCA. Because of the small number of variables per domain, all the items were retained in the domain-specific PCA. As described for the area based measure of relative SEP, only the first principal component was taken into consideration, resulting in the creation of 4 different area-based domain scores assigned to each household enumerated in the census of Lusaka. The scores were analysed both as continuous and categorical variables.

4.5.9 Data analysis

The socioeconomic profile of each CSA was first described through the simple frequency distribution of the original census variables of interest. Then, each CSA was characterised by looking at the median and range values of each index and the proportion of households falling into the lowest tercile of ABSM. The proportion of households living below the poverty line in each CSA was also provided.

Indices were compared with each other in terms of CSA ranking correlation and agreement. The ranking correlation among ABSM was assessed by looking at the size and significance level of the Spearman's correlation coefficient. A P value ≤ 0.05 was considered indicative of significant correlation in the CSA ranking observed for the indices compared. The indices agreement was also assessed by exploring how each CSA was ranked by the different ABSM. These two analyses are complementary: whereas the first analysis indicates how much the ABSM correlate with each other, the second one may suggest patter of living conditions at CSA level.

The effect of the community socioeconomic characteristics on the risk of TB could only be assessed at ecological level as it was not possible to link up the individuals involved in the prevalence survey with the people enumerated in the census dataset. The low number of TB cases and CSA further limited the use of a formal multilevel analytical approach.

The association between TB prevalence rates and each ABSM was assessed through a general linear regression analysis, with the crude TB prevalence rate treated as dependent variable and the ABSM as covariate. Each model was minimally adjusted for sex, age and type of setting (i.e. rural and urban) and weighted for the population size of the CSA. All socioeconomic indicators showing a statistically significant effect ($P<0.05$) on the TB rates that were included in a multivariable linear regression model. The analysis was conducted using STATA 9.0.

4.6 The case control study

4.6.1 Case and control definitions

The case and control definitions were the same as those used in the prevalence survey.

A *case* was defined as any person aged more than 15 years having at least one sputum sample culture positive for *M. tuberculosis* according to the microbiological diagnosis adopted in the prevalence survey. As described earlier, a change in the formal case definition deployed for the prevalence survey had implications for the case-control study which will be described later.

A *control* was defined as any person recruited in the prevalence survey (i.e. aged more than 15 years), with a negative sputum culture in the prevalence survey. To be considered negative, each culture was incubated for at least 8 weeks.

4.6.2 The frequency matching design

Controls were frequency matched with TB cases by area of residence (i.e. urban and rural) and by age-group (i.e. 15-24; 25-34; 35-44; 45-54; 55-64; >64 years).

In frequency matched case control studies, controls are not sampled completely at random from the 'non-case' population but they are selected from sub-set of the non-case populations so to reflect (or to match) the distribution of particular characteristics of the cases that may be related to the risk of disease and/or exposure. The frequency matching design is considered a less 'extreme' design compared to the individual matching, but it still a suitable way to make the comparison groups more similar. Likewise the individual matching, the frequency design can be considered a form adjustment for postulated confounding factors, where the adjustment is conducted *a priori*, before the data analysis,

and incorporated in the study design. The rational for the selection of these matching variables has been described earlier in the Section **4.2.3.**

Because of the complexities associated with the matching design - for example, the analysis of matched data, the possibility of overmatching, the impossibility of analysing the effect of the matching variable, and the chance of over-adjustment - methodologists advise to use this design only when its advantages are clear. In this study, the matching design was judged to be appropriate, especially considering the confounding factors that need to be controlled for. Matching is recommended for nominal variables such as neighbourhood or area of residence which represent a wide and indefinable range of environmental and genetic factors that are especially difficult, if not impossible, to quantify and thus control by other means. It is important to note that cases and control have been matched by urban and rural area of residence and not CSA, which made it possible to assess the role of the community socioeconomic characteristics on TB prevalence rates.

Sex was not treated as a matching variable and its potential confounding effect was just accounted for in the analysis by including this variable in each statistical model run. The reason of that is purely practical: to randomly extract matching controls using three matching criteria would have made the procedure with STATA a lot more complicate. Plus, gender is a relevant issue in TB epidemiology: matching by sex would have precluded any possibility to explore further the important role of gender in the association between household SEP and TB.

4.6.3 Sample size

Given an overall sample of 10,000 in the prevalence survey and an expected prevalence of 1% in the general population, approximately 100 cases of prevalent TB were expected.

Although in the case control study the exposure of interest was a measure of relative poverty, the sample size calculation was based on a measure of absolute poverty (i.e. % of people living below the poverty line) because this was the only indication I had of the extent of poverty exposure in this population.

Assuming an exposure to extreme poverty in the general population of the Lusaka province of 36% [14] a sample size of 150 cases and 150 controls (one case to one control) is required to detect an OR= 2.0 of difference between cases and controls in terms of socio-economic position with a study power of 80% and 5% significance.

In order to reduce the number of cases required, maintaining the same study power and level of significance, the case:control ratio was brought to 1:3, yielding to a final sample size requirement of approximately 100 cases and 300 controls.

Sample size calculation was computed with Epi 6.4 (CDC, Atlanta GA).

4.6.4 Study participants enrolment

(a) Eligibility criteria

For individuals to be included in the study they had to give written informed consent and have both questionnaires on the individual and household characteristics completed.

Cases whose culture was positive for any Mycobacterium other than M. *tuberculosis* were excluded from the study. For controls to be eligible they had to have a negative culture for *M. tuberculosis* AND not reporting cough for more than 2 weeks at the time of the recruitment. Despite the poor predictive value of coughing, the combination of these two criteria was considered enough to minimise the risk of recruiting false-negative TB people among the controls. Furthermore, in order to avoid the collection of data from individuals having the same households/living conditions characteristics, controls could not be eligible for this study if they happened to be an household member of any of the TB cases.

(b) Sampling strategy

Cases – By the start of my fieldwork (March 2006), 106 cases of *M. tuberculosis* had been already identified through the application of this microbiological case definition, whereas a further 297 positive cultures were still undergoing speciation to confirm the mycobacteria identification. Because the number of cases was thought unlikely to increase significantly and because 106 cases were already enough to satisfy the sample size requirement, it was agreed to consider these subjects as the final sample of cases to match to 318 controls by age group and area of residence as by the study design. Extra cases would have been extracted only in case of death, move or participation decline from anyone in this initial pool of cases. A case could be replaced only if another case meeting the same matching criteria was available (i.e. age group, area of residence).

The decision not to include any further subject to this initial pool of cases was mainly based on the need to maintain the age and area distribution of the cases stable so as to facilitate the

matching process. It was also dictated by budget and logistic constraints: the increase of the sample size would have increased the duration of the fieldwork, making it too expensive and also likely to overlap with the rainy season, during which many areas around Lusaka become inaccessible.

Controls – Consistently with the 1:3 case:control ratio, 318 controls were required. These subjects were a stratified random sample of disease-free participants from the prevalence survey. In line with the required sample size, for each case three controls were randomly selected. All the non-TB subjects enrolled in the prevalence survey were first stratified by area and age groups and then randomly selected in each category according to the age distribution and area distribution of the cases.

Controls who declined to participate, who were dead, or moved somewhere else at the time of the study recruitment were replaced by individuals meeting the same matching criteria (i.e. the same area and age group). These individuals were classified as "replacements". The list of cases and controls to replace was provided from the field workers at the end of every day of data collection. The extraction of the replacements used to occur on a weekly basis through the following procedure:

1. Controls to replace were first stratified by age group and area of residence to count of the replacements needed in each stratum;

2. The 7641 participants to the prevalence survey with no evidence of sputum culture positive were sorted by age group and area of residence;

3. The random selection of replacements was performed first by using the command "`set seed #`": this command is used to generate a random list of subjects from a given dataset, where # is a random number (usually a positive integer). This command has also the advantage to make the generation of this random list reproducible (provided the same number # is specified).

4. Finally, by using the command "`sample`" the subjects within each specified stratum were drawn. By adding the "`count`" option it was possible to sample the exact number of subjects needed in each age group and by area of residence.

5. The subjects so extracted were then listed by age group and area of residence to check whether they were actually reflecting the age group and area characteristics of the controls to replace.

The same procedure was repeated to identify substitutes to those replacements who were themselves unavailable.

The replacement procedure here described can be considered a type of oversampling strategy; however, selection bias is still possible. For this reason unavailable controls were compared to the final sample of controls recruited to assess whether they systematically differ by demographic characteristic or by living conditions. This comparison was possible because the prevalence survey dataset contained data on few SEP proxies (e.g. daily number of meals and type of durable assets owned).

4.6.5 Data collection

(a) The questionnaires

Two questionnaires were developed and administered to each person enrolled in the case-control study. One questionnaire was used to collect information on individual-level characteristics from the cases and the controls (respectively **Appendix C** and **D**) and one to collection of information on the socioeconomic characteristics of households (**Appendix E**). The design of the questionnaires was driven by the conceptual framework and the analysis plan. In terms of content, both these tools have been inspired by the questionnaires included in the latest LCMSIII: this decision was both pragmatic (i.e. to make sure that questions asked and the classifications of the answers were locally sound) and conceptual (i.e. to allow, whenever needed, comparison between the population included in this project and population enrolled in the LSMSIII as requested for the poverty mapping analysis described earlier).

In order to ensure confidentiality no data on name or family names were collected: all study participants and their households were identified by an ID barcode (the same used in the prevalence survey). Both questionnaires were translated from English into Pemba and Nianjia (the two most common dialects spoken in the study sites) and then back-translated. Inconsistencies in interpretation were discussed and reconciled.

The individual - level questionnaire - Respondents were selected cases or controls ('index' hereafter) and no proxy could be used in case of their absence. This questionnaire contained

demographic characteristics including age, sex, ethnic group and marital status. Consistently with the conceptual framework, the questionnaire also included sections on educational and occupational status, and also information on his/her food consumption, hazardous behaviours - such as exposure to smoking (both indoor and cigarettes smoking), drugs use and alcohol consumption, imprisonment, migration – and, data on contact with known cases of TB.

Table 4.2 presents the way potential mediators were measured, originally coded and then re-coded for analytical purposes.

Both cases and controls were also asked about BCG vaccination, HIV status and any other immunosuppressive conditions, including diabetes or cancer. BCG status was assessed through the upper arm deltoid examination. Information about HIV status was based on self-reported data. Interviewees' response was then checked against the lab results obtained during the prevalence survey: in case of discrepancy, the outcome from the prevalence survey (based on laboratory testing) was considered the most reliable.

The only difference between the case and control questionnaire was the section dedicated to TB status: cases were enquired about their current TB episode, whereas controls were asked about any possible episodes in the past. Data gathered included information on clinical and laboratory findings, treatment received and any exposure to stigma because of TB.

The household-level questionnaire - The ideal respondent was the head of the household or – if different from the head of the household – the main bread winner of the household (i.e. the person whose economic activity mostly contributes to the livelihood of the household). The questionnaire was administered to the head of the household, but normally the answers were collectively gathered by different members of the households that – because of their role in the house – were more familiar with the specific questions asked.

Definition of headship in the household varies across cultures and also across households. For this study household members were free to indicate who they considered the head of the household. If they named more than one person they were asked who they considered to be the one who most contributes to the resources of the household.

For the purpose of this thesis a *household* was identified as a group of individuals who normally sleep in the same house and that regularly share their meals and expenses. Other members of the family living away were not counted.

Consistent with the conceptualisation of household SEP, the questionnaire was designed to include the following information:

1. *Human resources of the household*, including education, occupation of the head of the household (or any other main bread winner) and household roster. The purpose of the household roster was to assess the structure of the family in terms of age, sex, education, occupation and health status, all reflecting household's resources in terms of human capital. All qualified household members, including the index, were listed into the household roster and identified by the barcode followed by a progressive number. To avoid confusion, the index was always enumerated with the household barcode + 0, and the head of the household with the household barcode + 1.

2. *Food availability and vulnerability*, including number of meals per day, number of meals containing proteins (i.e. meat, chicken, eggs, fish, etc) and number of months spent without having enough to eat. The purpose of this section was to assess the eating pattern of the household and specifically: to assess the frequency or regularity of food consumed by the household; to identify the consumption of specific items that can not only document the spending power of the household, but also the quality of food consumed; and, finally, to measure long-term food shortage in the household. Vulnerability was measured asking people the type of coping strategies implemented in the past 12 months, where coping strategy was defined as all those activities implemented to react to shocks and adverse events.

3. *Housing quality and asset ownership*, including dwelling-related indicators and number and types of assets owned. In this setting, a dwelling may consist of different rooms, located outside from the central compound. For this reason, a dwelling was defined as all enclosed living spaces used by the family on a routine basis. Building structures used primarily for storage or livestock were not considered part of the dwelling. To ensure objective measurement, the type of roof, floor and walls was recorded only under the direct observation of the fieldworkers. The list of durable assets was restricted as much as possible to a list of observable assets and could only include working conditions assets and assets that could be sold if the owner wished so. Assets ownership also included land and animals ownership.

4. *Access to infrastructure and facilities*: in this thesis this was measured just in term of physical access to facilities (i.e. walking distance from the households to the closest food market, public transport, clinic, primary and secondary school, drinkable water) and ownership of infrastructures (i.e. power, type of water sanitation and type of water supply). No data was collected about the quality and the real availability of these services.

Some of the information included in the household-level questionnaire was aggregated to derive variables easier to use in the analysis of the data. **Table 4.3** outlines the most relevant output variables and the methods of computation used.

(b) The interview procedure

The fieldworkers received periodically the list of cases and controls to interview, the GPS coordinates to locate the selected households and the list of the ID barcodes identifying each case/control and his/her household.

Use of proxy respondents was allowed only for the heads of the households. Cases/controls not at home at the time of the visit of the fieldworkers could not be replaced by any member of the household for the completion of the questionnaire. For each case/control not available, two further visits were requested before discarding them from the study.

Interviews were conducted in participants' homes. Once the selected household and the index were identified – fieldworkers were trained to go through the following 4-steps procedure for each interview:

1. *Screen the household for eligibility* - a selected household could be disqualified under the following conditions:

   - the index was not available for the interview (either because still absent after three attempt visits, or because moved somewhere else or because dead).
   - The household could not be traced back with the GPS coordinates.

2. *Introduce the study* – once the index was judged eligible, the fieldworkers had to read an information sheet about the study in front of the respondents and make sure they understood the scope and the methods of the study (particularly what the study was for, why their household was selected and why their participation was important).

Collect the informed written consent: the fieldworkers had to mandatorily collect the informed written consent from each index and the respective head of the household before proceeding with the interview. If a selected index was also the head of the household he/she had still to fill in both informed written consent forms.

Table 4.2: The hypothesised mediating factors: measurement and coding

Mediators	Question	Question N.	Response code	Analysis
Individual education/occupation status				
Literacy status	Can you read and write?	Q10_Write	No = 0 Yes = 1 Unknown = 9	Not recoded for the analysis
Occupational status	What was your main activity?	Q15_Activity	Unemployed = 0 Seasonal/Piece worker = 1 Student = 2 Employed = 3 Retired = 4 Self-employed = 5 Housewife = 6 Unknown = 9	Unemployed and all the other options* = 0 Employed = 1 Self-employed = 2 *Including seasonal/piece worker, student, housewife, retired
Food availability				
Daily number of meals	How many meals did you normally use to eat?	Q20_Meals	0 = 0 1 = 1 2 = 2 3 = 3 > 3 = 4 Unknown = 9	***Daily meals frequency*** ≥ 2 = 0 < 2 = 1
Weekly number of meals containing proteins	How many times in a week did you normally eat meat, fish, chicken or pork?	Q21_Luxfood	0 = 0 1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 > 5 = 6 Unknown = 9	***Weekly frequency of meals containing proteins*** 0 = 0 1 = 1 2 = 2 > 2 = 3

Table 4.2 – Continued

Mediator	Question	Question n.	Response code	Analysis
Biological-Behavioural risk factors				
BCG vaccination status	Have you ever been vaccinated for BCG?	Q25-BCG	No= 0 Yes = 1 Unknown	***BCG confirmed = 0*** If: - Q25_BCG = 1 & Q26_Scar = 2 OR - Q25_BCG = 0 & Q26_2= 2 ***BCG not confirmed = 1*** If: - Q25_BCG = 0 & Q26_Scar = 1 - Q25_BCG = 0 & Q26_Scar = 0 - Q25_BCG = 1 & Q26_Scar = 1 - Q25_BCG = 1 & Q26_Scar = 0
	Can I see if you have any BCG scar on your arm?	Q26_Scar	Examination not allowed = 0 No visible scar on both arms = 1 BCG confirmed by scar examination = 2	
Smoking	How would you classify your smoking habits?	Q39_smoke	Never smoked = 0 Daily smoke = 1 Occasional smoker = 2 Ex smoker = 3	Not smoker = 0 If Q39_smoke = 0 Smoker = 1 If Q39_smoke = 1, 2 or 3
Indoor pollution	Are you ever exposed to cooking fire indoor	Q42_Cooking	No = 0 Yes = 1 Unknown = 9	Not recoded for the analysis

Table 4.2 – Continued

Mediator	Question	Question n.	Response code	Analysis
Biological-Behavioural risk factors				
Alcohol consumption	How often do you drink any drink containing alcohol?	Q32_Alcohol	Never/rarely = 0 1-4 times/month = 1 Many times/week = 2 Every day = 3	Alcohol drinkers = 1 If Q32_Alcohol ≥ 0 & Q33_Drinks > 0 Not alcohol drinker = 0 If Q32_Alcohol = 0 & Q33_Drinks = 0 OR If Q32_Alcohol > 0 & Q33_Drinks =0
	When you drink, how many drinks containing alcohol might you have on a typical day?	Q33_Drinks	1-3 = 0 4-6 = 1 7-9 =2 ≥ 10 = 3 Unknown = 10	
Migration	In the 5 years before we first visited you, have you lived anywhere other than where you live today for a period of at least 6 months?	Q51_Migration	No = 0 Yes = 1 Unknown = 9	Not recoded for the analysis

Table 4.2 – Continued

Mediator	Question	Question n.	Response code	Analysis
Biological-Behavioural risk factors				
	Have you ever been tested for HIV?	Q29_HIVtest	No = 0 Yes = 1 Unknown = 9	***HIV negative = 0*** If Q29_1_HIVresult = 0 & HIV result from prevalence survey = Negative OR Q29_1_HIVresult = 0 & HIV result from prevalence survey = Missing OR Q29_HIVtest = 9 & HIV result from prevalence survey = Negative ***HIV positive = 1*** If: Q29_1_HIVresult = 1 & HIV result from prevalence survey = Positive OR Q29_1_HIVresult = 0 & HIV result from prevalence survey = Positive OR Q29_1_HIVresult = 1 & HIV result from prevalence survey = Missing OR Q29_HIVtest = 9 & HIV result from prevalence survey = Positive ***HIV status Unknown = 9*** If: Q29_1_HIVresult = 9 & HIV result from prevalence survey = Missing OR Q29_HIVtest = 0 & HIV result from prevalence survey = Missing
HIV status	Would you mind telling me the results?	Q29_1_HIVresult	Negative = 0 Positive = 1 Unknown = 9	

Table 4.2: Continued

Mediators	Question	Question N.	Response code	Analysis
Exposure to tuberculosis				
TB contact	Had you been in contact with anyone with TB before we first visited you?	Q44_TBcontact	No = 0 Yes = 1 Unknown = 9	
Attending public venues**	Do you ever go to video club?	Q47_Video	No = 0 Yes = 1 Unknown = 9	*Frequency of exposure* Low = 0 If: - Q47_Video = 0 OR - Q47_Video = 1 & Q47_1_Oftenvideo ≤ 1 & Q47_2_Lonvideo = 0
	If yes, how often?	Q47_1_Often video	< 1/Month = 0 1/Month = 1 > 1/Month = 2 Unknown = 9	Medium = 1 If: - Q47_Video = 1 & Q47_1_Oftenvideo ≤ 1 & Q47_2_longvideo ≥ 1 & Q47_2_longvideo ≤ 2
	For how long?	Q47_2_Longvideo	≤ 1 Hour = 0 1-2 Hours = 1 > 2 Hours = 2 Unknown 9	High = 2 If: - Q47_Video = 1 & Q47_1_Oftenvideo = 2 & Q47_1_Oftenvideo != 9 Unknown = 9 If: - Q47_Video = 9 OR - Q47_Video = 1 & Q47_1_Oftenvideo = 9 OR Q47_2_Longvideo = 9
Imprisonment	Have you ever been in prison?	Q46_Prison	No = 0 Yes = 1 Unknown = 9	Not recoded for the analysis

**Including bar, video shops, hair dressing shops and churches. In the table only the example of the video club is reported.

Table 4.3: Aggregated variables derived from the household-level questionnaire

Variable description	Name	Computation method
Mean age household members	M_age	Mean age aggregated by household
N. children/household	Num_child	Sum of number of individuals belonging to age 0-14
N. elderly people/household	Num_eld	Sum of number of individuals belonging to age >= 65
N. adults/household	Num_adult	Sum of number of individual belonging to age 15-64
Child dependency ratio	Child_dratio	Number of children/num of adultsx100
Aged dependency ratio	Eld_dratio	Number of elderly people/num of adultsx100
N. women/household	Num_wom	Sum of number of women per household
N. men/household	Num_men	Sum of number of men per household
Sex ratio	Sex_ratio	Ratio of females divided by menx100
N. economically active members	Eco_act	Number of individual above the age of 12 and not students or retired
N. economically inactive members	Eco_inact	Number of individuals above the age of 12, including students, full time housewives and those who are not available for work
Total land owned for agricultural activities	Land_size	Total number of hectares owned, after applying the following conversion measures: 1 hectare = 1 1 acre = 1/2.47 hectares 1 lima = 1/0.049 hectares
Having animals	Animals	Count of number of household declaring to have at least one animal
Households having chickens or ducks or guinea fowls or any other type of poultry	Poultry	Count of number of households having chicken or ducks or guinea fowls or any other type of poultry
Households having cattle or pigs or goats or sheep	Livestock	Count of households having cattle or pigs or goats or sheep
N. assets owned, regardless the type (including car, TV, fridge, radio, bicycle, motorcycle, domestic	Assets	Sum of the number of assets reported
N. of coping strategies implemented, regardless the type	Coping	Sum of the number of coping strategies reported
Crowding	Crowding	Ratio of number of people divided by number of sleeping rooms
N. windows/sleeping room	Wind_room	Ratio of number of windows divided by number of room
N. vents per sleeping room	Vents_room	Ratio of number of vents divided by number of room

The three key messages provided in the informed written consent forms were: a) the purpose of the study; b) the voluntary nature participation; and, c) the assurance of confidentiality. Staff were trained to make sure participants understood the study and felt free to withdraw if they wished so (for example, fieldworkers were trained to make sure that household heads could not coerce a index to join the study if he/she was not willing to do so).

Consent was obtained in writing unless they were unable, in which case a thumbprint was taken, in the presence of another adult.

The informed written consent was requested for:

- The permission to proceed with the interview and to collect both questionnaires (for a index to be enrolled in the study both the individual and the household questionnaire had to be completed);

- The collection of blood sample for the assessment of the infection status from the controls only (this procedure will be described in details in **Chapter 7**). It is important to highlight that controls could refuse the donation of blood, and still be able to take part in the study (if permission to collect both questionnaires had been given), whereas the opposite situation (giving consent to the blood collection, but not to the collection of data) was not a possible option.

3. *Proceed with the interview:* Once the 4 steps above had been completed, each page of each questionnaire was labelled with the barcode assigned to the case/control (and their household). The interview was conducted according to the instructions included in the fieldwork manual.

(c) Duration of data collection and recall period

The fieldwork started in March 2006 (when the data collection for the prevalence survey was almost finalised) and was conducted over a 12 months period, during which I was based in Lusaka, Zambia.

Because all the study participants were also originally enrolled in the prevalence survey, it seemed convenient to use the time of the interview for the prevalence survey as a recall period. In order to minimise any chance of recall bias, the interviews were scheduled so to make sure that people who were first interviewed in the prevalence survey were also the first ones to be interviewed for this project.

(d) Quality assurance and data management

Interviews were conducted by four research assistants divided in two teams, each per study area. All the research assistants were local, experienced, staff, formerly involved in the prevalence survey and other research projects in the two study sites. This ensured not only an extremely good knowledge of the sites, but also a better acceptance of the study from the local population.

The actual data collection was preceded by a two–weeks training period during which the fieldworkers received an overview of the intended purposes of the project and guidance on how to conduct the interviews and their responsibility in the field.

Both the individual and the household-level questionnaire were piloted on approximately ten eligible individuals (and their households) originally enrolled in the prevalence survey. Since people involved in the piloting were eventually excluded from the actual study, piloting was deliberately conducted only on individuals (and their households) classified as non-TB prevalent cases according to the prevalence survey (i.e. in order not to have exclude any of the identified cases).

Piloting was conducted under my supervision and was aimed to improve the quality and the efficiency of the data collection. More specifically, objectives of the piloting were to: 1) assess the feasibility of the study and the acceptability of the questions asked; 2) check that the instructions given to the fieldworkers were understood; 3) understand how to best adapt the data collection to the local conditions without altering the underlying intent of the questionnaires; 4) identify logistic problems. Following the piloting, the questions about employment status and health risky behaviours needed to be rephrased: in the first case, people did not consider self-employment (like selling fruits or soft drinks) a source of income and were all consistently reporting to be unemployed; in the second case, questions on alcohol and/or drugs consumption and having been in prison appeared to be sensitive so the fieldworkers helped me to rephrase them in order to sound less offensive or judgmental.

After finishing each interview, fieldworkers had to check whether the questionnaires were filled in correctly and legibly and that there was no missing information.

At the end of each day of work questionnaires were submitted to me for quality checks and to discuss any problem encountered in the field. Whenever possible, errors or missing information detected were corrected by revisiting the household. Interviewers were forbidden to change or repair data without consulting me or re-interviewing the household.

Task of the fieldworkers was also to inform me about any discarded household and the reason of exclusion.

(e) Data entry and data cleaning

Data were double entered into two separate files of a purposely designed Microsoft Access database. Data entry was conducted by myself and another staff member using precise rules and systematic procedures in order to minimise data-entry errors. The database was also programmed to contain data checks and macros to further minimise the chance of errors during the data-entry. Questionnaires presenting ambiguous data were discussed with me before finalising the data entry and, when possible, corrected. Once data entry was completed, the two files were compared and inconsistencies were checked against the source and, when possible, amended. Data were checked for outliers, missing data and inconsistencies via logical queries.

4.6.6 Change of the case definition and the study implications

As described in detail above, the microbiological definition of a positive culture for *M. tuberculosis* was initially based on the response of the isolate to the Niacin accumulation test and/or the hybridization profile identified through spoligotyping.

As anticipated in the sampling strategy section, by the start of my fieldwork (March 2006), 106 cases of *M. tuberculosis* had been already identified through the application of this microbiological case definition, whereas a further 297 positive cultures were still undergoing speciation to confirm the mycobacteria identification. (**Figure 4.7A**). At the end of my field work (March 2007) data from 106 cases and 318 controls had been collected.

The preliminary analysis of the prevalence survey data revealed some unusual results (including the lack of association between HIV and TB). These observations, in addition to some inconsistencies in the speciation with spoligotyping, led the prevalence study team to re-test all 403 cultures initially identified as positive for any type of mycobacteria using more specific laboratory testing. No further testing was performed on negative cultures. This assessment started in May 2007, when the fieldwork of this research project was already concluded since two months.

The new laboratory diagnosis was based on a nucleic acid amplification technology, the Genotype Mycobacteria CM assay (HAIN, Life Science), known to be highly specific for *M.*

tuberculosis [226]. This new approach revealed that some participants initially classified as cases were misdiagnoses probably caused by laboratory cross-contamination or by oral colonization/infection with mycobacteria other than tuberculosis. Further investigations revealed that the cross contamination was not in the culture part, but rather the result of DNA contamination in the spoligotyping process (Ayles personal communication).

The re-testing process with HAIN led eventually to the identification of only 79 cases of *M. tuberculosis* among the 403 positive cultures. Of these, only 52 were among the 106 cases recruited to the case control study (**Figure 4.7B**), meaning that the remaining 54 cases were actually positive for non-tuberculous mycobacteria and therefore not eligible for the present study.

The almost 50% reduction in the number of cases available for the case control study had serious implications for this dissertation project that were discussed with my supervisor and my advisory committee. Together we agreed the following actions:

- To restrict the analysis only to the 52 cases newly confirmed by HAIN. This would ensure consistency with the definition adopted for the prevalence survey and avoid any dilution of on the strength of the association between household SEP and case/control status through the inclusion of non-confirmed TB cases.

- To retain all the 318 controls originally enrolled in the study in order to maximize the study power. The study would no longer be considered strictly a frequency-matched case controls study, but still age-band and area of residence cannot be ignored as these 318 controls were more likely to be similar to the cases than the general population.

- Analytically, two approaches can be used: 1) standard logistic regression always adjusting by the two matching variables; 2) conditional logistic regression, a variant of logistic regression in which the two matching variables are replaced by a combination of them. These two approaches are equivalent and should provide almost identical results, but conditional logistic regression allows the reduction of the number of parameters included in the multivariable model, which is particularly desirable with a small sample size like this one.

- In order to further reduce the number of the parameters in the multivariable model the number of categories for each variable investigated was reduced, including the matching variables (for example age group re-classified into 3 levels: 15-29; 30-44; $\geq$ 45 years, instead of six). This would cause a loss of information, but would ensure more stable estimates of the association of interest.

- For the ecological analysis we decided to use all 79 cases detected in the prevalence survey instead of just the 52 available for the case control study. This choice allowed to slightly increase the power of the ecological analysis and also to ensure consistency between the prevalence proportions reported in the ecological study and the prevalence survey.

The results I present here refer to the 52 prevalent TB confirmed cases and the 318 controls originally selected. I will refer to the 106 original cases only to justify the sampling strategy and the profile (in terms of age and area of residence) of the 318 controls recruited.

4.6.7 The measurement of household socioeconomic position

Household SEP was measured through an assets-based index developed through PCA. In addition to an overall household SEP index, four more SEP indices were developed: one for each dimension considered in my conceptualisation of SEP. Likewise for the assessment of SEP at community level, the creation of different household SEP domains enabled me to assess the relative importance of different attributes of material SEP [227]. The use of these four different asset-based indices using PCA was considered less arbitrary than choosing one or more variables representative of these domains and it was considered useful to assess which of the SEP domains hypothesised was more strongly associated with prevalent TB.

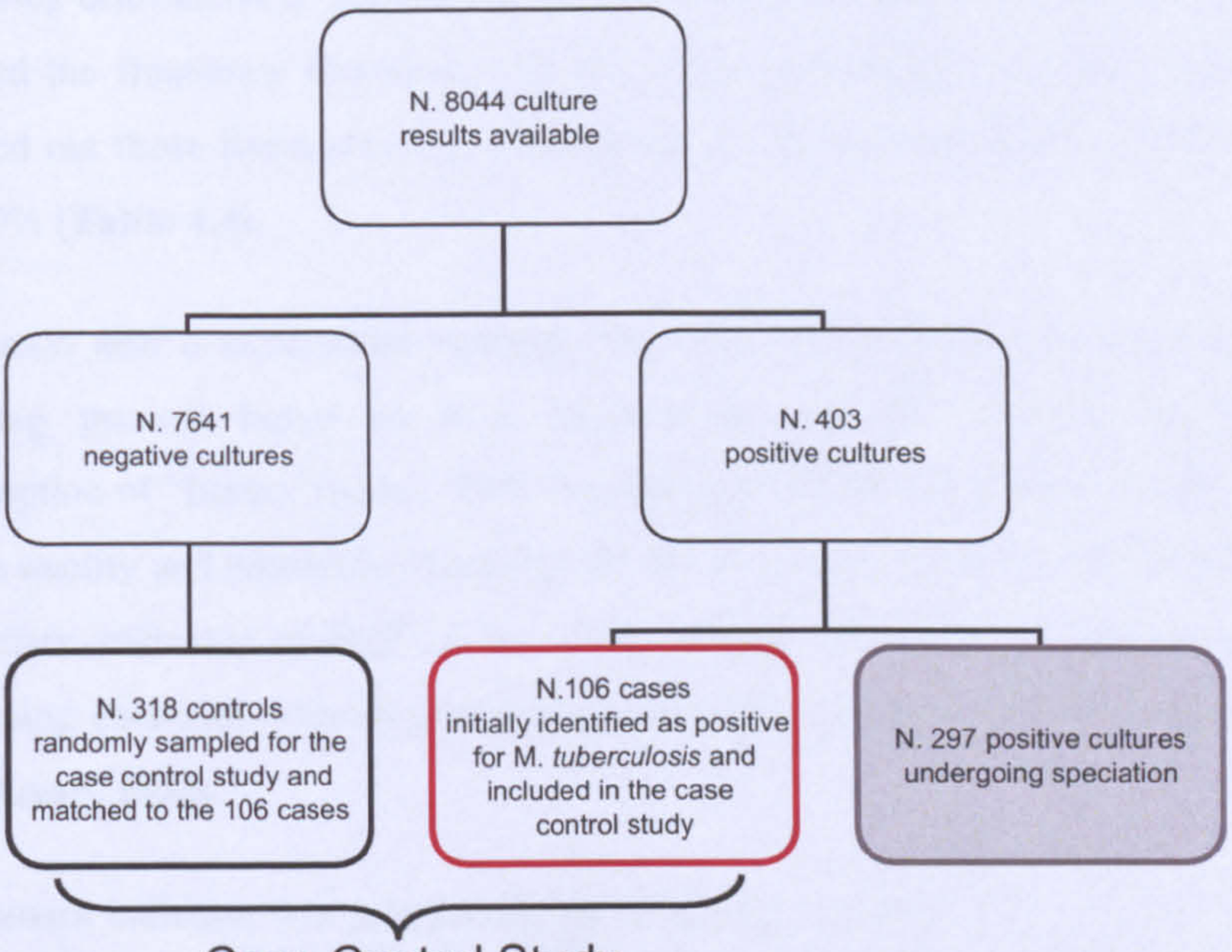
PCA was conducted only on data from the controls. This group of subjects approximate a sample of the general population it is possible that the weights generated from them will reflect those of the general population. Performing PCA on the cases alone, or cases and controls together, would not have achieved this. Once the weights were specified, they were also applied to the households of cases. The methodology was applied both for the construction of the overall SEP index and the SEP domain indices.

The development of the SEP indices required the preliminary conversions of all the indicators into dummy variables for each category of the categorical variables.

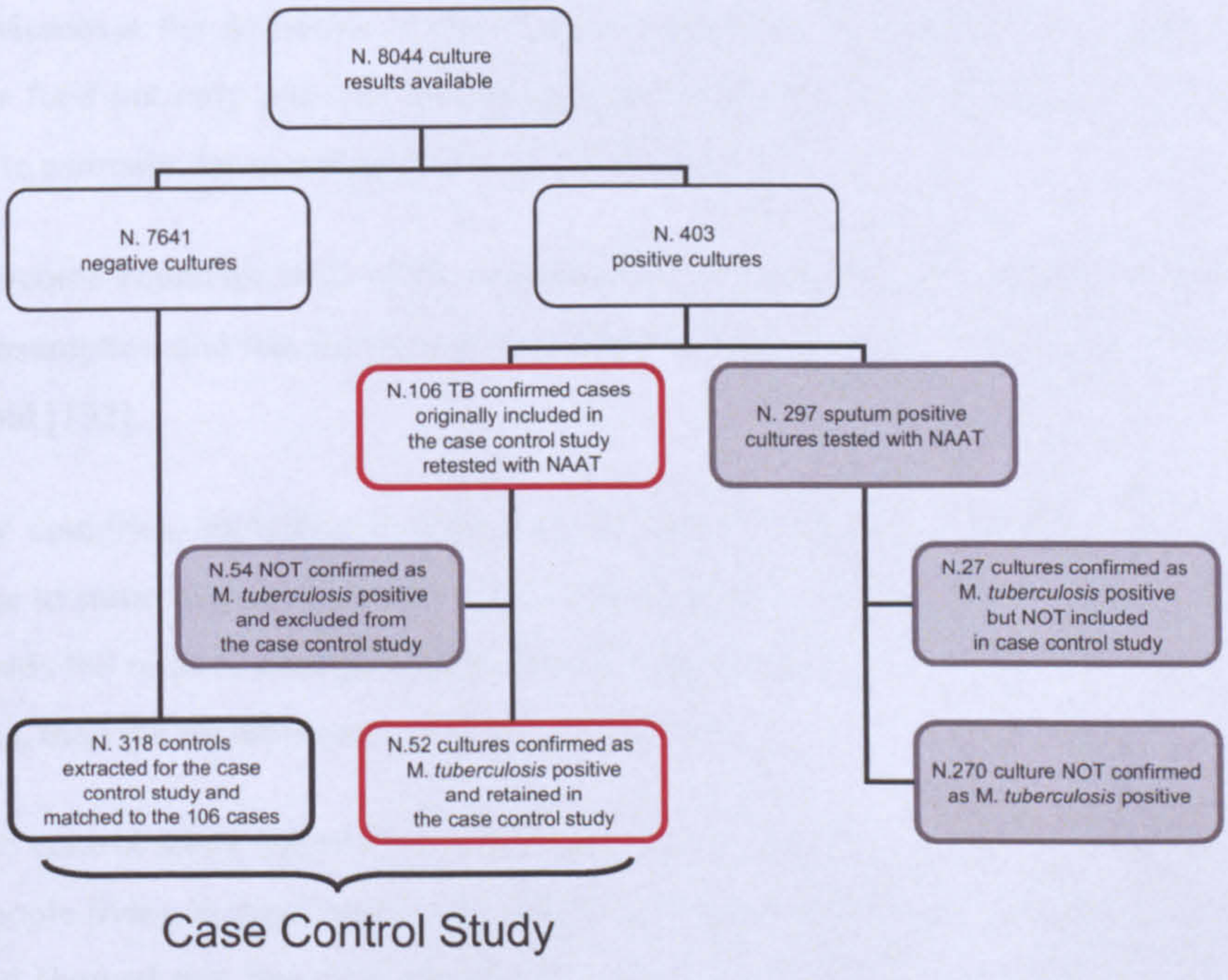
Although no best practice has been identified for selecting variables to proxy household SEP [224], experts suggest to implement a series of filters in order to ensure that the resulting index does not include too many variables and does not represent a distorted measure of SEP. The objective is to reduce the number of variables while losing as little information as possible. In this thesis, the socioeconomic indicators employed for the construction of the SEP indices underwent two selection processes: one before the run of PCA and the other after PCA (**Table 4.4**).

Figure 4.7: Number of cases eligible for the case control study: (A) during the fieldwork and (B) after the completion of the fieldwork

A: At the time of the start of the field work



B: After the field work completion



(a) Preliminary selection of the assets variables

The selection of the variables was based on the following screening procedure:

1. *Frequency distribution of assets:* As for the area-based measure of relative SEP, I first explored the frequency distribution of the assets considered inclusion in the PCA and screened out those items showing a frequency in the population below 20% and higher than 80% (**Table 4.4**).

2. *Correlation with a benchmark indicator:* the variables so selected underwent a further screening process based on their association with the variable on the weekly consumption of "luxury foods". This variable, defined as foods items locally considered of high quality and relatively expensive for the average household, was hypothesised the benchmark indicator of SEP in the study setting. In Zambia – like in many other developing countries – meals containing proteins (e.g. meat, eggs, fish, and cheese) can act as luxury foods.

This benchmark indicator was selected for the following reasons:

1. It is known that luxury foods cannot be consumed regularly by all the households, but are more frequently consumed in wealthier households than in poorer households. The use of number of daily meals can be meaningless because even poorest households can have 3 or more meals per day, but this doesn't necessarily translate into adequate food intake. Moreover the definition of meal can be subjective. By contrast, the consumption of costly food not only provides information on the quality of food intake, but also can be used to estimate the spending power of the household.

2. In low income countries most of the consumption expenditure of an household is due to food consumption and this measure is very likely to capture most of the expenditure in a household [192].

3. In many countries, including Zambia, the absolute poverty lines are constructed with reference to some minimum dietary requirements. The argument for this choice is that if households fail to have enough food to meet the minimum nutritional requirements of its members, then the members are considered to be poor.

4. The case control study questionnaire included a final free text question on what poverty is for people living in these two communities. Although not included in the analysis, data collected showed that the most frequently definition of poverty provided was "lack of food". These informal data were also supported by information included in a relatively

recent qualitative study [11] showing that in Zambia, especially in the rural areas, hunger – together with malnutrition, food insecurity and few meals each day – were universally used as characteristics of the poorer classes of people during wealth rankings.

5. Finally, the World Bank analysis [12] showed a clear gradient of chronic malnutrition across the income quintiles. Although these data were derived from children, to some extent it is reasonable to extrapolate this evidence to the adult population and assume that lack of food can work well as a proxy of household SEP.

The limitation in the use of food consumption data include the fact that food consumption patterns can be drastically altered during special events. Also, sometimes what is considered a luxury food changes by seasons and can vary between rural and urban communities.

Variables were filtered depending on the strength and direction of association with benchmark indicator. The significance of the association between the socioeconomic proxies and the selected benchmark indicators was tested with the chi square test: socioeconomic variables showing a P values ≤ 0.05 were considered significantly associated with the benchmark indicator and therefore suitable for inclusion in the PCA (**Table 4.4**). As suggested by Henry, variables showing a counterintuitive direction in the association with the benchmark indicator were excluded [224].

The above filtering process left 28 variables. Authors have suggested to limit the number of variables used in PCA to no more than 20-25 so to reduce the complexity of the resulting calculated component [224]. The inclusion of large number of variables increases the observed agreement of the asset index with consumption expenditure; however, the gain is not substantial above 20-25 indicators [228-229]. The number of variable eligible for PCA inclusion was further reduced by:

1. Excluding those variables with a high proportion of missing values. Missing values can cause the exclusion of a substantial number of households from the sample affecting the statistical power of the study and more importantly the representativeness of the results.

2. Excluding those variables closely related variables and likely to provide redundant information.

Variables were also screened in order to avoid the inclusion of an insufficient and unbalanced range of the indicators that can result in "clumping" or "truncation". Clumping (or clustering) occurs when an insufficient number of asset indicators is employed and households end up being grouped together in a small number of distinct clusters.

In presence of clumping, the amount of useful information about inequality that can be inferred from the index is inevitably limited. Truncation (or skewing) implies a more even distribution of SEP, but spread over a narrow range, making differentiating between socio-economic groups difficult (e.g. not being able to distinguish between the poor and the very poor) [223] [221].

(b) The index construction

Only the filtered variables (N. 21) were included in the PCA. The analysis was performed in STATA 9.0.

In order to identify the best variable combination to proxy the underlying concept of SEP I have repeated PCA several times. As for the community measure of relative SEP, the selection of the most appropriate variable combination was guided by size and direction of the assets weights, the Eigen value and the percent share of common variance.

In addition to the above criteria, the appropriateness of the choice of variables was also determined by examining the graphical distribution of the socioeconomic scores produced to assess any evidence of clumping or truncation. The final selection was based on the empirical observations I made and discussed with my field team during the fieldwork and my understanding of which variable was the most able to discriminate the SEP of the households. For example, in terms of housing quality, households with floor made of dirt were visibly quite poor; also critical was the type of roof material, with asbestos being apparently more common among wealthy households in the urban community and grass typical of the poorest households in the rural area.

These last procedures led to a combination of 11 variables (**Table 4.4**).

(c) The creation of the household SEP score

As already stated, only the first principal component extracted was considered sufficient to capture the underlying SEP of each household enrolled in this project. The weights corresponding to this first principal component were then used to estimate an 'imputed' SEP score for every household, including the cases. Because of the signs and size of the weights used, the bigger was this imputed score the greater was the household SEP.

Table 4.4: Variables selection process

N.	All variables considered#	Before PCA			After PCA
		I filter*	II Filter†	III Filter‡	Variables in the SEP index
		%	Chi square P value		
1	Floor in wood	0.3			
2	Floor in tile	0.0			
3	Floor in parquet	0.0			
4	Walls in mud bricks	0.0			
5	Having a motorcycle	0.2			
6	Having VIP latrine	0.3			
7	Head of the household unemployed	0.3			
8	Walls in pole	0.5			
9	Walls in mud	0.5			
10	Having domestic help	1.3			
11	Toilet facility – Shared flush toilet	1.9			
12	Drinking water source – River	2.1			
13	Drinking water source – Shallow well	3.5			
14	Having a car	4.0			
15	Drinking water source - Inside well	4.8			
16	Toilet facility - Flush toilet	6.2			
17	Toilet facility – Bush	6.2			
18	Land rented	12.0			
19	Having a fridge	13.8			
20	Walls in burnt bricks	18.9			
21	Drinking water source - Bore hole	**21.9**	**0.05**	**–**	
22	Roof in grass	**22.0**	**<0.001**		
23	Drinking water source- Piped inside the house	**22.4**	**<0.001**		
24	Drinking water source – Traditional well	**23.8**	**0.001**		
25	Main energy source for lighting - electricity	**25.0**	**0.001**		
26	Main energy source for lighting- candles	**29.4**	0.3		
27	N. contributors other than head of the household	**30.3**	0.3		
28	House rented	**31.0**	**0.001**	**–**	
29	Female head	**31.0**	0.2		
30	Drinking water source – piped outside the house	**31.3**	**<0.001**		
31	Having bicycle	**32.0**	**0.01**		
32	Roof in asbestos	**34.6**	**<0.001**		
33	Walls in concrete mud	**35.4**	**<0.001**		
34	Having animals	**39.4**	**0.001**		
35	Having TV	**39.7**	**<0.001**		
36	Main energy source for lighting – Kerosene	**42.7**	**<0.001**		
37	≥ 2 meals containing proteins per week§	**43.2**	-		
38	Roof in iron	**43.5**	0.2		
39	Household size (N. members ≥4)	**43.8**	0.2		
40	Head of the household self employed	**43.8**	**<0.001**	**–**	
41	Floor in dirt	**45.7**	**<0.001**		

Table 4.4: Continued

N.	All variables considered[#]	Before PCA			After PCA
		I filter*	II Filter†	III Filter‡	Variables in the SEP index
		%	Chi square P value		
42	> 3 coping strategies	**45.9**	**<0.001**		
43	< 30 minutes walking from clinic	**46.2**	**<0.001**		
44	Head of the household employed	**47.0**	**0.001**		
45	Walls in Cement	**54.0**	**<0.001**		
46	Education level Grade ≥ 2	**55.7**	**<0.001**		
47	< 30 minutes walking from market	**60.8**	**<0.001**		
48	< 30 minutes walking from transport	**63.5**	**<0.001**	–	
49	≥ 3 meals per day	**65.7**	**<0.001**	–	
50	< 3 months without enough to eat	**67.6**	**<0.001**	–	
51	House owned	**69.2**	**0.001**	–	
52	Having a radio	**73.8**	**0.004**		
53	Land owned	**73.8**	**0.01**	–	
54	< 15 minutes walking from source of water	**80.0**	**<0.001**		
55	Self perceived as poor	82.4			
56	Having pit latrine	85.4			
57	Being literate	87.0			

[#]Variables are sorted by frequency distribution. Variables surviving each filter process are indicated by the gray boxes.

*I Filter: Variables with a frequency distribution in the controls population lower than 20% and higher than 80% were screened out.

[†]II Filter – Variables not significantly associated with the benchmark indicator (i.e. weekly consumption of meals containing proteins in the household) were screened out ($P<0.05$).

[‡]Filter III – Variables with missing values or because providing redundant information were screened out. Variables excluded are indicated with the symbol –.

[§]Benchmark indicator.

4.6.8 Data analysis

The analysis of data consisted of four parts: a) the descriptive analysis of the study population; b) the assessment of prevalent tuberculosis risk factors; and c) the analysis of association between household SEP and prevalent TB d) the assessment of the population attributable fraction for any exposure found to be associated with prevalent TB.

(a) Descriptive analysis

I first conducted a descriptive analysis of the socioeconomic profile of the study population. In order to have a better representation of the households living conditions, as for the construction of the SEP index, the analysis was restricted to the controls only. Households enrolled in the study have been characterised by looking both at the frequency distribution of the socioeconomic variables collected and their ranking according to the SEP index. The analysis was conducted overall and by area of residence.

The household SEP index has been analysed as both a continuous and categorical variable. In the first case, the Mann-Whitney test was used to assess whether the SEP index score distribution was significantly different between the urban and the rural area. The SEP score distribution was also explored graphically to assess the extent of clumping and truncation and by looking at the median and range values. For the creation of a categorical index of SEP, households were first sorted in ascending order according their SEP score and then divided into terciles (33% cut-off) and classified as household belonging respectively to 'Low', 'Medium' and 'High' SEP. As for the ABSM, terciles, rather than the conventional quintiles, were used in order to avoid the problem of small cell-sizes for subgroups analysis and to simplify the interpretation of the SEP categories.

(b) The risk factors for prevalent TB

I assessed the prevalence of known TB risk factors among cases and controls and across different SEP groups (among controls only): the association between TB and the investigated risk factors was conducted through univariable and multivariable analysis. Odds Ratio and 95% confidence intervals were computed and statistical significance was tested through chi square test. Mantel-Haenszel test for trend was used to see whether there was any significant gradient in the risk factors distribution across the socioeconomic gradient ('High',

'Medium', 'Low' SEP), after adjusting for age. Finally, I explored the distribution of HIV prevalence by TB, SEP and sex.

(c) The association between household SEP and prevalent TB

Minimally adjusted analysis – I first explored the socioeconomic profile of the cases compared to the controls looking at the frequency distribution of the main socioeconomic indicators collected. I then investigated the association between household SEP index and prevalent TB. The index was analysed in the categorical format of household SEP terciles ('High', 'Medium' and 'Low').

Both analyses were performed by using conditional logistic regression. As described earlier, conditional logistic regression is a variant of logistic regression in which cases and controls are only compared in the same matched strata resulting from the combination of the matching variables. In this study, this matched set was made of six levels (3 age groups x 2 areas of residence). Both analysis were minimally adjusted by sex. Because of the use of conditional regression the analysis was also always adjusted by age group and area of residence. The results are presented in the forms of number and percentages of cases and controls in each SEP group. Odds Ratio and 95% Confidence Intervals are the primary measures of association.

The association between household SEP index and prevalent TB was analysed also using the index as continuous variable: in this case a linear logistic regression model was run minimally adjusted by sex, age group and area of residence. The association with prevalent TB was expressed as % of TB risk increase per unit decrease of the household SEP score.

The multivariable analysis – This analysis was run using as independent variable the household SEP index in categorical format. In this analysis the SEP score was categorised in two groups: "Low" and "High" using a 50% cut-off. The choice of two levels of SEP instead of three (i.e. household SEP terciles) was dictated by the need to reduce further the number of parameters in the multivariable model.

Statistical analysis was driven by the *a priori* defined conceptual framework. The analysis was conducted to properly take into account those variables that might confound the association of interest and to assess potential pathways of mediation.

In this study, the main confounding factors considered were age, area of residence and gender: the first two confounders were controlled *a priori* through the matching design. As discussed earlier in this chapter, gender was the only other variable treated as a confounder.

When looking at mediation, the aim of my analysis was to assess how much of the effect of the main exposure under study was captured/explained by the postulated mediating factors. Analytically, mediation can be suggested any time the adjustment for variables hypothesised to be on the causal pathway causes a reduction in the magnitude of the odds ratio between the outcome and the predictor of interest. Whatever is left after the observed reduction is that part of the exposure effect that is not mediated by the mediator [182]. As suggested by Victora et al when testing for mediation, the choice of factors to be included in the multivariable model is not purely based on statistical association, but rather based on the conceptual framework outlined in **Figure 4.1** and describing the hypothesised relationship between risk factors [182].

Regression models were first fitted accounting only for the hypothesised confounding factors; then, potential mediating factors were included in a hierarchical fashion (i.e. following the order outlined in the conceptual framework). All models were adjusted for gender even if there was no evidence of a significant impact of gender on the association between tuberculosis and SEP or any SEP domain. Odds ratios were the main measure of association. P-values were calculated from likelihood ratio tests comparing models with and without the factor of interest.

Five regression models were fitted (**Figure 4.8**):

Model 1 (minimally adjusted analysis): Household SEP adjusted by sex and the matching variables (i.e. age group and area of residence)

Model 2: as model 1 plus education/occupation-related variables

Model 3: as model 2 plus food intake-related variables

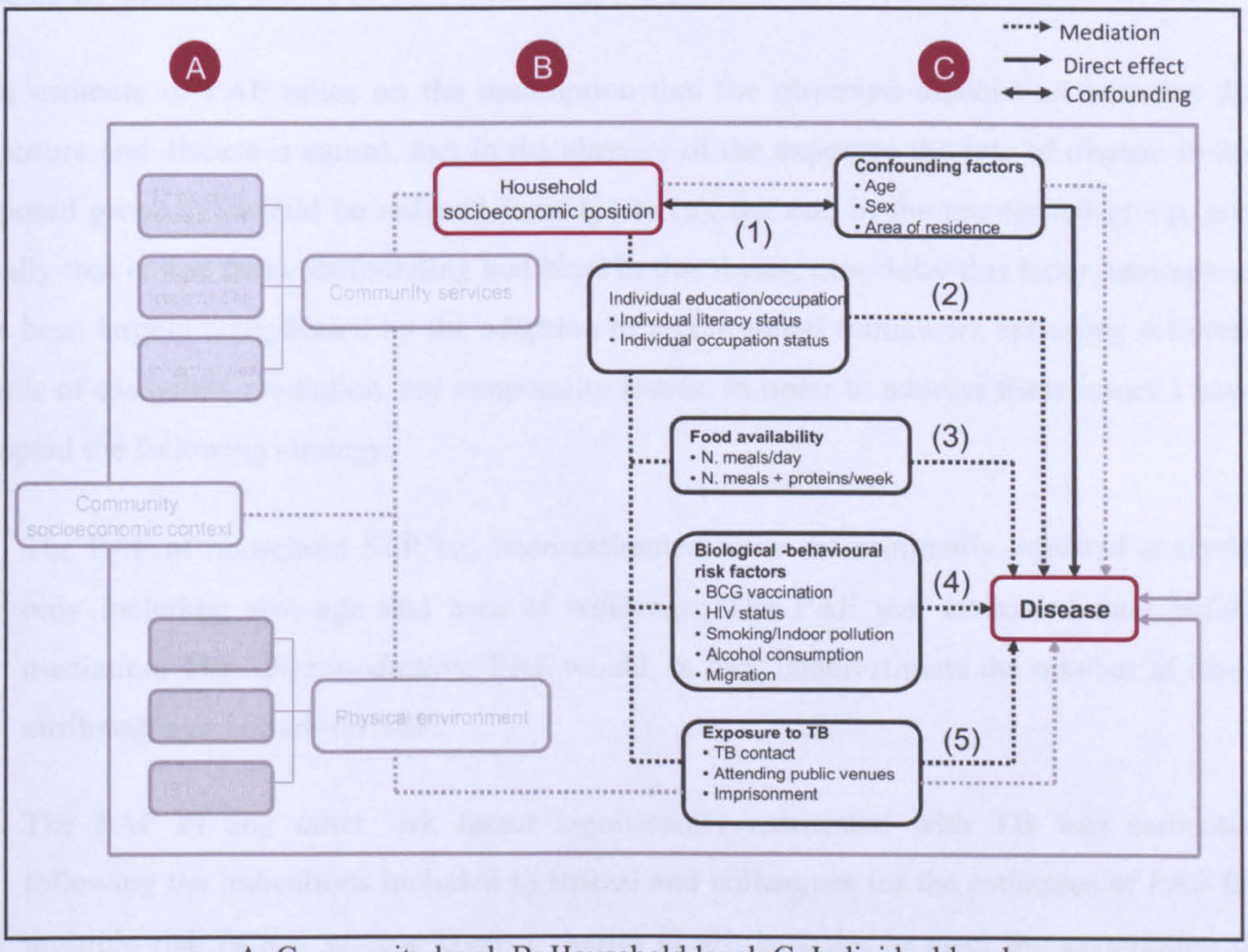
Model 4: as model 2 plus biological-behavioural risk factors for TB

Model 5: as model 2 plus TB exposure-related variables

The potential mediation effect on the main association of interest was assessed upon inclusion of groups of variables (i.e. food intake-related variables, behavioural-related variables) and not single one because the aim of the analysis was to verify possible mechanisms, rather than the effect the mediation effect of a specific exposure. The only

exception to this approach was done for HIV infection, whose potential mediating effect was studied in combination with the other biological and behavioural-related variables and on its own.

Figure 4.8: Conceptually driven strategy for the multivariable analysis of risk factors for prevalent tuberculosis disease



A: Community level; B: Household level; C: Individual level

The association between household SEP and prevalent TB was also explored by looking at the effect of the separate SEP domain indices. This analysis was conducted to understand which aspect of SEP was the one most strongly capturing the effect of SEP on the risk of prevalent TB. After a minimally adjusted analysis conducted separately on each of the 4 SEP domain indices, they were all included into a multivariable model including also sex, age and area of residence.

(d) The population attributable factor (PAF)

The Attributable Fraction has been defined as the proportion of cases of a disease occurring in a group of exposed people that would be avoided if the exposure were removed.

Following from the attributable fraction definition, the Population Attributable Fraction (PAF) is the proportion of cases of disease occurring in *the total population* that would be avoided if the exposure were removed [230]. PAF has been also referred to as the etiologic fraction [231].

In this thesis PAF was estimated for SEP and for any other risk factor significantly associated with TB.

The estimate of PAF relies on the assumption that the observed association between the exposure and disease is causal, that in the absence of the exposure the rate of disease in the exposed group (y) would be reduced from (y) to (z), the rate in the unexposed group, and finally that is free from confounding and bias. In this thesis, especially this latter assumption has been largely complicated by the adoption of a conceptual framework assuming different levels of causation, mediation and temporality issues. In order to address these issues I have adopted the following strategy:

1. The PAF of household SEP has been estimated from the minimally adjusted analysis, only including sex, age and area of residence. The PAF was estimated only *before* mediation. The *'after mediation'* PAF would, in fact, underestimate the number of cases attributable to household SEP.

2. The PAF of any other risk factor significantly associated with TB was estimated following the indications included in Bruzzi and colleagues for the estimates of PAF for multiple risk factors in case control studies [230]. In order to meet the assumption of non-confounding, I had to consider both the confounding effect of household SEP played on the risk factors identified at individual level and the reciprocal confounding effect played by these exposures on each other (due to the fact that they occupy the same level in the causation chain outlined in the conceptual framework). In order to that, I have firstly built a multivariable model including sex, age, area of residence SEP and, at the same time, any other significant risk factors for TB.

Both for SEP and for the other risk factors for TB, PAF was estimated using the command `'aflogit'` in STATA 9.0. This command gives the value of the adjusted PAF and confidence intervals for each risk factor considered in the multivariable model and allows the estimate of PAF for each exposure when all the others are present. The sum of the PAFs so estimated returns the number of cases of TB that could be attributed to this subset of risk factors considered, adjusting for the confounding effect of each other.

4.7 The infection study

4.7.1 The nested design

The infection study was *nested* into the case control study, as the study participants were the controls coming from the case control study for whom a blood sample for the assessment of TB infection was available. The restriction of the analysis only to the 318 disease-free controls enrolled in the case-control study had two major methodological implications:

1. In this investigation no formal *a priori* sample size calculation has been conducted as I could only rely on the controls from whom a test for TB infection was available. In these circumstances to estimate the sample size is not useful and to perform *post hoc* power calculation is generally not recommended [232].

2. Because the study participants came from an existing dataset, the data employed in this investigation are the same of the case control study one. This allowed the use of the same SEP index adopted in the case control study and therefore the comparison of the effect of household SEP on the risk of TB infection and TB disease.

4.7.2 The assessment of *Mycobacterium tuberculosis* infection

(a) Sample collection

Blood samples were collected from all the cases and the controls recruited to the case control study who gave informed written consent to be tested for TB infection at the time of the interview. The collection of samples from the cases was considered a sort of positive control of the infection status of the study participants. Eligible people were given all the information on the purpose of the test, the meaning of the results, and the amount of blood required. The field workers also discussed with them their right to decline the blood collection and the possibility that participants could be identified as having TB and, therefore, being stigmatised.

For safety issues, blood samples were collected in two clinics, one in each study area. Consenting individuals were asked to attend the clinic and to indicate a preferred date to have the blood collected. This approach was aimed to ensure a better monitoring of the number of people expected at the clinic every day and to minimise interference of the research with the daily routine of the study participants.

Study participants were given a *card* to present to the clinic reporting their name, age and sex, and the date agreed for the blood sample collection. In this way nurses could track the

number of patients coming to the clinic and *tested* against a list of *expected* patients provided by the fieldworkers at the end of each day of data collection. In case of loss of the card, nurses could still identify people coming to the clinic by their names.

No money or transport vouchers were used to encourage participation, but people living far from the clinics were offered free transport to the health care facilities using the vehicles used for fieldwork. This was the only 'incentive' provided in this study. Patients not coming to the clinic after three reminders were considered "drop-outs".

Blood samples collection was conducted by two professional nurses. Nurses were trained concerning the purpose of the study, their responsibilities and what to answer in case a participant needed further clarifications. For the training purposes, standard operating procedures were also developed, containing technical details on the test adopted and a detailed list of steps to follow to ensure the proper collection of the sample. All the phlebotomy equipment was provided by the project.

Blood samples were labelled with the ID barcode corresponding to the patient, handled as recommended by the manufacturer and stored at room temperature until transport to the laboratory, where they were incubated at 37°C. Following the manufacturer's instructions, transport to the laboratory was within 16 hours from sample collection. For this reason nurses and fieldworkers were asked to always note the time of the data collection and time of arrival to the lab for incubation.

(b) Blood testing

TB infection is traditionally assessed by Tuberculin Skin Test (TST), a tool in which tuberculosis extracts are injected and skin induration two days later is considered a sign of tuberculosis infection.

For the purpose of this study, infection status was assessed using QuantiFERON® - TB Gold (In Tube) [QFT], (Cellestis Ltd), an in vitro laboratory test using a whole blood specimen for the diagnosis *M. tuberculosis* complex infection (http//www.cellestis.com).

I decided to adopt this method instead of the traditional TST because of the following reasons:

- TST is prone to false positive results as a consequence of BCG vaccination and exposure to environmental bacteria [233] [234], both of which can be associated with SEP [235-237].

- QFT is unaffected by BCG vaccination and environmental mycobacteria, allowing more accurate assessment of TB infection [238-239]. Like for TST QFT sensitivity is known to be low and variable (75%-97%); however, QFT specificity is much higher than with TST with values above 95% in most of the studies [238].

- QFT avoids the second patient visit required with TST (for the reading of the induration response), saving time and money and resolving compliance issues.

- As a laboratory based essay, QFT is not subject to biases and errors of TST placement and reading.

- It was considered a good opportunity to verify the practical feasibility and limitations of the use of this method in a poor-resource setting.

QFT is based on the measurement of Interferon (IFN-γ) released by sensitised T cells after stimulation with TB antigens, including the Early Secreted Antigenic Targets (ESAT)-6, the culture filtrate protein (CFP)-10 and TB7.7. The test consists of three tubes: a Nil control tube, a TB antigen tube and an optional Mitogen Tube (used as positive control). For each tube 1 ml of blood is required.

In the lab, blood samples were stored and tested according to the manufacturer's instructions ((http//www.cellestis.com). The cut-off for the results interpretation was that suggested from the Manufacturer. Patients were classified as:

- **QFT positives** (likely to be infected with *M .tuberculosis*), if showing a concentration of IFN-γ in the sample well after stimulation with ESAT-6 and/or CFP-10 greater than or equal to 0.35 IU/mL (after subtraction of the value of the nil well), regardless the result of the positive control.

- **QFT negatives** (unlikely to be infected with *M. tuberculosis*), if the response to the specific antigens (after subtraction of the value of the nil well) was less than 0.35 IU/mL and if the IFN-γ of the positive control (after subtraction of the value of the nil well) was greater than or equal to 0.5 IU/mL.

- **QFT indeterminate** if the response to the specific antigen was less than 0.35 IU/mL and if the value of the positive control was less than 0.5 IU/mL, after the subtraction of the nil value.

The proportion of infection (i.e. QFT positivity) was defined as the number of QFT positives on the overall number of subjects tested.

(c) Data and results management

Data were entered to a Microsoft Exel spreadsheet that could be linked to the case-control study database and to the laboratory database. Study participants were classified as positive, negative, and indeterminate, but QFT results were also recorded as continuous variable (i.e. the concentration of IFN- γ, IU/mL).

Participants were informed that results would have been reported back to them only in case of a positive result.

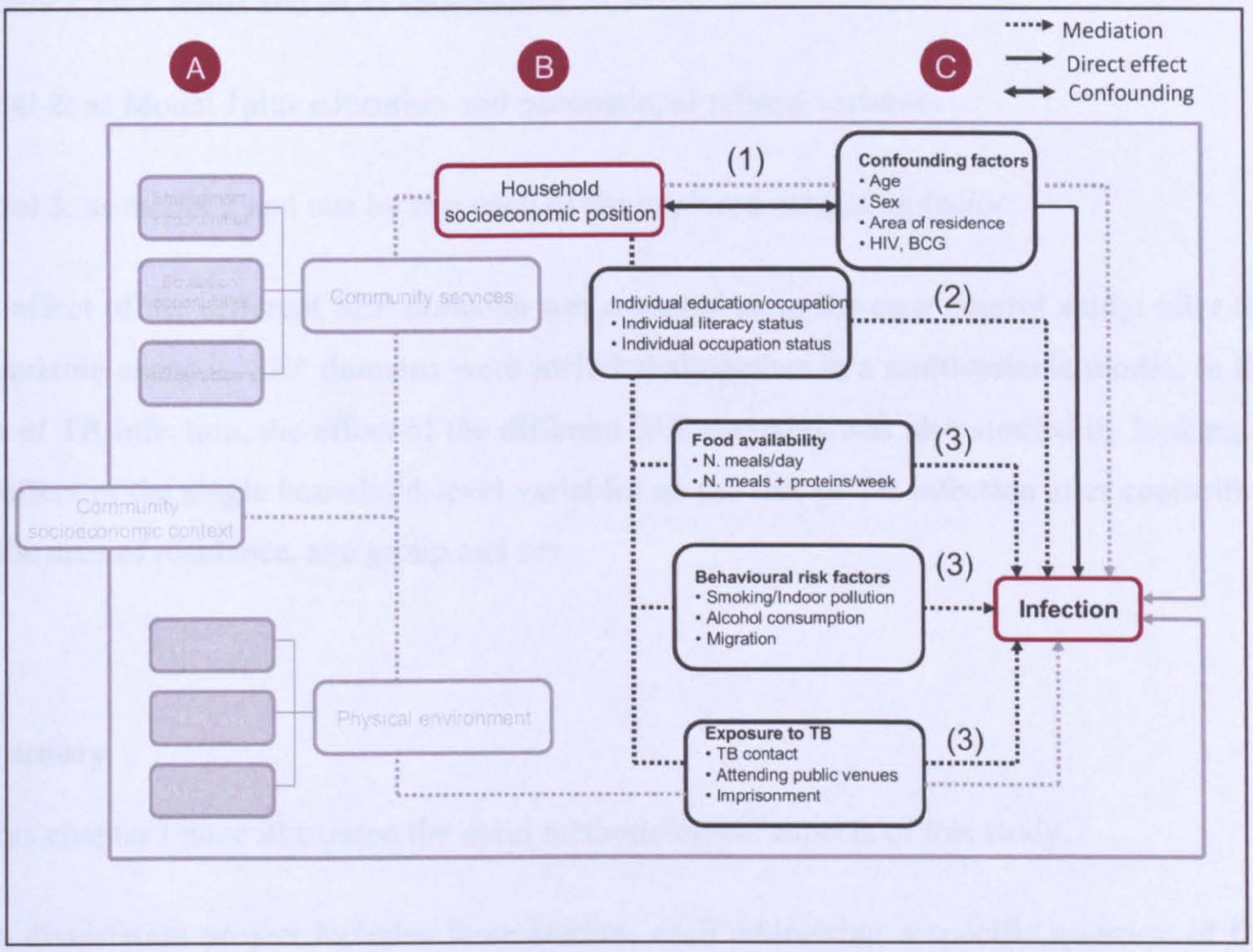
QFT positive patients were normally visited at home by fieldworkers and informed about their status with a letter. To ensure confidentiality, results were disclosed only to the directly involved patient. QFT positive patients were referred to the nearest health centre for further investigations depending on the response to questions contained in a diagnostic algorithm form and completed when the results were communicated. No further information was collected from these patients and it is unknown how many of them (if any) subsequently progressed from infection to TB disease.

4.7.3 Data analysis

I used the same analytical approach used in the case-control study. The association between SEP and TB infection was explored analysing the SEP index both as continuous and categorical variables. Odds Ratios (OR) with 95% CI were assessed through univariable and multivariable logistic regression analysis.

As for TB disease, I hypothesised the existence of a causal pathway in which the association between household SEP and TB infection was mediated by four different pathways. The conceptual framework driving the analysis is outlined in **Figure 4.9**.

Figure 4.9: Conceptually driven strategy for the multivariable analysis of the risk factors for tuberculosis infection



Differently from TB disease, for the TB infection study BCG vaccination and HIV status were considered potential confounding factors because both associated with SEP, the likelihood of TB infection and QFT positivity. These two postulated confounding factors were retained in the subsequent multivariable models only if showing evidence of confounding effect.

After checking for confounding, the mediation effect was assessed through the construction of several multivariable models containing SEP, the outcome of interest and, one by one, each risk factor was explored. The decision to include one potential mediator at the time (instead of groups of mediating variables as in the case control study) was due to the small sample size of the cross sectional study and the need to reduce the multivariable models instability through the inclusion of a smaller number of variables.

A reduction in the OR for SEP associated with TB infection upon inclusion of any of these variables was considered suggestive of mediation. Each model was always adjusted for age group, sex and area of residence. In summary, three set of models were built:

Model 1: Household SEP adjusted by confounding factors, including sex, age group, area of residence, HIV status and BCG vaccination.

Model 2: as Model 1plus education and occupational related variables

Model 3: as model 2 and one by one each of the explored mediating factor

The effect of the different SEP domains was assessed as in the case control study: after the univariable analysis, SEP domains were included altogether in a multivariable model. In the case of TB infection, the effect of the different SEP domains was also studied by looking at the effect of the single household-level variables on the risk of TB infection after controlling for the area of residence, age group and sex.

Summary

In this chapter I have illustrated the main methodological aspects of this study.

This dissertation project includes three studies, each addressing a specific question of this thesis. These three investigations are linked to each other by two fundamental elements: 1) they all originate from the latest population-based TB-HIV prevalence survey conducted in Zambia in 2006; 2) The *a priori* defined conceptual framework, which was used to generate the study hypotheses, to justify the selection of the SEP measures (both at community and household level) and to drive the analysis strategy.

The scope of this thesis is to explore in detail the association between SEP and TB in Zambia through a descriptive analysis of the spatial correlation between area-based measures of SEP and rates of prevalent TB (the ecological study) and two analytical studies looking at the association between household measures of SEP and the risk of TB disease and infection at individual level (the case control and the cross-sectional study respectively). These two latter studies, in particular, were designed not just to quantify the strength of the associations, but also to explore the causal pathway underlying these associations.

The next three chapters will illustrate the results gathered, starting with the findings of the ecological analysis.

5. Ecological analysis of the prevalence survey data

To be impoverished is to be an internal alien, to grow up in a culture that is radically different from the one that dominates the society. The poor can be described statistically; they can be analysed as a group. But they need a novelist as well as a sociologist if we want to see them. They need a Dickens to record the smell and texture and quality of their lives.

Harrington, M[6]. 1962

Introduction

In this chapter I will test the hypothesis that the segregation of material circumstances (both in terms of physical environment and community services) is associated with the prevalence of TB in the 11 CSA involved in the TB prevalence survey. For this purpose I will use an ecological analysis. The discussion of these findings will be preceded by a thorough description of the living conditions of these communities, which will help the reader to better contextualise the present study.

5.1 Chapters objectives

1. To summarise the main results of the population-based TB prevalence survey conducted in these two communities and, in particular, the distribution of prevalence rates across the CSA enrolled in the survey **(Section 5.2.1)**.

2. To describe the living conditions in the 11 CSA involved in the TB prevalence survey, emphasising those aspects that have been hypothesised to be determinants of TB prevalence at community level **(Section 5.2.2)**.

3. To describe the CSA socioeconomic position ranking according the different area-based socioeconomic measures (ABSM) **(Section 5.2.3)**.

4. To assess the ecological association between the CSA living conditions and the distribution of TB prevalence rates **(Section 5.2.4)**.

5.2 Results

5.2.1 Main results of the prevalence survey

This section provides an summary of the results of the prevalence survey. A more detailed discussion is available from Ayles et al [211].

Of the 8814 individuals enrolled in the prevalence survey, 8044 (91.3%) had an evaluable sputum culture. Of them, 7641 were classified as negative. Of the 403 positive cultures tested with HAIN technology, 324 were classified as positive for non-tuberculous mycobacteria and 79 positive for *M. tuberculosis*, yielding a crude prevalence rate of 980/100,000 and a cluster-adjusted prevalence of 870/100,000 (95% CI 570-1160/100,000). Of these 79 TB cases, 42 (53%) were classified as bacteriologically confirmed, 2 (2.5%) as bacteriologically unconfirmed, and 15 (19%) as sublinical cases of TB. The remaining 20 (25%) could not be classified as they did not complete their follow up.

The cluster-adjusted TB prevalence rate was almost as twice as high in the urban community (1200/100.000, 95%CI: 750-1640/100,000) compared to the rural one (650/100,000, 95%CI: 360-940/100,000). The crude and adjusted prevalence rates by CSA are shown in **Table 5.1**. The crude TB prevalence rates by CSA ranged between 1184.7/100.000 (in the urban CSA U-36) and 315.0/100,000 (in the rural CSA R-11). Among the rural communities the highest TB prevalence rates were observed for CSA R-13 and R-14, but still below the threshold of 1000/100,000 characterising all the urban CSA.

Table 5.1: Tuberculosis prevalence rates by CSA and study site

	TB prevalence rate*				TB prevalence rate*		
CSA	n/N[§]	Crude	Adjusted[†]	CSA	n/N[§]	Crude	Adjusted[†]
U-22	7/640	1093.8	1061.0	R-11	2/635	315.0	308.0
U-34	4/389	1028.3	1024.0	R-12	3/554	541.5	529.0
U-35	23/1512	1521.2	1484.0	R-13	8/831	962.7	901.0
U-36	17/1435	1184.7	1138.0	R-14	5/559	894.5	896.0
				R-15	6/771	778.2	789.0
				R-16	1/274	365.0	356.0
				R-17	3/428	700.9	686.0
Urban				Rural			
	51/3976	1282.7	1200.0		28/4052	691.0	650.0

*Rate per 100,000 population

[§] n= number of TB cases. N= number of subjects for whom a lab result was available.

[†]Prevalence rates were estimated through generalised linear models with random effects to control parameter estimates and standard errors for the effect of clustering at the standard enumeration area (i.e. the sampling unit) [211].

The prevalence of TB was highest in the age group 35-44 years (adjusted OR=3.31, 95% CI: 1.67-6.56, comparing 35-44 years with 15-24 years), but there was no significant difference between men and women.

HIV results were available for 7963 of the 8044 individuals with a valid TB culture: the adjusted HIV prevalence was 29.3% overall (95% CI: 26.8-31.9), with no significant difference observed between the rural and urban area (respectively equal to 28.9%, 95% CI: 25.2-32.7 and 28.2%, 95% CI: 24.3-32.1, P=0.7). As expected, the prevalence of TB was higher in HIV positive individuals (Adjusted OR=2.3, 95% CI: 1.4-3.7). The population attributable fraction for HIV was 36%.

Only 51 out of the 79 cases of TB (65%) reported cough and of these 34 (67%) had been coughing for more than three weeks. Having a cough for more than 21 days provided the strongest associations with prevalent TB (Adjusted OR=12.7, 95% CI: 7.0-22.9). All the other investigated symptoms (e.g. fever, shortness of breath, weight loss, night sweats, chest pain) were significantly associated with prevalent TB, but in the multivariable analysis only fever remained associated with the outcome (Adjusted OR=2.0, 95%CI: 1.2-3.4) [211].

None of the socioeconomic indicators enquired (based on the ownership a list of durable assets and the daily number of meals) was significantly associated with prevalent TB. However, for the assets ownership it was possible to detect a trend suggesting that higher SEP was protective for prevalent TB [211].

5.2.2 The profile of community living conditions

The data here presented were derived from the Lusaka Census dataset (year 2000).

Overall the number of households included in the analysis of the living conditions was equal to 5469, ranging between 147 in CSA R-16 and 975 in CSA R-13. The profile of living conditions will be provided first by describing the frequency distribution of all the socioeconomic indicators taken into account and then by looking at the ABSM developed.

(a) The socioeconomic indicators

Table 5.2 and **Table 5.3** summarise the CSA distribution of the socioeconomic indicators (respectively for the urban and the rural CSA) hypothesised to be relevant in the

epidemiology of TB inequalities. In order to make the reading simpler, results have been organised into domains.

Household structure – The household structure variables included in the living conditions profile analysis exhibited a fairly similar distribution between and within the investigated urban and rural CSA. Generally, in rural CSA the number of household members were higher. Rural CSA were also characterised by a higher proportion of households with an age dependency ratio higher than national average (e,g. 79%).

Low education opportunities – The percentage of households where no adult was able to read and write was extremely low (below 13%) for all the CSA considered, except for CSA R-14 and R-15 (26.6% and 30.7% respectively). Few households reported no adult ever attending school, implying relatively good access to education. However, between 26.0% and 51.0% of the households had no child (≤ 15 years) who had completed primary school (respectively for CSA R-16 and R-17). The proportion of households reporting no woman with a complete primary education was extremely variable and ranged between 18.7% for CSA U-22 and 62.5% for CSA R-15.

Low occupation opportunities – This domain showed large differences between urban and rural CSA, probably reflecting the different economies running in these two settings. Urban CSA appeared to be quite homogenous in terms of employment opportunities with low proportions of households with no adult involved in the formal sector or no adult having some sort of income (Range: 20-30% for both variables). By contrast, rural CSA showed much higher variability: the proportion of households with no adult employed in the formal sector ranged between 16.2% and 86.0% and between 23.2% and 87.1% when considering the proportion of households reporting no adult employed.

Low access to community services – Compared to urban CSA, rural CSA showed generally lower access to community services; however, this was not consistently observed for all the indicators considered. For example, in terms of access to electricity CSA U-35 and U-36 did not appear significantly different from most of the rural CSA. Despite being respectively urban and rural, CSA U-22 and R-16 looked very similar to each other and remarkably different compared to the remaining CSA. Both these CSA exhibited unusually high access to community services with almost all households having access to electricity, protected water supply and water piped directly into the house.

Housing quality and assets ownership – As for the domain assessing access to community services, it was not possible to identify a clear distinction between urban and rural CSA. Rural communities did not appear systematically worse off than the urban ones, especially

when looking at CSA R-16. Despite being rural, this CSA was characterised by housing characteristics similar to the urban CSA-22 and higher than what observed in CSA U-35 and CSA U-36. Similar observations can be made for CSA R-13, characterised by similar, if not better, housing quality compared to the urban CSA U-35 and CSA-36. In particular, CSA U-35 and CSA U-36 seemed to be characterised by poor wall quality and poor source of energy for lighting. The same gradient was observed among the rural CSA, starting with the better off CSA R-16, followed by CSA R-13 (showing an intermediate profile) and by the remaining rural CSA characterised by less assets ownership and lower housing quality.

Table 5.2: Distribution of the socioeconomic indicators of community living conditions in the urban CSA

	Urban			
CSA	**22**	**34**	**35**	**36**
N. households	230	305	945	839
Household structure index				
%Household size > 5	32 2 (26 1-38 2)	40.0 (34.5-45.5)	36 8 (33.7-39.9)	29.8 (26.7-32.9)
% Households with age dependency ratio> 79	41.3 (34 9-47.7)	48 2 (42.6-53.8)	55 6 (52.4-58.7)	53.1 (49 6-56.4)
% Households female headed	14.2 (9 6-18.7)	15.3 (11 2-19.4)	21 4 (18.8-24 0)	24.5 (21.5-27.4)
% Households headed by female widows	5.4 (2 2-8.5)	7.6 (4.5-10 8)	12 2 (9 9- 14 4)	11.4 (9.0-13.8)
Crowding index				
% Households overcrowded	37.8 (31.5-44.1)	33 1 28 8-38 4)	45.0 (41 8-48.1)	52.1 (48.7-55.5)
Education index				
% Households with no literate adults	1.7 (0.003-3 4)	7 2 (4 2-10 1)	9.9 (8.0-11.8)	13.5 (11.1-15.8)
% Households with illiterate children	31.3 (25.3-37.4)	29.2 (24 4-34 1)	43 6 (40.4-46 8)	45.3 (41.9-48.7)
% Households with no literate female adult	18.7 (13.6-23.8)	33.8 (28.4-39.1)	44.1(41.0-47.3)	45.2 (41.8-48.5)
% Households reporting no school attendance	2 2 (0.03-4.0)	5 2 (2 7-7.8)	8.4 (6.5-10.1)	10.5 (8.4-12.6)
% Household highest school attainment	2.6 (0.5-4.7)	12.8 (9 0-16 6)	23.0 (20.3-25 7)	26.0 (23.0-28.9)
Employment index				
% Households with no formal jobs	21.7 (16.4-27.1)	29.8 (26.5-35.0)	19.8 (17.2-22 3)	27.6 (24.6-30.7)
% Households with unemployed adults	21.3 (16.0-26.6)	29.5 (24 4-34.6)	20.3 (17.7-22.9)	27 8 (24.7-30.8)
Community services index				
% Households with no electricity	2.6 (0.5-4.7)	37.7 (32 2-43 2)	97.2 (96.2-98.3)	93.2 (91.5-94.9)
% Households with no private water supply	12 2 (7.9-16.4)	88.8 (85.3-92.4)	98 5 (97.7-99.3)	95.3 (93.8-96.7)
% Households with no protected water	0	55.1 (49.5-60 7)	35 2 (32.1-38.2)	4.6 (3.2-6.1)
% Households with no waste disposal	89.1(85.1-93.2)	99 7 (99-100.0)	79.4 (76.8-81 8)	74.4 (71.4-77.3)
Housing quality, assets ownership index				
% Households with no asset ownership	3.5 (1.9-5 9)	22.6 (17.9-27 4)	46 8 (43.6-49.9)	53.6 (50.2-57.0)
% Households with poor roof quality	47.8 (20.0-75.6)	20.3 (15.8-24.9)	32.6 (29 6-35.6)	40.4 (37.1-43.7)
% Households with poor wall quality	3.5 (1.9-5 9)	22.6 (17.9-27.3)	78 2 (75.6-80.3)	77.6(74.8-80.4)
% Households with poor floor quality	0.9 (0 3-2.0)	6.2 (35 0-89 6)	49 0 (45.8-52.2)	40.0 (36.7-43 4)
% Households with poor lighting energy	3.0 (0.08-5.3)	25.9 (20.9-30.8)	82 4 (80.0-84.9)	69.5 (66 4-72.6)
% Households with poor cooking energy	0	14.7 (10.7-18.7)	36.5 (33.4-39.6)	29.0 (25.9-32 0)
% Households with no toilet facilities	0	0.1 (0.02-2 6)	8.7 (6.9-10.5)	9.6 (7.6-11.6)

Table 5.3: Distribution of the socioeconomic indicators of community living conditions in the rural CSA

	Rural						
CSA	**11**	**12**	**13**	**14**	**15**	**16**	**17**
N. households	495	290	975	286	706	147	251
Household structure index							
%Household size > 5	44.6 (40 2-49 0)	48 3 (42 5-54.0)	39 1 (36.0-42.1)	41.9 (36 5-47 7)	39 5(35.9-43.1)	55 8 (47.6-63 9)	47.4 (41.2-43.6)
% Households with age dependency ratio > 79	58.2 (53.8-62.5)	59 3 (53 6-65.0)	50.4 (47 2-53.5)	56 6 (50.9-62.4)	55.9 (52.3-59.6)	54.4 (46 3-62.6)	57 8 (51.6-63.9)
% Households female headed	14 6 (11 4-17 7)	17 6 (13 1-22 0)	23 9 (21 2-26 6)	27 3 (22.0-32.6)	23 8 (20 6-26 9)	12 5 (7 0-18 0)	20 7 (15.6-25 8)
% Households headed by female widows	8 3 (5.7-10.9)	6 8 (3 6-9 9)	12 3 (10 1-14 5)	16 2 (11 5-20 9)	14 3 (11 5-17.0)	2 3 (0 03-4.9)	12 9 (8 5-17 4)
Crowding index							
% Households overcrowded	51 3 (46.9-55.7)	45.9 (40 1-51 6)	37.7 (34 7-40 8)	56.3 (50 5-62 1)	43 9 (40 2-47 6)	34.0 (26 2-41 8)	52 6 (46 3-58 8)
Education index							
% Households with no literate adults	9 1 (6 5-11 6)	19 3 (14 7-23 9)	9 0 (7 2-10 8)	26 6 (21 4-31 7)	30 7 (27 3-34 1)	0 7 (0 06-2 0)	19 5 (14 6-24 4)
% Households with illiterate children	50 9 (46 5-55.3)	49.3 (43 5-55.0)	38 9 (35 8-41 9)	38 8 (33 1-44 5)	49 1 (45 3-52 7))	26 6 (21 8-36 7))	51.0 (44 8-57 2)
% Households with no literate female adult	39 0 (34 7-43 3)	57 9 (52 2-63 6)	37 4 (34 3-40 4)	58 3 (52 6-64 1)	62 0 (58 4-65 6)	25 3 (18 1-32.5)	57 8 (51 6-63 9)
% Households reporting no school attendance	9 1 (6 5-11.6)	31 3 (25 7-36 4)	11.8 (9 8-13 8)	21 3 (16.6-26.1)	14 3 (11 7-16 9)	0 7 (0.07-2.0)	16 3 (11 7-20 9)
% Household highest school attainment	23 7 (19 9-27 4)	44 1 (38 4-49 9)	19 3 (16 8-21 8)	40.0 (34 3-45 7)	36.3 (32.7-39 9)	2 0 (0 03-4.4)	30 8 (25.0-36 6)
Employment index							
% Households with no formal jobs	16 2 (12.9-19 4)	45 9 (40 1-51 6)	66 4 (63 4-69.4)	86 0 (82.0-90.0)	36.2 (32 6-39.7)	23 9 (17 0-31 0)	33 5 (27.6-39.3)
% Households with unemployed adults	26 1 (22.2-29.9)	65 2 (59 6-70 7)	64 6 (61 6-67 6)	87 1(61 6-67.6)	43 8 (40 0-47 4)	23 2 (16 3-30 2)	54 2 (48 0-60.4)
Community services index							
% Households with no electricity	99.4 (98.7-100 0)	96.6 (94 4-98 7)	68.4 (65 5-71.3)	99 6 (99.0-100.0)	92.4 (90.5-94.4)	4.1 (0.8-7 3)	100 0
% Households with no private water supply	100 0	99.7 (98.9-100 0)	63 3 (60 2-66 3)	97 2 (95.3-99 1)	99.3 (98.7-99.9)	43 5 (35.4-51.6)	99.6 (98.8-100.0)
% Households with no protected water supply	89 1 (86 3-91 8)	98.3 (96 8-99 8)	45.7 (42 6-48.9)	80 4 (75 8-85 0)	82.3 (79 5-85.1)	6 1 (2 2-10.0)	74 5 (69 1-79 9)
% Households with no waste disposal	84 8 (81 7-88 0)	76.2 (71 3-81 1)	78.5 (75 9-81 0)	79 4 (74 6-84.0)	92 9 (91.0-94 8)	88 4 (83 2-93 7)	91.6 (88.2-95 1)
Housing quality and assets ownership index							
% Households with no asset ownership	36 8 (32.5-41 0)	41 0 (35 3-46 7)	31.4 (28 4-34 3)	42 7 (36 9-48 4)	41 1 (37 4-44 7)	5.4 (1 7-9 1)	57 0 (50 8-63 1)
% Households with poor roof quality	94.1 (92 1-96 2)	92 7 (98 7-95 7)	61 8 (58 8-64 9)	95 8 (93.5-98 1)	86 1 (83 6-88 7)	40 1 (32 1-48 1)	98 0 (96 3-99 7)
% Households with poor wall quality	87 3 (84.3-90.2)	78 6 (73 8-83 4)	47 9 (44 7-51 0)	92.7 (89 6-95 7)	78 6 (75 6-81 6)	2.7 (0 05-5 0)	90 0 (86.3-93 8)
% Households with poor floor quality	83 4 (80.6-87.1)	90 0 (86 5-93 5)	40.2 (37 1-43 3)	87 8 (83.9-91.6)	76 1 (72.9-79.2)	1.4 (0.5-3.2)	90 0 (86.3-93.8)
% Households with poor lighting energy	16 0 (12 7-19.2)	23.4 918.5-28.3)	37 4 (34 4-40 5)	26 9 (21.7-32.1)	17.8 (15 0-20.7)	2.0 (0.02-4.3)	23.1 (17 8-28.3)
% Households with poor cooking energy	93 5 (91 4-95.7)	95 1 (91 4-96 8)	28 9 (26 1-31 8)	90 6 (87 1-93.9)	89.4 (87 0-91 6)	6 8 (2 7-10.9)	98 8 (97 4-100.0)
% Households with no toilet facilities	34 3 (30 1-38.5)	11 0 (7 5-14 7)	10 7 (8.7-12 6)	16 1 (11 8-20.4)	25 8 (22 5-29 0)	4.1 (0 8-7 3)	7 6 (4.2-10 9)

Overall the data suggested the existence of a gradient characterised by at least three different levels of living conditions. Among the urban areas, CSA U-22 looked relatively wealthier compared to CSA U34 and even more compared to CSA U-35 and CSA U-36 (both very similar to each other and very poor).

(b) The area-based measure of relative SEP

The filtering process described in **Chapter 4** (page 128) resulted in the selection of the 7 variables as shown in **Table 5.4**. After the selection process, only variables accounting for education and occupation opportunities as well as assets ownership were included in the index. Overall, the index generated by PCA accounted for approximately 40% of the total variance of the variables included in the index, ranging between 34.7% and 41.6% when looking at the CSA-specific index (respectively for CSA R-11 and CSA U-36). The analysis of the weights of the overall and CSA-specific indices revealed two different patterns emerging:

1. The weights of the variables included in the area-based measure of relative SEP were similar to each other *within* each CSA and also *across* CSA. This could be interpreted as the fact that the selected variables were contributing in the same way in each CSA and also that they had the same importance in the definition of the area-based measure of relative SEP across all the 11 sites considered. This suggests that the indices were a balanced combination of variables and were also comparable with each other.
2. The CSA-specific weights were similar to the weights of the overall index suggesting that the overall index was adequately representing all the different CSA despite the evident different living conditions.

Figure 5.1 shows the boxplot of the area-based relative SEP by CSA. Because of the way census variables have been coded for this index, the higher is the SEP score, the lower is the relative SEP of the CSA.

In terms of relative SEP, CSA did not show significant heterogeneity: among them, CSA R-14 was the CSA with the lowest household SEP median score (4.0, Range: -1.6-4.6). Consistently with what was observed in the analysis of socioeconomic indicators, CSA R-16 and CSA U-22 were the CSA characterised by the highest standards of living according to the area-based measure of elative SEP (Median score of -1.7, Range: -1.6-2.9 for both). With the exception of CSA R-11, the wealthiest communities were all urban.

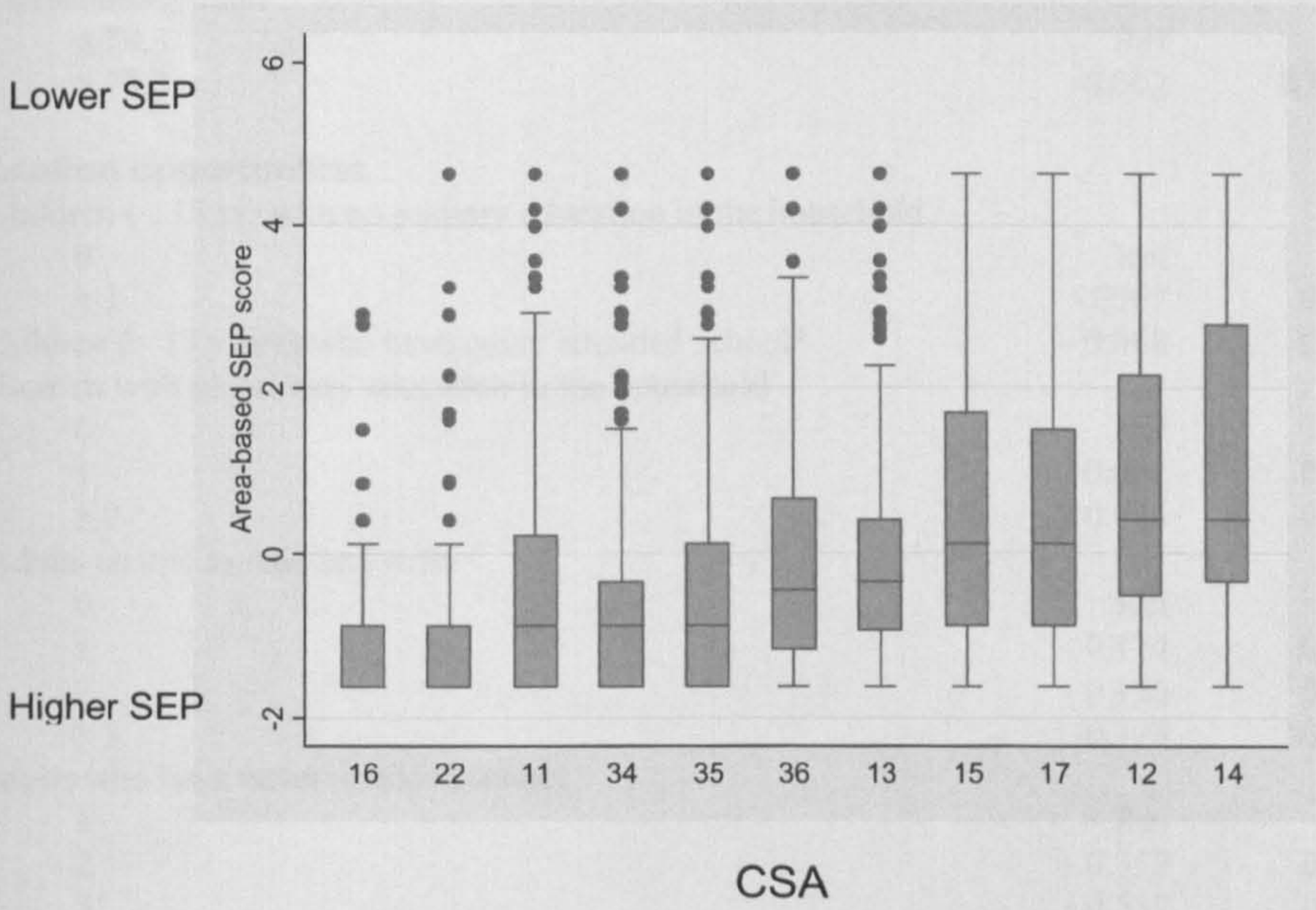
Table 5.4: Main characteristics of the area-based measure of relative SEP: overall and by CSA

Variables	Total	Rural CSA							Urban CSA			
		11	12	13	14	15	16	17	22	34	35	36
% HH* with no literate adults	0.45	0.45	0.43	0.42	0.46	0.50	0.29	0.46	0.37	0.41	0.42	0.47
% HH with no literate woman	0.37	0.40	0.38	0.37	0.38	0.40	0.29	0.42	0.26	0.37	0.30	0.33
% HH reporting no school attendance	0.45	0.49	0.49	0.46	0.45	0.45	0.29	0.50	0.39	0.42	0.43	0.44
% HH with low education attainment	0.47	0.48	0.52	0.49	0.48	0.50	0.43	0.53	0.45	0.44	0.44	0.43
% HH with no employed in formal sector	0.30	0.27	0.24	0.29	0.28	0.21	0.52	0.04	0.42	0.33	0.40	0.35
% HH with no member currently working	0.31	0.20	0.15	0.30	0.29	0.19	0.52	0.10	0.42	0.33	0.40	0.35
% HH with no asset ownership	0.22	0.21	0.26	0.23	0.20	0.9	0.12	0.24	0.30	0.29	0.17	0.21
N. observations	**5457**	**494**	**290**	**972**	**285**	**705**	**146**	**250**	**230**	**305**	**942**	**838**
Eigen value	**2.8**	**2.4**	**2.7**	**2.7**	**2.8**	**.6**	**2.5**	**2.5**	**2.7**	**2.8**	**2.9**	**2.9**
% Total variance explained	**39.8**	**34.7**	**38.2**	**39.4**	**39.9**	**37.8**	**36.2**	**35.4**	**39.2**	**39.9**	**41.1**	**41.6**

*HH= Households

CSA appeared less homogenous when the area-based index was analysed as categorical variable: the proportion of households grouped in the bottom tercile ranged between 8.3% (for CSA U-22) and 68.4% (for CSA R-14).

Figure 5.1: Boxplot of the area-based measure of relative SEP by CSA



(c) The area-based measure of absolute SEP

The reader is reminded that prediction of per capita consumption expenditure was obtained through the regression analysis of data contained in the LSMSIII.

The final regression model for the prediction of per capita consumption expenditure was run on a total of 1564 of the 1557 households available in the LSMSIII dataset. The model included 16 variables, representing all different area-based SEP domains taken into account with the exception of crowding and access to community services. The list of explanatory variables together with their coefficients and P values are illustrated in **Table 5.5**. The model showed a relatively good explanatory power ($R^2 = 0.60$).

Table 5.5: Estimates of regression coefficients for the poverty mapping

Socioeconomic variables by SEP domain	Coefficients	Standard Error	P-value
Household structure			
Average household size			
≤ 5	Ref		
> 5	-.133	0.145	0.003
Age dependency ratio			
≤ 79.3	Ref		
> 79.3	-0.002	0.0023	< 0.001
Education opportunities			
N. Children (< 15 yr) with no primary education in the household			
0	Ref		
≥ 1	-.0767	0.033	0.02
N. children (< 15 years) who have never attended school*	- 0.068	0.014	<0.001
N. Women with no primary education in the household			
0	Ref		
1	- 0.062	0.042	0.14
≥ 2	- 0.086	0.065	0.19
N. Adults unable to read and write			
0	Ref		
1	- 0.174	0.057	0.03
2	- 0.126	0.075	0.09
≥ 3	-0.128	0.194	0.17
N. adults who have never attended school			
1	Ref		
2	- 0.352	0.047	< 0.001
3	- 0.537	0.063	< 0.001
≥ 4	- 0.771	0.075	< 0.001
N. adults with primary school as highest grade achieved			
1	Ref		
2	0.396	0.077	< 0.001
3	0.415	0.077	< 0.001
≥ 4	0.574	0.0815	< 0.001
Occupation opportunities			
N. members employed in agricultural sector			
0	Ref		
1	0.033	0.066	0.6
N. members employed in the construction sector			
0	Ref		
1	-.0193	0.043	0.6
N. members employed in the service sector			
0	Ref		
1	0.094	0.045	0.03
2	0.085	0.055	0.12

Table 5.5 – Continued

Socioeconomic variables by SEP domain	Coefficients	Standard Error	P-value
Housing quality and assets ownership			
N. of assets owned			
0			
1	0.179	0.041	< 0.001
2	0.231	0.046	< 0.001
3	0.306	0.055	< 0.001
4	0.457	0.067	< 0.001
≥ 5	1.022	0.074	< 0.001
Type of roof			
Asbestos	Ref		
Iron	- 0.070	0.035	0.05
Grass	0.030	0.081	0.7
Other	0.219	0.058	<0.001
Type of lighting energy			
Kerosene	Ref		
Electricity	0.289	0.091	0.002
Candles	0.165	0.055	0.003
Other	- 0.332	0.123	0.007
Type of cooking energy			
Wood	Ref		
Charcoal	0.155	0.082	0.06
Electricity	0.215	0.115	0.06
Other	0.294	0.113	0.009
Type of toilet			
Flush Toilet	Ref		
Latrines	- 0.359	0.044	< 0.001
Other	- 0.262	0.075	0.001

*Included as continuous variable

The agreement between the prediction of consumption expenditure and consumption expenditure - Before applying the regression model to the census dataset I have first compared the prediction of consumption expenditure with the actual consumption expenditure provided in the LSMSIII. This allowed me to assess the goodness of the model. Both the actual and predicted values of per capita consumption expenditure derived from the LSMSIII data showed a normal distribution, but the predicted values appeared to be less skewed and more spread than the actual ones (**Figure 5.2**).

Despite the similar trends, the two variables showed a considerable disagreement in terms of household SEP ranking: overall, only 44% of all the households were classified in the same quintile by both variables, corresponding to a kappa value of 0.4 (P< 0.001). 657 households

(42.3%) were misclassified by one quintile, 189 (12.2) by two, 24 (1.6%) by three, and one by four quintiles (0.06%).

From the cross tabulation of the actual and predicted per capita consumption expenditure quintiles, it looked like the prediction of per capita consumption expenditure performed better in the classification of the households in the extreme quintiles (as shown by the 56.1% and 71.9% of households respectively classified in the lowest and highest quintile by both variables). For all the other quintiles the level of household misclassification was higher **(Table 5.6)**.

Table 5.6: Household socioeconomic position ranking: the agreement between predicted and the actual per capita consumption expenditure

	Per capita consumption expenditure* Quintiles					
Prediction Quintiles		I	II	III	IV	V
	I	175 (56.1)	91 (29.2)	27 (11.)	9 (2.9)	0
	II	90 (28.9)	98 (31.5)	79 (25.4)	39 (12.5)	5 (1.6)
	III	35 (11.3)	86 (27.8)	77 (24.9)	92 (29.8)	19 (6.1)
	IV	9 (2.9)	35 (11.2)	94 (30.1)	110 (35.3)	64 (20.5)
	V	1 (0.3)	1 (0.3)	24 (7.2)	61 (19.7)	223 (71.9)

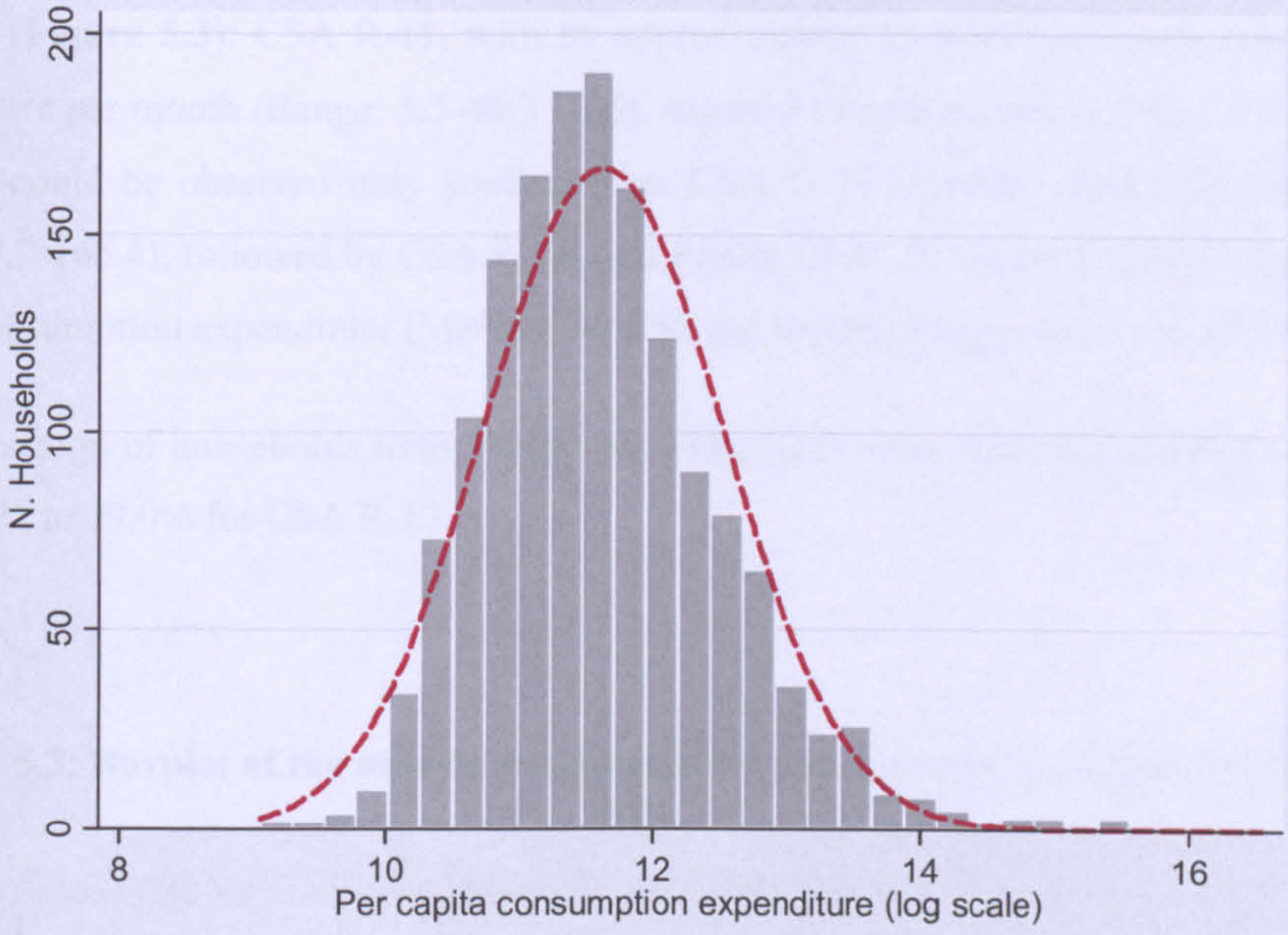
*From LSMSIII data [14]

The adequacy of the regression model was also assessed by assessing the proportion of households falling below the extreme and moderate poverty lines in the province of Lusaka according to the Central Statistical Office of Zambia [14]. The predicted measure of per capita consumption expenditure slightly underestimated the proportion of households below the extreme poverty line (23.0% compared to the 27.6% according to the actual measurement). Both the actual and the predicted measure of per capita consumption expenditure did provide a very similar estimate of the households living below the moderate poverty line (44.8% and 43.8% respectively).

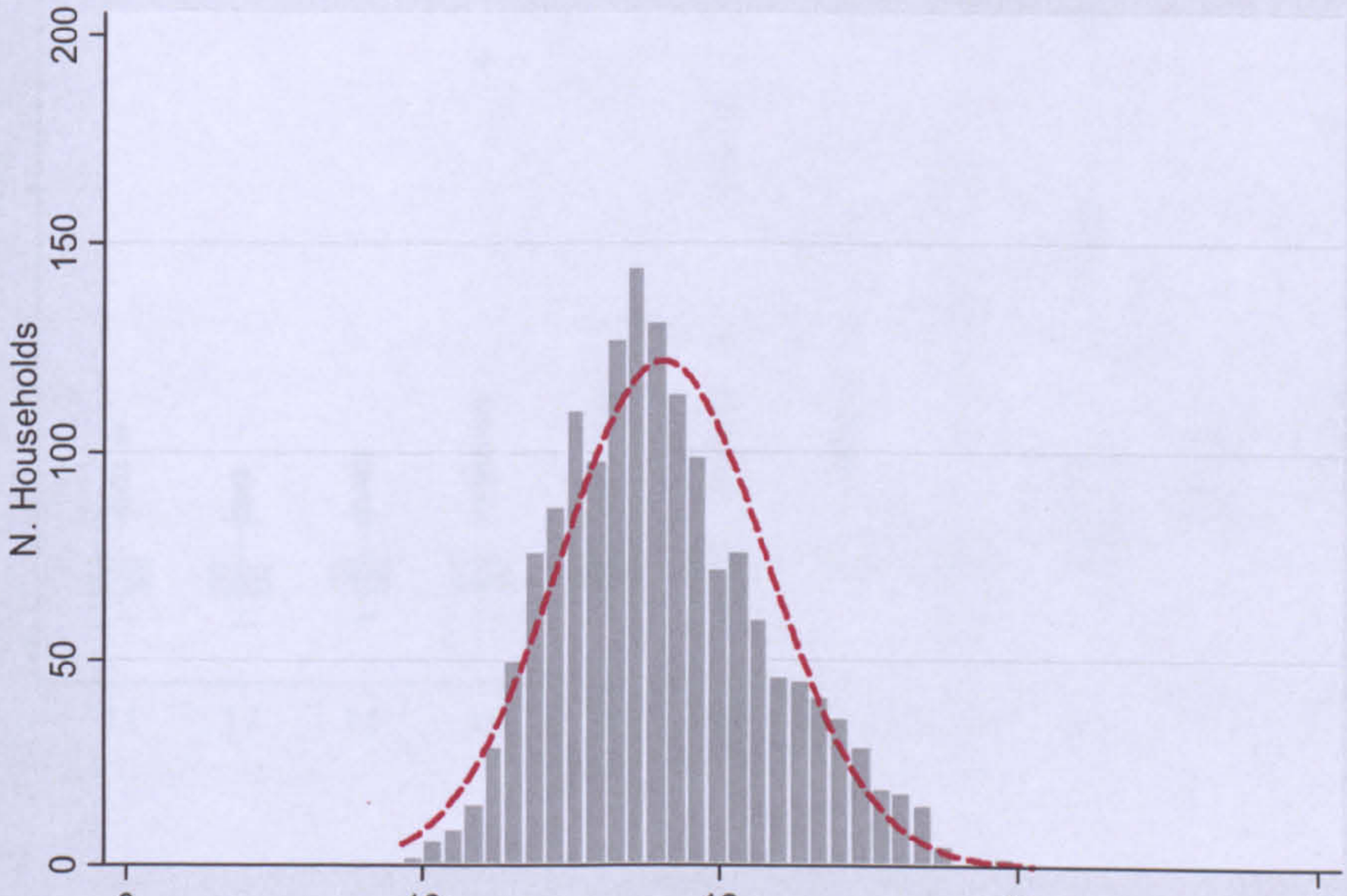
Figure 5.2: Distribution of household per capita consumption expenditure for the province of Lusaka: actual and estimated values

---

A) Per capita consumption expenditure*



B) Prediction of per capita consumption expenditure*



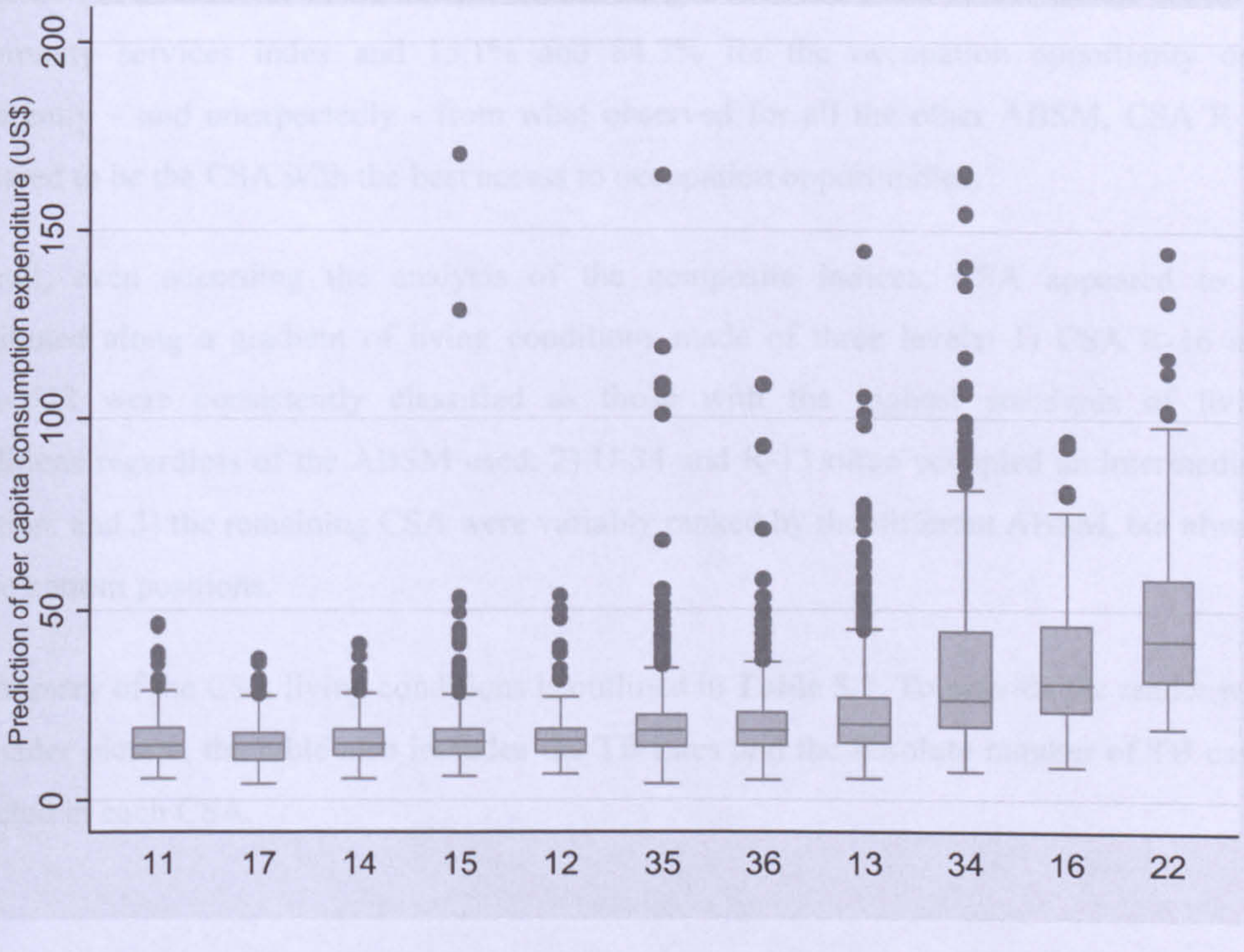
*From LSMSIII data

Application of the model to census dataset - The regression coefficients so estimated were applied to a list of equivalent variables contained in the census dataset to impute a measure of per capita consumption expenditure in this dataset.

The CSA looked very similar in terms of absolute SEP, but it was still possible to detect a gradient (**Figure 5.3**): CSA R-11, with its approximately 15 US$ per capita consumption expenditure per month (Range: 5.5-46.3 US$), resulted to be the poorest CSA. A significant increase could be observed only starting from CSA U-34 (median: 26.3 US$ per month, Range: 7.2-165.4), followed by CSA R-16, and finally CSA-22, the area with the highest per capita consumption expenditure (Median: 41 US$ per month, Range: 19.0-144.2 US$).

The proportion of households living below the extreme poverty line ranged between 0% for CSA U-22 to 29.0% for CSA R-17.

Figure 5.3: Boxplot of the area-based measure of absolute socioeconomic position by CSA



(d) The remaining area-based measures of SEP

In this chapter, I will present the results of the remaining 4 ABSM only as categorical variables. The main findings are summarised in **Figure 5.4** showing the proportion of households in the lowest terciles for each of these indices ranked in ascending order. As for the other two ABSM, even for these two composite indices the higher the proportion of households, the worse the living conditions in these communities. In the graphs, the indices have been grouped according to the extent of variability they showed across CSA.

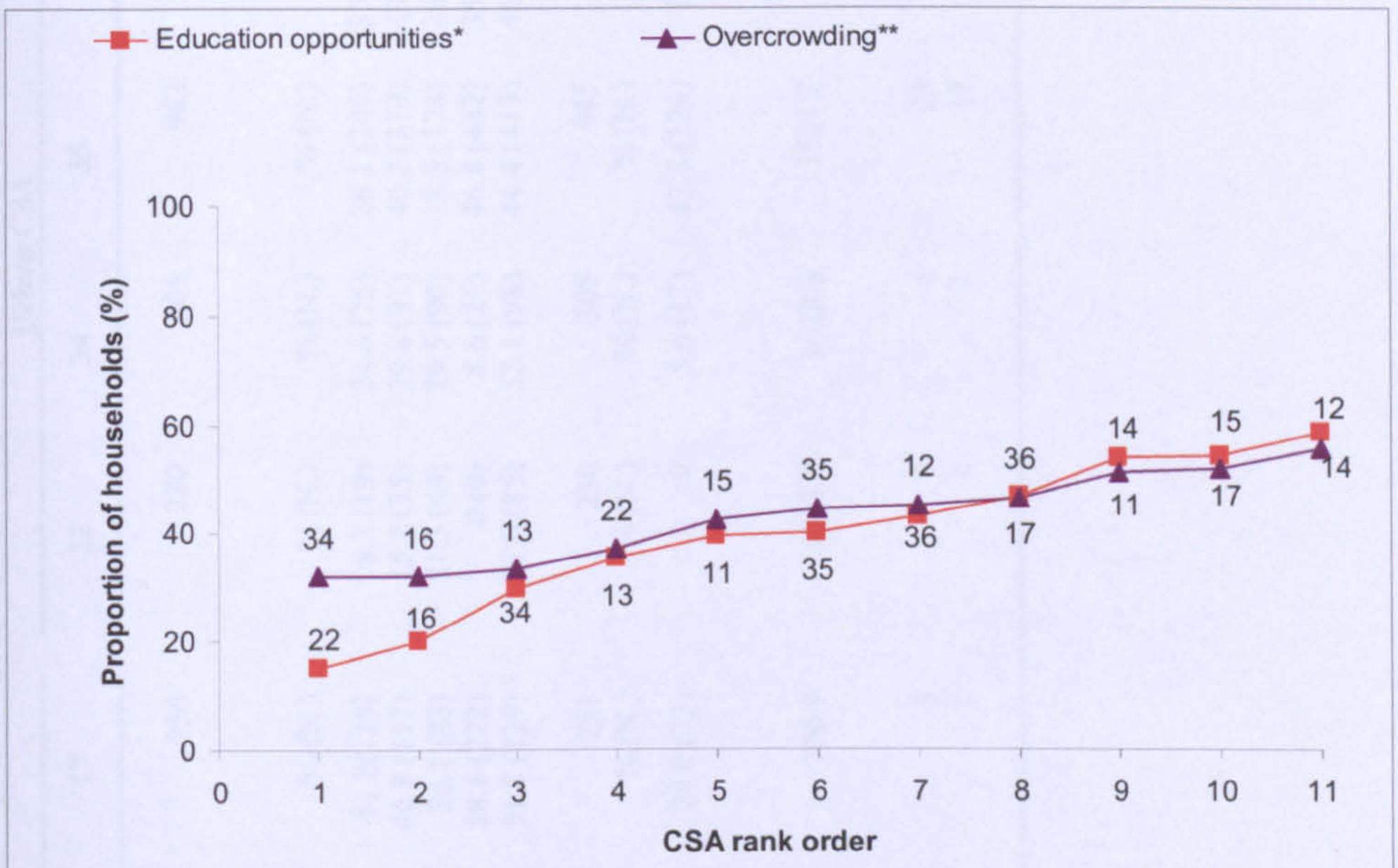
The crowding and the education opportunities index showed modest variability across CSA. CSA U-34 and R-14 were the communities with respectively the lowest (32.1%) and highest proportion of overcrowded households (55.5%). The proportion of households having poor access to education opportunities ranged between 15.2% for CSA U-22, confirmed to be one of the wealthiest communities, and approximately 60% for the rural CSA 12.

The occupation and access to community services indices showed a much steeper increase, suggesting a much higher heterogeneity between CSA compared to the previous indices. The proportion of households in the bottom terciles ranged between 0 and 96.2% for the access to community services index and 15.1% and 84.3% for the occupation opportunity one. Differently – and unexpectedly - from what observed for all the other ABSM, CSA R-11 appeared to be the CSA with the best access to occupation opportunities.

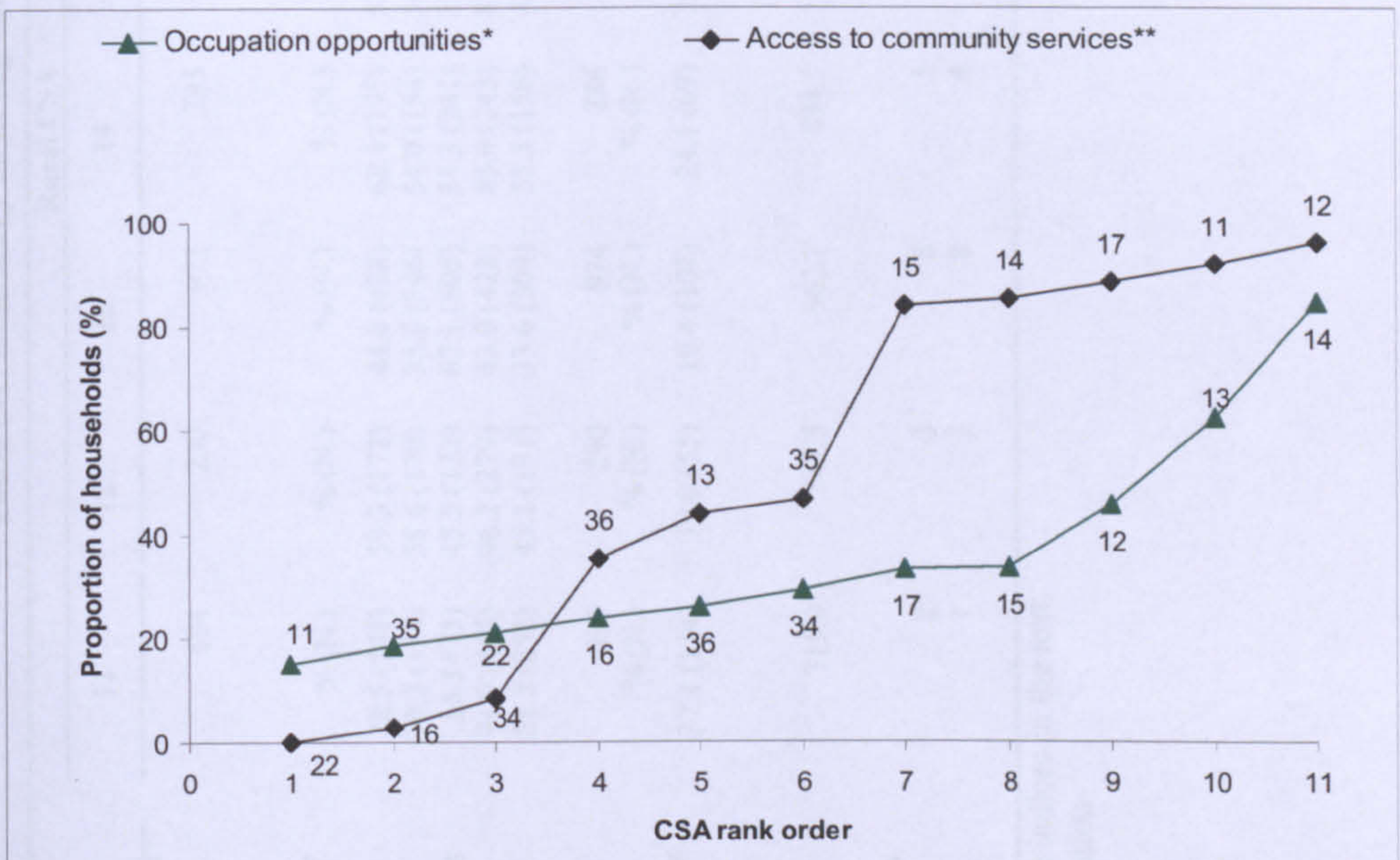
Overall, even according the analysis of the composite indices, CSA appeared to be distributed along a gradient of living conditions made of three levels: 1) CSA R-16 and CSA-U22 were consistently classified as those with the highest standards of living conditions regardless of the ABSM used; 2) U-34 and R-13 often occupied an intermediate position; and 3) the remaining CSA were variably ranked by the different ABSM, but always in the bottom positions.

A summary of the CSA living conditions is outlined in **Table 5.7.** To provide the reader with a broader picture, the table also includes the TB rates and the absolute number of TB cases detected in each CSA.

Figure 5.4: Other area-based measures of socioeconomic position



*Proportion of households in each CSA having poor access to education opportunities
**Proportion of overcrowded households in each CSA



*Proportion of households in each CSA having poor access to occupation opportunities
**Proportion of households in each CSA having poor access to community services

Table 5.7: Summary of living conditions and TB burden by CSA

	Rural CSA							Urban CSA			
	11	**12**	**13**	**14**	**15**	**16**	**17**	**22**	**34**	**35**	**36**
Area-based measures of SEP‡											
Total N. Households	494	290	972	285	705	146	250	230	305	942	838
Proportion of households in the bottom tercile of each ABSM:	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)
Relative SEP	28.5 (141)	59.3 (172)	44.0 (428)	68.4 (195)	52.2 (368)	11.6 (17)	51.2(128)	8.3 (19)	24.6 (75)	26.1 (246)	37.8 (283)
Poor education opportunities	39.3 (194)	58.6 (170)	35.6 (346)	54.0 (154)	54.2 (382)	19.9 (29)	46.8 (117)	15.2 (35)	29.8 (91)	40.2 (379)	43.3 (363)
Poor employment opportunities	15.1 (75)	45.5 (132)	62.1 (605)	84.3 (241)	33.4 (236)	23.9 (34)	33.1 (83)	21.3 (49)	29.5 (90)	18.8 (178)	26.3 (221)
Poor access to public services	91.9 (455)	96.2 (279)	43.9 (428)	85.0 (243)	83.7 (591)	2.7 (4)	88.4 (222)	0 (0)	8.6 (26)	46.8 (442)	35.4 (297)
Overcrowding	51.2 (253)	45.2 (131)	33.6 (309)	55.5 (156)	42.4 (294)	32.2 (47)	51.8 (129)	37.0 (85)	32.1 (98)	44.4 (419)	46.6 (353)
Total N. Households	495	290	974	286	706	146	251	230	305	945	839
	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)	% (N.)
Proportion of households living below the poverty line	27.1 (134)	19.0 (55)	10.4 (101)	24.1 (69)	21.6 (152)	2.0 (3)	29.0 (73)	0	5.6 (17)	13.3 (126)	10.8 (91)
TB prevalence rates											
Crude TB prevalence rates†	315.0	541.5	962.7	894.5	778.2	365.0	700.9	1093.8	1028.3	1521.2	1184.7
Absolute number of TB cases											
Detected in prevalence survey	2	3	8	5	6	1	3	7	4	23	17
Enrolled in case control study	1	2	8	4	3	0	1	4	1	17	11

‡See the description of the indices in the text.
†Rates per 100,000 population

5.2.3 The CSA ranking according to the different area-based measures of SEP

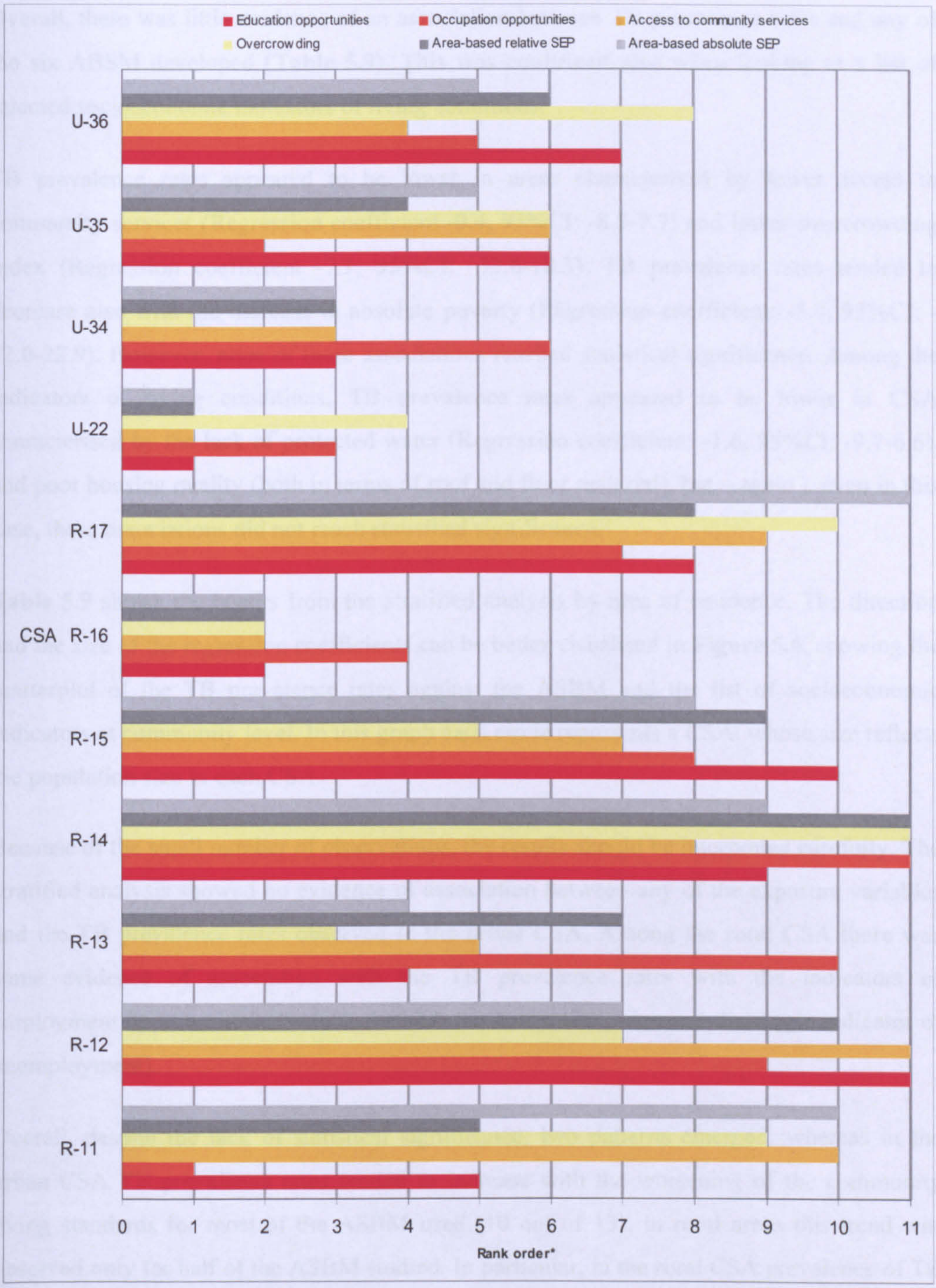
The ABSM showed a moderate to high correlation with each other (**Table 5.8**). The highest correlation was observed between the area bases measure of relative SEP and education opportunities and between the index of access to community services and the area based measure of absolute SEP (Spearman coefficient = 0.9, P< 0.001 for both). The occupation opportunities index was the only one not showing significant correlation with any of the other indices, except the area based measure of relative SEP (Spearman correlation coefficient = 0.77, P=0.05).

Table 5.8: Size and significance of Spearman correlation coefficients among ABSM

Indices	Education index	Occupation index	Access to services	Crowding	Relative SEP	Absolute SEP
Education opp.	1.0					
Occupation opp.	0.53 (0.08)	1.0				
Access to services	0.79 (0.003)	0.28 (0.4)	1.0			
Crowding	0.63 (0.03)	0.13 (0.7)	0.70 (0.01)	1.0		
Relative SEP	0.90 (<0.001)	0.77 (0.05)	0.76 (0.001)	0.62 (0.03)	1.0	
Absolute SEP	0.73 (<0.001)	0.20 (0.53)	0.9 (< 0.001)	0.8 (< 0.001)	0.71(0.01)	1.0

The agreement among indices was also explored graphically. **Figure 5.5** suggests that there was agreement in the way the different ABSM were ranking the CSA. The only exception to this was observed for the overcrowding and the occupation opportunities indices: these two indices often ranked CSA differently from the other indices. This was particularly evident for CSA U-22, U-34, U-35, R-11, R-13 and R-16.

Figure 5.5: Area based measure of SEP: internal coherence of CSA ranking



*Ranking in descending order from from the wealthiest to the poorest community.

5.2.4 The association between area-based measures of SEP and TB prevalence rates

Overall, there was little evidence of an association between TB prevalence rates and any of the six ABSM developed (**Table 5.9**). This was confirmed also when looking at a list of selected socioeconomic indicators of living conditions.

TB prevalence rates appeared to be lower in areas characterised by lower access to community services (Regression coefficient -0.4, 95%CI: -8.5-7.7) and lower overcrowding index (Regression coefficient -7.1, 95%CI: -32.6-18.3). TB prevalence rates tended to decrease also with the increase in absolute poverty (Regression coefficient: -5.0, 95%CI: -32.0-22.9). However, none of these associations reached statistical significance. Among the indicators of living conditions, TB prevalence rates appeared to be lower in CSA characterised by the lack of protected water (Regression coefficient: -1.6, 95%CI: -9.7-6.6), and poor housing quality (both in terms of roof and floor material), but – again – even in this case, these associations did not reach statistical significance.

Table 5.9 shows the results from the stratified analysis by area of residence. The direction and the size of the regression coefficients can be better visualised in **Figure 5.6**, showing the scatterplot of the TB prevalence rates against the ASBM and the list of socioeconomic indicators at community level. In this graph each circle represents a CSA, whose size reflects the population size in each CSA.

Because of the small number of observations, the results should be interpreted carefully. The stratified analysis showed no evidence of association between any of the exposure variables and the TB prevalence rates observed in the urban CSA. Among the rural CSA there was some evidence of association with the TB prevalence rates with the indicators of employment opportunities (P= 0.02 for both the composite index and the single indicator of unemployment).

Overall, despite the lack of statistical significance, two patterns emerged: whereas in the urban CSA TB prevalence rates tended to increase with the worsening of the community living standards for most of the ASBM used (10 out of 13), in rural areas this trend was observed only for half of the ASBM studied. In particular, in the rural CSA prevalence of TB seemed to decrease with the increase of 7 ABSM, including households living below the poverty line, having poor access to community services (like protected water and electricity), having poor housing quality and that were overcrowded.

Table 5.9: The association between area-based measures of SEP and TB prevalence rates

	Overall		Urban		Rural	
	Coefficient* (95%CI)	P value	Coefficient[§] (95% CI)	P value	Coefficient[§] (95%CI)	P value
Area-based measure of SEP						
Relative SEP[1]	7.0 (-7.5 ; 21.4)	0.3	-0.7 (-69.2 ; 67.8)	0.9	9.4 (-7.9 ; 26.7)	0.2
Absolute SEP[2]	-5.0 (-32.0 ; 22.9)	0.7	36.6 (-55.6 ; 128.7)	0.2	-13.9 (-45.7 ; 17.9)	0.3
Low education opportunities[3]	3.6 (-14.9 ; 22.2)	0.7	10.1 (-53.0 ; 73.3)	0.6	0.3 (-26.2 ; 26.8)	0.9
Low employment opportunities[4]	7.8 (-1.3 ; 17.1)	0.08	-42.0 (-110.8 ; 26.7)	0.1	9.3 (1.8 ; 16.8)	0.02
Low access to community services[5]	-0.4 (-8.5 ; 7.7)	0.9	10.1 (-9.9 ; 30.0)	0.2	-3.6 (13.5 ; 6.2)	0.4
Overcrowding[6]	-7.1 (-32.6 ; 18.3)	0.5	19.7 (-84.6 ; 124.1)	0.5	-14.8 (-44.1 ; 14.6)	0.2
Selected list of indicators of living conditions						
% HH[†] with no literate adult	4.1 (-19.0 ; 27.3)	0.7	9.3 (-158.7 ; 174.4)	0.8	3.6 (-24.2 ; 34.4)	0.7
% HH with no adult employed	8.2 (-2.7 ; 19.0)	0.1	-41.0 (-128.6 ; 46.1)	0.2	9.8 (0.22 ; 19.5)	0.02
% HH with no electricity	1.9 (-4.6 ; 8.5)	0.5	4.0 (-10.2 ; 18.3)	0.4	-1.3 (-13.6 ; 11.0)	0.4
% HH with no protected water	-1.6 (-9.7 ; 6.6)	0.7	3.2 (-26.3 ; 32.7)	0.7	-3.9 (-14.6 ; 6.7)	0.4
% HH with poor roofing quality[7]	-2.4 (-14.5 ; 9.8)	0.7	7.8 (-40.8 ; 56.4)	0.6	-5.6 (-20.5 ; 9.3)	0.4
% HH with poor flooring quality[8]	-0.04 (-8.1 ; 8.1)	0.9	8.9 (-11.5 ; 29.2)	0.2	-3.3 (-13.9 ; 6.9)	0.4
% HH with no asset	4.1 (-9.4 ; 17.7)	0.5	5.6 (-27.7 ; 38.8)	0.5	1.3 (-26.3 ; 29.0)	0.9

*Adjusted by age, sex and type of area of residence (i.e. rural and urban) and using CSA population size as weight. [†] HH= Households
[§]Adjusted by age and sex
[1]Based on the proportion of households classified as Low SEP in each CSA
[2]Based on the proportion of households living below the poverty line in each CSA
[3]Based on the proportion of households having low education opportunities n each CSA
[4]Bases on the proportion of households having low employment opportunities in each CSA
[5]Based on the proportion of households having low access to community services in each CSA
[6]Based on the proportion of households classified as overcrowded (i.e. more than 3.6 people per sleeping room) in each CSA
[7]Proportion of households with roof made of grass or iron
[8]Proportion of households with floor made of mud

Figure 5.6: Scatterplot of TB prevalence rates against the area-based measure of SEP, by urban and rural area

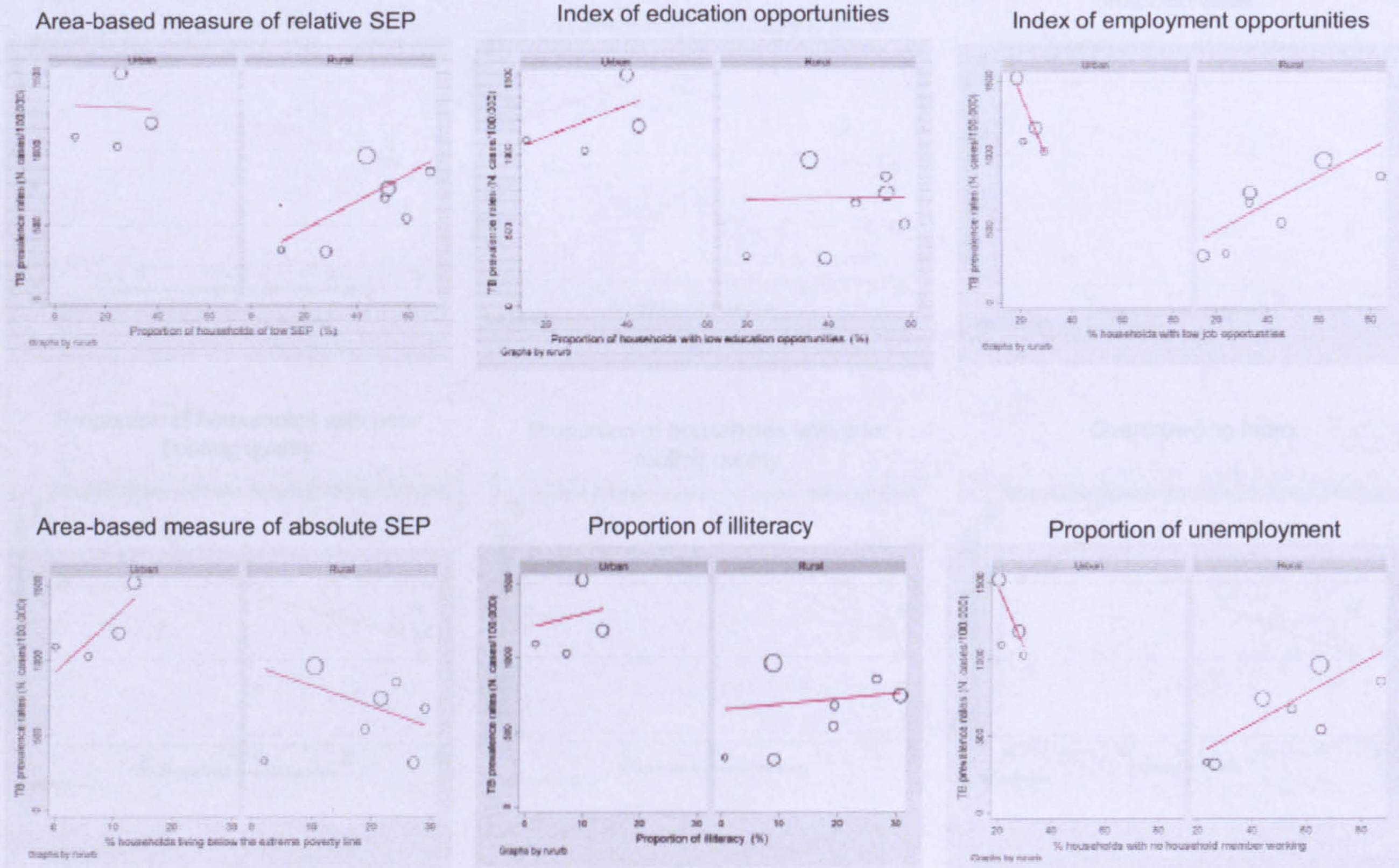
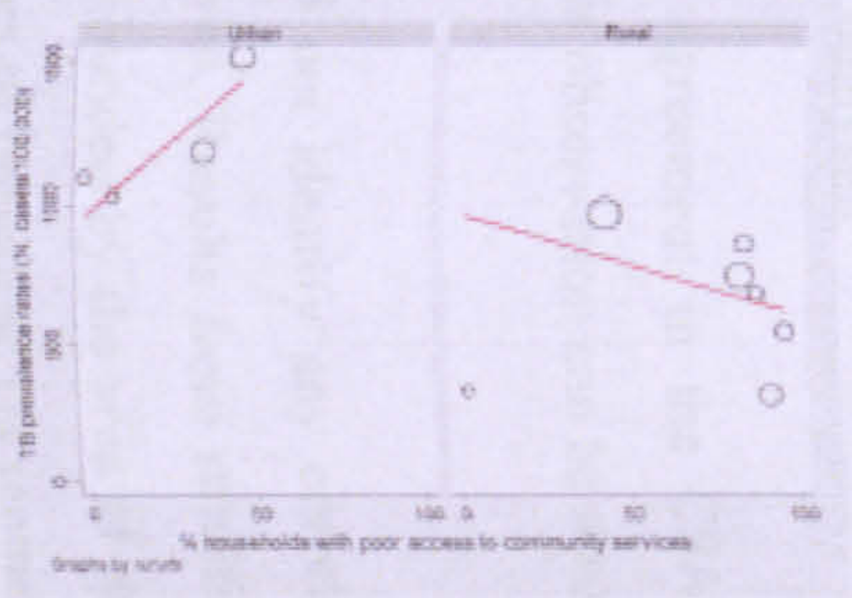
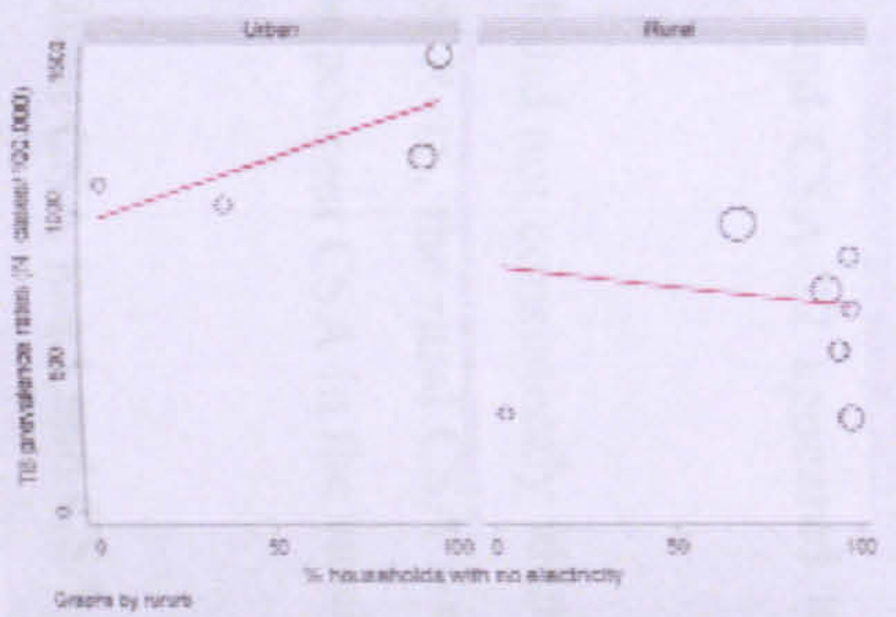


Figure 5.6: Continued

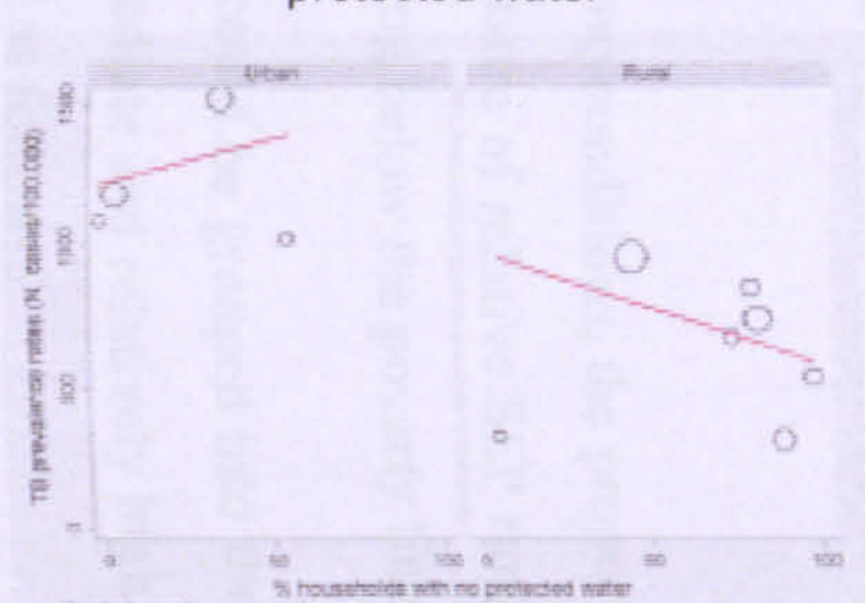
Access to community services index



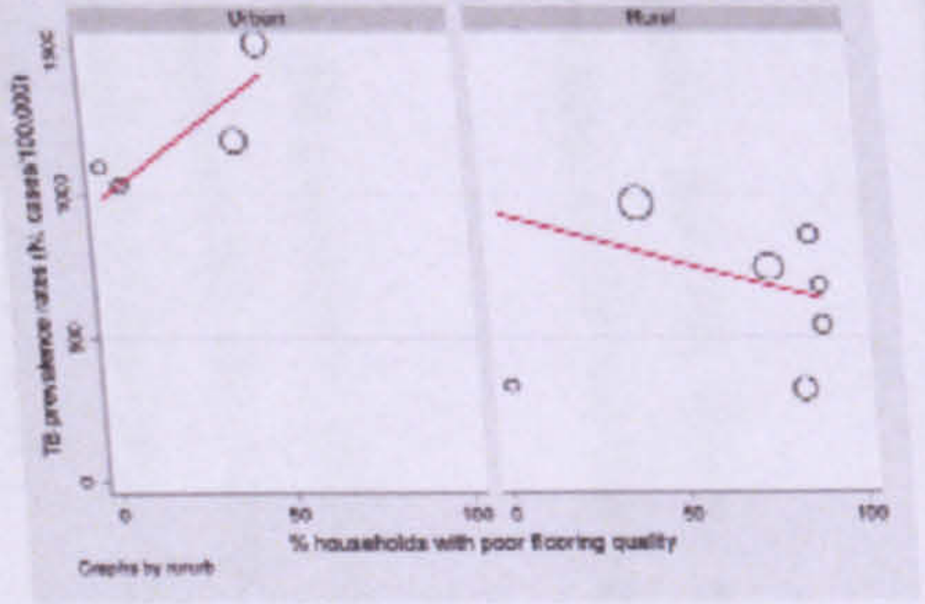
Proportion of households with no electricity



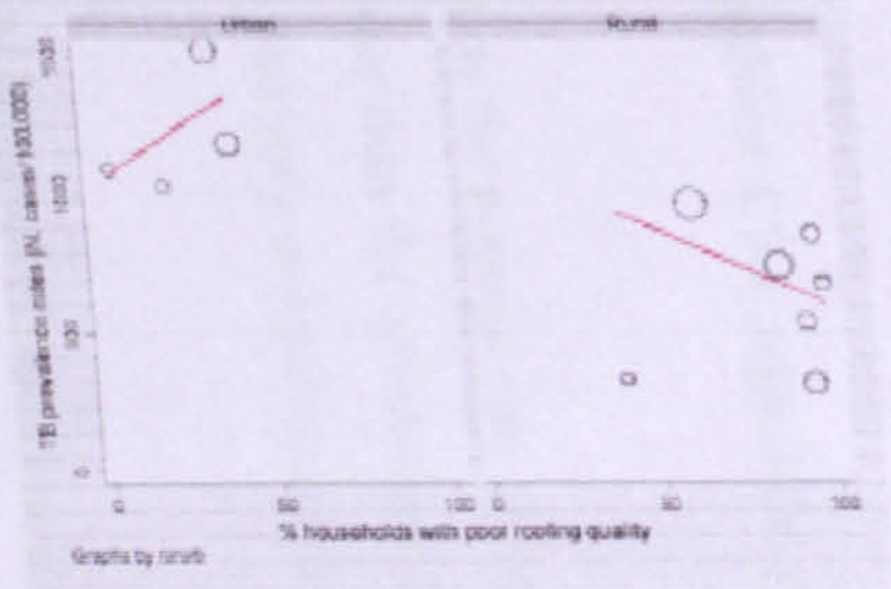
Proportion of households with no access to protected water



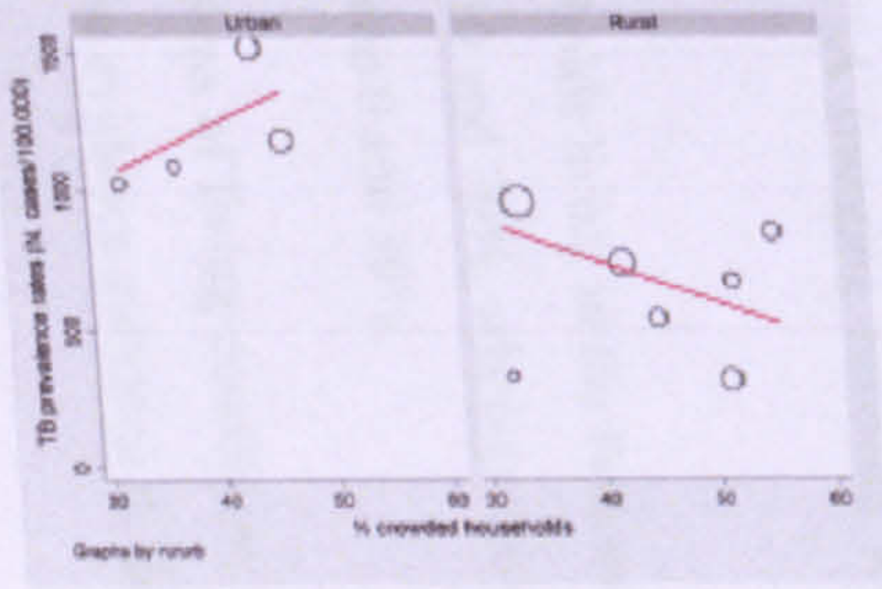
Proportion of households with poor flooring quality



Proportion of households with poor roofing quality



Overcrowding Index



Summary

The prevalence survey conducted in these two nationally representatives communities of Zambia showed an overall cluster-adjusted TB prevalence rate of 870/100,000. The TB prevalence rates were approximately 1000/100.000 in all the 4 urban CSA and consistently higher than the TB rates observed in the 7 rural CSA (ranging between 300/100,000 to 700/100,000).

In terms of living conditions, the proportion of households living in the bottom tercile of the area-based measure of relative SEP ranged between 10 and 70%, whereas the proportion of households living below the poverty line ranged between 0 and 30%.

Overall, CSA could be grouped into three broad patterns of living conditions, characterised by low, intermediate and relatively high standards. The evidence gathered in this chapter can be summarised as follows:

1. CSA R-16 and CSA-22 appeared to be the wealthiest CSA regardless of the ABSM employed.

2. The ABSM did not consistently identify any CSA as the poorest one. However, with the exception of R-16, the rural CSA were ranked lower than the urban ones by most of the indices. The poorest CSA in the rural area looked had very similar characteristics to each other.

3. In this gradient CSA R-13 and CSA U-34 occupied an intermediate position for most of the ABSM analysed.

The observed agreement in the CSA ranking suggests that the indices were working coherently. This observation can be also considered evidence of the fact that the indices were built correctly.

This study did not identify any correlation between TB prevalence rates and any of the ABSM employed. Results from stratified analysis suggested that the overall findings may have been confounded by the area of residence. Although not statistically significant, in the urban CSA rates of prevalent TB seemed to increase with the worsening of living conditions for most of the ABSM studied. This was also observed among rural CSA, especially when looking at employment type of ABSM; however, unexpectedly, in the rural communities, more than a half of the ABSM showed evidence of a negative association between TB prevalence and worsening of living conditions. Because of the small number of observations this results should be interpreted with extreme caution.

6. The Case Control study results

There can be no doubt that the socioeconomic position of individuals, groups, and places is defining characteristics of their level of health and disease.
While it is important to keep in mind the salience of socioeconomic position in determining the health status of individuals and populations, advancing our understanding of the reasons for these effects and their policy implications require more than simply pointing to this association.

Lynch, J[7]. 2000

Introduction

In this chapter I will try to answer the main research question of this thesis: is household SEP associated with prevalent TB? And if so, what is the possible causal pathway linking them together.

After a descriptive analysis of the socioeconomic indicators collected and the household SEP profile of the study population, this chapter will illustrate the main findings of this study, and in particular: 1) the main risk factors for prevalent TB and the confounding effect played by household SEP on these associations; 2) the role of household SEP in the epidemiology of TB prevalence and the possible mediation effect played by a list of known risk factors for TB; and finally 3) the population attributable risk factor for household SEP and for any other risk factor significantly associated with prevalent TB.

6.1 Chapters objectives

1. To describe the population enrolled in the case control study (**Section 6.2.1**).

2. To describe the socioeconomic profile of the study population (**Section 6.2.2**).

3. To explore the distribution of prevalent tuberculosis risk across household SEP groups of the controls population and between cases and controls (**Section 6.2.3**).

4. To assess the association between household SEP and prevalent TB looking both at the single socioeconomic indicators and the composite household SEP index **(Section 6.2.4)**.

5. To investigate which household SEP domain is mostly contributing to this association **(Section 6.2.4)**.

6. To undertake a multivariable analysis driven by an *a priori* defined conceptual framework to assess the mediation role of individual-level risk factors on the association between household SEP and prevalent TB **(Section 6.2.4)**.

7. To assess the population attributable fraction (PAF) for household SEP and other significant risk factors identified in the study **(Section 6.2.5)**.

6.2 Results

6.2.1 The population enrolled in this study

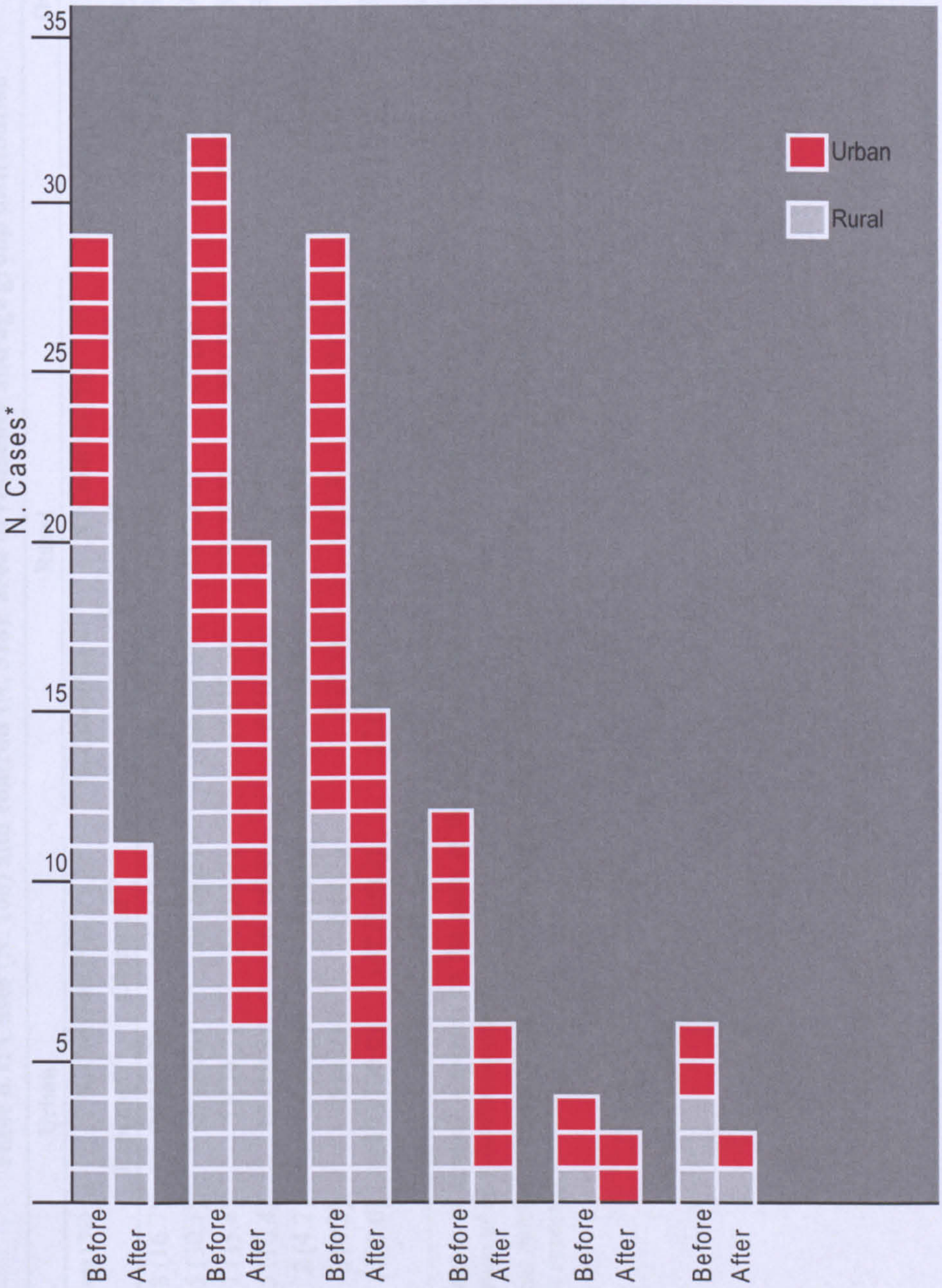
The final population recruited for the case control study included 52 cases and 318 controls.

Cases - The change in the microbiological definition of *M. tuberculosis* cases brought the total number of cases suitable for the case control study from106 to 52. The number of cases loss by age group and area of residence is outlined in **Figure 6.1**. This table shows that most of the cases excluded were from the rural area. According to clinical algorithm (**Figure 4.6**), of the 52 cases eventually included in the analysis 32 (61.5) were bacteriologically confirmed cases of TB, 2 (3.8%) were bacteriologically unconfirmed cases of TB, 12 (23.1%) were asymptomatic cases of TB, and for 6 subjects follow up could not be completed. None of the 52 cases included in the analysis needed to be replaced.

Controls – All the controls initially matched to the 106 cases were retained in the study. Their age group and area distribution, therefore, reflects the profile of these cases rather than the 52 eventually included in the analysis (**Table 6.1**). Of these 318 controls, 89 needed to be replaced (28.0%): 44 (30.0%) in the urban area and 45 (26.3%) in the rural area, respectively (**Figure 6.2**). In both areas, most of the controls were replaced because they had moved somewhere else (respectively 47.7% in the urban area and 82.2% in the rural area). Only a small percentage was replaced because they declined participation (18.2% and 2.2% respectively in the urban and rural area).

Overall, most of the controls replaced (29.1%) fell into the 15-24 age group, followed by the age group 25-34 years (28.1%) and 35-44 years (27.0%). The control replacement did not alter the original frequency matching design (**Table 6.1**).

Figure 6.1: Number of cases by area and age group: before (N.106) and after (N. 52) the change in the microbiological definition of *M. tuberculosis* in the prevalence survey



*Each box is a case.

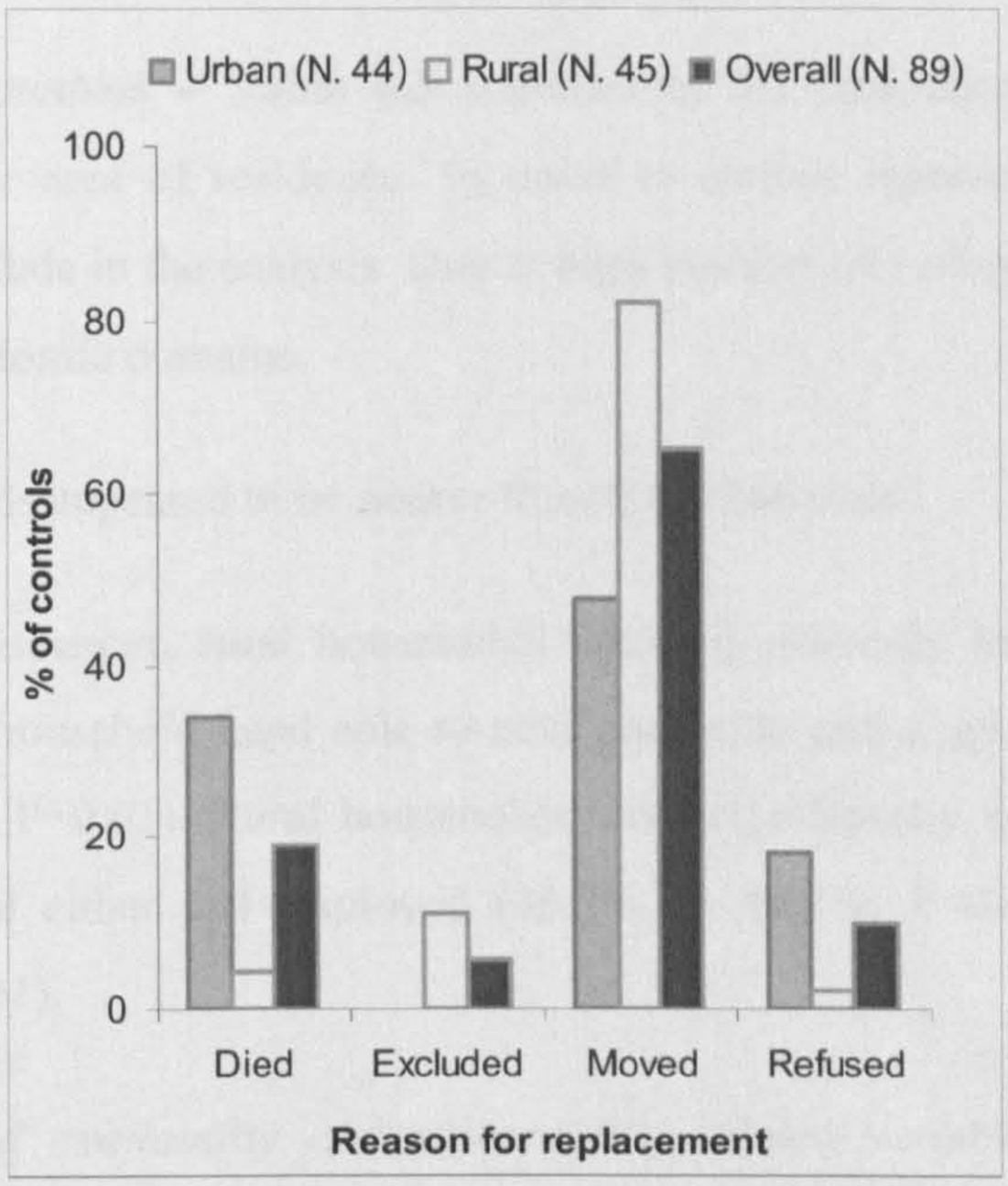
Table 6.1: Cases (N. 106) and controls (N. 318): area of residence and age group distribution

Age group	Urban			Rural			Overall		
	N. Cases (%)	N. Controls (%)		N. Cases(%)	N. Controls (%)		N. Cases(%)	N. Controls (%)	
		Expected[(a)]	Obtained[(b)]		Expected[(a)]	Obtained[(b)]		Expected[(a)]	Obtained[(b)]
15-24	8 (16.7)	24 (16.3)	25 (17.0)	20 (35.0)	60 (35.1)	62 (36.3)	28 (26.4)	84 (26.4)	87 (27.4)
25-34	15 (30.1)	45 (30.6)	39 (26.5)	16 (28.0)	48 (28.1)	45 (26.3)	31 (29.2)	93 (29.2)	84 (26.4)
35-44	17 (35.4)	51(34.7)	55 (34.7)	11 (19.3)	33 (19.3)	34 (19.9)	28 (26.4)	84 (26.4)	89 (28.0)
45-54	5 (10.4)	15 (10.3)	16 (10.9)	6 (10.5)	18 (10.5)	18 (10.5)	11 (10.4)	33 (10.4)	34 (10.7)
55-64	2 (4.2)	6 (4.2)	6 (4.1)	1 (1.7)	3 (1.7)	3 (1.8)	3 (2.8)	9 (2.8)	9 (2.8)
> 64	2 (4.2)	6 (4.2)	6 (4.1)	3 (3.3)	9 (5.3)	9 (5.3)	5 (4.7)	15 (4.7)	15 (4.7)
Total	49 (100.0)	147 (100.0)	147 (100.0)	57 (100.0)	171 (100.0)	171 (100.0)	106 (100.0)	318 (100.0)	318 (100.0)

(a) It refers to the number of controls expected on the basis of the age group and area of residence distribution of the original 106 cases (i.e. before the change of the microbiological definition of *M. tuberculosis* cases).

(b) It refers to the actual number of controls interviewed after replacement. In case of correct replacement the number of controls actually recruited in the study should reflect the number of controls expected.

Figure 6.2: Frequency of controls replacement, overall and by area of residence



When the controls included in the analysis (N. 318) and the excluded ones (N. 89) were compared, they did not appear to be significant different for any of the socio-demographic variables considered. The only exception was for radio ownership: excluded controls less likely than the participant controls to have a radio (OR= 1.8, 95%CI: 1.1-3.0, P=0.02) (**Table 6.2**).

Table 6.2: Main socio-demographic characteristics of the recruited controls (N. 318) and the excluded ones (N. 89)

Variable†	Controls			
	Recruited	Excluded		
	N. (%)	N. (%)	OR (95% CI)	P value
Being female	177 (55.7)	42 (47.2)	0.7 (0.4-1.1)	0.1
Not having a radio	78 (24.5)	33 (37.1)	1.8 (1.1-3.0)	0.02
Not having a TV	190 (59.7)	65 (73.0)	1.6 (0.6-2.5)	0.6
Not having a fridge (N.1)*	273 (86.2)	78 (88.8)	1.3 (0.6-2.6)	0.5
Not having a bicycle	212 (66.7)	66 (74.2)	1.4 (0.8-2.4)	0.2
Having ≤ 2 meals/day	97 (30.5)	20 (22.5)	0.7 (0.4-1.1)	0.1

†Comparison between recruited and excluded controls only allowed for those variables in common both to the case control study and the prevalence survey.
*N. missing values

6.2.2 The socioeconomic profile of the sampled population

The socioeconomic variables – **Table 6.3** summarises the socioeconomic profile of the sampled population by area of residence. In order to endure representativeness only the controls have been include in the analysis. Due to high number of variables results have been organised by socioeconomic domains.

Overall rural households appeared to be poorer than the urban ones.

In terms of *human resources*, rural households were significantly less likely than urban households to have a household head able to read and write and employed in some formal job (99.3% vs 75.9%, P=0.01). Rural households were significantly more likely to have a head of the household either self-employed (36.7% vs 49.7%, P<0.001) or unemployed (4.1% vs 0.6%, P<0.001).

When looking at *food availability and vulnerability* related variables, rural households appeared to be more food insecure and more vulnerable than urban households: rural households were significantly more likely to have less than 2 meals per week containing proteins (63.2% vs 42.2% urban households, P < 0.001). Although not statistically different, rural households were also more likely to have been for 3 or more than 3 months in the 12 months prior the interview without enough to eat (32.7% vs 25.9%, P = 0.2). Compared to urban households, rural households were statistically more likely to have implemented more than 3 coping strategies (57.8% vs 41.5%, P = 0.004).

In terms of *housing quality and assets ownership*, rural and urban households looked considerably different. Urban households reported a significant higher frequency of ownership for most of the assets investigated, with the few exception of land and animals. Urban households were also more likely to live in houses built with more solid and expensive materials like asbestos for the roof (55.8% vs 14.6%, P <0.001), concrete for the floor (59.2% vs 16.3%, P <0.001), and cement for the walls (75.5% vs 37.4%, P <0.001), provided with electricity (41.5% vs 14.1%, P <0.001), running water in the house (42.2% vs 9.2%, P < 0.001) and flush toilet (10.9% vs 4.6%, P = 0.05). Rural household were significantly more crowded than urban households with approximately 30% of them reporting more than 4 people per bedroom compared to 12.2% observed among the urban households.

The *access to community services* related variables revealed that rural households were also significantly more distant to any of the facility investigated.

Table 6.3: Household socioeconomic profile by area of residence (N.318)

Household SEP domains	Urban	Rural	
	N. (%)	N. (%)	P value
Human resources			
Female head of the household	46 (31.3)	52 (30.4)	0.8
> 3 children under 5 years of age	44 (29.9)	66 (38.6)	0.1
Head of the household literate	146 (99.3)	129 (75.9)	< 0.001
Head of the household unemployed	1 (0.7)	7 (4.1)	0.05
Head of the household self employed	54 (36.7)	85 (49.7)	0.02
Head of the household employed	83 (56.5)	71 (41.5)	0.008
Food availability and vulnerability			
Having ≤ 2 meals/day	45 (30.6)	57 (33.3)	0.6
Having ≤ 2 meals with proteins/week	62 (42.2)	108 (63.2)	< 0.001
≥ 3 Months with not enough to eat	38 (25.9)	56 (32.7)	0.2
≥ 3 Coping strategies	71 (41.5)	85 (57.8)	0.004
Housing quality and assets ownership			
Having < 2 assets	54 (36.7)	103 (60.2)	<0.001
Having a fridge	32 (21.9)	12 (7.0)	< 0.001
Having a TV	81 (55.1)	47 (27.5)	< 0.001
Having a radio	122 (83.0)	118 (69.1)	0.004
Having a bicycle	59 (40.1)	47 (27.5)	0.01
Having animals	15 (10.2)	118 (69.0)	< 0.001
Having land	22 (33.3)	126 (96.2)	0.001
> 4 people per bedroom	18 (12.2)	51 (29.8)	< 0.001
Walls - Burn bricks	15 (10.1)	50 (29.2)	< 0.001
Walls – Concrete bricks	87 (59.2)	18 (16.4)	< 0.001
Roof – Asbestos	82(55.8)	25 (14.6)	< 0.001
Roof – Iron	64 (43.5)	75 (43.9)	.< 0.001
Roof – Grass	1 (0.7	71 (41.5)	< 0.001
Floor – Cement	111 (75.5)	64 (37.4)	< 0.001
Floor – Dirt/Earth	35 (23.8)	107 (62.6)	< 0.001
Having electricity	61 (41.5)	24 (14.0)	< 0.001
Having flush toilet	10 (6.8)	11 (6.4)	0.9
Having private piped water	62 (42.2)	15 (8.8)	0.001
Distance from community services			
> 15 minutes walking from water	12 (8.2)	51 (29.8)	< 0.001
> 30 minutes walking from clinic	47 (32.0)	121 (70.8)	< 0.001
> 30 minutes walking from transport	14 (9.5)	107 (62.6)	< 0.001
> 30 minutes walking from market	55 (37.4)	152 (88.9)	< 0.001

The SEP index and SEP domains indices – Overall 11 socioeconomic variables were included in the household index of relative SEP (**Table 6.4**). The size of the weights assigned to each socioeconomic variable suggest that they correlated well enough with each other and with the SEP index and that their sign was coherent with the expected characteristics of relative SEP. For example, having a traditional well, the floor made in dirt and grass for the roof received a negative weight because of their negative correlation with the other indicators of high SEP in the PCA and the overall SEP index.

The SEP score exhibited an Eigen value of approximately 4 and accounted for almost 33% of the total variance of the variables that were included in the PCA (**Table 6.4**).

Table 6.4: Main characteristics of the household socioeconomic index

	Household socioeconomic index
N. Observations	318
N. Indicators	11
Weights	
Head of the household employed	0.217
≥ 2 meals containing proteins/week	0.253
< 3 coping strategies	0.283
Having electricity	0.354
Radio	0.220
Public piped water	0.186
Traditional well	-0.249
Floor – Dirt	-0.414
Roof material –Grass	-0.366
Roof material –Asbestos	0.325
< 30 min. walking distance from nearest market	0.354
Eigen Value	3.6
% common variance	32.6

Differently from the community SEP score, in the household SEP one, the higher the score of the SEP index the higher was the SEP of the households recruited in the study. Among controls, the SEP score median was equal to 0.1 ranging between -3.9 and +3.2. The rural households were characterised by significantly lower SEP score values compared to the urban households ($P < 0.001$, according to the Mann-Whitney test) (**Figure 6.3**). However, the analysis of the median SEP score across the SEP terciles (**Figure 6.4**) suggested that only the 'low' SEP and 'medium' SEP households in the rural area were significantly poorer than the 'low' and 'medium' SEP households of the urban area ($P < 0.001$ for both the SEP index categories). Households classified as 'high' SEP in the urban and rural area showed a very similar SEP score ($P = 0.8$).

Figure 6.3: Boxplots of the household socioeconomic score, overall and by area of residence

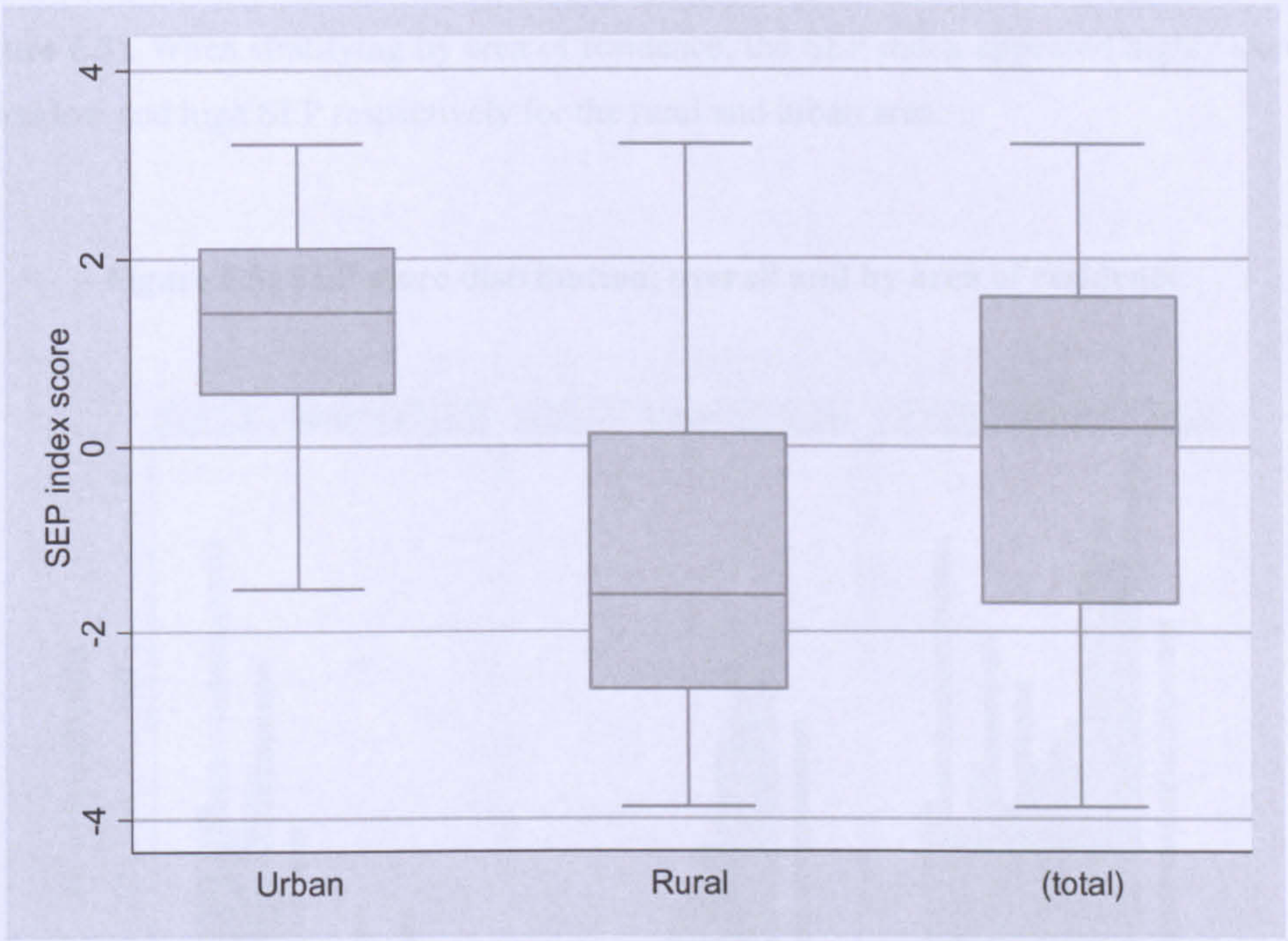
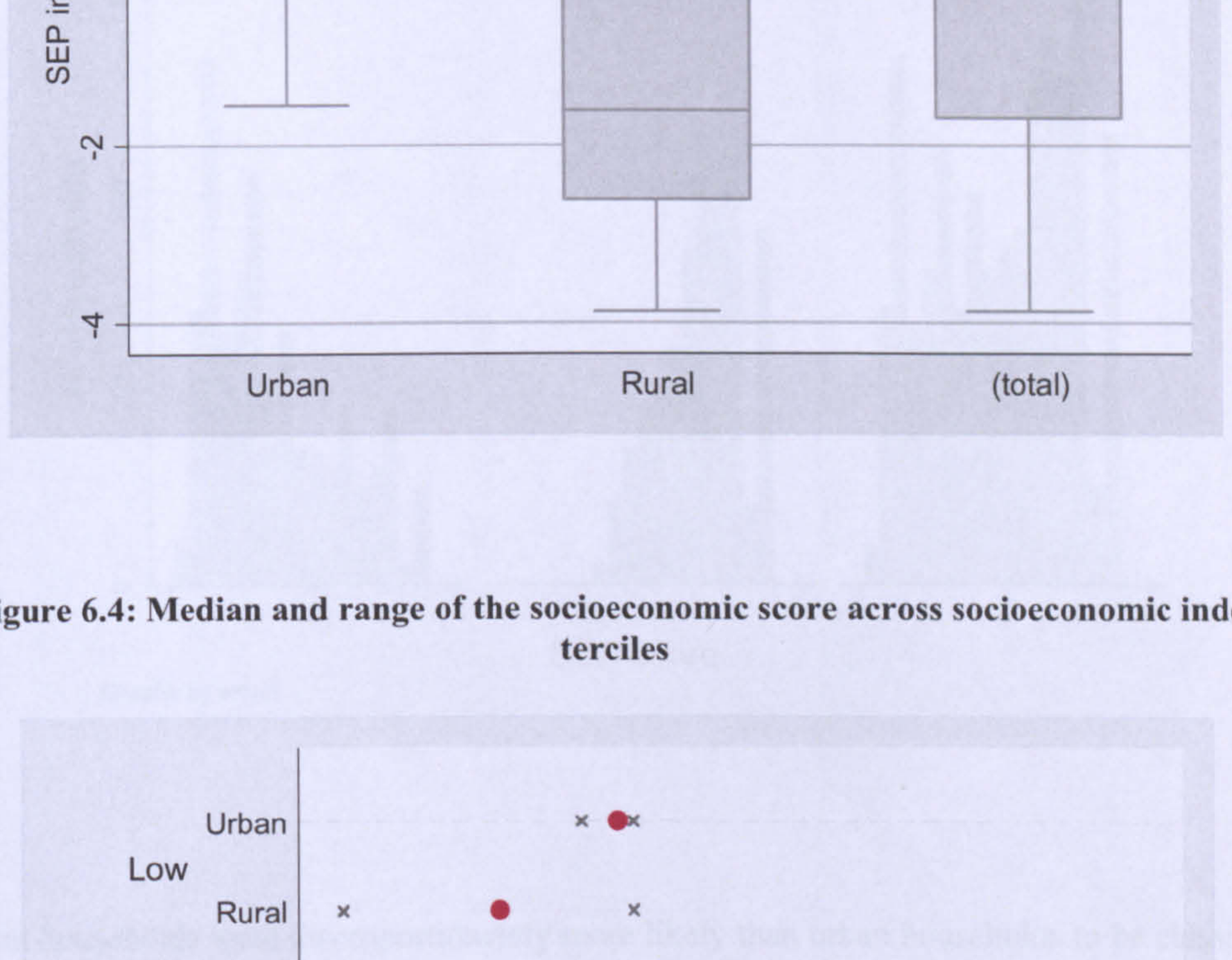


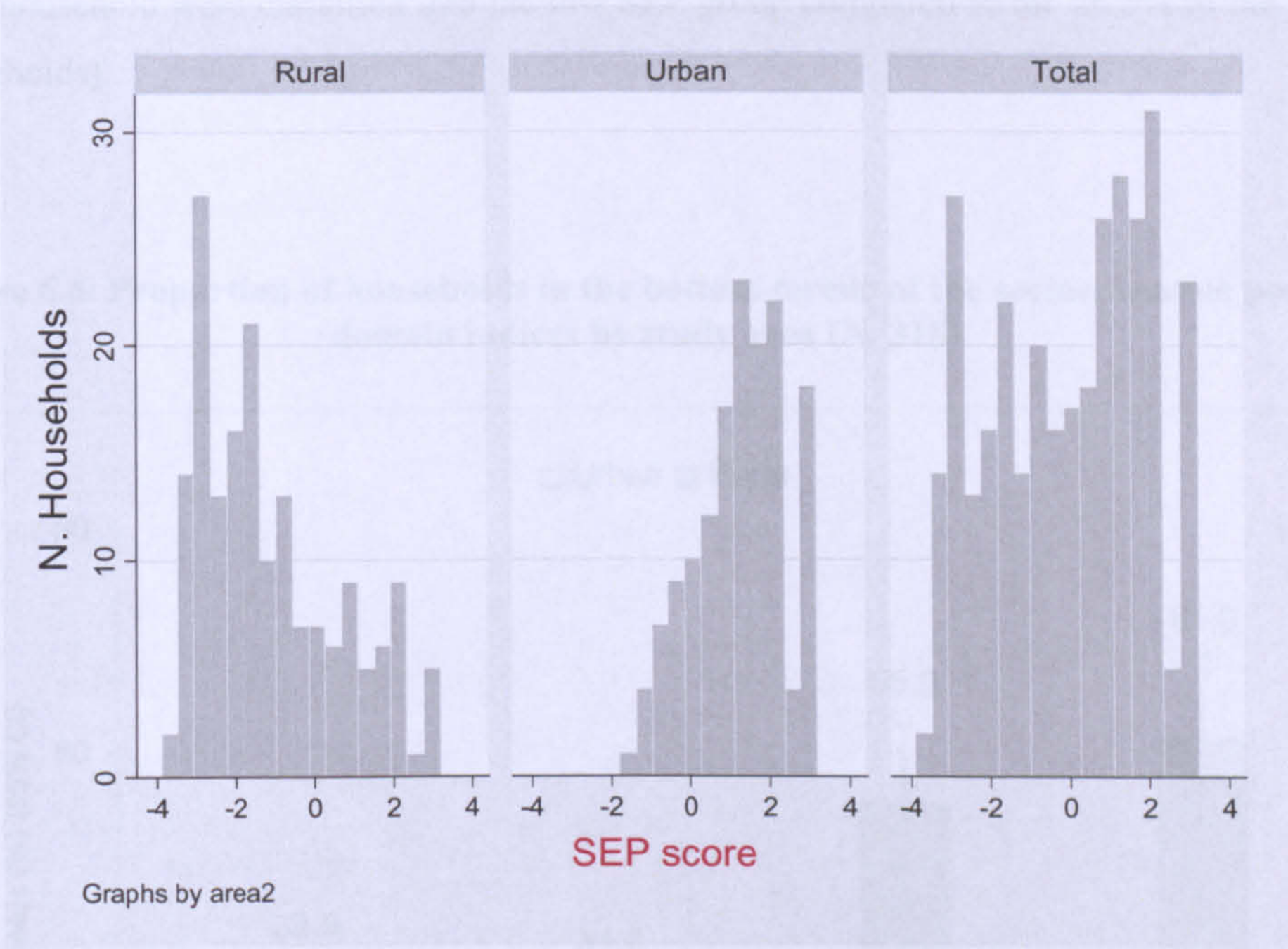
Figure 6.4: Median and range of the socioeconomic score across socioeconomic index terciles



Median = Red dots in the graph; Range = black crosses in the graph

In terms of overall score distribution, the SEP showed two peaks, identifying the rural households (on the left hand side of the graph) and the urban ones (on the right hand side) (**Figure 6.5**). When stratifying by area of residence, the SEP index appeared highly skewed toward low and high SEP respectively for the rural and urban area.

Figure 6.5: SEP score distribution, overall and by area of residence



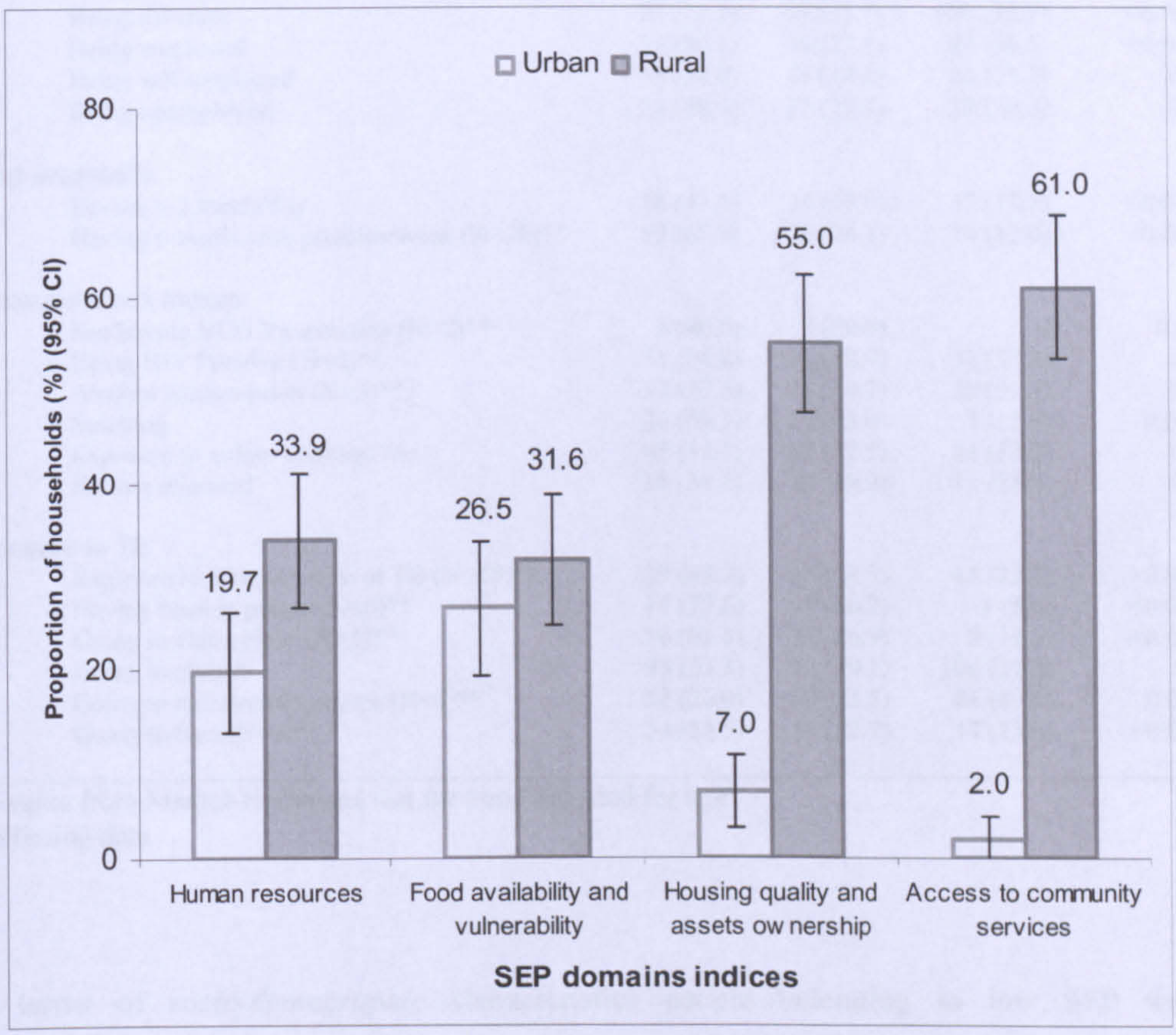
Rural households were disproportionately more likely than urban households to be classified as low SEP (95.4% vs 4.8% respectively, P<0.001), whereas the majority of the urban households occupied the top tercile (76.3% compared to 23.7% of rural households).

A similar pattern was observed when looking at the proportion of households classified as 'low' SEP in the urban and rural area according to the household SEP domain indices (**Figure 6.6**): rural households were significantly more likely to be classified in the low SEP group regardless of the SEP domain considered. The biggest difference between the rural and the urban area was observed for the access to community services domain (with respectively 2.0% and 61.0% of the urban and rural households in the bottom tercile) and the housing quality and assets ownership domain (with respectively 7.0% vs 55.0% of the urban and rural households in the bottom tercile).

Rural households were significantly more likely to be classified as 'low' SEP even when looking at the human resources index, but the gap with the urban households was less pronounced (with respectively 19.7% vs 33.9% of the urban and the rural household in the bottom tercile).

The food availability and vulnerability index was the only one for which the proportion of household in the bottom tercile across areas was not significantly different (31.6% of the rural household were classified into the low SEP group compared to the 26.5% of the urban households).

Figure 6.6: Proportion of households in the bottom tercile of the socioeconomic position domain indices by study area (N. 318)



6.2.3 Risk factors for prevalent tuberculosis

(a) The risk factors distribution across socioeconomic groups

The study of the risk factors for prevalent TB starts with the analysis of the distribution of these exposures across household SEP terciles. It is important to stress again that for the sake of representativeness, the analysis was restricted to the controls only. The results adjusted by age are outlined in **Table 6.5**.

Table 6.5: Tuberculosis risk factors distribution by SEP among the 318 controls

Risk factor	N. Socioeconomic group (%)			P value*
	Low	Medium	High	
Socio-demographic characteristics				
Being females	56 (31.6)	51 (28.8)	70 (39.6)	0.2
Being illiterate	77 (28.3)	89 (32.7)	106 (38.9)	<0.001
Being employed	31 (20.1)	36 (23.4)	87 (56.5)	<0.001
Being self-employed	36 (34.9)	41 (39.8)	26 (25.2)	0.8
Being unemployed	51 (40.5)	37 (29.4)	38 (30.2)	0.8
Food availability				
Having ≤ 2 meals/day	46 (47.4)	34 (35.0)	17 (17.5)	<0.001
Having 0 meals with proteins/week (N=38)**	53 (63.9)	20 (24.1)	10 (12.0)	<0.001
Behavioural risk factors				
Not having BCG Vaccination (N=2)**	8 (80.0)	2 (20.0)	0	0.01
Being HIV Positive (N=4)**	31 (34.8)	25 (28.9)	33 (37.1)	0.8
Alcohol consumption (N=3)**	33 (37.5)	27 (30.7)	28 (31.8)	0.2
Smoking	26 (59.1)	11 (25.0)	7 (15.9)	0.001
Exposure to indoor cooking fire	85 (34.1)	81 (32.5)	83 (33.3)	0.1
Having migrated	15 (34.1)	18 (40.9)	11 (25.0)	0.3
Exposure to TB				
Exposure to a known case of TB (N=45)**	29 (43.9)	22 (33.3)	15 (22.7)	<0.001
Having been in prison (N=6)**	14 (77.8)	3 (16.7)	1 (5.6)	<0.001
Going to video clubs (N=1)**	16 (61.6)	7 (26.9)	2 (11.5)	<0.001
Going to church	95 (33.8)	82 (29.2)	104 (37.0)	0.9
Going to hair dressing shops (N=1)**	52 (26.0)	67 (33.5)	81 (40.5)	0.003
Going to bars (N=1)**	34 (51.5)	15 (22.7)	17 (25.6)	<0.001

*P-value from Mantel-Haesenzel test for trend adjusted for age
**Missing data

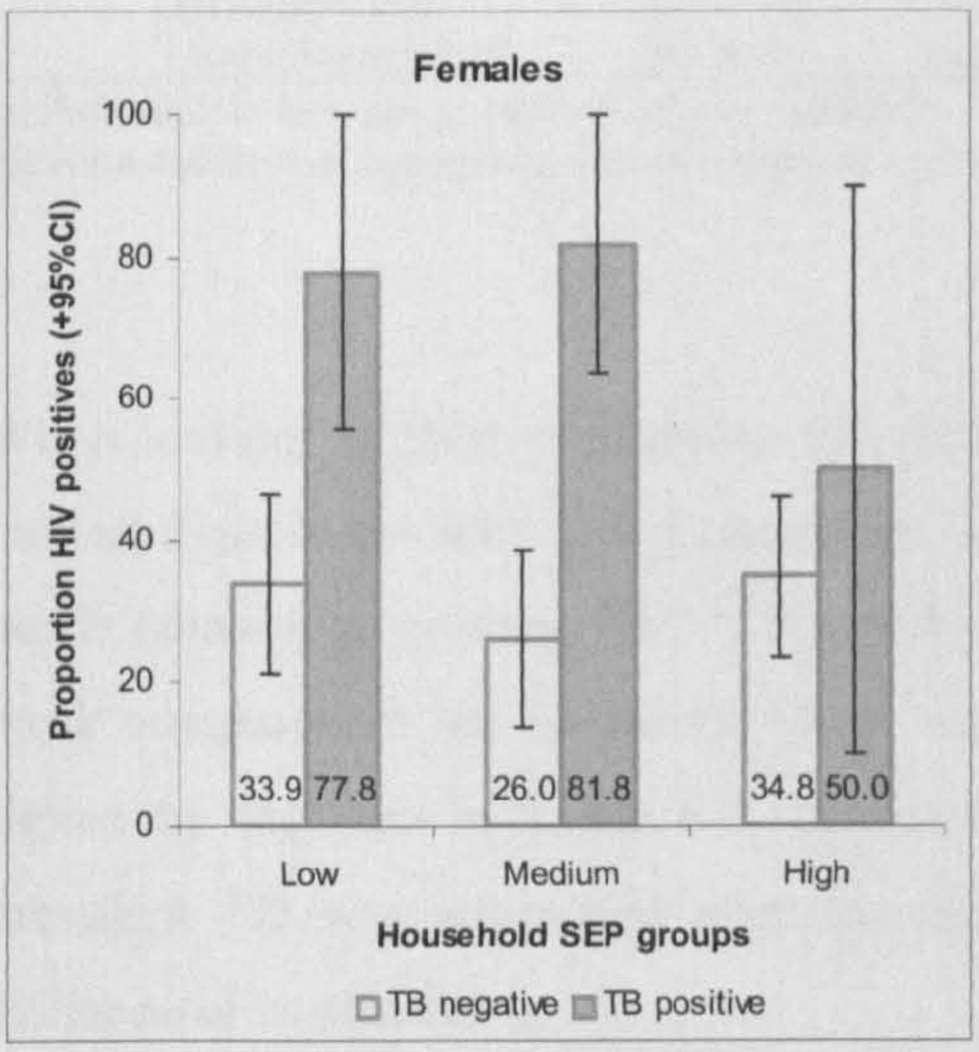
In terms of *socio-demographic* characteristics people belonging to low SEP were significantly more likely to be illiterate and less likely to have some kind of formal employment. In this population, female controls were not significantly more likely than male controls to be classified as Low SEP, suggesting that in this population SEP is gender-neutral (P=0.2).

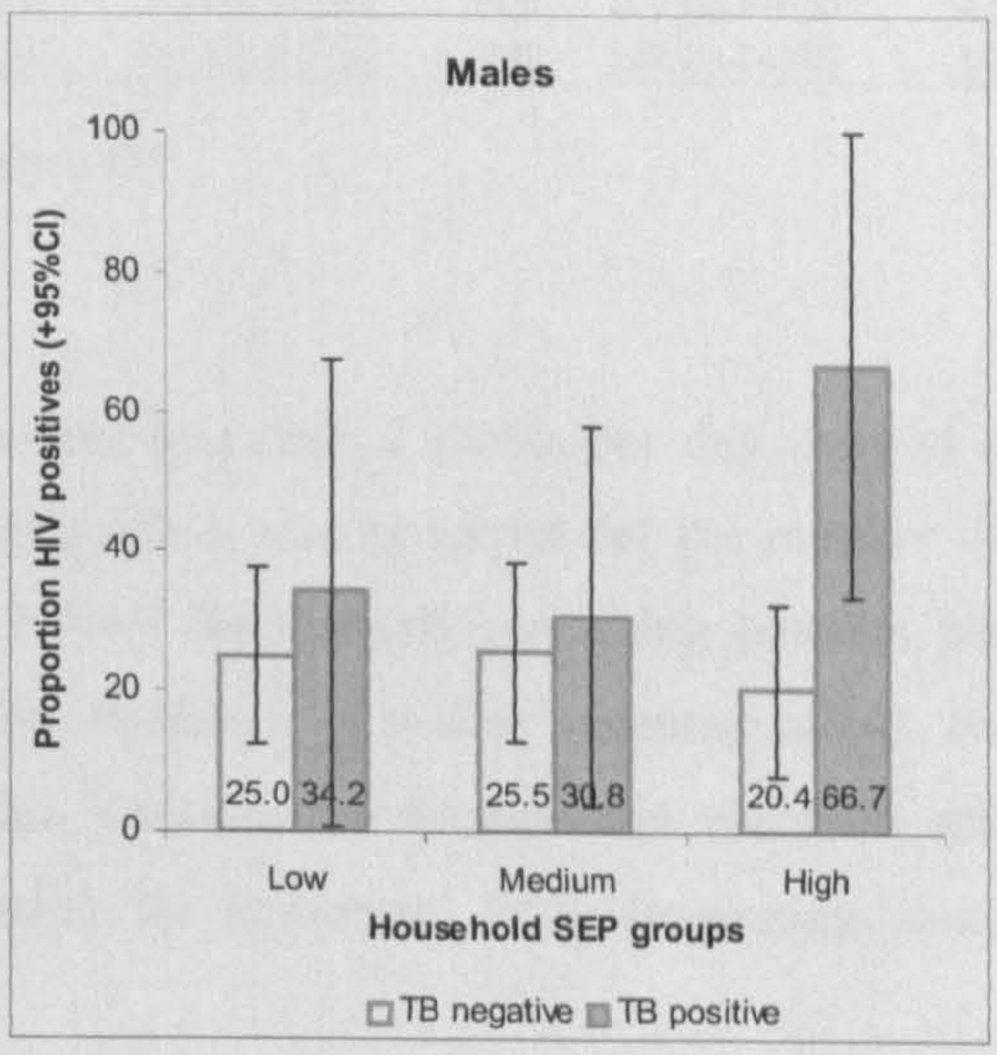
A significant trend was observed for both the *food-related risk factors* considered: low SEP people were more likely to have 2 or less than 2 meals per day and less likely to eat food containing proteins (P<0.001 for both variables explored).

Among the *biological-behavioural risk factors*, only smokers and people with no BCG vaccination were more likely to belong to low SEP households (P<0.001), By contrast, no socioeconomic gradient was observed for the HIV prevalence distribution or alcohol consumption. No evident trend between HIV and SEP groups was observed even when the analysis was repeated by TB status and separately for males and females (**Figure 6.7**).

In terms of *exposure to TB*, people attending video clubs, hair dressing shops and bars, to have been in prison and to have been exposed to a known case of TB were significantly more likely to have a low household SEP (P<0.001).

Figure 6.7: HIV frequency distribution by household SEP and TB status in males and females





(b) The risk factors distribution between cases and controls

The association between prevalent TB and the investigated risk factors was explored before and after adjustment for SEP. Because of the high number of risk factors, results have been grouped by type of exposure.

In terms of *socio-demographic characteristics*, being female was not significantly associated with prevalent TB. No association was also observed for any of the education/occupation risk factors considered **(Table 6.6).**

Table 6.6: Socio-demographic risk factors for prevalent tuberculosis

Socio-demographic factors	N. Cases (%)	N. Controls (%)	Adj OR[(a)] (95% CI)	P value	Adj OR[(b)] (95% CI)	P value
Being a female	24 (46.1)	177 (55.7)	0.7 (0.4-1.3)	0.3	0.8 (0.4-1.4)	0.4
Being illiterate	5 (9.6)	46 (14.5)	1.3 (0.5-3.8)	0.6	1.4 (0.5-3.8)	0.7
Highest grade achieved						
1-4	4 (8.2)	39 (13.4)	1.0		1.0	
5-7	24 (49.0)	111 (37.1)	1.8 (0.6-5.8)	0.3	2.1 (0.6-6.6)	0.2
8-9	15 (30.1)	8 (27.4)	1.6 (0.5-5.4)	0.4	1.8 (0.5-6.3)	0.3
10-12	3 (6.1)	56 (18.7)	0.4 (0.09-2.2)	0.3	0.7 (0.2-3.9)	0.7
College	3 (6.1)	11 (3.7)	1.9 (04-10.0)	0.5	3.6 (0.6-21.2)	0.1
Economic status						
Employed	16 (31.0)	89 (28.0)	1.0		1.0	
Self-employed	17 (32.7)	103 (32.4)	1.0 (0.4-2.0)	0.9	0.7 (0.3-1.5)	0.3
Unemployed/other	19 (36.5)	126 (39.6)	1.2 (0.6-2.6)	0.6	0.9 (0.4-2.0)	0.8

(a) Adjusted by sex, age group and area of residence
(b) Adjusted by sex, age group, area of residence and household SEP

When looking at *food availability* variables, having less than 2 meals per day showed a modest association with TB. Conversely, a strong effect was observed for the number of meals containing proteins (OR = 3.1, 95%CI: 1.1-8.7 for 0 meals containing proteins per week compared to the baseline). There was little evidence of a dose response effect. As shown by the data in **Table 6.7**, the association between the food-related variables and prevalent TB was attenuated after the adjustment for household SEP, suggesting some evidence of confounding.

Table 6.7: Food availability related risk factors for prevalent tuberculosis

Food availability factors	N. Cases (%)	N. Controls (%)	Adj OR[(a)] (95% CI)	P value	Adj OR[(b)] (95% CI)	P value
≤ 2 meals/day	23 (44.2)	97 (30.5)	1.8 (0.9-3.4)	0.07	1.3 (0.6-2.6)	0.5
N. meals with proteins/week						
> 2	6 (11.5)	80 (25.2)	1.0		1.0	
2	6 (11.5)	61 (19.2)	1.3 (0.4-4.3)	0.6	1.1 (0.3-3.6)	0.9
1	24 (46.1)	94 (29.6)	3.8 (1.4-10.0)	0.006	2.7 (1.0-7.4)	0.05
0	16 (30.8)	83 (26.1)	3.1 (1.1-8.7)	0.03	2.0 (0.6-6.0)	0.2

(a) Adjusted by sex, age group and area of residence
(b) Adjusted by sex, age group, area of residence and the composite measure of household SEP

Among the *biological-behavioural risk factors* (**Table 6.8**), prevalent TB was strongly associated with the lack of BCG vaccination as well as with HIV infection. The association with lack of BCG vaccination was attenuated after adjusting for household SEP, whereas the association with HIV was pretty much unaffected by the adjustment for household SEP. The risk of prevalent TB was also significantly increased in people consuming more than 3 drinks containing alcohol every time he/she drinks. However, this association was no longer significant after controlling for household SEP. Having migrated somewhere in the 5 years prior the interview was also significantly associated with prevalent TB and this association persisted after the adjustment for household SEP. Both cigarette smoking and the exposure to indoor smoking pollution were not significantly associated with prevalent TB.

Table 6.8: Biological-behavioural risk factors for prevalent tuberculosis

Behavioural factors	N. Cases (%)	N. Controls (%)	Adj OR[(a)] (95% CI)	P value	Adj OR[(b)] (95% CI)	P value
Not having BCG (N.6)*	9 (18.7)	10 (3.2)	7.7 (2.8-20.8)	<0.001	6.1 (2.2-17.1)	0.001
Being HIV positive (N.4)*	29 (55.8)	89 (28.3)	3.1 (1.7-5.8)	0.001	3.2 (1.5-7.2)	<0.001
Alcohol abuse†	23 (44.2)	88 (27.7)	1.8 (1.0-3.4)	0.05	1.6 (0.9-3.1)	0.1
Cigarette smoking	10 (19.2)	44 (13.8)	1.5 (0.7-3.5)	0.3	1.5 (0.6-3.3)	0.4
Indoor smoking pollution	44 (84.6)	249 (79.0)	1.4 (0.6-3.1)	0.4	1.2 (0.5-2.7)	0.7
Migration‡	21 (40.4)	44 (13.8)	5.2 (2.7-10.2)	<0.001	5.3 (2.7-10.7)	<0.001

*Missing values
(a) Adjusted by sex, age group and area of residence
(b) Adjusted by sex, age group, area of residence and household SEP
†Drinking more than 3 drinks containing alcohol every time that he/she drinks
‡ Migration was defined as having lived anywhere else for more than 6 months in the 5 years before the interview

Finally, among the *TB exposure* related variables, having been in contact with anyone with TB in the 12 months prior the interview showed a significant association with TB. There was some evidence that this association was confounded by household SEP. Going to church showed a significant protective effect on the risk of prevalent TB. The attendance at any of other venues enquired was not associated with prevalent TB (**Table 6.9**).

Table 6.9: TB exposure related risk factors for prevalent tuberculosis

TB exposure factors	N. Cases (%)	N. Controls (%)	Adj OR[(a)] (95% CI)	P value	Adj OR[(b)] (95% CI)	P value
Known contact with TB case (N. 48)*	19 (38.8)	66 (24.2)	2.8 (1.3-5.6)	0.005	2.4 (1.2-5.0)	0.01
Attending						
Video clubs (N.1)*	3 (5.8)	26 (8.2)	1.0 (0.3-3.6)	1.0	0.9 (0.2-3.2)	0.8
Bars (N.2)*	12 (23.5)	66 (20.8)	1.2 (0.6-2.4)	0.7	1.1 (0.6-2.3)	0.8
Hairdressing shops (N.1)*	35 (67.3)	200 (63.1)	1.1 (0.6-2.1)	0.8	1.1 (0.6-2.2)	0.8
Churches	38 (73.1)	281 (88.4)	0.4 (0.1-0.8)	0.006	0.4 (0.2-0.8)	0.01

*Missing values
(a) Adjusted by sex, age group and area of residence
(b) Adjusted by sex, age group, area of residence and household SEP.

6.2.4 The association between household socioeconomic position and prevalent tuberculosis

(a) Socioeconomic profile of cases and controls

As for the comparison of household urban and rural living conditions, the household socioeconomic profile of the cases and the controls has been or grouped by household SEP domain (**Table 6.10**).

In terms of *household human resources*, cases were more likely than controls to have their head of the household unemployed (OR= 3.3, 95%CI: 1.3-8.5, P=0.01). No other significant difference was observed in terms of household human resource.

Cases and controls showed major differences when looking at the household *food availability and vulnerability* – related variables (**Table 6.10**). Approximately 50% of the cases lived in households normally having 2 or less than 2 meals per day compared to 32.1% of the controls (OR=1.8, 95%CI: 1.0-3.4, P=0.05). 31% of the cases belonged to households reporting no weekly consumptions of meals containing proteins compared to approximately 24.2% of the controls (OR=4.5, 95%CI: 1.5-13.5, P<0.001). Compared to controls (16.2%), cases (38.3 %) were also more likely to live in households reporting three or more than 3 months with not enough to eat (OR= 4.2, 95%CI: 2.0-8.9, P < 0.001) and the implementation of more than 3 coping strategies to react to hardship in the 12 months prior the interview (57.7% of the households of cases compared to 37.4% of the control households, OR= 2.5, 95%CI: 1.2-5.2, P=0.001).

Households of cases and controls appeared different in terms of household *assets ownership*, however – as shown in **Table 6.10** - the number of assets was more important than the type of asset owned: only not having a radio was significantly associated with prevalent TB (OR

= 2.1, 95% CI: 1.1-4.0, P = 0.02), but cases were more likely to have less than two assets compared to the controls (OR = 2.0, 95%CI: 1.0-3.7, P = 0.03). In terms of *housing quality*, cases were more likely than controls to live in houses with walls made with mud bricks (OR= 2.2, 95%CI: 1.1-4.5, P=0.03), with no electricity (OR=3.0, 95%CI: 1.3-3.7, P<0.001) and with private piped water (OR=3.5, 95%CI: 1.4-9.0, P=0.08). There was some evidence that they were also more likely to live in households with floor made of dirt (OR = 1.8, 95%CI: 0.9-3.5, P = 0.07). Cases and controls did not differ in terms of household crowding. Cases and controls did not show any significant difference in terms of distance from the facilities investigated.

(b) The association with the household SEP index

The minimally adjusted analysis (i.e. adjusting for sex, age group and area of residence) revealed that cases were significantly more likely to be in the low and medium SEP group than controls (OR= 6.2, 95%CI: 1.2-19.2 and OR= 3.4, 95%CI: 1.5-7.6 respectively for the low and medium SEP group compared to the baseline; overall P < 0.001) (**Table 6.11**).

This association was also observed in the urban area (OR = 4.0, 95%CI: 0.8-23.4 and OR=3.8, 95%CI: 1.6-9.8 for the low and medium SEP group compared to the baseline), but not in the rural area (OR = 3.8, 95%CI: 0.5-30.0 and 1.2, 95%CI: 0.1-13.7 for the low and medium SEP group compared to the baseline). There was no evidence of interaction between area and household SEP (Test for interaction, P=0.5) (**Table 6.11**).

When analysed as binary variable, household SEP was still associated with prevalent TB with the cases 2.7 times more likely to fall in the low SEP category compared to the controls after adjusting for sex, age group and area of residence. The association was still modestly significant in the urban area (OR = 2.9, 95%CI: 0.9-5.0), but not in the rural area (OR = 5.6, 95%CI: 0.7-43.8) (**Table 6.11**).

The linear regression model including the household SEP index as continuous variable (together with age, sex and area of residence) showed a 3.5% increase in the risk of prevalent TB for each unit decrease of the household SEP score (Regression coefficient: -0.035, 95%CI: -0.06;-0.01, P = 0.05). The percentage of increase was even higher when the analysis was restricted to the urban area (6.5% increase), whereas only a modest association was observed for the rural area (**Table 6.11**).

Table 6.10: Household socioeconomic profile of cases (N. 52) and controls (N. 318)

Household SEP domain	Cases	Controls		
	N. (%)	N. (%)	OR† (95%CI)	P value
Human resources				
Female head of the household	17 (32.7)	98 (30.8)	1.1 (0.6-2.2)	0.7
> 3 children under 5 years of age	12 (23.1)	110 (34.6)	0.6 (0.3-1.1)	0.09
Head of the household literate	47 (90.4)	275 (86.8)	0.9 (0.3-2.9)	0.9
Head of the household activity				
Employed	20 (38.5)	154 (48.4)	1.0	
Self-employed	23 (44.2)	139 (43.7)	1.5 (0.7-2.8)	0.3
Unemployed	9 (17.3)	25 (7.9)	3.3 (1.3-8.5)	0.01
Food availability and vulnerability				
N. meals/day				
> 2	27 (51.9)	216 (67.9)	1.0	
≤ 2	25 (48.1)	102 (32.1)	1.8 (1.0-3.4)	0.05
N. meals with proteins/week				
> 2	5 (9.6)	83 (26.1)	1.0	
2	7 (13.5)	65 (20.4)	1.8 (0.5-5.8)	0.3
1	24 (46.1)	93 (28.5)	4.7 (1.7-13.0)	0.003
0	16 (30.8)	77 (24.2)	4.5 (1.5-13.5)	0.007
N. months with not enough to eat/year				
0	17 (36.2)	189 (62.4)	1.0	
1-3	12 (25.5)	65 (21.4)	2.4 (1.1-5.4)	0.03
> 3	18 (38.3)	49 (16.2)	4.2 (2.0-8.9)	< 0.001
N. coping strategies/year				
0	13 (25.0)	101 (31.8)	1.0	
1-3	8 (17.3)	98 (30.8)	0.9 (0.3-2.2)	0.8
> 3	30 (57.7)	119 (37.4)	2.5 (1.2-5.2)	0.01
Housing quality and assets ownership				
Having < 2 assets	32 (61.4)	157 (49.4)	2.0 (1.0-3.7)	0.03
Not having a fridge (N.1)*	45 (86.5)	273 (86.2)	1.3 (0.5-3.0)	0.6
Not having a TV	33 (63.5)	190 (59.7)	1.4 (0.7-2.7)	0.3
Not having a radio	19 (36.5)	78 (24.5)	2.1 (1.1-4.0)	0.02
Not having a bicycle	40 (76.9)	212 (66.7)	1.3 (0.5-3.0)	0.6
Not having animals	13 (25.0)	133 (41.8)	1.7 (0.8-3.9)	0.2
Not having land (N. 52)*	8 (38.1)	49 (24.9)	2.1 (0.5-8.8)	0.3
N. people per bedroom				
≤ 2	25 (48.1)	171 (53.8)	1.0	
3-4	14 (26.9)	78 (24.5)	1.2 (0.6-2.5)	0.5
> 4	13 (25.0)	69 (21.7)	1.5 (0.7-3.0)	0.3

Table 6.10 – Continued

Household SEP domain	Cases	Controls		
	N. (%)	N. (%)	OR (95%CI)	P value
Walls material				
Concrete	16 (30.8)	115 (36.2)	1.0	
Mud bricks	31 (59.6)	134 (42.1)	2.2 (1.1-4.5)	0.03
Burnt bricks	5 (9.6)	69 (21.7)	0.8 (0.3-2.4)	0.7
Roof material				
Asbestos	21 (40.4)	107 (33.6)	1.0	
Iron	22 (42.3)	139 (43.7)	1.0 (0.5-2.0)	0.9
Grass	9 (17.3)	72 (22.6)	1.0 (0.3-3.1)	0.9
Floor material				
Cement	25 (48.1)	175 (55.0)	1.0	
Dirt/Earth	27 (51.9)	143 (44.9)	1.8 (0.9-3.5)	0.07
Not having electricity	45 (86.5)	233 (73.3)	3.0 (1.3-3.7)	0.01
Having latrines	48 (92.3)	268 (84.3)	2.0 (0.7-5.8)	0.2
Water supply				
Piped inside the house	6 (11.5)	77 (24.9)	1.0	
Piped outside the house	46 (88.5)	233 (75.2)	3.5 (1.4-9.0)	0.08
Distance from community services				
> 15 minutes walking from water	9 (17.3)	63 (19.8)	0.9 (0.4-2.1)	0.9
> 30 minutes walking from clinic	31 (59.6)	168 (52.8)	1.7 (0.9-3.2)	0.1
> 30 minutes walking from transport	38 (73.1)	197 (61.9)	1.3 (0.6-2.8)	0.6
> 30 minutes walking from market	31 (59.6)	207 (65.1)	0.9 (0.5-1.7)	0.7

†Adjusted by sex, age group and area of residence

*Missing values

(c) The multivariable analysis of household socioeconomic position and prevalent tuberculosis

The multivariable analysis was driven by the conceptual framework presented in **Figure 4.8.** For the multivariable analysis, the SEP index has been categorised into "Low" and "High" SEP. As shown in **Table 6.12**, only the inclusion of food intake-related variables caused a major reduction in the magnitude of the odds ratio showing some evidence of mediation effect of these variables on the association between SEP and TB (OR = 1.8, 95% CI: 0.7-4.2 after mediation, P= 0.2).

The inclusion of TB exposure related variable caused a reduction in the OR of SEP as well. However, the reduction was less prominent and SEP was still significantly associated with prevalent TB after mediation (OR= 2.3, 95%CI: 1.1-5.2, P=0.04). There was no evidence that the association between household SEP and prevalent TB was mediated by education and occupation variables (OR=2.7, 95%CI: 1.2-6.1, P=0.01) as well as the biological-behavioural related variables (OR = 2.6, 95%CI: 1.1-6.3, P=0.04). This was observed also when HIV infection status was included alone in the model.

Table 6.11: Household socioeconomic position and prevalent TB – The multivariable analysis results

	SEP index adjusted OR (95% CI)	P-value
Model 1		
High household SEP	1.0	0.01
Low	2.7 (1.2-5.9)	
Model 2 – Education/Occupation		
High	1.0	0.01
Low	2.7 (1.2-6.1)	
Model 3 – Food availability/vulnerability		
High	1.0	0.2
Low	1.8 (0.7-4.2)	
Model 4 – Biological – behavioural risk factors		
High	1.0	0.04
Low	2.6 (1.1-6.3)	
Model 5 – TB exposure		
High	1.0	0.04
Low	2.3 (1.1-5.2)	

Model 1: SEP adjusted by sex, age group and area of residence (minimally adjusted)
Model 2: SEP adjusted by sex, age group, area of residence, education and occupation
Model 3: as in model 2 plus food intake related variables
Model 4: as in model 2 plus behavioural risk related variables
Model 5: as in model 2 including TB exposure related variables

Table 6.12: Household socioeconomic position and prevalent TB – results from the minimally adjusted analysis, overall and by area of residence

	Overall				Urban area				Rural area			
	Cases N (%)	Controls N. (%)	OR (95%CI)	P value	Cases N. (%)	Controls N. (%)	OR (95%CI)	P value	Cases N. (%)	Controls N. (%)	OR (95%CI)	P value
Categorical variable*												
Household SEP												
Low	18 (34.6)	105 (33.0)	6.2 (2.0-19.2)	0.002	2 (6.2)	5 (3.4)	4.7 (0.7-29.4)	0.09	16 (84.2)	100 (58.5)	3.9(0.5-31.1)	0.2
Medium	24 (46.1)	99 (31.1)	3.4 (1.5-7.6)	0.003	22 (66.7)	55 (37.4)	3.7(1.6-8.8)	0.02	2(10.5)	44 (25.7)	1.2 (0.1-13.9)	0.8
High	10 (19.2)	114(35.0)	1.0		9 (27.3)	87 (59.2)	1.0		1 (5.3)	27 (15.8)	1.0	
Low	29 (55.8)	155 (48.7)	2.7 (1.2-5.5)	0.01	11 (33.3)	27 (18.4)	2.9 (0.9-5.0)	0.09	18 (98.7)	128 (74.8)	5.6 (0.7-43.8)	0.1
High	23 (44.2)	163 (51.3)	1.0		22 (66.7)	120 (81.6)	1.0		1 (5.3)	43 (25.2)	1.0	
Continuous variable*												
Household SEP	Regression coefficient** (95%CI)			P value	Regression coefficient** (95%CI)			P value	Regression coefficient** (95%CI)			P value
Low	-0.035 (-0.06; -0.01)			0.005	-0.065 (-0.11; - 0.01)			0.01	-0.022 (- 0.05; -0.003)			0.08

*Household SEP variable analysed as categorical (both in tercile and dichotomous format) and continuous.
** Regression coefficients from linear regression controlling for sex, age and area of residence

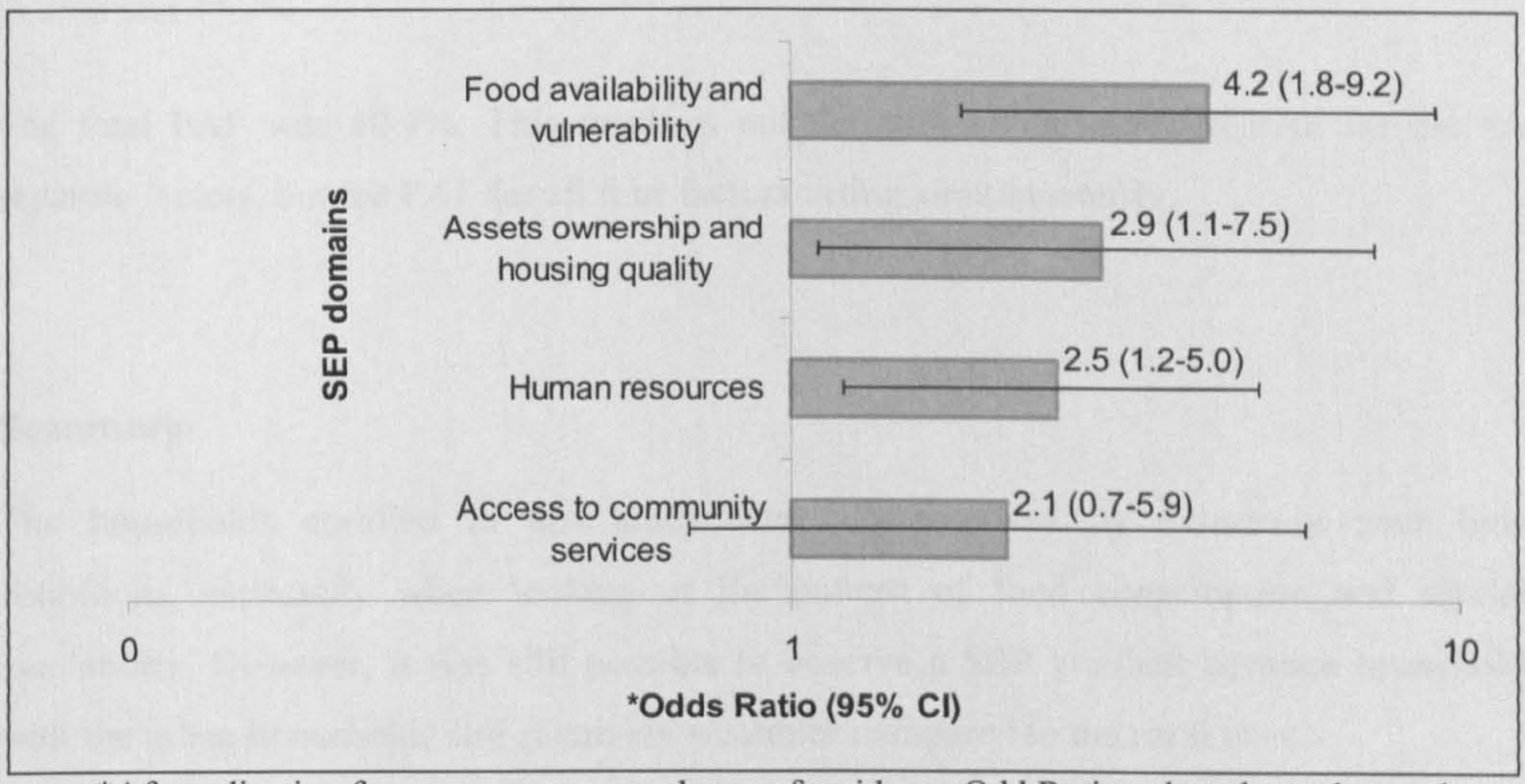
(c) The association between household SEP domain indices and prevalent TB

In **Figure 6.9** SEP domains are ranked by the strength of their association with prevalent TB.

All the indices of SEP domains were significantly associated prevalent TB, expect the one about access to community services. The odds of being in the lower SEP group of the food availability and vulnerability index were 4 times greater for the TB cases than the controls. This SEP domain showed the strongest association with prevalent TB, followed by the assets ownership and housing quality domain (OR = 2.9, 95%CI: 1.1-7.5), the human resources (OR = 2.5, 95%: 1.2-5.0) one and access to community services (OR = 2.1, 95%CI: 0.7-5.9).

When included in a multivariable model, together with sex, age group and area residence, only the food availability and vulnerability domain as well as the human resources one remained significantly associated with prevalent TB (respectively OR=2.9, 95%CI: 1.4-6.0 and OR=1.9, 95%CI: 1.0-3.7). Instead, the association between housing quality domain and prevalent TB became not significant with an OR = 1.1 (95%CI: 0.5-2.4).

Figure 6.8: The association between indices of SEP domains and prevalent TB



*After adjusting for sex, age group and area of residence Odd Ratios plotted on a log scale.

6.2.5 The population attributable fraction

The PAF for the household SEP index was 30% (95%CI: 5-47.5%).

The SEP index was then included in a multivariable model together with sex, age group, are of residence and all the other identified risk factors for prevalent TB, including having 2 or less than 2 meals per week containing proteins, not having BCG vaccination, being HIV positive and having migrated somewhere in the 5 years prior the interview and having been in contact with someone with TB. Attending church was not included in the analysis as it showed a protective effect.

The model was repeated excluding variables with $P > 0.05$ and the PAF was estimated only on the variables left in the model using the `aflogit` command of STATA 9.0. The adjusted OR (95%CI) used for the PAF computation and the resulting PAF values are presented in **Table 6.13.** Despite showing the strongest effect, not having BCG vaccination exhibited the smallest PAF (11.2%) because of the low frequency of unvaccinated people in this population. Apparently 23.4% of the cases in this setting could be attributed to migration. In this population the highest proportion of TB cases could be attributed to HIV and the weekly number of meals containing proteins with an adjusted PAF respectively equal to 35.8 and 41.7%.

The total PAF was 80.4%. This result is not the sum of the adjusted PAF for the four separate factors, but the PAF for all four factors acting simultaneously.

Summary

The households enrolled in this study were characterised by extremely poor living conditions, especially when looking at the pattern of food consumption and services availability. However, it was still possible to observe a SEP gradient between households, with the urban households still relatively wealthier compared to the rural ones.

In this population, lack of BCG vaccination, HIV positivity, migration and having been in contact with a TB case were risk factors for prevalent TB even after controlling for the possible confounding effect of household SEP. Attending regularly a church showed a significant protective effect against prevalent TB even after controlling for household SEP.

Table 6.13: Population attributable fraction

	% Exposed among cases	Adj.* OR (95% CI)	P-value	Adj**. PAF	95% CI
Household SEP[§]					
Low	55.8	2.6 (1.2-5.5)	0.01	29.3%	(0.05-47.4)
High	44.2	1.0			
Weekly N. meals + proteins					
0	30.8	2.7 (0.9-8.5)	0.08	41.7%	3.0-64.6
1	46.2	3.2 (1.2-9.1)	0.02		
2	11.5	0.9 (0.2-3.4)	0.9		
>2	11.5	Ref			
Not having BCG	18.8	5.8 (1.8-18.6)	0.03	11.2%	2.0-19.0
Being HIV positive	55.8	3.9 (1.9-7.0)	<0.001	35.8%	15.3-51.4
Migration	40.4	4.2 (1.9-9.3)	<0.001	23.4%	9.0-35.6
Total				**80.4%**	**57.2-91.0**

[§] PAF for household SEP estimated before mediation. The OR is derived from minimally adjusted analysis (i.e. controlling for sex, age group and area of residence).

* The model includes all the variables included presented in the table, plus sex, age group and area of residence. In the multivariable analysis the effect of household SEP was attenuated to non significance level because of the inclusion of the food-related variables (i.e. mediation effect). As a result, it was no longer possible estimating PAF for this variable.

**PAF estimates derived after controlling for the reciprocal confounding effect of all the variables included in the model.

This study showed a strong association between household SEP and prevalent TB, with the TB cases being approximately 3 times more likely than controls to belong to low SEP households in the minimally adjusted analysis. This result was confirmed also when using household SEP as continuous variable: each unit decrease of the household SEP score was associated with a statistically significant 3.5% increase in the risk of prevalent TB.

There was some evidence that this association was partly mediated by food availability as suggested by the substantial reduction in the size and level of significance of the OR between household SEP and TB after the inclusion of food related variables. There was some evidence of mediation even when the TB contact pathway was tested, but this effect was less prominent than the one played by food-related variables. In this thesis most of the biological-behavioural related variables, including HIV, were not associated with SEP so it is not surprising that effect of SEP resulted not mediated by these risk factors.

The role of food related variables was confirmed when looking at the association between SEP domain indices and prevalent TB: the food availability domain was the one showing the strongest association with TB even when included in a multivariable model together with the other domain.

The PAF analysis showed that before mediation household SEP could explain approximately 30% of the cases recruited in this study. Among the other exposures significantly associated with TB, the highest PAF estimates were observed for HIV (PAF = 36%) and lack of adequate intake of meals containing proteins explain (PAF = 42%). These figures derive from a multivariable model developed for the estimates of adjusted PAF for multiple risk factors and suggest that after controlling for the reciprocal confounding effect of these exposures, HIV and inadequate food intake were explaining most of the cases recruited in this study.

7. The Infection study results

Most of the new recruits to the industrial labor had known poverty in their former rural surroundings, but their life there had been relatively free from stresses and physiological hardship. [...] When they moved into industrial areas in search of prosperity, adventure and comfort, they found instead exploitation and other forms of poverty. [...] Intense crowding in workshops and in unsanitary living quarters provided all that was required for the rapid spread of infection...

Dubos, R[8]. 1987

Introduction

The aim of this chapter is to illustrate the association between household SEP and TB infection.

Since I used the same SEP indicator adopted in the main case control study, this chapter will also provide insights on the mechanism through which household SEP may affect the risk of TB infection compared to TB disease. The results of this part of the thesis dissertation have been recently published in the June issue of the American Journal of Hygiene and Tropical Medicine (**Appendix F**).

7.1 Chapters objectives

1. To describe the TB infection status of the cases and controls recruited for this thesis **(Section 7.2.1)**.

2. To compare the main socio-demographic characteristics of the individuals recruited in the cross sectional study with those who were excluded to assess any potential selection bias **(Section 7.2.2)**

3. To describe the risk factors for TB infection **(Section 7.2.3)**

4. To describe the socioeconomic profile of a sample of TB-infected disease-free people living in two communities in Zambia and quantify the association with the association with the household SEP index **(Section 7.2.4)**.

5. To assess the role of confounding and mediating factors on the association between household SEP and TB infection and to highlight differences compared to the effect on TB disease (**Section 7.2.4)**.

6. To explore which of the 4 domains of household SEP is mostly associated with the risk of TB infection (**Section 7.2.4**).

7.2 Results

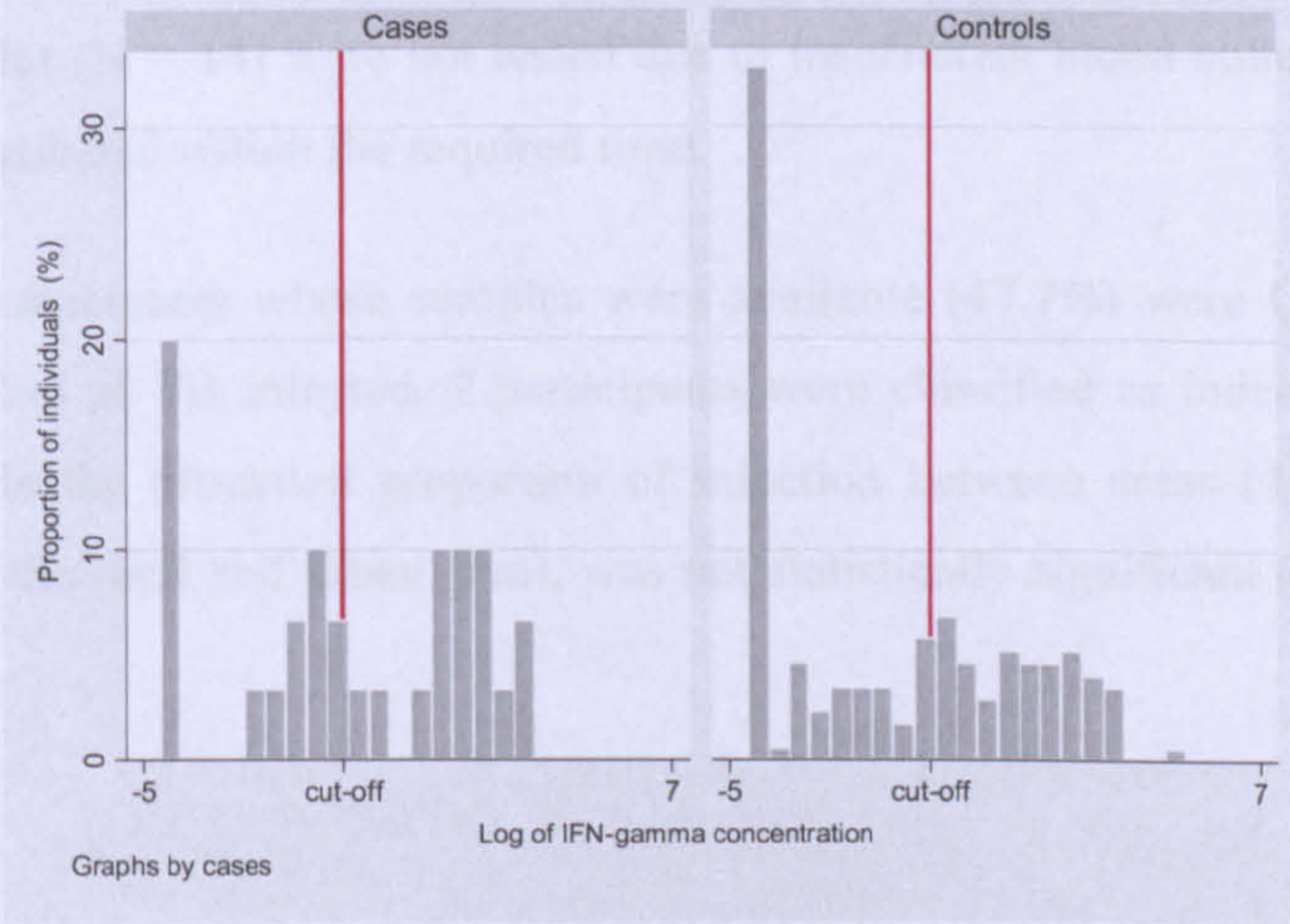
7.2.1 Tuberculosis infection status among cases and controls

The TB infection status has been analysed both quantitatively and qualitatively for cases and controls.

The QFN quantitative results are outlined in **Figure 7.1**, showing the distribution of the concentration of interferon-γ produced both from cases and controls. The red line indicates the concentration cut-off below which samples are classified as QFN negative. Compared to the QFT negative cases, the QFT negative controls did not appear to produce a significantly lower concentration of interferon gamma (P=0.5, from Mann Whitney test for the median concentration of interferon-γ produced by cases and controls).

The tall bars at the left-hand side of both graphs represent those individuals whose interferon-gamma production was below the detection capacity of the test.

Figure 7.1: Interferon-γ concentration production in cases and controls



The qualitative results will be discussed in the following sections separately for cases and controls:

(a) Infection status of the cases

Among the 52 confirmed cases, 45 people gave consent to blood testing (86.5%). Of them, only 37 (82.2%) provided a blood sample and QFT results were available only from 30 of these subjects (81.1%). 7 samples could not be tested because of insufficient blood collection (in 5 cases) and because of mislabelling of the blood tubes (in 2 cases). Among the 30 blood samples available, 19 resulted QFT positive (63.3%) and 11 were QFT negative (36.7%) **(Figure 7.2)**.

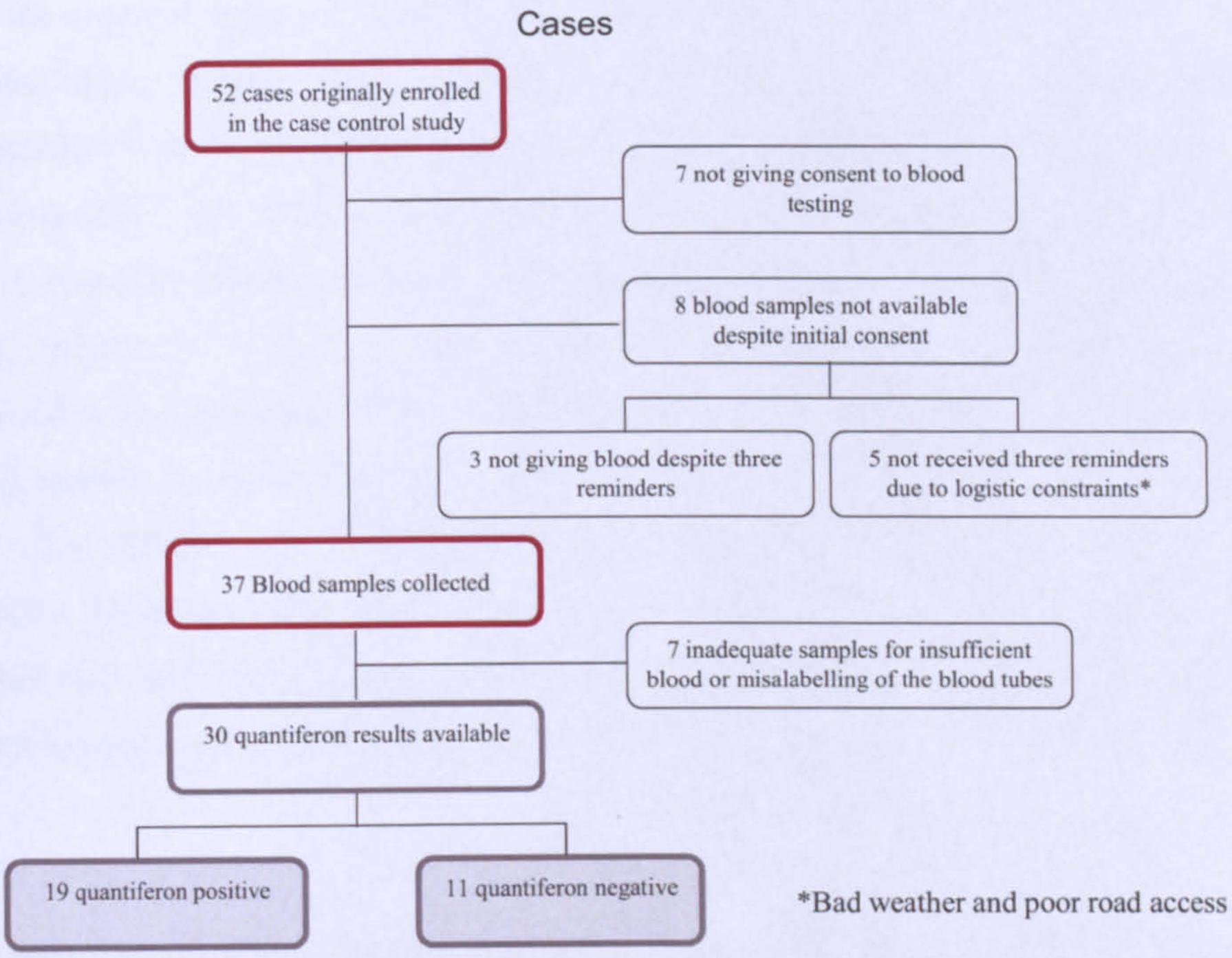
No attempt has been made to analyse the difference between the positive and the negative cases for the following reasons: firstly, the cases with QFT negative results are likely to be due to the sensitivity characteristics of the test, rather than biological features of these individuals. This is obviously a limitation of the test and it will be discussed further in the discussion. Secondly, given the small number of cases with a QFT result any further analysis of the potential differences between QFT positive and negative cases would have limited utility.

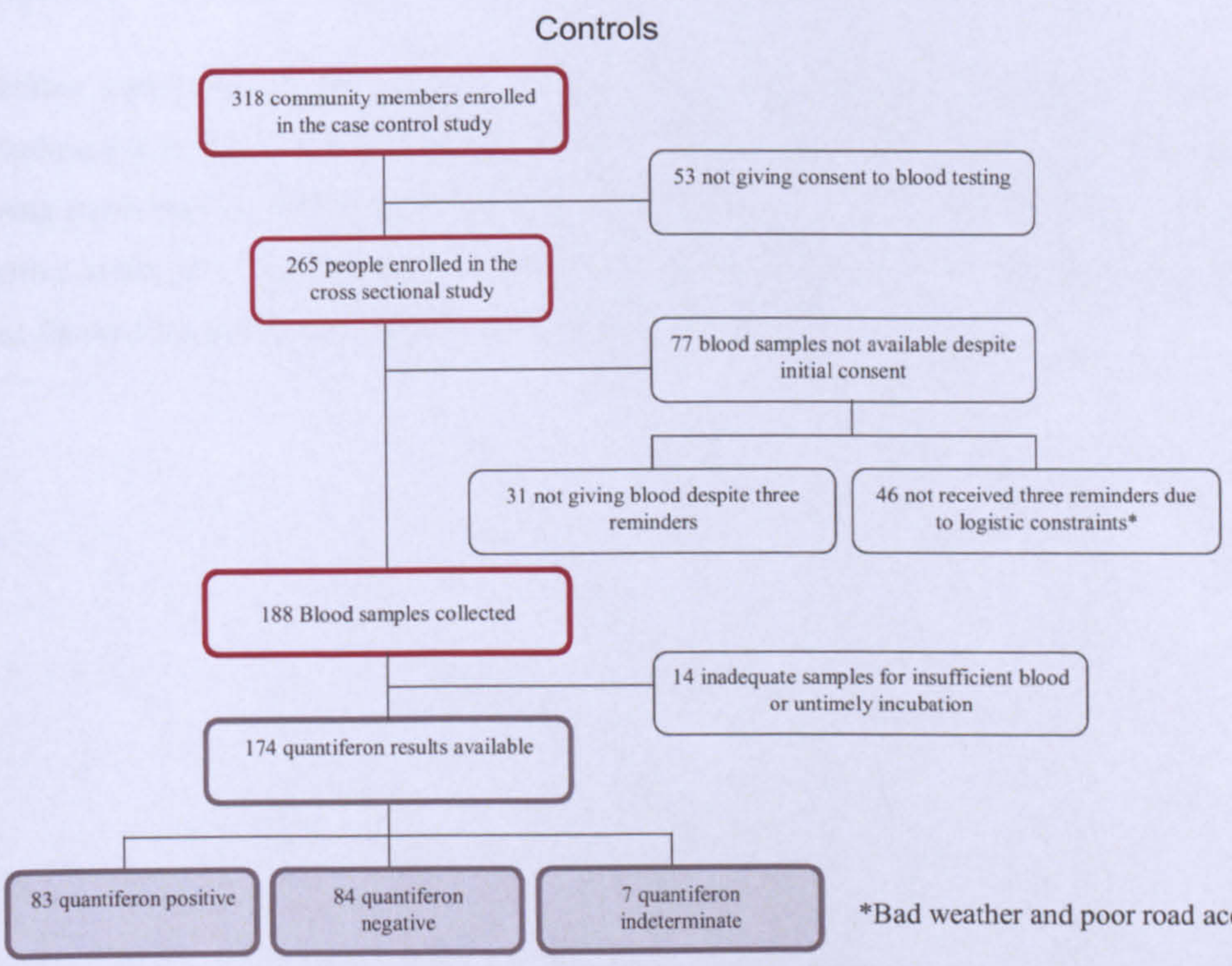
(b) Infection status of the controls

Consent to blood testing was given by 265 out of the 318 participants (83.3%) originally enrolled as controls in the case control study. Blood samples were collected from 188 of these subjects and QFT testing performed on 174 of the 188 (92.4%) available samples. The remaining samples (N = 14) were not tested due to insufficient blood collection or because they were not incubated within the required time.

83 out of 174 participants whose samples were available (47.7%) were QFT positive and therefore classified as TB infected. 7 participants were classified as indeterminate (4.0%). The difference in the estimated proportion of infection between areas (45.9% and 51.9% respectively for the rural and urban area), was not statistically significant (P = 0.5) (**Figure 7.2**).

Figure 7.2: Flowchart of the assessment of tuberculosis infection





7.2.2 Population recruited in the cross sectional study

Only the controls originally enrolled in the case control study were included in the cross sectional study (**Figure 7.2**). Compared to the controls for whom blood samples were unavailable (N=144), the subjects included in the final analysis (N=174) were more likely to be literate (OR = 2.9, 95%CI: 1.5-5.6, P < 0.001), not attend video clubs (OR = 0.3, 95%CI: 0.1-0.7, P=0.005) belong to households with higher SEP (OR = 2.0, 95%CI: 1.2-3.6 and OR = 4.4, 95%CI: 2.5-7.9, P < 0.001, respectively for medium and high household SEP compared to low household SEP), come from the urban area (OR = 5.6, 95%CI: 3.4-9.2, P < 0.001), to have electricity (OR = 2.0, 95%CI: 1.2-3.4, P = 0.008), and live closer to the clinic (OR = 2.4, 95%CI: 1.4-4.0, P<0.001) and to public transport (OR = 5.1, 95%CI: 5.0-8.6, P<0.001). However, after these variables were included in a logistic regression model together with sex and age, only coming from the urban area was independently associated with an increase odd of having given blood (OR = 3.8, 95% CI: 2.2-6.7, P<0.001).

7.2.3 Risk factors for tuberculosis infection

Table 7.1 shows the results of the univariable analysis on the 83 QFN positive and 84 QNF negative subjects enrolled in the cross sectional study. Exposures of interest have been grouped into confounding and mediating factors as from the conceptual framework.

Infection status was not associated with any other sociodemographic variables considered, including gender (P = 0.6) and age (P = 0.9). Among the food related variables, only people having more than two meals containing proteins per week were more likely to have a QFT positive result (P = 0.04). None of the behavioural risk factors nor the TB exposure related ones showed an association with the likelihood of infection.

Table 7.1: Risk factors for tuberculosis infection, univariable analysis

Exposures		QFT positives n/ N (%)	Unadjusted Odds Ratio (95% CI)	P-value
Being female		47/91 (50.0)	1.2 (0.6-2.2)	0.6
Age group (years)	15-29	29/58 (50.0)	1.0	0.9
	30-44	37/76 (48.7)	0.9 (0.5-1.9)	
	≥ 45	17/33 (51.5)	1.1 (0.4-2.5)	
Urban area of residence		55/106 (51.9)	1.3 (0.7-2.4)	0.5
Education/occupation characteristics				
Being literate		78/152 (51.3)	2.1 (0.7-6.4)	0.2
Occupational status	Employed	25/51 (49.0)	1.0	0.9
	Self-employed	24/48 (50.0)	1.0 (0.5-2.1)	
	Unemployed	34/68 (50.0)	1.0 (0.5-2.3)	
Food availability				
N. meals with proteins/week[(c)]	0	14/40 (35.0)	1.0	0.04
	1	25/46 (54.3)	2.2 (0.9 -5.3)	
	2	17/37 (46.0)	1.6 (0.6-3.9)	
	> 2	27/44 (61.4)	3.0 (1.2-7.2)	
≤ 2 meals/day	≤2	21/44 (47.7)	0.9 (0.4-1.8)	0.8
Biological-behavioural factors				
Not having BCG vaccination		3/5 (60.0)	1.6 (0.2-9.7)	0.6
Being HIV positive		24/49 (49.0)	1.0 (0.5-1.9)	0.9
Alcohol consumption [(a)]		25/49 (51.0)	1.1 (0.6-2.1)	0.8
Smoking[(b)]		11/23 (47.8)	0.9 (0.4-2.2)	0.8
Exposure to cooking fire indoor		65/129 (50.4)	1.1 (0.5-2.3)	0.8
Migration [(c)]		10/20 (50.0)	1.1 (0.4-2.6)	0.9
Exposure to tuberculosis				
Known contact with TB		16/32 (50.0)	1.0 (0.5-2.2)	0.9
Attending video clubs		4/7 (57.1)	1.4 (0.3-6.3)	0.7
Attending bars		20/37 (54.1)	1.2 (0.6-2.6)	0.6
Attending hair dressing shops		60/114 (52.6)	1.5 (0.7-2.8)	0.3
Attending church		70/144 (48.6)	0.7 (0.3-1.8)	0.5
Having been in prison		5/9 (55.6)	1.2 (0.3-4.7)	0.8

[(a)] Drinking more than 3 drinks containing alcohol every time he/she drinks
[(b)] Including current or past smoking
[(c)] Migration as been defined as having lived anywhere else for more than 6 months in the 5 years previous the interview

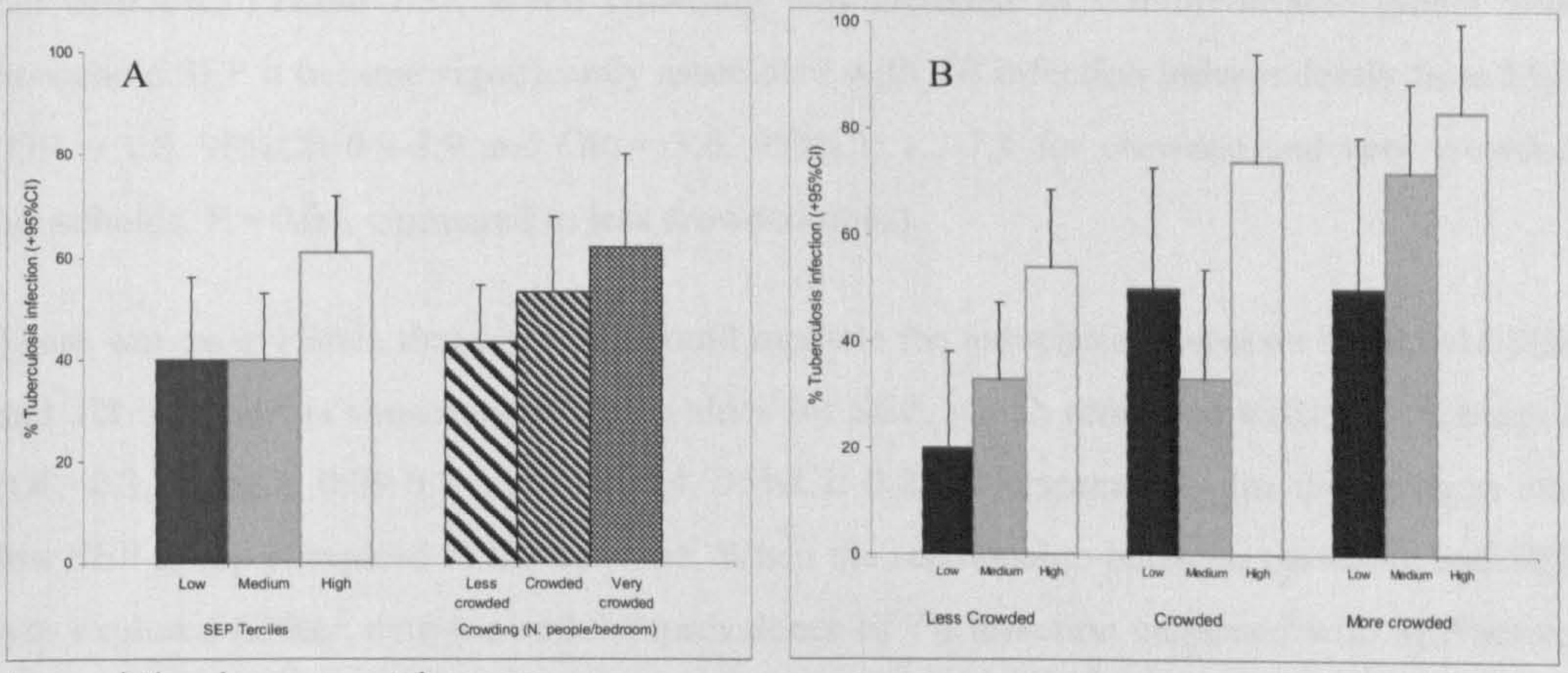
7.2.4 The association between household SEP and tuberculosis infection

The household SEP index – The univariable analysis showed that higher household SEP was associated with higher frequency of TB infection with an OR of 0.4 (95%CI 0.2-0.9) for the medium SEP group and of 0.4 (95%CI: 0.2-0.8) for the high SEP category compared to the reference group (**Figure 7.3**).

As for the case controls study, the multivariable analysis was driven by the conceptual framework outlined in **Figure 4.9**. There was little evidence that the association between TB infection and SEP was confounded by any of the postulated confounding factors. After the adjustment for sex, age group, area of residence, HIV and BCG, the OR for household SEP was essentially unvaried (OR= 0.3, 95%CI: 0.1-1.0 and OR = 0.4, 95%CI: 0.2-0.8 respectively for the Medium and High SEP category compared to the baseline).

In terms of potential mediation, the association between household SEP and TB infection remained unchanged when educational and occupational variables were included in a multivariable model also including sex, age group and area of residence (OR= 0.4, 95%CI= 0.1-1.0 and OR=0.4, 95%CI: 0.2-0.8 respectively for the medium and high SEP group compared to the baseline). None of the remaining postulated mediating variables appeared to explain any part of the association between household and TB infection (**Table 7.2**).

Figure 7.3: The association between household socioeconomic position, crowding and tuberculosis infection



A. association shown separately;
B. Association between household SEP and TB infection by level of crowding
Crowding defined according to the number of people sharing the same sleeping room. Three levels of crowding were identified: less crowded (< 3 people sharing the same sleeping room), crowded (3-4 people sharing the same sleeping room), more crowded (> 4 people sharing the same sleeping room). The cut-offs were decided according to the average number of people per sleeping room in the sample (i.e. approximately 3 people per sleeping room).

The SEP domain indices – After adjusting for age group and area of residence, subjects classified as high SEP by the 'food availability and vulnerability' and the 'access to community services' SEP domain indices appeared to be significantly more likely to be TB infected (respectively OR = 1.8, 95%CI: 1.0-3.5; OR = 2.7, 95% CI: 1.0-7.1).

When SEP domains were included in multivariable model only people classified as wealthier according to the community access domain had a significant higher risk of TB infection (OR= 2.7, 95%CI: 1.0-7.7).

These observations were confirmed even when looking at single socioeconomic indicators. Of the 4 domains considered, variables from the domains 'food availability and vulnerability' (e.g. weekly number of meals containing proteins and number of coping strategies) and 'access to community services' (e.g. having electricity, walking distance from the market) were the only ones significantly associated with TB infection in the multivariable analysis. By contrast, variables considered as indicators of 'human resources' and 'housing quality' were not generally associated with TB infection (**Table 7.3**). The analysis of single socioeconomic indicators showed that even when the association did not approach statistical significance, prevalence of TB infection tended to be always higher in wealthier categories for each of the SEP proxy analyzed (**Table 7.3**).

Data showed some evidence of a modest association between crowding and TB infection: households with more than 4 people per bedroom showed a higher frequency of TB infection (OR=1.9, 95%CI: 0.9-4.3) compared to the less crowded households (i.e. less than 3 people per bedroom) (**Table 7.3**). When crowding was included in a multivariable model with household SEP it became significantly associated with TB infection independently from SEP (OR = 1.8, 95%CI: 0.9-3.9 and OR = 3.0, 95%CI: 1.2-7.4 for crowded and very crowded households, P = 0.01, compared to less crowded ones).

There was no evidence that crowding could mediate the association between household SEP and TB infection as shown by the OR values for SEP, which remained virtually unchanged (OR=0.3, 95%CI: 0.09-0.8 and OR=0.4, 95%CI: 0.2-0.8 respectively for the medium and low SEP group compared to the baseline. When the relationship between crowding and SEP was explored further, data showed that prevalence of TB infection increased with SEP across all level of crowding (**Figure 7.3**).

Table 7.2: The association between household socioeconomic position and TB infection, the multivariable analysis

		Adjusted OR (95%CI)	P value
Food availability			
SEP index	Low	0.5 (0.1-1.5)	0.09
	Medium	0.4 (0.2-0.9)	
	High	1.0	
Weekly number of meals containing proteins	0	1.0	0.02
	1	2.2 (0.9-5.5)	
	2	1.3 (0.5-3.5)	
	>2	2.2 (0.8-6.0)	
SEP index	Low	0.4 (0.1-1.0)	0.01
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Having <2 Meals/day		1.2 (0.6-2.7)	0.5
Biological - behavioural risk factors			
SEP index	Low	0.4 (0.1-1.0)	0.02
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Alcohol consumption		1.2 (0.6-2.4)	0.6
SEP index	Low	0.4 (0.1-1.0)	0.02
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Smoking		1.2 (0.4-3.4)	0.7
SEP index	Low	0.4 (0.1-1.0)	0.02
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Exposure to cooking fire indoor		1.1 (0.5-2.6)	0.8
SEP index	Low	0.4 (0.1-1.1)	0.02
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Migration		1.2 (0.5-3.3)	0.7
SEP index	Low	0.3 (0.09-0.8)	0.005
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Exposure to TB			
SEP index	Low	0.2 (0.08-0.9)	0.01
	Medium	0.4 (0.2-0.9)	
	High	1.0	

Known contact with TB		1.4 (05-3.3)	0.5
SEP index	Low	0.4 (0.1-1.2)	0.04
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Attending churches		0.6 (0.2-1.6)	0.3
SEP index	Low	0.4 (0.1-1.0)	0.01
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Attending hairdressing shops		1.8 (0.9-3.8)	0.1
SEP index	Low	0.4 (0.1-1.4)	0.02
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Attending bars/video shops		1.9 (0.6-5.5)	0.2
SEP index	Low	0.3 (0.1-0.9)	0.01
	Medium	0.4 (0.2-0.8)	
	High	1.0	
Having been in prison		1.9 (0.4-8.9)	0.4

Table 7.3: The association between proxies of household socioeconomic position and TB infection

Exposures[(a)]		QFT positives n/N (%)	Unadjusted Odds Ratio (95% CI)	P-value	Adjusted Odds Ratio (95% CI)[†]	P-value
Human resources						
Occupation of the head of the household	Self employed	33/62 (53.2)	2.3 (0.6-8.3)	0.2	4.1 (0.9-18.7)	0.07
	Employed	46/93 (49.5)	1.9 (0.6-6.9)		3.9 (0.6-17.0)	
	Unemployed	4/12 (3.3)	1.0		1.0	
Food availability and vulnerability						
N. meals containing proteins/week	0	12/34 (35.3)	1.0	0.2	1.0	0.04*
	1	23/47 (48.9)	1.7 (0.7-4.3)		1.9 (0.7-5.1)	
	2	23/40 (57.5)	2.5 (1.0-6.3)		3.2 (1.1-9.0)	
	>2	25/46 (54.3)	2.2 (0.9-5.4)		2.9 (1.0-8.1)	
N. of meals/day	≤ 2	21/49 (42.9)	1.0	0.2	1.0	0.2
	> 2	62/118 (52.5)	1.5 (0.7-2.9)		1.5 (0.7-3.1)	
N. of coping strategies	> 3	21/51 (41.2)	1.0	0.07	1.0	0.03*
	1-3	25/55 (45.4)	1.2 (0.5-2.6)		1.3 (0.6-3.0)	
	0	37/61 (60.7)	2.2 (1.0-4.7)		2.5 (1.1-5.9)	
Assets ownership and housing quality						
Having animals	Yes	19/48 (39.6)	1.0	0.09	1.0	0.1
	No	64/119(53.8)	1.8 (0.9-3.5)		0.5 (0.2-1.2)	

Table 7.3: Continued

Exposures[(a)]		QFT positives n/N (%)	Unadjusted Odds Ratio (95% CI)	P-value	Adjusted Odds Ratio (95% CI)[†]	P-value
Crowding (N. people/bedroom)	0-3	37/85 (45.5)	1.0	0.09*	-	
	3-4	35/47 (53.2)	1.5 (0.7-3.0			
	> 4	21/35 (60.0)	1.9 (0.9-4.3)			
Floor material	Dirt/earth	28/64 (43.7)	1.0	0.2	1.0	0.2
	Cement	55/103 (53.4)	1.5 (0.8-2.7)		1.7 (0.8-3.5)	
Roof material	Grass	10/20 (50.0)	1.0	0.08	1.0	0.1
	Iron sheet	30/74 (40.5)	0.7 (0.2-1.8)		0.7 (0.2-2.2)	
	Asbestos	43/73 (58.9)	1.4 (0.5-3.9)		1.6 (0.5-5.3)	
Type of water sanitation	Latrines	67/143 (46.8)	1.0	0.07	1.0	0.08
	Flush toilets/other	16/24 (66.7)	2.3 (0.9-5.6)		2.3 (0.9-5.9)	
Access to infrastructures and facilities						
Having electricity	No	49/112 (43.7)	1.0	0.03	1.0	0.03
	Yes	34/55 (61.8)	2.1 (1.1-4.0)		2.2 (1.1-4.6)	
Type of water supply	Piped outside house	52/109 (47.7)	1.0	0.1	1.0	0.7
	Piped inside house	30/57 (52.6)	1.2 (0.6-2.3)		1.1 (0.6-2.4)	
Walking distance from the market (Min.)	> 15	31/80 (38.7)	1.0	0.07	1.0	<0.001
	0-15	52/87 (60.0)	2.3 (1.2-4.6)		2.7 (1.3-5.6)	

[(a)] SEP proxies are grouped by dimension considered in the conceptual framework. Results presented only for those variables with a P value ≤ 0.2 in the univariable analysis
[†] Multivariable analysis including sex, age group, area of residence. Because of the association of crowding with TB infection after adjustment for SEP index, the association between the household socioeconomic proxies and TB infection was also adjusted by crowding.
*Test for trend.

Summary

Unexpectedly TB infection was more common among disease free individuals of higher, rather than lower, SEP. This observation was made for most of the socioeconomic indicators collected, not just the composite SEP index.

None of the traditional risk factor variables that I have explored appeared to mediate the association between SEP and prevalent TB, suggesting that in this setting TB transmission may occur through the exposure to risk factors for TB other than those traditionally explored and associated with poverty, like alcohol consumption, inadequate nutrition, smoking and contact with a TB case.

The analysis of SEP proxies grouped into different conceptual domains confirmed further the association between higher SEP and TB infection, as showed by the fact people classified as wealthier in terms of better 'access to infrastructures and facilities' were also those at higher risk of TB infection.

This study also confirmed the importance of crowding in the epidemiology of TB and it also showed a strong independent effect of both crowding and SEP on the risk of TB infection.

8. Sensitivity analysis

Don't ask me what poverty is because you have met it outside my house.
Look at the house, and count the number of holes. Look at my utensils and the clothes that I'm wearing. Look at everything and write what you see. What you see is poverty.

A poor man, Kenya[9]. 1997

Introduction

Asset-based indices aim to rank households according to their SEP by creating a socioeconomic score resulting from the observed measures of households' material living conditions, especially ownership of durable goods and housing characteristics. The choice of assets included in the index and the method used to estimate the assets weights can influence the household SEP ranking and our assessment of health inequalities [240].

The aim of this chapter is to explore how these two aspects might have affected the findings presented in this thesis and the understanding of TB inequalities.

8.1 Chapter objectives

1. To investigate the importance of the type and number of assets and the weighting strategy in the development of assets-based index **(Section 8.2)**.

2. To develop a list of alternative household SEP indices based on the above issues **(Section 8.3)**.

3. To explore the extent of agreement of these indices in the household SEP ranking **(Section 8.4)**.

4. To assess the robustness of the main findings of this thesis by comparing them to the results obtained gathered using the alternative indices **(Section 8.4)**.

8.2 The robustness of assets-based indices

Assets-based indices have become increasingly common in the assessment of health inequalities; however, there is debate concerning the extent to which assets-based indices can provide robust measurements of health inequalities. In this section I will discuss the two aspects making the reliability of assets-based indices most questionable and how I have tried to address these issues.

8.2.1 The choice of type and number of assets

There is disagreement in the literature on the extent to which the use of alternative lists of assets items can affect the direction and extent of observed health inequalities. Whereas Filmer and Pritchet found that household ranking was robust to the choice of the items included in the asset index [85], Houweling and colleagues [227] found that the household ranking into SEP groups was sensitive to this choice. Most importantly, they found that the magnitude of inequality for their outcome of interest was sensitive to the use of different indices sometimes to an important extent (ranging up to a 60% change in observed inequality).

Of all the issues discussed around the impact of using different sets of items, two seemed to be particularly relevant for this thesis:

(a) The urban bias

In Demographic Health Surveys, urban households dominate the top quintiles of SEP indices, whereas in most settings rural households area often ranked much lower [86]. A possible explanation is that the items most commonly used in assets indices are often most frequent in urban settings. As a result, the SEP score of urban households is frequently overestimated not because urban households are actually wealthier compared to rural households, but simply because they are more likely to have access to consumer durables, modern housing materials and especially public services [192]. Thus, it is possible that the exclusion of these variables from the indices reduces the difference in the SEP classification between rural and urban households and helps the separation of poor households in the rural area, that otherwise would be mostly classified as poor. The urban bias can lead to misleading observations of health inequalities. Some authors, for example, have argued that the association between wealth and HIV in some African countries can be biased by the use

of the Demographic Health Surveys wealth index including assets (such as radio, bicycle, phone) that can be purchased in modern cash economy that is more common in urban societies. As a result, people owning other type of assets based on a more traditional economy more common in rural settings are mistakenly classified as poor. If this is the case, a legitimate question would be whether wealthy people are truly more vulnerable to HIV or whether it is their participation in modern, rather than traditional, forms of economy that make these people be assessed as apparently wealthier? [241]

The urban bias can potentially have caused an overestimate of the extent of TB inequalities observed in this thesis, by artificially increasing the SEP differences between urban and rural households. To explore this possibility I have built a SEP index excluding from the PCA all those items typically more common in urban settings.

(b) The inclusion of assets that have a direct effect on the health outcome

Houweling and colleagues have also hypothesised that observed extent of health inequalities can be attributed to the inclusions of assets that may have a *direct* effect on the outcome of interest (for example, water and sanitation for diarrhoea in children), apart from their indirect effect as indicators of SEP [227]. In this circumstance, a variable can be 'double counted': once when included in the composite index and once as risk factor, making the results of multivariable analysis unclear [192].

This argument is particular relevant in this thesis where food consumption variables were included in the household SEP index and at the same time used as risk factors for TB. Because of the well known association between nutrition and TB [38], any detected association between SEP and prevalent TB could reflect the association between nutrition and TB, rather than a direct association between SEP and TB [227]. However, this study represents a particular case where a strong marker of SEP, probably the strongest in this setting, is also a strong determinant of the health outcome under investigation. As a result:

1. The exclusion of food-related variables from the list of risk factors would have been difficult to justify and somewhat in contrast with the objective of this thesis, which is not only to quantify the association between household SEP and prevalent TB, but also to try to understand the nature of this association. Excluding food related variables from the list of possible mediators would have precluded the understanding of a possible mechanism through which SEP affects the risk of TB.

2. The exclusion of food related variables from the PCA was not deemed feasible due to the strong evidence suggesting how relevant food availability in the definition of household SEP in Zambia. Given this strong association excluding food related variables would have been theoretically more problematic than including them.

I have tried to address this issue both within the study design and through the sensitivity analysis. In designing the study I have adopted the following strategies:

1. Similarly to another study [113], I have decided to include in the household SEP index the food-related variable defined at household level, whereas as mediator I have used the food-related variable at individual level. Although correlation between household and individual level food-related variables is likely, I have considered them to be conceptually different: whereas food availability at household level is supposed to be a marker of household SEP, individual food availability is seen as a potential direct determinant of health.

2. The deconstruction of the overall composite SEP index into 4 different SEP domains, three of which do not include food related variables can be considered a further attempt to interpret better the association between household SEP and TB.

As part of the sensitivity analysis, I have explored the effect of a possible direct effect bias by developing a SEP index excluding all the food-related variables.

8.2.2 The choice of weighting strategy

One of the critical aspects for the construction of an asset index is the choice of the weighting strategy.

Authors have developed a wide range of alternatives to PCA for the construction of asset indices, ranging from the simple count of the assets to sophisticate modelling strategies [240]. Comparative studies have showed that depending on the methods adopted, the asset indices can return different household SEP ranking and a different extent of health inequalities [240, 242-243]. Following from this, it was felt appropriate to adopt at least one additional strategy to develop an asset index for this thesis. Among the alternatives, I have decided to adopt the method suggested by Stifel and Cristiansen [244], according to which the weights are derived from the regression of per capita consumption expenditure on a list of assets indicators included in a different dataset and then applied to the dataset in use for

this study. Details on this methodology and the rational of this choice are provided later in this chapter.

8.3 Methods

Following the discussion above, four alternative indices were built:

1. One assets index, PCA-based, excluding variables likely to be more common in the urban setting (**Index I**).

2. One assets index, PCA-based, excluding food-related variables accounting for the potential of direct effect bias (**Index II**).

3. Two assets indices where the assets' weights were derived through regression analysis. Of them one index included food related variables (**Index III**) and the other did not (**Index IV**).

8.3.1 The urban bias index

In order to assess the effect of the potential urban bias on the household SEP ranking and the association between SEP and prevalent TB, I developed an asset index that did not include electricity and those variables related to access to community services (i.e. distance from facilities and infrastructure from the household).

Among the different socioeconomic indicators probably electricity is known to be the one best differentiating urban households from rural ones [192]. Furthermore, electricity may not be an ideal asset to rank households SEP because rural households may not have because it is not provided by the community where they live, rather than because they can not afford it. As shown in **Chapter 6**, electricity was more common in the urban households than the rural ones (41.5% vs 14.1%, P<0.001). It can be argued that the exclusion of this item from the asset index may help in reducing the SEP disparities observed among urban and rural households. It is also possible that the exclusion of electricity could favour a better discrimination of the households SEP within the urban site because a higher proportion of households in this setting will be identified as having low SEP.

The data collected in this study have shown that the rural households were significantly more distant than the urban ones from all the community services investigated and were also

significantly less likely to have water directly piped into their houses. Therefore, like electricity, the exclusion of other variables indicating community services, such as the type of water sanitation and distance from schools, market, transport and clinics was thought able to reduce the urban bias.

For the purpose of this analysis I did not create a completely new index, but I have simply excluded the above items from the list of 21 variables initially selected for inclusion in the PCA that led to the creation of the SEP index used in this thesis (**Table 8.1**). Overall 7 items were removed from the original list of 21, including: having electricity, having a TV, having water piped inside the house, having water piped outside the house (e.g. public tap water), distance from the nearest clinic, market and source of water. The remaining 14 variables were included in a PCA and processed as described earlier in this thesis. No further screening after PCA was conducted for **Index I.**

8.3.2 The direct effect bias index

In order to minimise the chance of any direct effect of food related variables on the association between household SEP and TB, it was not sufficient to eliminate all the food related variables from the list of 21 variables initially selected for the SEP index used in this thesis. These variables had been screened depending on their strong association with the weekly consumption of food related variables used as benchmark indicator of SEP in the study setting. I had, thus, to develop a completely new index.

As for the SEP index, originally used in this thesis, the development of this index required two screening processes of the socioeconomic indicators available: one before PCA and one afterwards (**Table 8.2**). Before PCA only socioeconomic indicators presenting a frequency distribution higher than 20% and lower than 80% were considered eligible (Filter I in **Table 8.2**). From this list of variables, I first excluded food consumption related variables (Filter II) and finally those with missing values and providing redundant information (Filter III). This filtering process left a total of 20 variables suitable for PCA. PCA was run several times before identifying the best combination of variables to be included in the final version of Index II. Even for this index, the size and sign of the weights as well as the Eigen values and the percentage of common variance guided the inclusion/exclusion of the variables during the different PCA's attempts. Clumping and truncation were also used as indication of the appropriateness of the variables chosen.

Table 8.1: Variables included in Index I: the selection process

N.	All variables considered	Before PCA			After PCA
		I filter*	II Filter†	III Filter‡	Variables included in Index I
		%	Chi square P value		
1	Drinking water source – Bore hole	**21.9**	**0.05**		
2	Roof in grass	**22.0**	**<0.001**		
3	Drinking water source- Piped inside the house	**22.4**	**<0.001**		
4	Drinking water source – Traditional well	**23.8**	**0.001**		
5	Main energy source for lighting - electricity	**25.0**	**0.001**		
6	Main energy source for lighting- candles	**29.4**	0.3		
7	N. contributors other than head of the household	**30.3**	0.3		
8	House rented	**31.0**	**0.001**	–	
9	Female head	**31.0**	0.2		
10	Drinking water source – piped outside the house	**31.3**	**<0.001**		
11	Having bicycle	**32.0**	**0.01**		
12	Roof in asbestos	**34.6**	**<0.001**		
13	Walls in concrete mud	**35.4**	**<0.001**		
14	Having animals	**39.4**	**0.001**		
15	Having TV	**39.7**	**<0.001**		
16	Main energy source for lighting – Kerosene	**42.7**	**<0.001**		
17	≥ 2 meals containing proteins per week§	**43.2**	-		
18	Roof in iron	**43.5**	0.2		
19	Household size (N. members ≥4)	**43.8**	0.2		
20	Head of the household self employed	**43.8**	**<0.001**	–	
21	Floor in dirt	**45.7**	**<0.001**		
22	> 3 coping strategies	**45.9**	**<0.001**		
23	< 30 minutes walking from clinic	**46.2**	**<0.001**		
24	Head of the household employed	**47.0**	**0.001**		
25	Walls in Cement	**54.0**	**<0.001**		
26	Education level Grade ≥ 2	**55.7**	**<0.001**		
27	< 30 minutes walking from market	**60.8**	**<0.001**		
28	< 30 minutes walking from transport	**63.5**	**<0.001**	–	
29	≥ 3 meals per day	**65.7**	**<0.001**	–	
30	< 3 months without enough to eat	**67.6**	**<0.001**	–	
31	House owned	**69.2**	**0.001**	–	
32	Having a radio	**73.8**	**0.004**		
33	Land owned	**73.8**	**0.01**	–	
34	< 15 minutes walking from source of water	**80.0**	**<0.001**		

*I Filter: Variables with a frequency distribution in the controls population lower than 20% and higher than 80%.

†II Filter – Variables not significantly associated with the benchmark indicator (i.e. weekly consumption of meals containing proteins in the household) were screened out ($P<0.05$).

‡Filter III – Electricity and access to community services variables (indicated with the red box) were screened out as well as variables with missing values or providing redundant information (indicated with the symbol –).

§Benchmark indicator.

Table 8.2: Variables included in Index II: the selection process

N.	All variables considered	Before PCA			After PCA
		I Filter*	II Filter†	III Filter‡	Variables included in Index II
		%			
1	Drinking water source - Bore hole	**21.9**		–	
2	Roof in grass	**22.0**			
3	Drinking water source- Piped inside the house	**22.4**			–
4	Drinking water source – Traditional well	**23.8**			–
5	Main energy source for lighting - electricity	**25.0**			
6	Main energy source for lighting- candles	**29.4**		–	
7	N. contributors other than head of the household	**30.3**		–	
8	House rented	**31.0**		–	
9	Female head	**31.0**		–	
10	Drinking water source – piped outside the house	**31.3**			–
11	Having bicycle	**32.0**			–
12	Roof in asbestos	**34.6**			
13	Walls in concrete mud	**35.4**			–
14	Having animals	**39.4**			–
15	Having TV	**39.7**			–
16	Main energy source for lighting – Kerosene	**42.7**			–
17	≥ 2 meals containing proteins per week§	**43.2**			
18	Roof in iron	**43.5**		–	
19	Household size (N. members ≥4)	**43.8**		–	
20	Head of the household self employed	**43.8**		–	
21	Floor in dirt	**45.7**			
22	> 3 coping strategies	**45.9**			
23	< 30 minutes walking from clinic	**46.2**			–
24	Head of the household employed	**47.0**			
25	Walls in Cement	**54.0**			
26	Education level Grade ≥ 2	**55.7**			–
27	< 30 minutes walking from market	**60.8**			–
28	< 30 minutes walking from transport	**63.5**		–	
29	≥ 3 meals per day	**65.7**			
30	< 3 months without enough to eat	**67.6**			
31	House owned	**69.2**		–	–
32	Having a radio	**73.8**			
33	Land owned	**73.8**		–	
34	< 15 minutes walking from source of water	**80.0**			–

*I Filter: Variables with a frequency distribution in the controls population lower than 20% and higher than 80%.

†II Filter – Food related variables were screened out (indicated with the red box).

‡Filter III – Variables with missing values or providing redundant information were screened out (indicated with the symbol –).

8.3.3 The regression based asset indices

These indices have been developed following the strategy suggested by Stifel and Cristiaensen [244]. In their method the assets weights are estimated through the regression of expenditure data on a list of asset indicators included in a dataset containing data on expenditure. These weights were then applied to a list of identically defined indicators in the dataset which lacks per capita consumption expenditure data.

The regression modelling applied for the development of these indices is the same underlying the poverty mapping methodology described in **Chapter 4.**

In the poverty mapping the regression modelling was used to predict expenditure data and estimate the proportion of households living below the poverty line in each CSA. In the sensitivity analysis, the regression technique should be considered more as an alternative weighting strategy to PCA. In the case of poverty mapping, the two datasets used were the LSMSIII and the census for the Lusaka Province (2000). In the sensitivity analysis, the LSMSIII is still the database from which the regression model originates, but the regression coefficients so estimated are applied to the assets included in the case control study dataset.

According to the literature this method has been employed in one case to track poverty change in time in the absence of data on consumption expenditure [244] on at least nine datasets to assess the agreement between expenditure and asset indices and in one case to assess the socioeconomic profile of TB patients [22].

(a) Meaning of the SEP index based on consumption expenditure prediction

The household SEP index resulting from the regression modelling is not meant to provide insight on an aspect of household SEP different from that captured by the household SEP index built with PCA. Rather, they are thought to measure the same material and "econometric" dimension of household SEP. Furthermore, because housing characteristics, assets ownership, and the other indicators included in the regression model are generally considered 'slow moving' indicators of SEP [240], it is appropriate to assume that more than predicting the whole aggregate of consumption expenditure, the SEP index generated through regression is actually capturing a more stable, long-term, component of expenditure, that is also what the SEP index used in this thesis is meant to measure.

(b) The implementation of data regression in this thesis

As for the poverty mapping analysis, the LSMSIII was the dataset containing consumption expenditure data, but for the purpose of the sensitivity analysis the regression coefficients were applied to the assets included in the case control study dataset.

The imputation of per capita consumption expenditure data for the case control study required two steps, both of which were repeated both for Index III and Index IV. The only difference is that for Index IV, all food related variables were excluded from the regression analysis:

1. A model of consumption was first estimated using the data included in the LSMSIII survey. The analysis was restricted only to the Lusaka province households, where the study sites were located. In order to be extrapolated to the case control dataset only variables common to both the case control study and the LSMSIII could be used. This is also the only reason why I could not use the prediction of consumption expenditure estimated for the poverty mapping analysis illustrated in Chapter 4, which was based on variables in common to the LSMSIII and the census dataset. The outcome was represented by per capita consumption expenditure, whereas the set of assets in common to the two dataset were used as independent (explanatory) variables. In order to achieve near normality (and hence satisfy the assumption of linear modelling), per capita consumption expenditure was log transformed (on a natural scale).

The linear regression accounted for the survey's sampling weights. Each categorical variable was replaced by a set of binary indicators. Variables showing a linear trend in the association with the outcome (with the Likelihood Ratio Test) were included in the models as continuous variables. Continuous variables showing departure from trend have been firstly grouped into categories and then converted into binary variables.

In order to select the most predictive socioeconomic indicators and to maximise the explanatory power of the final model (as captured by R-squared statistic), I first ran a set of 4 separate models, one for each of the 4 SEP domains, using step-wise regression. This approach allowed the reduction of the assets in the index and, at same time, a balanced selection of the items included. Variables showing a significance P value ≤ 0.1 in each sub-model were finally included in an overall model for the imputation of per capita consumption expenditure.

Before applying the estimated coefficients to the variables of the case control study I first assessed the goodness of the regression model by comparing the households SEP

ranking and the proportion of households living below the poverty line, according the actual and the predicted per capita expenditure indices.

2. The second step consisted of applying the estimated regression coefficients to the indicators included in the case control study. The result was an imputed value of the log of per capita consumption expenditure for each household enrolled in the case control study.

8.3.4 Sensitivity analysis

The analysis consisted of three parts: a) descriptive analysis of the population socioeconomic profile according to the new indices; b) the assessment of indices agreement by comparing the household SEP ranking; and finally c) the assessment of the sensitivity of the results gathered in the case control study to the use of different household SEP indices.

(a) Descriptive analysis

First I characterised the socioeconomic profile of the study population using the different household SEP indices using only the population of controls.

As the original SEP index, the four "new" indices built for this part of analysis were explored both as continuous and categorical variables. Indices were explored graphically: histograms were used to assess the extent of clumping and truncation and box plots were used to characterise the SEP scores distributions in terms of median and interquartile range. The SEP scores were also grouped into terciles, yielding to the generation of three socioeconomic groups: "Low", "Medium" and "High" SEP.

Both the estimates of the per capita consumption expenditure (with or without the food related variables) were used to estimate the proportion of control households living below the moderate and extreme poverty line using the same method described in **Chapter 4, section 4.5.9.**

These proportions were estimated overall, separately for the urban and rural area, and for the cases and the controls.

(b) The indices agreement on the household SEP ranking

The indices illustrated in this chapter were compared to the SEP index used in this thesis.

The agreement between indices was assessed by looking at the household misclassification among terciles and tested with Kappa statistic, a test providing the expected probability that two observations are classified in the same group by chance. A kappa value of less than 0.5 is generally considered an indication of poor agreement [221]. In this study the assessment of agreement was explored not just by looking at the proportion of households falling in the same SEP index tercile, but also the proportion of households that were misclassified by one or two SEP terciles.

The agreement between indices was also verified graphically through Bland Altman plots. Traditional scatter plots measure correlation between two measures rather than agreement. Since two measures showing high correlation can have poor agreement the use of the normal scatter plots can lead to misleading interpretations [245].

The Bland Altman approach is based on the plot of the mean of the two measures for each subject against the difference of the two measurements: in this case the subjects are the households and the measurements are the score values generated by the two SEP indices compared. The plot of the mean against the difference allows to assess better the existence of outlying observation and to explore the relationship between the empirical values of the two measurements.

The Bland Altman plots are based on the creation of the *limits of agreement*: if the differences between the two measurements are normally distributed we should expect that 95% of them will be included between the mean of the difference ± 1.96 the standard deviation of the difference. The limits of agreement define the range within which most of the differences between the two methods should lie. The calculation of the 95% limits of agreement is based on the assumption that that the two measurements use the same scale. For this reason, the Bland Altman plots could be drawn only for the comparison between the original SEP index and respectively Index **I** and Index **II** (i.e. both created though PCA as the original index and therefore having the same scale). Finally, Bland Altman plots assume that the compared measurements follow a normal distribution. To meet this latter criteria, it was considered appropriate to log transform both indices before starting to analyse them.

The interpretation of the Bland Altman plots is based on the observation of the size of the limits of agreement (i.e. too widely spaced limits of agreement are index of bad agreement)

and the number of outliers that should not be much greater than 5% of the total of the observations.

The agreement among indices was also explored by assessing the proportion of households in the lowest tercile of each index across the urban and rural sites and the 11 CSA. This latter analysis allowed the assessment of how robust the CSA ranking was to the use of different indices and, most importantly, to assess whether the households enrolled in the case control study and living in these CSA were a representative sample of all the households included in the census.

(c) The assessment of results sensitivity

After comparing the household ranking according to the alternative indices, I explored the association between the SEP indices and prevalent TB. For this purpose, for each new index, I replicated both the minimally adjusted and multivariable analysis conducted for the case control study (details on the analysis are discussed in **Chapter 4**). The results so obtained were then compared to the original ones.

8.4 Results

For better clarity whenever possible the results will be grouped according to the weighting strategies employed for the construction of the indices.

8.4.1 Descriptive analysis of the socioeconomic position indices

(a) The PCA-based asset indices

Table 8.2 presents a summary of the main features of the two PCA-based asset indices. Despite the use of a different filtering strategy **Index II** included very similar socioeconomic items to **Index I** and the SEP index originally used in this thesis.

Socioeconomic indicators have been weighted similarly in **Index I** and **Index II** both in terms of size and sign. Indices explained respectively 34.5% and 43.6% of the common variance.

The two indices did not significantly differ from each other and from the original SEP index in terms of median SEP score (**Table 8.3**); as for the original SEP index, rural households showed a significantly lower median score compared to urban households for both indices ($P<0.001$ according to Mann-Whitney test) (**Figure 8.1**). Compared to the original SEP index, the exclusion of variables more typical of urban settings from **Index I** did not appear to reduce significantly the gap between rural and urban households.

In terms of SEP score distribution, the indices showed very similar distributions (**Figure 8.2**). As the SEP index, both **Index I** and **Index II** showed a bimodal pattern with two peaks probably indicating the urban and rural households populations and both indices did not appear to be particularly efficient in characterising the households with extreme SEP scores. **Index I** showed slightly less truncation than the main household SEP index suggesting that the removal of urban setting related variables allowed a better distinction of the SEP of rural households, especially among the very poor ones. Compared to **Index I** and the original SEP index, **Index II** showed a higher level of clumping and truncation both for the urban and rural score distribution suggesting that the removal of food related variables may have reduced the discriminatory capacity of the SEP index.

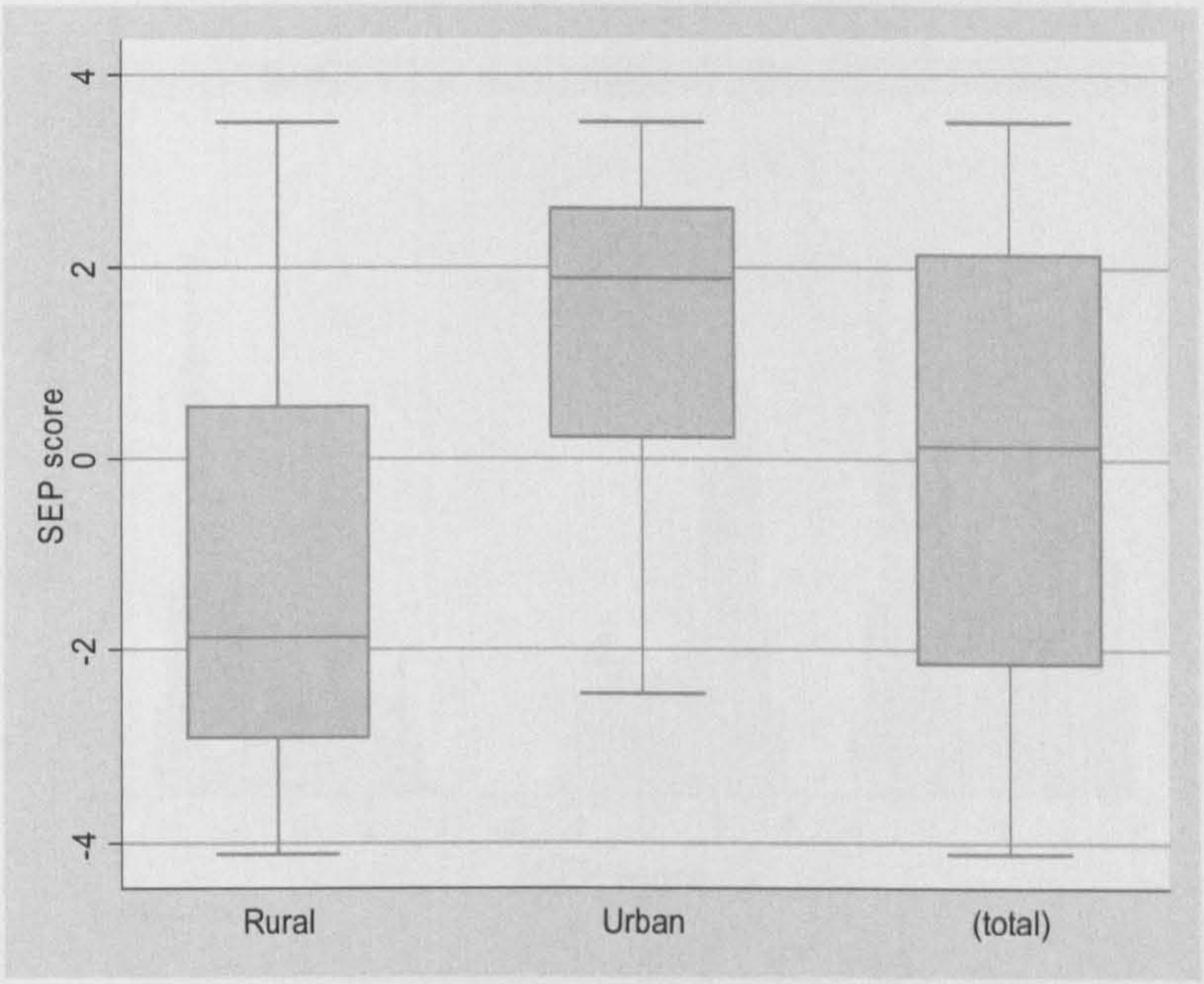
Table 8.3: Main characteristics of Index I and Index II

	Index I*	Index II**
N. Observations	318	318
N. Indicators	14	10
Weights		
Education level Grade ≥ 2	0.184	—
Head of the household employed	0.194	0.202
≥ 2 meals containing proteins per week	0.203	
< 3 coping strategies	0.205	0.242
Lighting source – Electricity	—	0.354
Lighting source – Kerosene	0.303	—
Radio	0.186	0.201
Bicycle	0.080	—
Having animals	-0.230	—
Source of drinking water – traditional well	-0.180	—
Wall – Concrete	0.338	0.386
Floor – Dirt	-0.379	-0.416
Floor – Cement	0.380	0.418
Roof material –Grass	-0.290	-0.310
Roof material –Asbestos	0.242	0.279
< 30 min. walking distance from nearest clinic	—	0.247
Eigen Value	5.2	4.3
% common variance	34.5	43.6
Median SEP score (Range)		
Overall	0.13 (-4.1-3.5)	0.07 (-3.2 – 3.2)
Rural	-1.9 (-4.1-3.5)	-1.6 (-3.2-3.2)
Urban	1.9 (-2.5-3.5)	1.6 (-2.5-3.2)

*Index I: accounting for potential urban bias
**Index II: accounting for potential direct effect bias

Figure 8.1: Boxplot of the two PCA-based assets socioeconomic position indices, overall and by study area

Index I

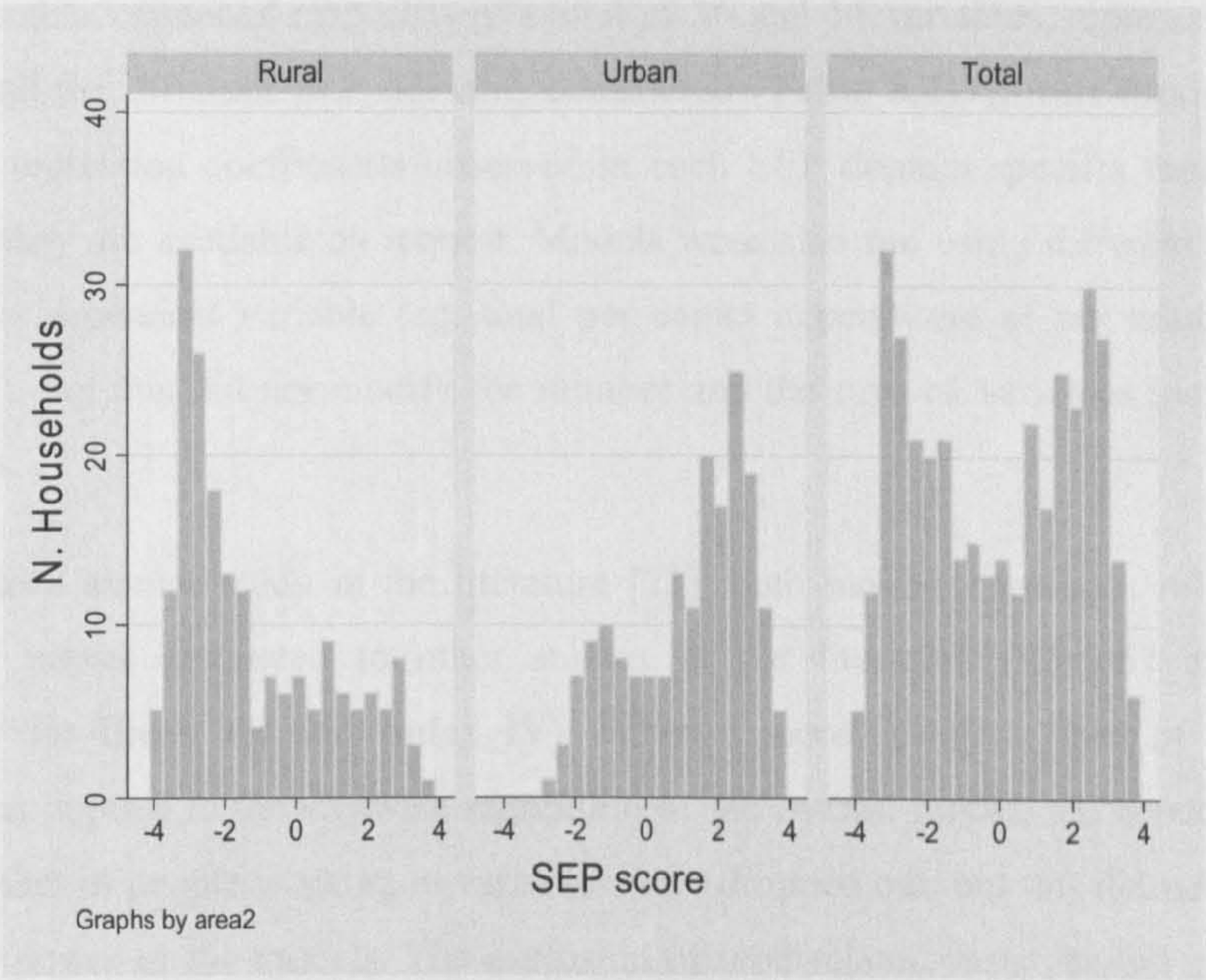


Index II

Index I accounting for the potential urban bias
Index II accounting for the potential direct effect bias

Figure 8.2: SEP score distribution of two PCA-based assets socioeconomic position indices, overall and by study area

Index I



Index II

Index I accounting for potential urban bias
Index II accounting for potential direct effect bias

(b) The Regression – based assets indices

Both models for the imputation of per capita consumption expenditure for the **Index III** and **Index IV** were run on a sample of 1555 households out of the 1567 available in the LMSIII. The final models contained respectively a total of 16 and 14 variables, representing a good balance of all the different SEP domains considered (**Table 8.2**). In this thesis, I will not present the regression coefficients observed in each SEP domain specific regression sub-model, but they are available on request. Models were also run using different equivalence scale for the dependent variable (eg. total per capita expenditure or per adult equivalent expenditure), but this did not modify the number and the type of variables included in the final models.

Compared to a similar study in the literature [22], both models showed a relatively high explanatory power compared to other studies in the literature (R^2=0.57 and R^2=0.55 respectively for **Index III** and **Index IV**). When a more stringent level of significance (P<0.05) was applied to the stepwise regression of the overall model, the economy activity and the number of people working in variables were dropped out, but this did not change the explanatory power of the models. The exclusion of food related variables did not appear to affect the explanatory power of the models nor the sign and size of the coefficients making **Index III** and **Index IV**, resulting in the creation of two fairly similar indices. For this reason, the descriptive analysis results will be presented only for **Index III**.

The distribution of the predicted and actual pro capita consumption expenditure (both in natural logarithmic scale) showed a fairly similar trend (**Figure 8.3**). However, when the actual per capita expenditure variable in the LSMSIII and both **indices III** and **IV** were grouped into quintiles and compared in terms of households ranking, there was a relatively modest agreement between them with approximately 46% of the households classified in the same quintile (kappa=0.3, P<0.001). **Index III** and **IV** seemed to perform better at identifying the poorest and wealthiest households in the survey with 57.6% and 66.3% of households in the bottom and top quintile (**Table 8.6**). The explanation of the poor agreement can be attributed to the relatively low explanatory power of the model (only 60%): although a particular effort was made to ensure that the model included variables representative of the different household SEP domains considered, the number of variables eventually included is still reductive and probably not adequate to express all the aspects of household SEP.

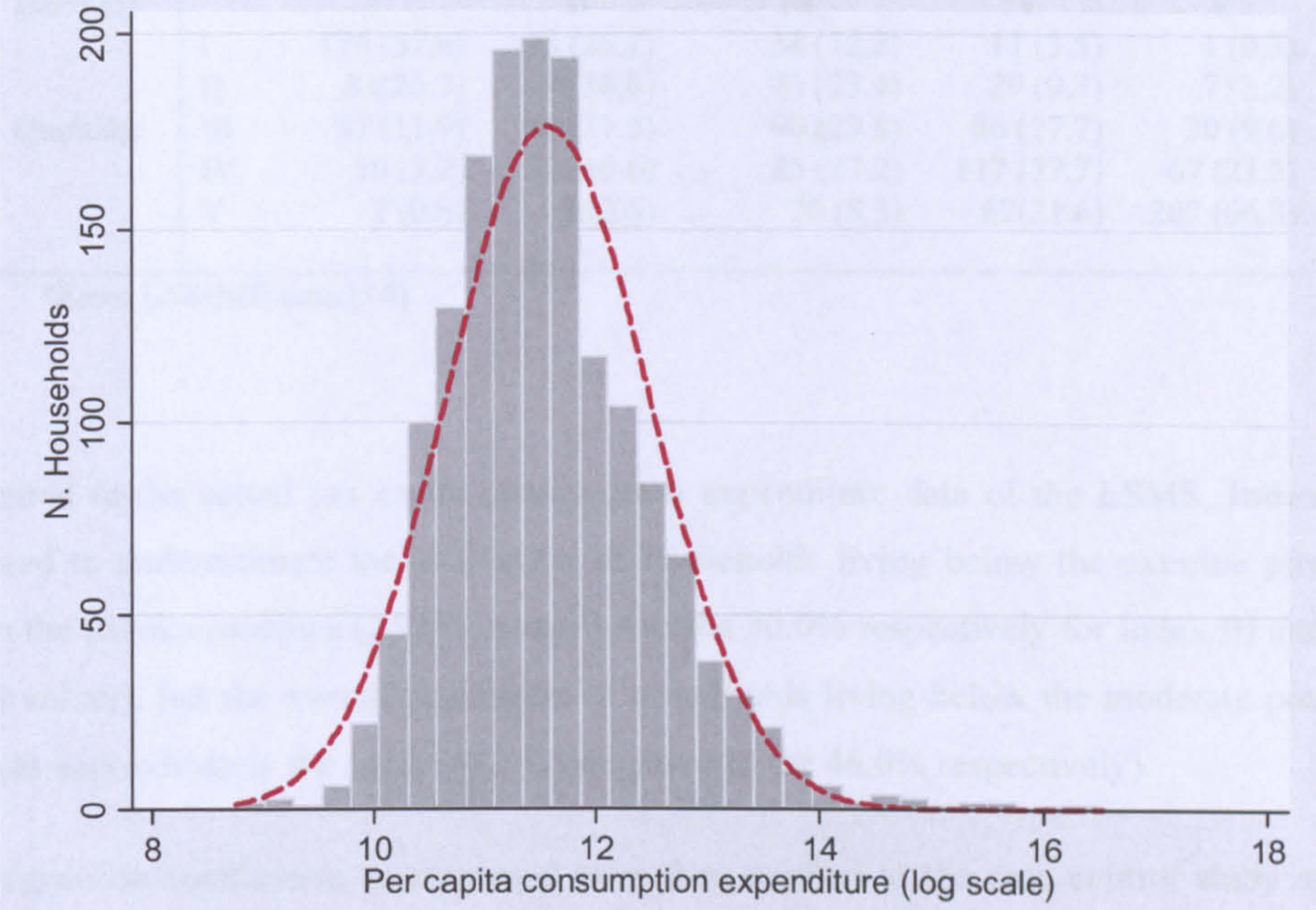
Table 8.4: Estimates of regression coefficients for the prediction of per capita consumption expenditure

Socioeconomic variables by SEP domain	Index III			Index IV		
	Coefficients	Standard Error	P-value	Coefficients	Standard Error	P-value
Human resources						
Economic activity of the head of the household						
Not working	Ref			Ref		
Working	0.046	0.102	0.6	0.091	0.103	0.4
Not working, but economically active	0.015	0.154	0.9	-0.109	0.127	0.4
Not working, but economically inactive	- 0.151	0.124	0.2	0.032	0.156	0.8
N. of people working in the household	- 0.042	0.021	0.04	-	-	-
N. of people economically inactive in the household	- 0.013	0.007	<0.001	- 0.124	0.008	<0.001
N. of children < 14 years old in the household						
0	Ref			Ref		
1	- 0.105	0.042	0.01	-0.094	0.433	0.03
2	0.030	0.120	0.8	0.065	0.12	0.6
Food availability and vulnerability						
Number of coping strategies						
< 3	Ref			Ref		
3-6	- 0.118	0.366	0.001	- 0.159	0.037	<0.001
> 6	- 0.110	0.048	0.02	- 0.165	0.048	0.001
Daily number of meals						
1	Ref					
2	0.058	0.065	0.3	-	-	-
3	0.062	0.066	0.3			
> 3	0.266	0.09	0.003			
Weekly number of meals containing proteins	0.056	0.01	<0.001			
Self-reported poverty						
Not poor	Ref			Ref		
Moderately poor	- 0.285	0.057	<0.001	- 0.0362	0.057	<0.001
Poor	- 0.323	0.065	<0.001	- 0.440	0.064	< 0.001
Housing quality and assets ownership						
Type of roof						

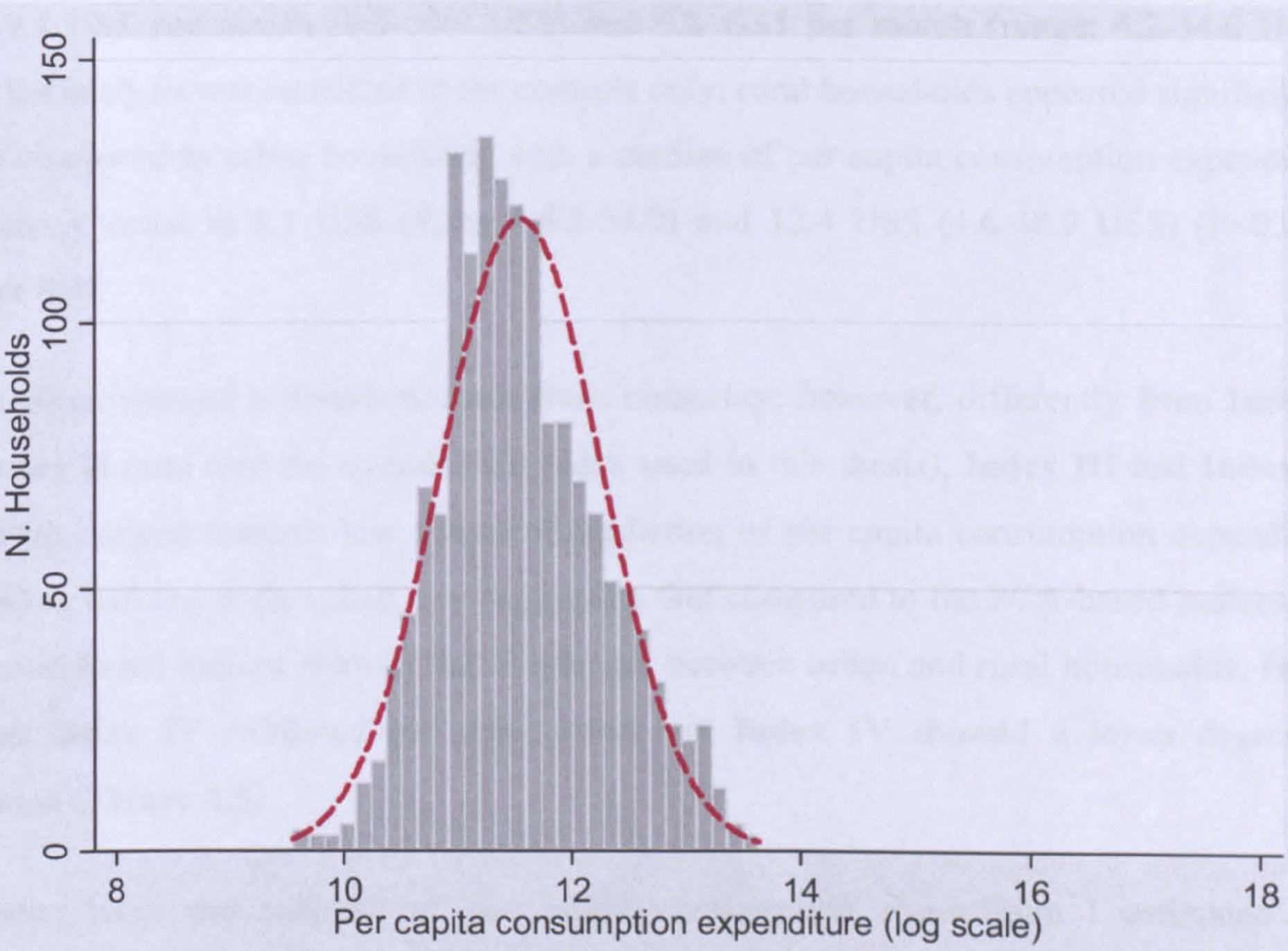
Asbestos	Ref			Ref		
Iron sheet	-.0.067	0.038	0.08	- 0.085	0.391	0.03
Grass	-0.180	0.092	0.05	- 0.194	0.093	0.04
Type of walls						
Concrete brick	Ref			Ref		
Mud brick	- 0.011	0.075	0.8	- 0.074	0.076	0.3
Burnt brick	-0.388	0.130	0.003	- 0.485	0.131	<0.001
Type of water sanitation						
Flush toilets	Ref			Ref		
Latrines	- 0.405	0.053	<0.001	- 0.422	0.054	<0.001
Other	- 0.462	0.110	<0.001	- 0.450	0.112	< 0.001
Having land for agriculture purpose	0.190	0.054	<0.001	0.217	0.054	< 0.001
Access to community services						
Type of water supply						
Piped inside the house	Ref			Ref		
Piped outside the house	0.277	0.054	0.001	0.313	0.055	< 0.001
Other	0.425	0.047	0.3	0.053	0.048	0.27
Energy source for lightening						
Kerosene	Ref			Ref		
Electricity	0.347	0.064	<0.001	0.393	0.064	< 0.001
Other	0.028	0.059	0.6	0.065	0.060	0.30
> 15 minutes walking distance from the nearest market	- 0.319	0.047	<0.001	0.354	0.047	< 0.001
> 30 minutes walking distance from the nearest clinic	- 0.143	0.048	0.003	- 0.157	0.049	0.001

Figure 8.3: Distribution of household per capita consumption expenditure for the province of Lusaka: actual and estimated values

A) Actual per capita consumption expenditure*



B) Prediction of per capita consumption expenditure* (Index III)



*from LSMSIII data [14]

Table 8.5: Households SEP ranking: the agreement between prediction and the actual per capita consumption expenditure

		Per capita consumption expenditure* N. Households per quintile (%)				
Index III		I	II	III	IV	V
	I	179 (57.6)	83 (26.7)	38 (12.2)	11 (3.5)	1 (0.3)
	II	83(26.7)	120(38.6)	73 (23.4)	29 (9.3)	7 (2.2)
Quintiles	III	37 (11.9)	67 (21.5)	90 (29.8)	86 (27.7)	30 (9.6)
	IV	10 (3.2)	33 (10.6)	85 (27.2)	117 (37.7)	67 (21.5)
	V	2 (0.6)	8 (2.6)	26 (8.3)	67(21.6)	207 (66.3)

*From LSMSIII data [14]

Compared to the actual per capita consumption expenditure data of the LSMS, **Index III** appeared to underestimate the proportion of households living below the extreme poverty line in the Lusaka province (25.8% compared to the 30.0% respectively for Index III and the actual values), but the overall proportion of households living below the moderate poverty line was approximately the same (47.5% compared to the 46.0% respectively).

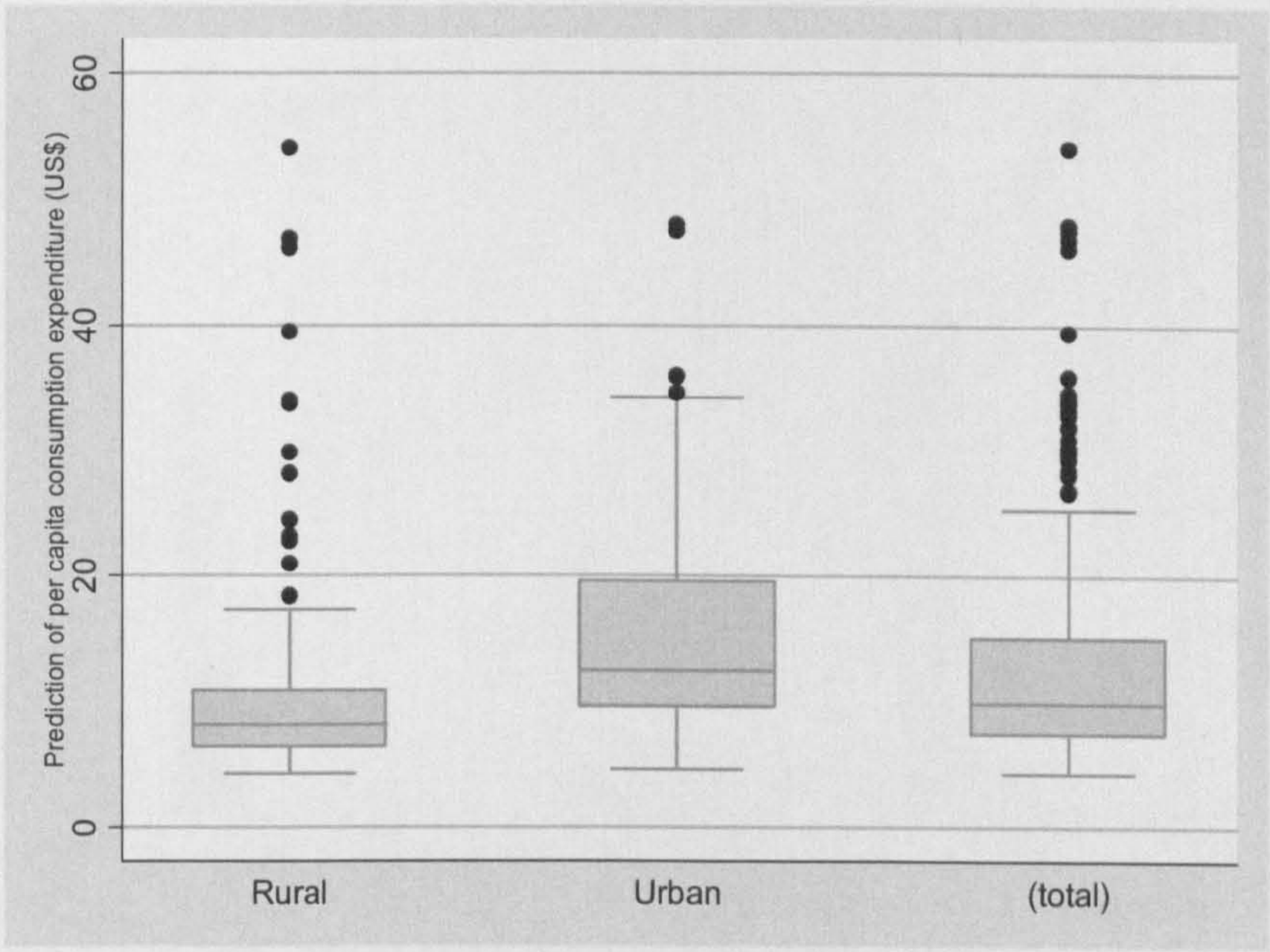
The regression coefficients so estimated were then applied to the case control study assets resulting in the estimate of two new variables for the prediction of per capita consumption expenditure for all the households enrolled in the study. The median per capita consumption expenditure did not significantly differ between cases and controls and was respectively equal 9.4 US$ per month (5.0-59.9 US$) and 9.8 US$ per month (range: 4.2-54.0 US$). When the analysis was restricted to the controls only, rural households appeared significantly poorer compared to urban households with a median of per capita consumption expenditure respectively equal to 8.1 US$ (Range: 4.2-54.0) and 12.4 US$ (4.6-48.0 US$) ($P<0.001$) **(Figure 8.4)**.

Both indices showed a distribution far from normality; however, differently from **Index I** and **Index II** (and also the overall SEP index used in this thesis), **Index III** and **Index IV** were both skewed towards low values of prediction of per capita consumption expenditure even when looking at the urban area suggesting that compared to the PCA-based indices, the regression-based indices showed less disparities between urban and rural households. **Index III** and **Index IV** exhibited the same trend, but **Index IV** showed a lower degree of truncation (**Figure 8.5**).

Following from the estimate of per capita consumption expenditure I estimated that respectively 60.7% and 76.1% of the controls were living below the extreme and moderate poverty line. As for what concern the cases, almost 90% of them were living below the moderate poverty line, whereas 78.8% of them were classified as extremely poor.

Figure 8.4: Boxplot of the two regression-based assets indices, overall and by study area

A) Index III

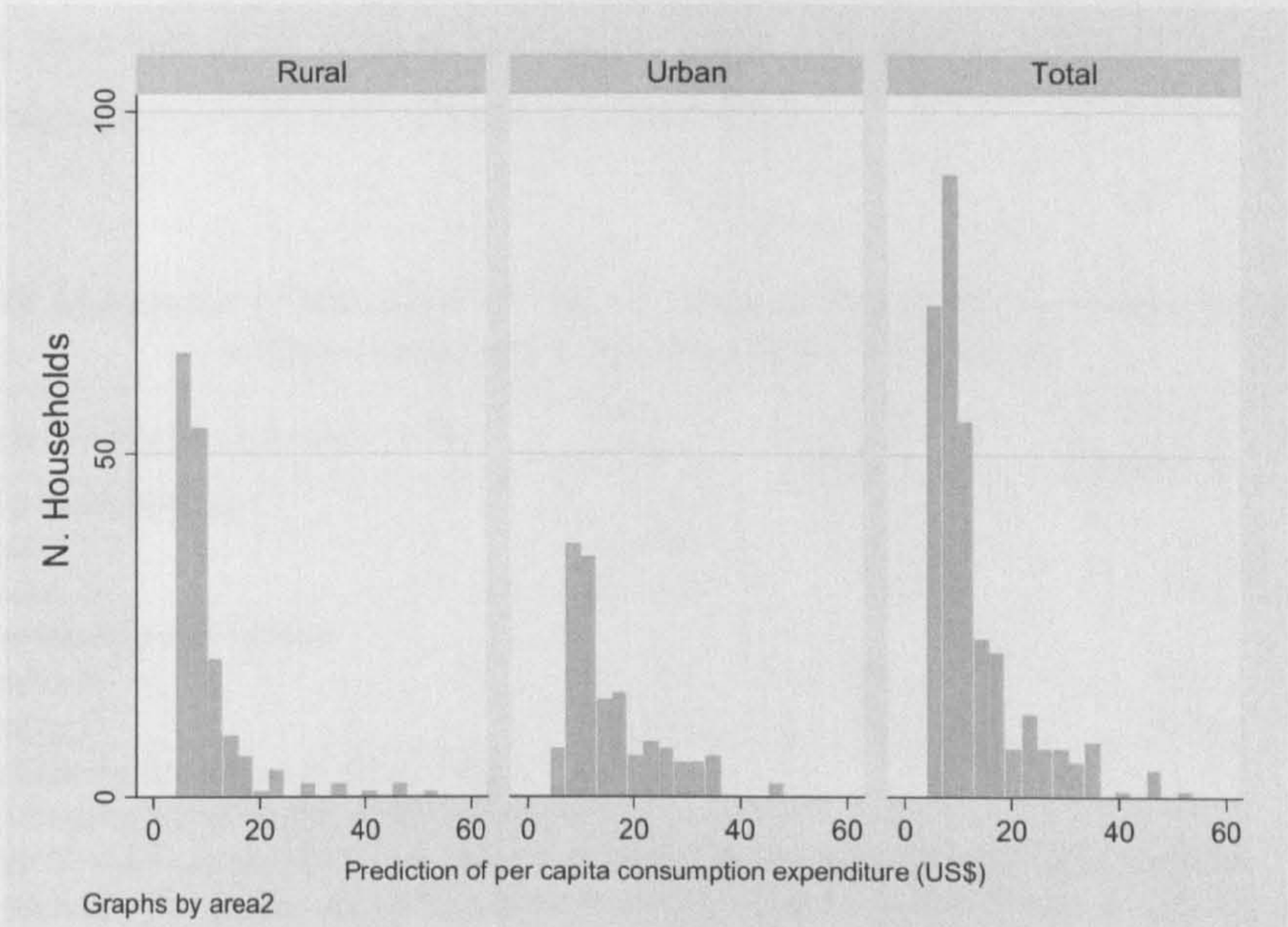


B) Index IV

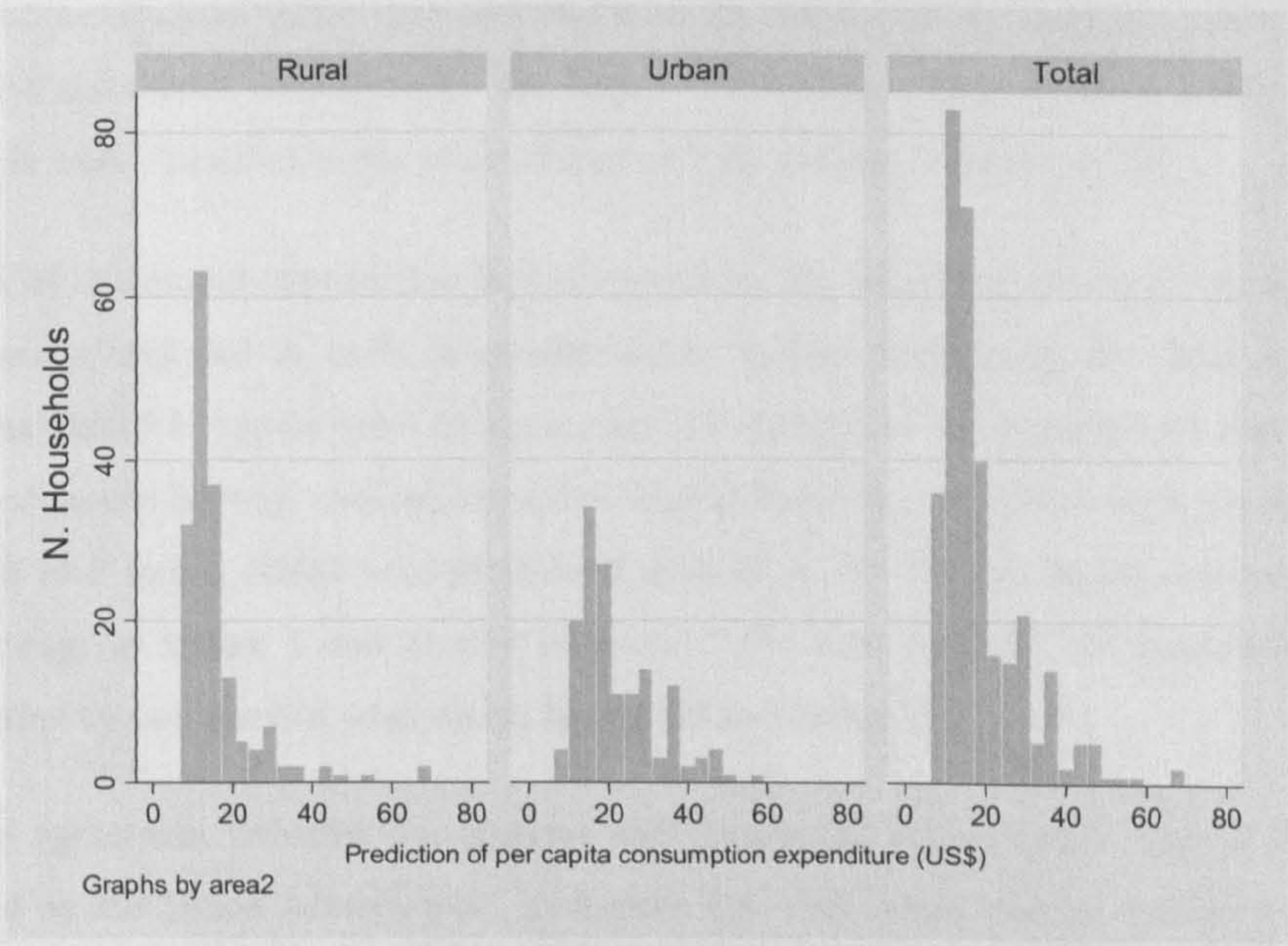
Index III: prediction of per capita consumption expenditure (including food related variables)
Index IV: prediction of per capita consumption expenditure (excluding food related variables)

Figure 8.5: Distribution of the prediction of per capita consumption expenditure (in US$) among the 318 control households enrolled in the case control study

A) Index III



B) Index IV



Index III: prediction of per capita consumption expenditure (including food related variables)
Index IV: prediction of per capita consumption expenditure (excluding food related variables)

8.4.2 Agreement among household socioeconomic position indices

(a) Level of household SEP misclassification

Table 8.6 shows the level of agreement of the indices created for the sensitivity analysis with the original household SEP index used in this thesis.

Table 8.6: Movement of households across terciles of household socioeconomi position indices compared to the one used in the thesis

% Households moving between terciles	Same tercile	Moving 1 tercile	Moving 2 terciles	Kappa
PCA-based asset indices				
Index I	82.7%	16.2%	0%	0.76*
Index II	81.8%	17.3%	0%	0.74*
Regression-based asset indices				
Index III	67.2%	30.2%	2.5%	0.50*
Index IV	64.8%	33.6%	1.6%	0.47*

Index I accounting for potential urban bias
Index II accounting for potential direct effect bias
Index III: prediction of per capita consumption expenditure including food related variables
Index IV: prediction of per capita consumption expenditure excluding food related variables
*P < 0.001

All the four indices showed a similar and high level of agreement with the reference index. The lowest level of agreement observed was observed for **Index IV** with approximately 65% of the households classified in the same tercile of the original index (kappa = 0.47), whereas the highest agreement was observed for **Index I** with approximately 83% of the controls households were classified in the same terciles of both indices (kappa = 0.76).

The level of agreement appeared to be influenced by the weighting strategy, more than the type of assets included in each composite index: indices built using the same weighting strategy exhibited a similar level of agreement. Compared to the regression-based indices, PCA-based assets indices showed overall a higher level of agreement with the reference household SEP index, which was itself built with PCA. No household moved two terciles when looking at **Index I** and **II**. By contrast, 2.5% and 1.6% of the households were misclassified by two terciles when using **Index III** and **Index IV**.

The good agreement between the original SEP index and Index I and Index II was also confirmed by the Bland Altman plots in **Figure 8.6**. The proportion of outliers (i.e. those above and below the limits of agreements defined by the mean of the difference of the two measurements ± 1.96 SD) was about 5% for both the indices considered and respectively

equal to 6.6% for **Index I**, 6.0% for **Index II**. In both comparisons, the outliers distribution appeared fairly random.

In order to explore which SEP domains the prediction of per capita consumption expenditure was best capturing I have also assessed the extent of agreement between **Index III** and **Index IV** and each of the SEP domain indices (**Table 8.7**).

Table 8.7: Movement of households across terciles of household socioeconomic domain indices and the indices of predicted per capita consumption expenditure

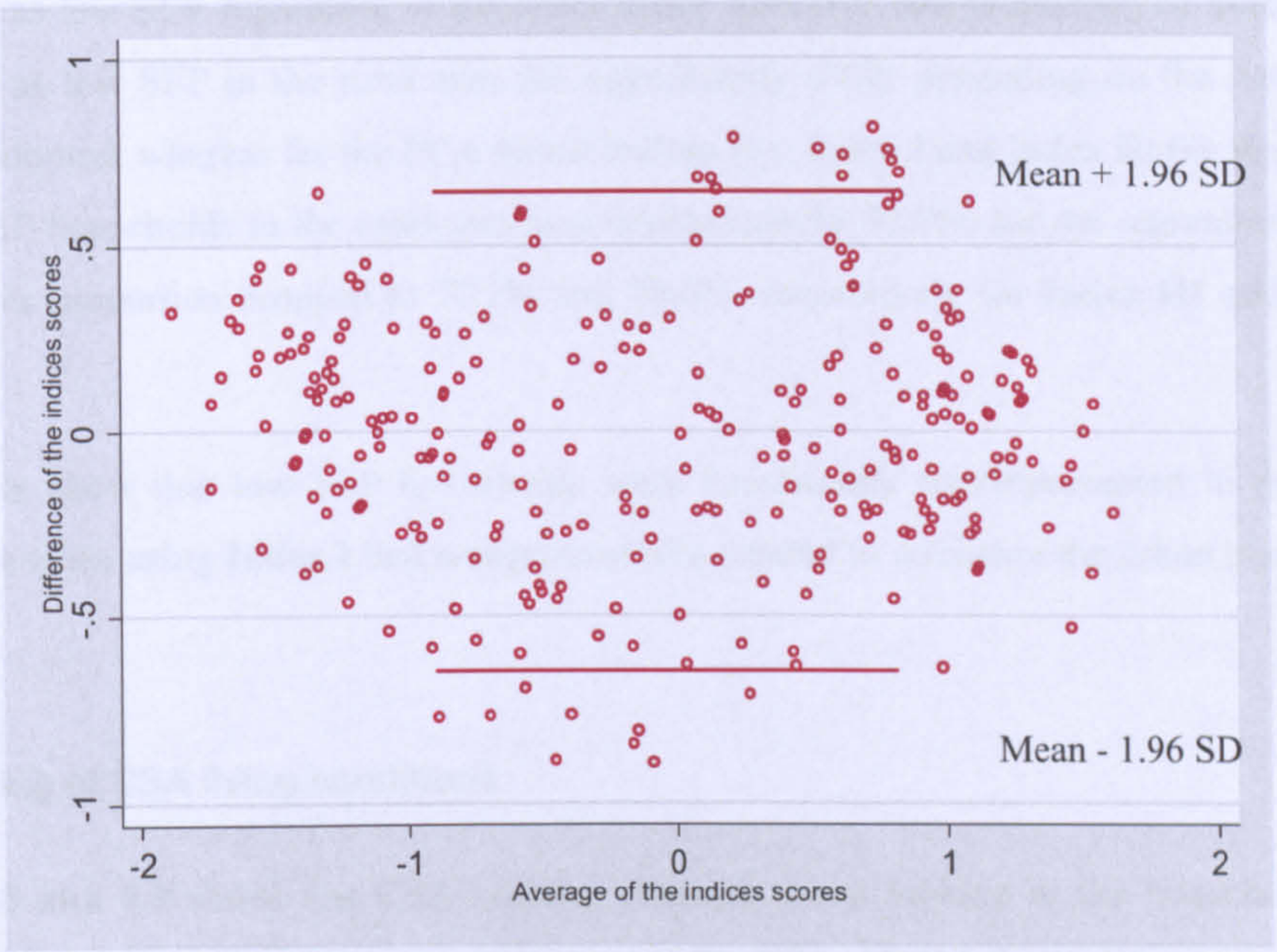
% Households moving between terciles	Same tercile	Moving 1 tercile	Moving 2 terciles	Kappa
Index III				
Human resources	45.9	35.5	18.6	0.18*
Food availability/vulnerability	50.6	40.8	8.8	0.25*
Assets ownership and housing quality	65.6	30.6	3.8	0.48*
Access to community services	59.6	34.4	8.9	0.39*
Index IV				
Human resources	46.5	34.6	18.9	0.19*
Food availability/vulnerability	45.6	42.8	18.2	0.18*
Assets ownership and housing quality	66.0	30.8	3.1	0.49*
Access to community services	60.4	34.0	5.7	0.40*

Index III: prediction of per capita consumption expenditure (including food related variables)
Index IV: prediction of per capita consumption expenditure (excluding food related variables)
*P< 0.001

The level of household SEP misclassification was much higher when comparing **Index III** and **Index IV** with the SEP domain indices than with the overall SEP index. However, assets ownership and housing quality domain showed a significantly higher level of agreement with the prediction of consumption (approximately 66% of households in the same tercile of **Index III** and **Index IV** and a kappa value of 0.5), suggesting that a considerable part of the consumption expenditure pattern observed for these households could be captured by this domain of SEP. **Index III** and **Index IV** also appeared to correlate well with the SEP domain describing access to community services (approximately 60% of the households in the same tercile corresponding to a kappa value of 0.4), whereas the lowest level of agreement with the food availability and vulnerability SEP domain index.

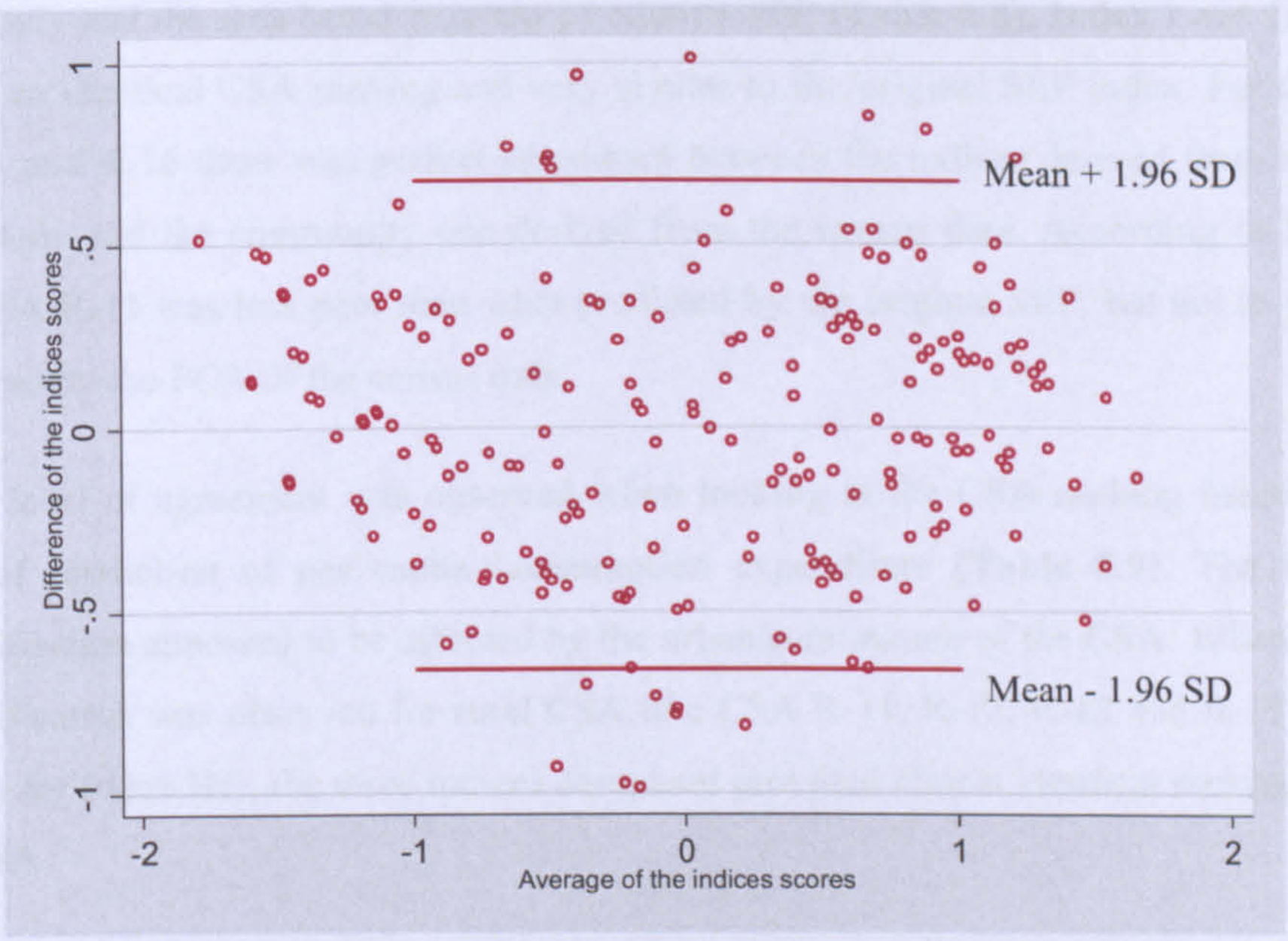
Figure 8.6: Bland Altman plots of the correlation between alternative household socioeconomic position indices with the index used in the thesis

A) Original SEP index and Index I*



*Accounting for the potential urban bias

B) Original SEP and Index II*



*Accounting for the potential direct effect bias

(b) Frequency of low SEP households across study sites

As for the original SEP index, rural households were disproportionately more likely to be classified as low SEP regardless of the index used. However, the proportion of households classified as low SEP in the rural area did significantly differ depending on the weighting strategy adopted: whereas for the PCA-based indices (eg. Index I and Index II) the frequency of low SEP households in the rural area was approximately 90.0%, for the regression based indices this proportion dropped to 77.1% and 78.0%, respectively for **Index III** and **Index IV**.

These data show that low SEP households were consistently overrepresented in the rural area, even when using **Index I** that was purposively created to minimise the urban bias.

(c) Ranking of CSA living conditions

Table 8.8 and 8.9 shows the CSA ranking obtained when looking at the household SEP indices derived from the case control study and 2 of the ABSM derived from the census dataset.

CSA are ranked very similarly when comparing the SEP indices derived from the case control study and the area-based measure of relative SEP **(Table 8.8)**. **Index I** and **Index II** provided an identical CSA ranking and very similar to the original SEP index. For CSA R-15, U-22, and R-16 there was perfect agreement between the indices derived from the case control study and the community one derived from the census data. According to **Index I** and **II** CSA R-11 was less poor than what predicted by the original SEP, but not as wealthy as appeared by the PCA of the census data.

A lower level of agreement was observed when looking at the CSA ranking based on the indices of prediction of per capita consumption expenditure **(Table 8.9)**. The level of misclassification appeared to be affected by the urban/rural nature of the CSA. Whereas high misclassification was observed for rural CSA like CSA R-11, R-12, R-13 and R-17 (in this case only for Index **III**), the three indices compared provided almost identical ranking for the urban CSA.

Table 8.8: CSA ranking according to the area based measure of relative socioeconomic position and the PCA-based assets indices

	CSA ranking order*			
	Census data based	Case control study based		
CSA	SEP Index[a]	SEP Index [b]	Index I	Index II
R-14	1	3	2	2
R-12	2	1	1	1
R-17	3	5	5	5
R-15	4	4	4	4
R-13	5	6	6	6
U-36	6	8	7	7
U-35	7	7	8	8
U-34	8	9	9	9
R-11	9	2	3	3
U-22	10	10	10	10
U-16	11	11	11	11

*CSA ranked from the poorest to the wealthiest according to the proportion of households classified as low SEP.
(a) PCA-based assets index derived from the census data. It represents the area-based measure of relative SEP.
(b) PCA-based assets index derived from the Case Control study data. It represents the household measure of relative SEP used in this thesis.
Index I accounting for the potential urban bias
Index II accounting fro the potential direct effect bias

Table 8.9: CSA ranking according to the area based measure of absolute socioeconomic position and the regression-based assets indices

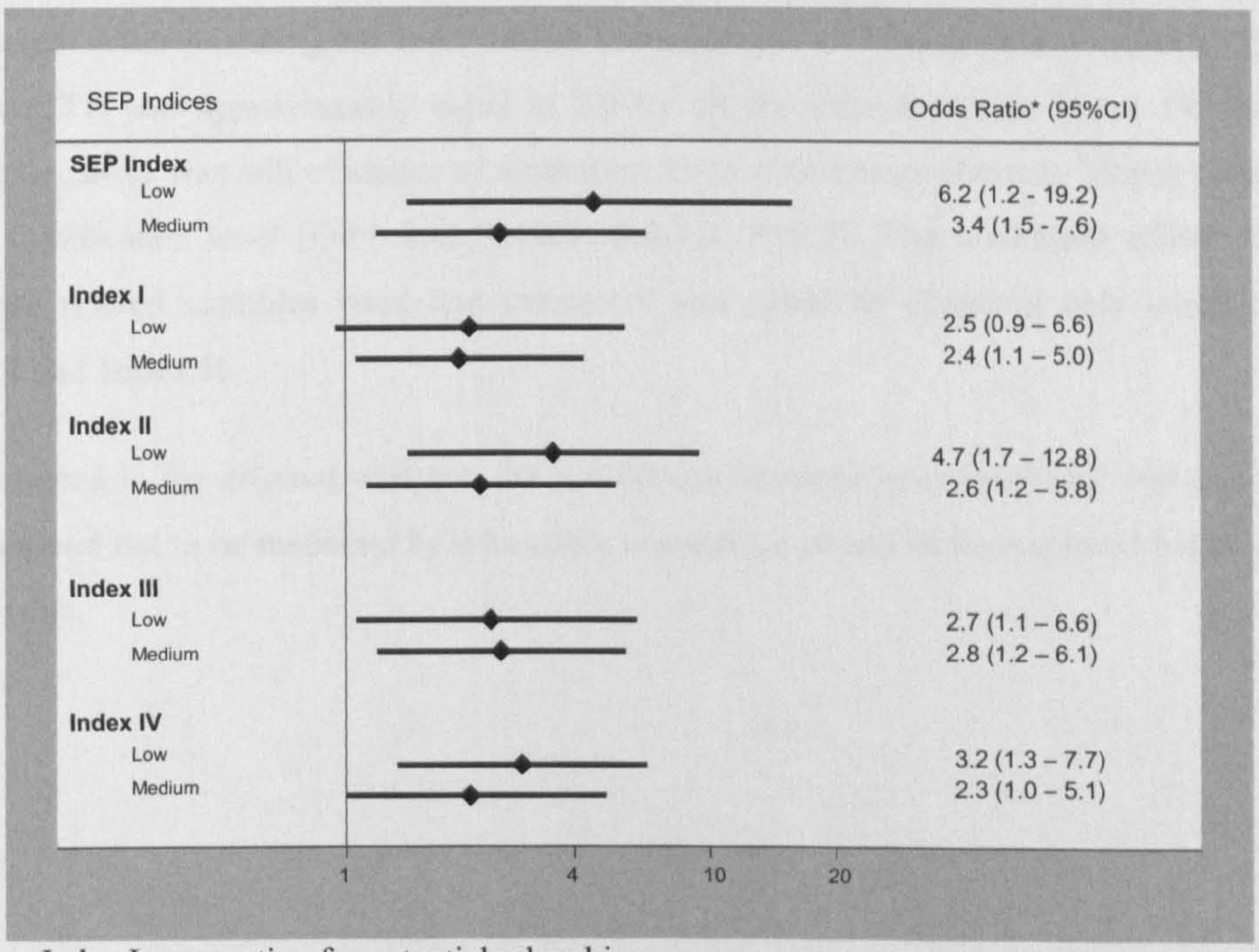
	CSA ranking order		
	Census based	Case control study based	
CSA	SEP Index[a]	Index III	Index IV
R-11	1	3	6
R-17	2	5	2
R-14	3	2	3
R-15	4	4	4
R-12	5	1	1
U-35	6	7	7
U-36	7	8	8
R-13	8	6	5
U-34	9	9	9
R-16	10	11	10
U-22	11	10	11

*CSA ranked from the poorest to the wealthiest according to the proportion of households living below the poverty line.
(a) It represents the area-based measure of absolute SEP estimated through the poverty mapping analysis of the Census data.
Index I accounting for the potential urban bias
Index II accounting fro the potential direct effect bias

8.4.3 Association with prevalent tuberculosis: the robustness of the results

The results of the minimally adjusted analysis are outlined in **Figure 8.7** showing the OR of low and medium household SEP compared to high. Low household SEP was strongly associated with prevalent TB according to all the indices considered, regardless of the inclusion of assets or the weighting strategy chosen.

Figure 8.7: The association between household socioeconomic position and prevalent TB: sensitivity of the results



Index I: accounting for potential urban bias
Index II: accounting for potential direct effect bias
Index III: prediction of per capita consumption expenditure (including food related variables)
Index IV: prediction of per capita consumption expenditure (excluding food related variables)
*Odds ratio minimally adjusted by sex, age group and area of residence. High SEP = Reference (not displayed in the graph)

The magnitude of the OR between low SEP and prevalent TB varies by indices, with the strongest association observed for the SEP index originally employed in this thesis, followed by **Index II** order, **Index IV**, **Index III** and **Index I**.

Differently from the household ranking, the extent of TB inequalities across SEP groups did not seem to be affected by the weighting strategy, for example **Index II** (PCA-based) showed results more similar to **Index IV** (regression-based), than **Index I**. In this regard, it is

worth noting that **Index II** and **Index IV** - those not including food related variables - were those showing the strongest association with prevalent TB after the original SEP index, suggesting small evidence for potential direct effect bias. By contrast, **Index I** showed the weakest association. A similar pattern was also observed when looking separately at the area of residence (data not shown).

The robustness of the results was confirmed also when looking at the multivariable analysis (**Table 8.10**): consistently with the SEP index originally used in this thesis, the multivariable analysis confirmed the mediation effect of food related variables for all the indices considered. After adjusting for food related variables, the OR between household SEP and prevalent TB was approximately equal to 2.0 for all the indices, except **Index IV**. In this latter case, there was still evidence of mediation, even if this association in Model 1 did not reach significance level (OR= 1.6, 95%CI: 0.8-3.3, P=0.2). The mediation effect of TB exposure related variables were less consistent and could be observed only when using **Index I** and **Index II**.

As presented in the original analysis, the association between household SEP and prevalent TB appeared not to be mediated by education, occupation or any of the explored behavioural risk factors.

Table 8.10: The assessment of mediation: sensitivity of the results

	Original results		Sensitivity analysis results							
			PCA based indices				Regression based indices			
	SEP index		Index I		Index II		Index III		Index IV	
	Adjusted OR (95% CI)	P-value	Adjusted OR (95% CI)	P-value	Adjusted OR (95% CI)	P-value	Adjusted OR (95% CI)	P-value	Adjusted OR (95% CI)	P-value
Model 1- Minimally adjusted										
High	1.0	0.01	1.0	0.01	1.0	0.01	1.0	0.02	1.0	0.2
Low	2.7 (1.2-5.9)		2.5 (1.2-5.3)		2.5 (1.2-5.2)		2.3 (1.1-4.9)		1.6 (0.8-3.3)	
Model 2 – Education/Occupation										
High	1.0	0.01	1.0	0.01	1.0	0.01	1.0	0.02	1.0	0.2
Low	2.7 (1.2-6.1)		2.6 (1.2-5.5)		2.6 (1.2-5.6)		2.4 (1.1-5.2)		1.6 (0.8-3.4)	
Model 3 – Food availability										
High	1.0	0.2	1.0	0.09	1.0	0.09	1.0	0.08	1.0	0.4
Low	1.8 (0.7-4.2)		1.9 (0.9-4.5)		2.0 (0.9-4.5)		2.0 (0.9-4.8)		1.4 (0.6-3.1)	
Model 4 – Biological/Behavioural risk factors										
High	1.0	0.04	1.0	0.008	1.0	0.03	1.0	0.07	1.0	0.2
Low	2.6 (1.1-6.3)		3.2 (1.3-7.6)		2.5 (1.1-5.8)		2.2 (0.9-5.2)		1.8 (0.8-4.5)	
Model 5 – TB exposure										
High	1.0	0.04	1.0	0.04	1.0	0.03	1.0	0.02	1.0	0.2
Low	2.3 (1.1-5.2)		2.2 (1.1-4.8)		2.3 (1.1-4.9)		2.4 (1.1-5.3)		1.6 (0.8-3.6)	

Model 1: SEP adjusted by sex, age group and area of residence (minimally adjusted)
Model 2: SEP adjusted by sex, age group, area of residence, education and occupation
Model 3: as in model 2 plus food intake related variables
Model 4: as in model 2 plus behavioural risk related variables
Model 5: as in model 2 including TB exposure related variables

Summary

The results of the sensitivity analysis suggest that the household SEP ranking originally described in this thesis was slightly affected by the choice of different SEP indices. This was confirmed both graphically and when looking at the frequency of household SEP misclassification across SEP terciles: depending on the index considered, between 16% and 33% of the households were ranked differently from the SEP index used in this thesis. More than the different inclusion of assets, the weighting strategy appeared to be predictive of the extent of agreement between indices in terms of household SEP ranking. PCA-based assets indices exhibited a higher level of agreement with the original SEP index (which was itself based on PCA), compared to the regression-bases assets indices.

The removal of urban-specific assets from one of the SEP index did not reduce the differences observed between urban and rural households. Regardless of which index considered, low SEP households were still disproportionately more likely to be found in the rural area. This gap was slightly reduced by using regression-based assets indices, but still 80% of the low SEP households were rural.

The use of regression analysis allowed the computation of the proportion of households recruited in this study living below the extreme and moderate poverty lines confirming the initial observation made in this thesis of a study population living under severe material constraints.

Despite the high proportion of households living below the poverty line, there was still a detectable difference between cases and controls with the cases significantly more likely to belong to low SEP households compared to the healthy controls regardless of the SEP index used. This is the second most important result of the sensitivity analysis and suggests that, at least in this specific setting, the existence of TB inequalities seems to be robust to the choice of household SEP measurement strategy. Noticeably, TB cases were more likely to be living in low SEP households even when excluding food-related variables from the SEP index.

The choice of different SEP index did not even appear to affect the observed causal pathway, with food related variables still playing an important mediation role in the association between household SEP and TB.

The only observation possibly affected by the choice of alternative SEP indices was the magnitude of the TB inequalities observed. However, it is difficult to say whether the differences observed are epidemiologically relevant.

9. Discussion

If social epidemiologists are to gain clarity on causes of and barriers to reducing social inequalities in health, adequate theory is a necessity, not a luxury. The old adage still stands: 'if you don't ask, you don't know and if you don't know, you can't act'. Ultimately, it is theory which inspires our questions, which enables us to envision a far healthier world than the one in which we live, and which gives us the insight, responsibility, and accountability to translate this vision to a reality. Who shall create this theory? The task is ours.

Krieger, N[10]. 2001

Introduction

The first edition of the British Journal of Tuberculosis in 1907 included a paper entitled "Civilization and tuberculosis" which opened with the following statement: "The tubercle bacillus is an index by inversion of the real progress of the human race. By it the claim of civilisation to dominate human life may fairly be judged" [246]. This thesis was essentially an attempt to see whether 100 years later this statement is still true, whether it is valid for a developing country like Zambia, and - if so – how, in this setting, the association between socioeconomic conditions and TB can be explained and possibly addressed.

In this chapter, I will discuss the results of this research project. The chapter has been structured as follows: after an overview of the study findings, I will discuss the strengths and limitations of the research. I will then explain the results in the light of the previous research and describe what this thesis adds to our understanding of the association between household SEP and TB and which knowledge gaps remain unsolved. I will conclude by discussing the study implications, both in terms of TB control policies and future research agenda.

9.1 Study results overview

To provide a better synthesis of the main study results the overview has been organised by study component in the text below and by study objective in **Table 9.1**.

9.1.1 The ecological study

The combined analysis of the Census dataset for the province of Lusaka (year 2000), together with the 2002-03 Living Conditions Monitoring Survey (LSMSIII), revealed that all the 11 census standard areas (CSA) involved in the latest Zambian population-based TB-HIV prevalence survey were generally characterised by very poor living conditions. Across CSA, the proportion of households belonging to low socioeconomic position (SEP) ranged between 10 to 70%, whereas the proportion of households living below the extreme poverty line reached 30% in some CSA.

Compared to rural CSA, urban CSA were consistently characterised by significantly wealthier living conditions: the only exception was represented by CSA R-16 that, despite belonging to the rural site, is located in the part of Chongwe crossed by a main road and characterised by urban-like features (like the food market, a primary school, the health care centre etc.). Urban CSA appeared to bear a greater burden of TB compared to the rural ones with TB prevalence rates that in some cases were up to three times higher than the average TB prevalence rates observed in the rural areas (**Table 9.1**).

At ecological level, the available data showed no evidence of significant spatial correlation between the TB prevalence rates and any of the area-based SEP indicators developed. However, there was a trend for an increase of TB prevalence with the worsening of community living conditions. This observation was largely confirmed when the analysis was repeated by study site (i.e. urban and rural). Among the rural CSA, TB prevalence rates were positively and significantly correlated to all the area-based indices of employment opportunities to the extent that it can be concluded that at least in terms of structural measures of community SEP, rural CSA with lower SEP have higher TB prevalence. Nonetheless, the area stratified analysis also showed that among rural CSA, most of the area-based indices showed an inverse association with the rates of prevalent TB in the rural CSA (**Table 9.1**). Because of the small number of communities included in the analysis results should be not over interpreted.

9.1.2 The case control study

The case control study showed that, at household level, prevalent TB was significantly associated with household SEP. TB cases were approximately 3 times more likely than controls to belong to low SEP households after adjusting for age, sex and area of residence. The exploratory analysis of the household SEP domains showed that most of this association was mainly driven by the food availability and human resources domain.

The individual risk factors analysis revealed that in this population, TB prevalent cases were significantly more likely to have a diet poor in proteins, to be not BCG vaccinated, to be HIV positive and to have migrated in the 5 years before. Surprisingly, attending church showed a significantly protective effect against prevalent TB. These associations all persisted after controlling for the confounding effect of household SEP.

When the household SEP and individual risk factors were hierarchically combined according to an a priori defined conceptual framework, the data revealed that the association between household SEP and TB was at least partially mediated by inadequate nutrition. There was some evidence of mediation also when the pathway of contact with known TB cases was explored, but this effect was less prominent than inadequate nutrition. Data did not seem to suggest any mediation from HIV or any other biological and behavioural factors. This should not be surprising since in this population the risk of HIV was not found to be associated with household SEP.

This study allowed the identification of the risk factors more strongly associated with prevalent TB (those with high OR) and those able to explain the most of the cases observed (those with high PAF): from the PAF computation it resulted that, *before* considering any mediation effect, household SEP was able to explain approximately 30% of the cases of TB enrolled At individual level, four variables appeared to contribute to the number of cases observed after accounting for confounding: BCG vaccination, HIV infection, inadequate nutrition (here defined as the weekly number of meals containing proteins) and migration.

Despite the strength of the association between the lack of BCG vaccination and prevalent TB only 11% of the cases enrolled in this study could be attributed to this exposure, due to the small number of unvaccinated people. HIV and having less than 2 meals containing proteins were equally important risk factors for prevalent TB. However, when the frequency of these two risk factors was also taken into account, it appeared that in this setting, inadequate nutrition (PAF=42%) could explain more cases of prevalent TB than HIV (PAF=36%) (**Table 9.1**).

Table 9.1: Summary of the main study findings by research objectives

Study objectives	Results
To describe the community and household living conditions of the people involved in the latest Zambian population-based HIV-TB prevalence survey	• Among the 11 CSA involved in the prevalence survey, the proportion of households classified as low SEP ranged between 8.3% and 68.4%. • In these communities, the proportion of households living below the extreme poverty line ranged between 0% and 29%. • Urban CSA were generally wealthier than rural ones.
To describe the ecological association between community living conditions and TB prevalence rates detected in the prevalence survey	• 79 cases of prevalent TB were detected in the prevalence survey, yielding a crude prevalence rate of 980/100,000 and a cluster adjusted prevalence rates of 870/100.000 (95%CI: 570-1160/100,000). The cluster adjusted prevalence rate of TB was almost as twice as high in the urban area (1200/100,000, 95%CI: 750-1640/100.000) compared to the rural one (650/100,000, 95%CI: 360-940/100,000). • Although not statistically significant, TB prevalence rates seemed to increase with worsening of living conditions both among the rural and urban CSA. Nonetheless, in the rural areas TB prevalence rates appeared to decrease with the worsening of living conditions for 7 of the 13 indicators of community living conditions.
To estimate the association between SEP and cases of TB disease detected within the prevalence survey	• Low household SEP was significantly associated with prevalent TB (OR=6.2, 95%CI: 2.0-19.2 and OR = 3.4, 95%CI: 1.8-7.6 respectively for low and medium household SEP compared to the baseline). Low household SEP was associated with prevalent TB in the urban area, but not in the rural one. • Prevalent TB was also associated with having no meals containing proteins/week (OR=3.1, 95%CI: 1.1-8.7), lack of BCG (OR= 7.7, 95%CI: 2.8-20.8), HIV infection (OR = 3.1, 95%CI: 1.7-5.8), having migrated (OR = 5.2, 95%CI: 2.7-10.2), having been in contact with a TB case (OR=2.8, 95%CI: 1.3-5.6). All these associations persisted after controlling for household SEP. • PAF for household SEP was 30% before taking into account any mediation effect. Adjusted PAF was equal to 42% for number of meals with proteins/week, 36% for HIV, 23% for migration and 11% for not having BCG.

Table 9.1: Continued

Study objectives	Results
To quantify the association between household SEP and TB infection among a disease-free population	• High household SEP increased significantly the risk for TB infection (OR= 0.4, 95%CI: 0.2-0.9 and OR= 0.4, 95%CI: 0.2-0.8 for high and medium household SEP compared to low SEP). • The analysis of single socioeconomic indicators showed that even when the association did not approach statistical significance, prevalence of TB tended to be always higher in the higher SEP category of the socioeconomic indicator investigated.
To assess which aspects or domains of household SEP are the most important driving force accounting for the associations under study.	• The association between low household SEP and prevalent TB was mostly driven by the food availability and vulnerability household SEP domain and the human resources household SEP domain. • The same analysis showed that the association between high household SEP and TB infection was mainly driven by the access to community services domain of household SEP.
To investigate the mediation effect of individual risk factors included in the hypothesised causal pathway between household SEP and TB (both disease and infection).	• The association between low household SEP and prevalent TB appeared to be mediated by food availability related variables. There was some evidence of mediation also from the TB contact related variables. • In the association between low household SEP and prevalent TB, there was no evidence of mediation from the biological-behavioural related variables or the human resources ones. • None of the postulated individual risk factors seemed to mediate the association between high household SEP and TB infection.
To verify to which extent the study findings are robust to the choice of the household SEP indicators	• The association between household SEP and TB was robust to the choice of household SEP indicator. • There was some evidence that the choice of different household SEP indicator may affect the extent of the TB inequalities observed.

Apart from the ecological analysis and the PAF estimates, all the results presented so far were confirmed by the sensitivity analysis, suggesting that the association between household SEP and prevalent TB was robust to the choice of household SEP measurement. In particular, the sensitivity analysis helped to rule out the possibility that the results observed were biased by the inclusion in the SEP composite index of food-related variables, which are known risk factors for TB. The sensitivity analysis showed that, although always significant, the magnitude of the TB inequalities observed was indeed sensitive to the choice of SEP indicator. The choice of different SEP index did not appear to affect the observed causal pathway, with food related variables consistently playing an important mediation role in the association between household SEP and TB **(Table 9.1)**.

9.1.3 The cross-sectional study

Contrary to the study hypotheses TB infection was associated with higher, rather than lower, household SEP. TB infection was consistently associated with markers of higher household SEP even when looking at the single proxies of household SEP (like housing quality variables), suggesting that this result was unlikely to be spurious. The analysis of the different household SEP domains reinforced further this observation: households closer to community services and infrastructures as well as those characterised by a more frequent and better food intake were also those more likely to have a TB infected household member. Once included in a multivariable model, only easier access to community services and infrastructures remained significantly associated with increased odds of TB infection. As for TB prevalent disease, the study was designed to explore two levels of influence: the household and the individual. When these two levels were hierarchically included in the multivariable analysis, it resulted that the effect of higher household SEP was not captured by any of the postulated mediating factors operating at individual level. This was not surprising since none of the exposures investigated in this study (and commonly thought to be associated with TB infection) was associated with the outcome. The only other risk factor associated with TB infection was household crowding, confirming the importance of this risk factor in the epidemiology of TB transmission. Interestingly, the association between crowding and TB infection was not confounded by household SEP **(Table 9.1)**.

Because of a number of limitations that will be discussed later in this chapter these results should be interpreted with caution.

9.2 Added value and weaknesses of the study

In order to better understand the meaning and the potential impact of the results summarised above, it is essential to first consider what the added value of the current study is, compared to previous research in this field, and what design and conduct limitations may hamper the study validity.

9.2.1 Strengths

To my knowledge, this study is the first analytical study on the association between household SEP and TB, nested into an HIV-TB prevalence survey. As mentioned in the background chapter, the focus on prevalent cases of TB has the considerable advantage to minimise the bias from studying incident cases of TB (that spontaneously go to the health care centres), often resulting in the underestimate or even the observation of a reverse association between SEP and TB [7] [22]. Because of their prolonged infectiousness, prevalent cases are epidemiologically more relevant for TB transmission than incident cases. As a result, a study on prevalent cases of TB like this one offers the opportunity to understand better how SEP sustains TB transmission and can be potentially interrupted.

Because of the nested design of the infection study within the case control study, this research project offered a unique opportunity to compare the effect of household SEP in two different stages of TB (i.e. disease and infection). As a result, any observed difference is more likely to arise from a genuine differential role of household SEP rather than from methodological differences when comparing across studies, such as the involvement of different study populations or the adoption of different household SEP data or measurement strategies.

Michael Marmot has suggested that much social epidemiology has spent too much time looking for an association between socioeconomic factors and health and is happy to conclude that, like in this case for example, 'poverty causes TB' without asking why [104]. Probably, the most relevant added value of the current study is exactly this: the attempt to understand the mechanism underlying the link between household SEP and TB, rather than simply quantifying this association. In this attempt I was guided by a conceptual framework characterised by multiple levels of causation, including the individual level (accounting for the risk factors commonly investigated in traditional TB epidemiology), and the household level (including those social determinants acting beyond and above the individual risk factors). It is the introduction of this conceptual framework, the most innovative element of this study and the one that makes the study results able to suggest interventions that can

hopefully advance TB control policies and reduce TB inequalities. The conceptual framework was inspired by the material/neo material theory of diseases distribution: to date, only few studies have tried to test this model with an analytical approach and even less have formally employed it in the interpretation of TB inequalities. This model was chosen for pragmatic reasons (limited resources hampered the possibility to test the psychosocial model), but also because of my understanding of the determinants of TB inequalities in this study population.

Finally, this study is also one of the few where the association between these social determinants and prevalent TB and the underlying mechanism has been explored using a range of different SEP measurement strategies. The sensitivity analysis suggests that my results were robust to the choices of household SEP measurement available in this research.

9.2.2 Limitations

In this section I will describe the main 'design' limitations of this research. Minor limitations will be discussed throughout this chapter whenever appropriate.

(a) Reverse causality

A major methodological limitation of this study was its cross-sectional nature. Despite the strength of the evidence reported, it is difficult to make conclusive inference on the effect of household SEP on the risk of TB infection or developing TB because in this research I could not accurately measure the effect of this exposure *before* the onset of TB infection and disease. This temporality issue is complicated by the fact that TB disease is known to have a strong impoverishing effect on individuals and households [5, 183]. As a result, in cross-sectional studies it is almost impossible to say whether affected cases have TB because they are "poor" or they are "poor" because they have TB. Because of the "vicious cycle" between low SEP and TB [184], the risk is that what researchers are actually observing is the impoverishing effect of TB rather than the effect of socioeconomic factors on the risk of TB. The main consequence of this phenomenon, also called *reverse causality* or *endogeneity*, is that in the lack of a clear chronological sequence between the exposure and the outcome, one cannot infer a causal role for socioeconomic factors on the risk of TB.

The only design able to overcome the issue of reverse causality is a longitudinal study following a cohort of patients, in which SEP is measured before the onset of TB. Since TB is

a relatively rare disease, a study of this type would require a very large sample size and considerable follow up period. Both these requirements are incompatible with the budget and the time normally allowed to a PhD thesis. An alternative approach would be a trial where poverty reduction strategies are randomly allocated to communities or people to see how these interventions translate in the prevention of TB, but - again - time and budget limits make also this design unsuitable for the purpose of a PhD.

In this thesis I could only try to minimise the reverse causation bias by choosing on SEP measurement strategy based on assets rather than econometric types of measurements. Differently from income and expenditure, assets are considered to be "slow moving" [85]; in other words, even important changes in the household SEP may leave these assets virtually unchanged in the medium-long term [69]. If this is the case, it can be assumed that the household SEP I measured in this study was likely to reflect the living conditions of these households as they were *before* the onset of TB in the case.

Another approach would be to collect information on the duration of the symptoms and to restrict the study to patients reporting the shortest duration. In this way the time allowed for the impoverishment effect of TB to take place would be reduced and so the reverse causality bias. This option was not deemed to be appropriate: not only the prevalence survey showed that almost 70% of the cases detected had symptoms for more than three weeks before being diagnosed [247], but also the already small number of cases available made this sub-group analysis unfeasible. The sample size limitation is probably the second most important limitation of this study.

(b) Sample size

Conclusions concerning the results of the case control study can be hampered by the small sample size available. This limitation was somewhat inevitable as TB is a relatively rare disease. Even high TB burden countries rarely present with prevalence rates higher than 1200/100.000 population. As a result, studies based on prevalent cases of TB often rely on a small number of cases, which is hardly bigger than a few hundreds even in prevalence surveys involving large study population [248]. This is consistent with what was observed in the latest prevalence survey in Zambia, where overall 79 TB cases were identified.

This number was already below the minimum sample size requirement (i.e. approximately 100), but the number of cases included in the case control study was reduced even further to 52 as the consequence of change in the microbiological case definition. This decision was

made in order to ensure a case definition consistent with the one adopted in the prevalence survey and to make sure that only "real" cases of TB were eventually included in the case control study; however, it could severely affect the power of this study. Surprisingly, this did not happen: although some measures of association are likely to have lost precision, they turned out to be quite large and thus unlikely to be explained by chance. These observations can be only explained by assuming that the original sample size calculation had underestimated the strength of the association under study (with an OR for household SEP well above the predicted value of 2.0). In addition – low study power should be of less concern as - despite the risk of imprecise estimates – it is unlikely to introduce biases.

Another problem related to the loss of cases is the observation that most of the cases excluded came from the rural area. In principle, this could have resulted in the reduction of the strength of the association between household SEP and TB (because rural households are poorer in this study). However, the individuals excluded were actually non-TB cases and their exclusion has most likely reduced the chance of any kind of dilution effect in the association between household SEP and TB disease. Since these two things balance each other it is unlikely that this issue has influenced my study results. It is also to be said that the exclusion of individuals mainly from the rural area has nothing to do with their household SEP, but it is probably just a reflection of the order in which samples were processed and therefore contaminated in the lab.

For the infection study, it is difficult to judge how underpowered the study was since no sample size calculation was undertaken. Formal *a priori* calculation of the sample size may be useful when planning a new study, but in investigations like the infection study the analysis was performed on data that were already available and could not be modified. In these circumstances, authors are recommended to indicate clearly the conditions that determined the study size [232] and are discouraged from performing *post hoc* power calculation [249].

Despite the potential power limitations, the nested design of the infection study represents an unusual, but particularly convenient strategy to address another important research question (i.e. the association between household SEP and TB infection) with virtually no extra resources investment required. The results obtained showed an unexpected effect of household SEP on the risk of TB infection. Low study power can lead to misleading conclusion and these results should be considered far from being definitive; however, it is unlikely that in this case the limited study power could provide a plausible explanation for these unexpected findings. Further limitations deriving from the cross sectional study design will be discussed later in this section.

The sample size limitations may also explain the results of the ecological analysis: the lack of any significant association between the area-based measures of living condition and the rates of prevalent TB can be due to the small number of CSA included in the analysis (e.g. only 11). The study power was further limited by the fact that these communities were all close and relatively similar to each other.

(c) Potential selection bias arising from the replacement of controls

In this study approximately 30% of controls needed to be replaced because the ones originally selected were not available for the case control study. The comparison of the controls excluded with those eventually recruited in the study did not show large difference between the two groups. However, because very few variables were available for the comparison, selection bias cannot be ruled out. Should the controls included in the study be systematically wealthier than those who could not participate, this would result in an over-estimate of the strength of the association between household SEP and prevalent TB.

Death controls could be attributed to TB, but this hypothesis is impossible to verify. Even if we assume that they died of TB, for this to be a bias for the study we should also imply that the likelihood of dying of TB is different across different SEP groups. Again, this is not known. In Zambia, for example, TB mortality is increased by HIV infection however in this study we could not find any association between household SEP and HIV status.

(d) The gap between the conceptual treatment and the empirical measurement of household SEP

Someone may argue that after the discussion of structural determinants of health inequalities, in this thesis the definition and the measurement of household SEP is mainly based on assets and consumption which are outcome of the stratification process, but that in themselves do not generate or reinforce inequalities. The lack gap between the conceptual definition of household SEP and its empirical measurement is well known problem in studies on health inequalities from developing countries [250]. An explanation of this gap can be mainly attributed to the fact that the bulk of definitions of SEP or SES derive mainly from industrialised countries, whereas definitions and conceptualisations specifically addressing the issue in developing countries are a lot scarcer. Consider occupation, for example: this variable is often considered one of the key dimensions of socioeconomic stratification in

industrialised countries and as a result a great deal of attention has been given both to the classification of occupation status and its measurement. As opposite, very little attention has been given to occupational status in studies on health inequalities in developing countries, primarily due to somewhat constrained labour market conditions [250]. One exception is represented by a study from Bills and colleagues that tried to develop an occupational scale for Brazil based on occupational status and education, but this approach has not been widely adopted [251]. In a review aimed to assess the SES definitions and measurement methods adopted to measure fertility inequalities in developing countries, authors noted that occupation was included relatively infrequently.

The use of assets and food consumption as an indicator of SEP is an important difference between in how researchers measure SEP in developing countries compared with measures in industrialised countries. As for other studies of health inequalities in developing countries, even in this thesis, the use of assets data reflects the difficulty of obtaining accurate income data, the lack of meaningful definition of occupational status and education level and the widespread use of LSMS surveys that collect assets and housing data, but no information about income or consumption.

In the thesis I have acknowledged the importance of employment status of the head of the household. However, using only this construct would have resulted in a poor discrimination between households: only a small proportion of this study population declares to be unemployed (as shown from only the 8% of unemployment among the controls recruited in this study) and among those declaring to be employed or self-employed is difficult to judge what this exactly mean. In Zambia most of the people run some kind of income-generating activities (which collectively fall under my category of "self-employment"), but these activities are rarely within the formal sector and it is difficult to judge people's *position* from it. In order to overcome the poor definition and measurement of occupational status I have tried to develop an index of household SEP following from this question: what are the most relevant attributes that define household SEP in this setting? Starting from this question, I have developed a multidimensional index of household SEP based on four domains: each domain was selected according to my understanding of how people in this setting rank household SEP and my understanding of which aspect of SEP can be relevant to TB epidemiology. Although I could not measure them directly in this project, I have assumed that these domains of household SEP were all determined by socioeconomic factors at context level. In agreement with the conceptual framework of CSDH, household SEP was still conceptualised as a structural determinant of TB inequalities, but – as often happens in developing countries – I had to measure this determinant using an asset-based approach.

(e) The methodological and conceptual limitations of using an asset-based index for the measurement of household SEP

In this thesis the results interpretation may be challenged both methodologically (because of issues related to the use of PCA) and conceptually (because of the uncertainties around the meaning of the assets-based index).

Methodologically, the use of PCA has been the subject of at least three main criticisms:

1. It suffers from an underlying lack of theory to motivate either the choice of variables or the estimates of the weights.

2. The use of just the first principal component is questionable since even this component frequently explains a low proportion of the common variance of the assets included in the PCA (often less than 20%) [85] [223] [225]. As a result, the generated index could capture only a small of the household SEP, resulting in a misclassification of the household SEP ranking. Although many authors have suggested to look beyond the first principal component, this would not have been particularly useful in this case: the aim of the PCA was, in fact, the generation of a single composite index of household SEP to be used as main exposure in this study. Considering multiple components would have made the interpretation of the results actually more confusing. Furthermore, it is unlikely that using the subsequent components would greatly increase the explanatory power of the index [192].

3. Finally, significant debate concerns the fact that PCA is a method designed for use with continuous, normally distributed, data and its application to discrete data in a SEP index is therefore questionable [252]. Alternative strategies to PCA include the use of equal weight, the use of the inverse of the proportion of households that own an asset as its weight [252], the latent variable approach [113, 253] and Multiple Correspondence Analysis (MCA), particularly appropriate for the analysis of categorical data commonly collected for most assets. While the use of equal weight has been often considered too simplistic (as it assumes that different indicators can have equal meaning in terms of SEP), the inverse approach and MCA have been equally criticised as considered statistically too difficult and unsuitable for most researchers. In a recent paper, Howe and colleagues have tried to compare wealth indices constructed using different weighting strategies (i.e. equal weight, PCA, inverse of the proportion approach, and MCA) and different ways of coding variables (i.e. binary versus ordinal) [221]. They found that the way variables were coded was far more important than the weighting strategy adopted in predicting the extent of agreement among indices. They also found

that neither the use of a different weighting strategy or the inclusion of variables coded differently led to a significant improvement in the ability of these indices to proxy consumption expenditure: all the indices showed a similar level of misclassification between quintiles when the households ranking of the wealth indices was compared to the one derived from per capita expenditure. The authors also showed that the indices generated by PCA and MCA presented with high agreement and demonstrated very similar agreement with consumption expenditure, suggesting that the use of discrete data with PCA can be of limited concern. Given the limitations of all the alternatives presented in their analysis and the apparent small advantages offered by more sophisticated statistical approaches (e.g MCA), the authors concluded that there seemed to be little reason to adopt any of these alternatives over PCA. They still advocated the use of ordinal variables treated as continuous terms when dealing with PCA; however, should this not be possible, the use of PCA even with discrete data seemed justifiable.

Conceptually, it remains uncertain what SEP aspects an asset-based index actually measures and why it is associated with health [192, 243, 254]. Obviously these indicators reflect important aspects of living conditions, but their meaning in terms of SEP is not clear yet. In this respect, the use of consumption expenditure or income data has the advantage of having a much more straightforward meaning.

Authors in favour of the asset-based approach have described the asset index as 'an acceptably reliable proxy for consumption' [84]. In reality, several studies have shown the agreement between these two measures varies and is largely dependent on the dataset, the setting and the measure of correlation used (i.e. household ranking misclassification, correlation coefficients, R^2 from regression), but it is typically very weak [86-88, 90, 229, 243]. A recent systematic review concluded that the variability in the level of agreement observed is too large to make legitimate the assumption that the underlying concept measured by the asset based index is the same of consumption expenditure [255]. Other authors [243] have argued that a large part of the disagreement observed can be attributed to the fact that normally an asset-based index does not include proxies of direct consumption of food items, resulting in the exclusion of the most important component of the aggregate consumption. This interpretation contrasts with an analysis from Howe indicating that the asset index does not act as a good proxy of consumption expenditure not even when sub-aggregate of consumption expenditure not containing food-related indicators are taken into account [192].

The main implication of such modest agreement between the assets based indices and the measures of consumption expenditure is that the quantification of health inequalities may be

different depending on whether authors use an asset-based index or consumption expenditure. Howe and colleagues have shown that such differences vary across studies and that their direction is difficult to predict [192].

Some authors have been invited to be aware that their findings may be sensitive to the choice of indicator adopted and – whenever possible – to employ in parallel both methods [243]. As in many other studies on health inequalities, this was not possible in this research project. To partially overcome the lack of consumption expenditure data I have used the regression analysis technique. Nevertheless, the relatively low explanatory power of the regression model (i.e. 60%), suggests that this prediction of consumption expenditure was just an approximation to the real data. This is probably due to the fact that - despite the relatively high number of variables included in the model - they still represent a limited list of possible items. Thus, they are unable to capture completely what contributes to a household's per capita consumption expenditure. The utility of the regression technique is further limited by the fact that the analysis is based on data collected in 2002 that is almost 5 years before the assets data collection for this study. The disagreement between real and predicted household consumption expenditure apparently did affect the household SEP ranking of the LSMSIII. Therefore, it is difficult to predict what extent and directions of TB inequalities would have been observed if consumption expenditure data were available.

Despite the more straightforward meaning of consumption expenditures it is known that, because of reliability and limited resources issues, the use of consumption expenditure may not be ideal for studies conducted in developing countries [69]: Furthermore, an asset-based index approach offers the advantage to represent a proxy of long term or 'permanent' aspect of economic status; whereas, consumption expenditure reflects a more dynamic aspect of SEP, more variable in time [69]. This specific feature of the asset-based index is probably the strongest justification for using this approach over consumption expenditure: as discussed earlier, the longer term of SEP captured by the asset-index was considered a strategy to partially address the issue of reverse causality.

To better clarify the meaning of the asset-based index I have made very specific assumption about the household SEP, which I viewed as a multidimensional phenomenon resulting from four different domains. For each of these domains I have explicitly described the pathway through which they were meant to affect the risk of TB (**Chapter 4**). This has allowed to quantify not only the overall effect of household SEP, but to disaggregate this concept into actionable dimensions.

(f) Additional specific limitations of the infection study

This research project presents additional limitations that are specific to the infection study.

The participants to the infection study are the controls recruited in the case control study. They, therefore, represent a group of individuals with no evidence of TB disease randomly selected from the participants in a population-based TB prevalence survey, presumably including all SEP strata of the population. Nonetheless, because of the original matching to the TB cases (and not to the general population), it is more appropriate to say that the participants to the cross-sectional study approximate, rather than represent, a random sample of the community.

Because controls were age-matched to TB cases, the controls sampled (and therefore the study population for the cross sectional study) could be older than the general population. However, the effect of age was always taken into account in the analysis and, therefore, it is unlikely that the possible oversampling of older age groups could have confounded the results.

The matching by area of residence could have caused an over-representation of study participants from the urban area (i.e. since cases are more frequent in this area). However, it is unclear how this could explain the results observed: even if study participants were more likely to be from the urban area and, therefore, presumably wealthier, this should not affect the distribution of the TB infection by household SEP: TB infection would be still more common among the poorest of this selected group of wealthy individuals. This matching bias could have explained the lack of any association between household SEP, but it does not explain the lack of an association opposite to what commonly expected.

The subjects who accepted to participate were more likely to come from the urban area compared to the rural one. This raises issues on the generalisability of these findings and, most importantly, about the possibility of selection biases (as people living in urban areas are generally wealthier compared to the residents in urban ones). As mentioned in the method, I tried to verify the extent of this possibility, by comparing those controls enrolled in the study with those for whom no blood testing result was available (for any reason). The results showed that the only significant difference was that controls included in the study were more often coming from the urban area. Once area of residence was taken into account, these two populations did not significantly differ for sex, age, socioeconomic position and any other risk factor associated with TB infection (i.e. known contact with TB cases) or impairing the immune system and therefore the response to Quantiferon (i.e. HIV status, food intake, alcohol consumption).

A recent study from Zambia suggested that the accuracy of QFT in populations characterised by HIV prevalence remains to be confirmed: in this study QFT sensitivity results were markedly different between TB-HIV positive and TB-HIV negative patients (63% vs 84% respectively) and immunesuppression was associated not just with indeterminate, but also to false negative results (20% versus 3% respectively in the two groups) [256]. Despite the limited number of cases under study, these observations are consistent with what was observed also in the cross-sectional study, where almost 37% of the confirmed TB cases was QFT negative. These observations can be also applied to the controls, however, this can represent a potential bias for this study only if the HIV status of these people (and therefore their likelihood to be diagnosed as QFT positive) was associated in some way with houshehold SEP. At least in this study, this does not seem to be the case so the bias should not be an issue.

It has been suggested that also malnutrition may reduce the expression of interferon-gamma and other mycobactericidal substances [257-258] and therefore the test sensitivity. If this were to be the case, this study may have missed infected individuals among the poor (i.e. false negatives) and may have been more likely to detect TB infection among better nourished, wealthier, individuals potentially biasing our results towards a lower risk of infection among the poor (as we noted in this study). However, a better immunological response probably can only help in providing an interpretable result, but does not necessarily translate into a positive QFT result.

The lack of an association between TB infection and TB-contact is particularly surprising especially considering the well known importance of this risk factor in the epidemiology of TB infection [28, 152, 155, 157, 171, 259]. Certainly, the small sample size can explain this observation; however, in this specific case the importance of this exposure may have been underestimated because of at least three different reasons: 1) these data were based only on the respondent recall; 2) because of fear of stigmatisation people may have denied any contact with TB cases or may have hidden the TB status of their household members; 3) the lack of association could be also explained by the way the question was formulated: in the questionnaire I have asked the study participants whether they had been in touch with anyone known to have TB. This means that they may have been in touch with someone with TB and simply not know that he/she was a TB case. Finally, even among those who declared to have been in contact with a TB case, no effort was made to enquire the nature, the frequency or the duration of the contact, making the results even more difficult to interpret.

9.3 The study results in relationship to previous findings

9.3.1 The ecological study

The comparison of the results of this study with previous findings is limited by the surprisingly low number of ecological analyses undertaken in developing countries in this field. Studies available [128, 132-133] (all conducted in poor urban areas of Brazil and South Africa), however, seem to support the observations made in the urban CSA represented in this study, where TB prevalence rates increased with the worsening of living conditions. As in this setting, these studies documented an association between TB rates and community illiteracy rates [128, 132], unemployment rates [128, 133], and measures of absolute poverty [128, 132] In the case of crowding, one of the studies in South Africa found only a weak association between this indicator and TB incidence rates. The authors hypothesised a 'saturation level' for crowding (expressed as average number of people/room in the area), according to which once a certain level of household members density is reached, a further increase of crowding does not necessarily increase the risk of intra-household transmission of M. *tuberculosis*.

In these studies the association between these socioeconomic indicators and TB rates reached statistical significance probably because they were designed to cover retrospectively a broad time interval and large geographic areas. This ensured an adequate number of outcome events and a sufficient number of communities, enough diversified from each other to detect a gradient. By contrast, in this study, I could only rely on eleven communities under which were very close and similar to each other. Moreover, prevalent cases were all identified in one point in time. All these features may explain the lack of significant association observed in the ecological analysis.

Although not significant the association between poor living conditions and high TB rates was largely observed for most of the ABSM indices used even when the analysis of repeated by rural and urban site. Because of the very small number of communities under study these results should be interpreted with extreme caution, however, among the rural communities, more than half of the ASBM showed evidence of an unexpected negative association between TB prevalence and worsening of living conditions, suggesting that the poorer the community the lower were the rates of TB.

Several studies, including the prevalence survey [211], have consistently documented lower TB rates in rural settings compared to the urban ones [196, 260-261], but I am not aware of any study formally addressing the reasons of this phenomenon. As a result, it is difficult to judge whether the observations made in this study simply depend on the way the area-based

indices of SEP have been constructed or whether in rural CSA, there are patterns of living conditions that, although poor, offer some kind of protective advantage from TB. This latter hypothesis seems to be supported by the observation that the ABSM for which the negative association was observed are all indicators of living conditions that are not ideal for TB transmission: communities with poorer housing quality, having no electricity, with poor access to community services probably are also community presenting with physical and environmental profile less suitable to TB transmission. In other words, in terms of TB transmission dynamics, poor communities may be more protected from high prevalence of TB. A study conducted in the Netherlands could also provide another possible explanation of these results: in this study mortality and hospitalisation rates were unexpectedly low in areas characterised by low household income and unexpectedly in apparently wealthier area. When they analysed the reasons of these 'exceptions to the rule', they found that healthy deprived areas where characterised by lower percentages of immigrant, people living alone and concentration of enterprises [262]. Although likely to be very different, it is possible that even in this ecological analysis compositional factors may account for this apparent paradox of the protective effect of living in deprived areas.

It is interesting to note that from the analysis of TB prevalence rates by CSA it emerged that some rural CSA (e.g. R-13 and R-14) were characterised by TB prevalence rates as high as in the urban ones. This could be explained by the fact that these areas, although classified as rural, include standard enumeration areas which are considered as urban in the Census database. Another explanation is the existence of 'hot spots' or clusters of disease in these communities. This second option is supported by two other studies from Peru and the Philippines in which authors have documented the existence of TB "hyperendemic" areas whose epidemiological profile in prevalence surveys can be masked by their proximity to areas with lower TB rates [131, 134].

Although no final conclusion can be drawn from this ecological analysis (both because of the low study power and the inherent limitation of the ecological fallacy), this piece of analysis can still provide useful formation. The analysis of the geographic distribution of the prevalent cases of TB, for example, has been extremely useful to characterise the profile of those areas where most of transmission seem to occur and may drive the implementation of targeted control strategies. Furthermore, the ecological analysis of the prevalence survey data offered me a great opportunity to test methods and datasets for the development of area-based measure of SEP. As far as I am aware, this is the first study trying to use LSMSIII and the Census 2000 for the Lusaka province to explore TB prevalence rates disparities across areas characterised by different poverty profiles. The only other study using a similar

approach is one addressing inequalities in paediatric mortality and birth weight in Lusaka [263]. In this latter case, however, areas of interest were simply divided in High-cost, Medium-cost, and Low cost-zones depending on the predominant type of housing in these areas. In this study I have tried to make a step forward in the characterisation of these communities and the understanding of the effect of community characteristics on TB. In particular, three methodological aspects made this ecological analysis quite unique compared to previous research: 1) communities have been characterised using composite indices rather than single socioeconomic indicators. Although single indicators have been shown to be just as effective in measuring area-level socioeconomic gradients as the composite indices [217, 264-265], their value in the characterisation of a community is more subject to changes making the comparison over time more difficult; Furthermore, the use of composite indices makes the choice of the most appropriate indicator less arbitrary; 2) As suggested by Messer and colleagues [219], only those items having comparable loading factors both within and across the 11 CSA have been included in the final area-based measure of relative SEP. This screening strategy allowed the creation of a an area based index of relative SEP working equally well in urban as rural CSA despite the evident differences in living conditions in these two settings; 3) finally, CSA have been described also through area-based domains of SEP. The selection of these domains has been theoretically driven and they have been created to represent community aspects suitable for practical interventions.

Despite the inconclusive results of the ecological analysis, the accumulated experience in large dataset handling, theorisation and construction of area-based measures of SEP represents a resource that can be applied in future studies better designed to address the role of contextual variables in the epidemiology of TB in a developing country.

9.3.2 The case control study

(a) The role of household SEP

The existence of a socioeconomic gradient in the distribution of prevalent TB was somewhat expected: the cases included in this study are more likely to be detected in prevalence surveys because they are more likely to delay their diagnosis and treatment, behaviours often resulting from limited material resources. It could be almost said that they are 'prevalent' because they are poor. What was less expected is that this association was so strong, so robust, and as strong and independent from HIV.

It may be argued that the strength of this association is biased (or overestimated) by the inclusion in the household SEP index of food related variables, known to be risk factors for TB [227]. In fact, very similar results were seen when these variables were removed from the household SEP indicator as shown by the sensitivity analysis. It is worth mentioning once again that the removal of the food –related variables from the household SEP index, although useful to avoid the indirect effect bias, would have been difficult to justify on theoretical grounds: in Zambia, both in the urban and rural areas, malnutrition, failure to meet basic nutritional requirement and food insecurity are consistently mentioned as key defining characteristics of extremely poor households. The importance of including food-related variables is also supported by two observations: 1) only the household SEP domain based on food-related variables was able to attenuate to non significance level the urban/rural household SEP disparities; 2) removing these food related variables from the indices resulted in a higher level of clumping and truncation in the household SEP distribution. This evidence suggest that food-related variables are in this setting a sort of 'universal' indicator of household SEP transcending the urban and rural differences and that the non consideration of these variables would have resulted in less discriminatory power of the household SEP index.

Although the association between household SEP and prevalent TB appeared to be robust to the specific household SEP measurement strategy, three aspects emerged:

1. The household SEP ranking was slightly affected by the choice of different household SEP measure. This is consistent with what was observed in several other studies [227, 240, 266]. In this study, household misclassification ranged from 16% to 33% depending on the index. As also shown by Howe [192], the extent of misclassification was affected by the weighting strategy adopted more than the assets included in the index: PCA based indices (as the index used in the main analysis) showed a lesser degree of misclassification compared to regression based indices. Probably, in this study, the extent of household SEP misclassification was not enough to change the effect of household SEP. It cannot be ruled out, however, that a higher extent of misclassification would have changed the results observed in this study.

2. The extent of TB inequalities was sensitive to the choice of household SEP measure, with OR for the low household SEP category ranging between 2.5 and almost 5 depending on the index used. Differently from the household SEP ranking, it did not seem that either the choice of assets or the weighting strategy had a major role in driving the size of the inequalities. This is consistent with what documented in other studies

showing that the size and direction of such sensitivity can be hardly predicted on the ground of the type of index used [227, 240, 266].

Some authors have argued that, compared to consumption expenditure, the use of assets-based indices is more likely to result in the over-estimation of health inequalities largely because of a spatial effect: moving from consumption expenditure to assets indices results in the increase of the socioeconomic disparities between urban and rural households, with rural households more likely to be in the bottom positions of the ranking. If the health outcome is more frequent among the rural households for whatever reason this will result in an overestimate of the health inequalities [243]. In this study the use of assets-based indices is unlikely to have produced such over-estimation because TB is significantly more frequent in the urban areas. As a result, individuals who are more likely to be TB cases should be more concentrated among the urban, relatively wealthier, households resulting in the attenuation, rather than the amplification, of the TB inequalities. This also should explain why the associations observed using different types of indices appear largely similar.

The sensitivity of health inequalities to the strategy of SEP measurement has been documented elsewhere for different type of health outcomes, including under-5 mortality [227], immunisation coverage [227], health services utilisation [243], fertility [240, 266], however, the epidemiological meaning of such inequalities are not clear yet. Someone has argued that in the vast majority of the cases the change in the size of health inequalities is never big enough to justify speculations: the order of the magnitude tends often to stay the same, and even when it doesn't the confidence intervals, like in this study, largely overlap [227]. The judgement of the epidemiological meaning of these results is made more difficult also by the fact that these differences were evident only when the indices scores were grouped into terciles. By contrast, OR were almost identical (around 2.5) when the indices were used as dichotomous variables.

3. Regardless of the index used, the association between household SEP and prevalent TB was always comparable if not bigger than the one observed for HIV (which will be discussed later). This result suggests that in this study household SEP and HIV were at least equally important as determinants of prevalent TB and that this observation is independent from the SEP measurement strategy.

The exploratory analysis of the household SEP domains showed that the food availability and vulnerability domain was the one mostly driving the association between household SEP and prevalent TB. In this case the association probably results from the *direct* link between

malnutrition and TB, rather than from the *indirect* association with household SEP [227]; however, is it really a limitation in the case of the exploratory analysis? More than quantifying the extent of TB inequalities, the purpose of the exploratory analysis was to disaggregate the association between household SEP and prevalent TB into addressable pieces and identify the one most suitable for intervention purposes. This approach suggested that among the different aspects contributing to the SEP of a household in these two Zambian communities, the one related to inadequate nutrition could be the one most likely to have an effect on prevalent TB if addressed. The exploratory analysis also showed that the effect of household SEP is likely to be multidimensional since also the housing quality and human resources domains were associated with prevalent TB. The domain about the access to community infrastructures was not significantly associated with prevalent TB. A possible explanation is that cases and controls were matched by urban/rural area of residence and therefore probably overmatched in terms of infrastructures and facilities shared.

In the prevalence survey, from which this study population originates from, they found no association between prevalent TB and any of the socioeconomic indicator enquired (e.g. assets ownership and daily number of meals) [211]. Interestingly, this is consistent with what also emerged from the case control study: assets ownership was significantly associated with prevalent TB only when the items were grouped into a cumulative variable describing the overall number of items owned rather than the type. This could be because the ownership of the single asset does not vary across households as much as the total number of assets owned; as a result, enquiring about single assets may of limited use in differentiating household SEP. The lack of association with assets ownership and prevalent TB was also documented in a prevalence survey conducted in Harare, Zimbabwe: authors enquired eight different types of asset and observed a protective effect for each type of asset owned, but the association was never statistically significant [267].

As in this study, the Zambian prevalence survey could not detect any association between the individual daily number of meals and prevalent TB. A possible explanation is that this index of food intake is unlikely to be a sensitive indicator of household SEP: first of all, it cannot be assumed that the food shared by someone is equal for each household member. Male, economically active members, for example, are known to eat more often and better than other members [192]; secondarily the question about number of meals is likely to reflect the local food habits as much as the food availability; thirdly this question presumes the existence of a definition of meal which may be culturally dependent; finally in Zambia even extremely poor households can have up to three meals per day, but this does not necessarily translate into adequate food intake.

No association between socioeconomic indicators and prevalent TB was also observed in the latest prevalence survey conducted in Chiapas, Mexico [161] and the prevalence survey conducted in Karachi, Pakistan in 2002 [135]. In Mexico, the study was deliberately conducted in communities classified as poor or very poor and therefore the population recruited may have been socioeconomically too homogenous to detect a difference between cases and non cases. In Pakistan, only 18 TB cases were detected and it is likely that the study did not have sufficient power to detect any association with socioeconomic indicators.

In the next section I will discuss the most relevant individual level risk factors identified in this setting. For each of these exposures, I will comment on their mediation effect in the association between household SEP and TB and, independently from that, their role as risk factor for prevalent TB.

(b) The role of nutrition

The mediation analysis suggested that the association between household SEP and prevalent TB was largely captured by the food-related variables to the extent that almost no effect of household SEP was left once the effect of food availability was taken into account. Inadequate nutrition might, therefore, be considered as the '*active ingredient*' of the main association under study in this thesis [200]. The identification of a postulated mediating factor is important because it supports the biological plausibility of the association between household SEP and prevalent TB, thus providing evidence that this association may be causal [200].

This observation was robust to the type of household SEP indicator used suggesting that the attenuation in the association between household SEP and TB was not due to the fact that food related variables were accounted for twice (i.e. in the index and as covariate) in the multivariable model.

Among the food related variables, the one on the weekly consumption of proteins was the strongest social determinant of TB at individual level. The important role of food was evident both from the OR and from the PAF estimates. This observation is biologically plausible: the key role of protein deprivation in TB infection and development has been confirmed in several studies under different animal models, doses and route of infection [38] and is also supported by human data.

In guinea pigs protein malnutrition affects the immune response by significantly reducing the production of interleukins, interferons and tumor necrosis factor alpha [268]. A diet poor in proteins also alters the absolute and the relative number of total T-lymphocytes and various subpopulation cells, including CD2+, CD4+, CD8+ and Fc receptor bearing [38]. It is important to note that animal model experiments show quite clearly that the dramatic effect of a protein-deprived diet on the T-cells mediated resistance is substantially and rapidly reversible. If reproduced in humans, of course, this would be particularly relevant for the possible use of food support programs as TB prevention strategies [38].

Among human studies, a 20 year follow up study of a representative sample of the US population showed that having body mass index, average skin-fold thickness or upper-arm muscle area in the lowest decile of the population increased the adjusted hazard of TB from six to ten fold, controlling for other known risk factors for TB [269]. There are also three other studies that, despite their ecological design, present fairly convincing evidence that a low protein diet play a direct effect on TB incidence. In Denmark, during the I World War, TB rates climbed while the country was exporting massive quantity of its meat, fish, poultry and diary products to the extent that the local diet was poor in proteins. When Germany blocked Denmark exportations, however, food items rich in proteins became more available for the local population and TB rates dropped rapidly. In Norway, the high TB morbidity rates among the recruits of the Naval Training School in the early 20th century dropped soon after the diet was fortified with margarine, cod liver oil, whole wheat bread, fresh fruits and vegetables, and milk [38]. Finally, another study showed that during the II World War British prisoners in the German detention camps exhibited lower TB prevalence rates compared to the other detainees and this could be only explained by the fact that British prisoners were receiving Red Cross food supplements providing 30 g proteins and 1000 Kcalories/day [38].

(c) The role of HIV

This study has largely confirmed the observation on the association between prevalent TB and HIV made also in the prevalence survey, which reported an adjusted OR of 2.3 (95%CI: 1.4-3.7) [211]. The results are also consistent with findings in three other studies conducted in Zimbabwe and South Africa reporting an association between HIV and prevalent cases of TB ranging between 2.0 and 4.1 [25, 267, 270].

These results confirm the impression that although HIV is an important determinant of prevalent TB, its impact remains much lower than that observed for incident TB, for which

measures of association range between 6 and 10 [31]. The common explanation for this is that the rapid TB disease progression associated with an early onset of symptoms among HIV positive patients favours an early TB case detection and treatment start, resulting in a more rapid reduction in the infectiousness period. Consequently, TB-HIV co-infected cases will be less likely to be identified within a prevalence survey. This contrasts with what was observed in another cross-sectional survey in South Africa that reported a large burden of HIV-associated undiagnosed TB in a community [271]. Clearly, further research is needed to fully understand the role of HIV on prevalent TB.

Interestingly, this study did not confirm the hypothesis that the association between household SEP and prevalent TB is mediated by HIV. It also showed the association between HIV and TB was not confounded by household SEP. This observation is largely due to the fact that in this study there was no socioeconomic gradient in the distribution of prevalent HIV infection: no significant trend was observed even when stratifying either by age or by TB status.

The case control study was not designed to detect an association between household SEP and prevalent HIV so these results should be not surprising; however, even the prevalence survey was not designed for this purpose, nonetheless some of the socioeconomic indicators showed evidence of a statistically significant association between lower SEP and HIV [211]. The lack of association also contrasts with what is known about the HIV epidemic currently affecting Zambia: the start of the epidemic was characterised by strong social-status differentials in HIV seroprevalence, with higher rates of infection found among those with higher educational achievement compared to those less-well educated [272]. More recent data have revealed a dominant trend of declining HIV prevalence among younger women since the early 1990s: this decline seems to have involved only women with higher educational status, whereas rates are stable or rising in less-educated groups. This striking diversity in prevalence trends by level of education seems to prove that the epidemic is now mainly involving lower socioeconomic groups [273]. Evidence of this shift has been revealed in several other African countries, but the phenomenon seems to be particularly evident in Zambia [273]. Nowadays, HIV prevalence rate among the poor continues to rise and it is mainly attributed to poor access to information and the structural barriers for poor people to adopt healthy behaviours [273].

Given the strong evidence for HIV inequalities in Zambia, it seems more plausible to believe that the lack of any mediation effect from HIV could be due to some inherent limitation of the study. The alternative explanation is that, in this setting, TB prevalence in the whole

population is independently driven by two forces: a high burden of low household SEP and a relatively smaller burden, on a population level of HIV.

In interpreting the role played by HIV in this study compared to household SEP, one may argue that the economy of Zambia constantly declined starting from 1970, but TB rates increased only when the HIV epidemic started (end of 1980s'). This is true, but it is also true that the most evident contraction of the Zambian GDP can be observed only in the early 90's: real GDP fell by 11% in 1994 and a further 5% in 1995 as a result of drought and the end to most of the government subsidies for agriculture. During the period 1991-1998 poverty increased in the urban areas, as a consequence of the reduction in the employment in the parastatal sector. In the same years, the HIV/AIDS epidemic further decreased the productivity of the labour force in the urban communities. Finally, between 2001 and 2003 Zambia has experienced its last major drought. As a result, the cereal production declined by 29% in year 2000-2001 compared to the previous year and was equally bad in the following crop season due to more extended drought that affected larger parts of the country [12]. At the same time, maize prices increased up to 5 times the five-year average and in some regions of the country maize was not available on the market [12]. Apparently these events did not affect the levels of acute malnutrition in young children (which stayed far below the level observed in case of famine of ongoing conflict), but the Zambian DHS documents a constant increase of chronic malnutrition over the 1990s [12]. In addition, the dietary energy supply in Zambia (Kcal/day per person) has dropped in the past several decades and index of domestic food production has not shown any sign of increase between 1990 and 2000 [12]. This brief overview does not allow any conclusion, but suggests that while no one can dismiss the effect of HIV in fuelling the TB epidemic in Zambia (as in many other African countries), the interplay between TB secular trends, HIV epidemic and socioeconomic conditions is likely to be complicate.

(d) The other risk factors for prevalent TB

BCG – This study showed a very highly protective effect of BCG. This is somewhat surprising given the extremely variable efficacy of BCG (0-80%) demonstrated in clinical trials [203]. Given the very small number of subjects lacking BCG vaccination these results should be interpreted with caution, especially considering that the attribution of BCG status was made simply by scar reading and it was therefore prone to considerable observer variation [274]. In this particular case, it is possible that scars from other causes have been mistakenly attributed to BCG vaccination, explaining the extremely low proportion of

unvaccinated subjects. However, BCG vaccination coverage in Zambia is known to be extremely high (>80%) [275]. The very low proportion of unvaccinated subjects also explains why, despite the strength of the association, only 11% of the TB cases in this study could be attributed to the risk factor.

Migration – After not being vaccinated with BCG, the strongest risk factor for prevalent TB was migration (hereby defined as having lived somewhere else in the past 5 years for a period of at least 6 months). The comparison with previous findings is limited by the fact that most of the literature about migration and TB refers either to foreign-born people or the impact of migration patterns from developing to industrialised countries. In this case, none of the migrants was born abroad and migration has to be considered more a risk factor increasing someone's mobility and therefore his/her likelihood of TB exposure.

It is unclear why people who reported migration in this study had a higher prevalence of TB. Neither lower household SEP nor higher likelihood of HIV infection seem to be an explanation.

In this study migration did not act as a mediator of the association between household SEP and prevalent TB. At the same time, household SEP was not a confounding factor for the association between migration and prevalent TB suggesting that those who migrated did not necessarily belong to poorer households. This is somewhat consistent with recent statistics showing that in the last decade migration patterns in Zambia have not been driven by an economic pull [276]: in Zambia, cross-border migration seems to be very low involving only 0.1% and 0.3% of the unskilled an semi-skilled Zambians. The urban to urban and rural to rural have become the predominant migratory pattern of Zambia, but interestingly, there was a reduction in the immigration rate in the major cities, particularly in the Copperbelt and the Lusaka province, and an increase in the migration towards less economically developed provinces [276].

This study also showed that the association between migration and prevalent TB is also unlikely to be confounded by HIV. This was a plausible hypothesis considering the known association between HIV and migration and considering that, in Zambia, migrant labour is now mainly in the informal setting where migratory workers are less likely to be reached by HIV/AIDS prevention programs and services [12]. Nonetheless data contradict this possibility.

Another hypothesis is that people who went through a migration experience may have been exposed to a significant amount of psychological stress and depression-related symptoms which ultimately may have increased their vulnerability to TB. The association between

migration and mental health is well documented in the literature [277] and it is plausible in the context of this study, but it cannot be confirmed with the available data.

TB Contact – This study has confirmed the importance of known contact with TB cases that is documented in the literature and suggests the hypothesis that at least part of the association between low household SEP and TB may be mediated by an increased exposure to TB cases. Since most of the transmission seems to happen in the community through casual contacts [267, 274], the mediation pathway observed may suggest that low household SEP increase someone's risk of TB disease not just influencing where someone lives, but also where someone works and socialises. This mediation pathway is supported by studies proving that TB transmission occurs through networks that are socially constructed [207, 278].

Further speculations on the relevance of this risk factor for TB disease are hampered by the observation that contact with a known case of TB was no longer significant when the other risk factors for prevalent TB were included in the same multivariable model. This may suggests that the contact with a contagious case of TB is a necessary, but not sufficient cause for TB disease.

Attending churches – This exposure was associated with a significant protective effect against prevalent TB. Interestingly this association was also found in another recent prevalent survey [267]. Such protective effect apparently contradicts the evidence from recent studies identifying churches as potential 'hot-spots' for TB transmission [209] and settings of TB outbreaks [208]. It is possible that, in religious communities like the ones in Zambia, churches represent areas of aggregation particularly prone to TB transmission both because of the frequency and duration of attendance and because of singing activities. Conversely, people attending churches may be less likely to progress to TB disease because of healthier behaviours and because of the advantages conferred by the higher social capital these people may share (in the form of food support, health messages, education, etc). In conclusion, attending churches may favour TB transmission, but may also be a proxy of factors reducing the vulnerability to disease progression.

(e) The interpretation of the main PAF results

From the analysis of the PAF results three main issues emerged:

1. This study shows that, before taking into account any mediation effect, household SEP can explain at least 30% of the TB prevalent cases observed in this study. Because of

reverse causality issues this estimate may be overestimated. However – as mentioned earlier in this chapter – the asset-based approach may have helped to minimise the reverse causality bias and, therefore, in this case the overestimate of PAF. The computation of PAF was conducted only with the index used in the main analysis; however, given the robustness of the association between household SEP and prevalent TB, it is unlikely that the use of a different index would have produced dramatically different results. The only other study that tried to estimate the PAF for indicators of household SEP in relation to TB found a PAF of 50%, but given the large differences in the study design, setting, and the way household SEP was conceptualised and measured, the results are not comparable [279].

2. In this study PAF for HIV was equal to 36%, which exactly matches what reported from the Zambian prevalence survey [211] and it is consistent with what observed in prevalence survey recently conducted in Zimbabwe (PAF = 33%) [267], and that reported in a study from WHO for the whole African region (PAF for the African region = 28%) [31]. The similarity between the PAF for HIV and household SEP can be interpreted by saying that removing HIV from the this population would have a similar level of effect on the overall TB prevalence as increasing the level of SEP of all the households to that seen in the higher SEP group in this study.

3. The highest proportion of TB prevalent cases observed in this study (42%) can be attributed to the consumption of an insufficient number of meals containing proteins. The PAF observed for this exposure is consistent with the PAF for malnutrition (47%) in the African region reported by the WHO study above [31].

Combining the OR and PAF evidence, it can be concluded that whereas at individual level HIV is still one of the strongest predictor of TB disease, at population level malnutrition may be able to explain a larger proportion of cases than HIV.

The high consistency between the PAF results from this study and the findings from the literature support these conclusions; however, when it comes to PAF, results should be interpreted with caution because of the several unverifiable assumptions on which the measurement of PAF is based. These assumptions can also be applied to the PAF findings generated by other studies.

First of all, PAF computation assumes the existence of a causal relationship between the outcome and the exposure. Particularly in the case of inadequate nutrition, no causal relationship can be conclusively inferred because of the potential for reverse causation to explain the association.

Another assumption in the computation of PAF is the lack of confounding factors. People reporting inadequate nutrition are likely to differ from the general population with respect to other risk factors for TB, so that even if their nutrition was adequate their risk of TB may be still higher compared to the general population. In the analysis, the PAF of inadequate nutrition has been estimated after accounting for the confounding effect of the other TB risk factors operating at the same level of the causative chain. However, it could be that not all the actual confounding factors were properly taken into account so that some residual confounding effect cannot be excluded.

By definition, PAF represents "the proportion of the total disease experienced by a population that would not have occurred if the effect associated with the risk factor of interest was absent" [230], or, as reported by Fielding and colleagues, "the proportion of cases of a disease occurring in a population that would be avoided if the exposure were removed" (unpublished reference). Applied to this study, one may conclude that the rate of TB in the exposed group would be greatly reduced if their nutritional status would be improved to the level of the reference group for the nutrition related variables. In reality, this may be not immediately the case. Greenland, for example, distinguishes between "excess cases" (those cases that would not have occurred had the exposure under study not occurred); and the "aetiologic cases" (those cases that occurred earlier because of the exposure, but that would have happened anyway). In this latter circumstance, the exposure of interest is a contributory cause of the outcome, but not the only one [280]. Following from this distinction and given the uncertainties linked to the role of inadequate nutrition, in this study it would be safer to consider these 42% of TB cases due to inadequate nutrition as *aetiological* cases, rather than *excess* cases: they are attributable to inadequate nutrition, but it remains unknown whether even in the lack of this exposure these cases may have happened the same at a later stage of their life because of other risk factors, especially HIV. Furthermore, the estimate of PAF for inadequate nutrition assumes that all the other risk factors remain constant. But the improvement of nutrition may be accompanied by the change (decreases or increases) in another risk factor (e.g. diabetes).

In conclusion, PAF results strongly point in the direction of a major role of inadequate nutrition in the aetiology of prevalent TB, but this evidence is based on many assumptions.

9.3.3 The cross-sectional study

(a) The role of household SEP

To the best of my knowledge this study is only one documenting an association between high household SEP and TB infection. This result, although in contrast with what already published on this association, is plausible. Historically, the industrial revolution in 19th Century in Europe and the associated urbanisation were accompanied by an initial increase in TB disease and death due to TB [196, 261, 281-283]. One possible dynamic underlying these findings is that in modern day developing countries relatively wealthier living conditions could increase TB infection: living in apparently better equipped households (e.g. having electricity), made with more solid materials is a marker of wealth, but it could also mean poor air quality if houses are built without ensuring adequate ventilation. Similar hypotheses have been formulated by Glynn and colleagues [7] and more recently by Harling et al, according to whom in unequal communities individuals belonging to higher SEP groups are more likely to regularly come into contact with those of lower SEP, who are more likely to have active TB and, therefore to be infectious. This would result in a higher prevalence of TB infection among high SEP people compare to what observed in more equal societies [178].

Similar interpretation can be given to the findings on the distance from main infrastructures: households closer to facilities are likely to be wealthier, but also likely to reflect more urban-type setting, characterized by greater population density and higher chance of human interaction, fostering TB transmission. This has been demonstrated in previous studies, showing that in high TB prevalence settings, especially densely populated settings like this one, extensive TB transmission can occur via complex social networks that are likely to be as important as household contact in maintaining transmission [207] [278].

One may argue that the pattern observed in this study reflects the higher prevalence of HIV, and thus perhaps of TB infection, in relatively more educated individuals that has been described in Zambia in previous studies [272-273, 284]. However, I did not observe a higher prevalence of TB infection among HIV positive individuals; furthermore in this study population HIV prevalence was not associated with higher SEP.

(b) The role of household crowding

This study also confirmed the importance of crowding in the epidemiology of TB infection. As in other studies [36, 285-286], the data suggested a strong independent effect of both crowding and SEP on the risk of TB infection. Crowding and SEP perhaps represent two key forces, acting in different contexts and explaining different cases of tuberculosis infection: a) those infected in the household through overcrowding in poor households; and b) those

infected in the community because of dynamics reflecting higher SEP described above. Thus, there may be aspects of SEP that act over and above the role played by household overcrowding in fostering tuberculosis transmission.

(c) The lack of association between TB infection and the postulated mediating factors

The lack of association between TB infection and any of the other known risk factors is of concern because it raises issues on the validity of this study. On the other hand, the lack of significant association does not mean necessarily lack of association: looking at the measure of associations for all these variables, the impression is that they all go in the direction expected, but they simply did not reach the significant level. This seems to suggest that the most plausible explanation for these results is an insufficient study power, rather than biases related to the design of the study. The fact that the study size did not affect the association with household SEP can be explained by assuming that in this case the association was strong enough to overcome the low power of the study.

9.4 Meaning of the results and policy implications

Up to here I have reviewed in detail each finding of this study, trying to highlight what this study adds to our understanding to TB inequalities and why some of the results should not be over-interpreted. Before moving to the policy implications, it would be helpful to try to put together all these pieces and see what are the most important messages upon which to inform policy recommendations.

9.4.1 Assembling the jigsaw

I have attempted to investigate the nature and the extent of TB inequalities at different level of analysis: the community, the household and the individual. In this section I will try to integrate all the evidence generated at each level and provide a coherent picture.

Overall, urban communities appeared consistently wealthier but were also affected by higher TB prevalence rates compared to rural areas. Among urban communities, however, there was some weak evidence of a socioeconomic gradient in TB prevalence. By contrast, in the rural communities there was some weak evidence of a protective effect of poor living conditions against prevalent TB.

The same pattern was observed at household level with prevalent TB cases almost three times more likely to belong to households of lower SEP. This observation was stronger in the urban setting. Testing for interaction between household SEP and area of residence was not significant, but this test is not very sensitive and it can be affected by the limited sample size available. At individual level, the association between household SEP and prevalent TB was mainly mediated by inadequate nutrition. At the same time, households with higher SEP were characterised by an increased risk of TB infection among their members. No difference was observed in the frequency of TB infection by rural and urban area, however, households with higher SEP appeared to be characterised by housing quality, population density and likelihood of interaction more typical of an urban type of setting. No mediation effect was detected for TB infection.

Once the broad picture is assembled, apparently two paradoxes emerge:

1. *At individual level, TB disease is strongly associated with lower SEP. At the same time, at ecological level, TB rates are significantly higher in the urban areas rather than the rural ones, although urban areas appear to be consistently wealthier than the rural ones.*

Because of the low number of communities results should be interpreted with extreme cautions, however, four possible explanations can be given for this observation:

The rural/urban bias – As much for the household SEP, most of the area-based measure of SEP have been created using an asset-based approach. This may have introduced an urban/rural bias even in the characterisation of the CSA: most of the items included in the community based assets indices are associated with urbanisation, participation to a cash-based economy (more common in urban settings) and globalisation [241]. As a result, in urban areas, households tend to be scored higher compared to households in rural communities simply because they have more frequent access to these items, rather than because they are actually wealthier. It follows that the rural/urban disparities observed may reflect more my socioeconomic measurement approach rather a real socioeconomic gradient. If this is the case, urban CSA do not actually have higher SEP than rural CSA and TB prevalent rates are not actually higher in wealthier areas, but in areas characterised by different living conditions.

The existence of the urban/rural bias does not mean that assets-based index do not well correlate with wealth, but simply that the use of this approach may exaggerate the disparities between urban and rural households as well as urban and rural communities. The importance of this bias varies depending on the time of outcome: for example, as already explained in this chapter, this is unlikely to be an issue for this thesis. The increase of socioeconomic

disparities between rural and urban household could be a problem if, for example, the health outcome of interest was more prevalent among the rural households, known to be poorer. If this was the case, this would result in an overestimate of the health inequalities observed. In this study the use of assets-based indices is unlikely to have produced such over-estimation because TB is significantly more frequent in the urban areas. As a result, individuals who are more likely to be TB cases should be more concentrated among the urban, relatively wealthier, households resulting in the attenuation, rather than the amplification, of the TB inequalities.

Wealthier does not necessarily mean healthier – Even assuming urban CSA to be wealthier than rural ones, their living standards can be actually lower. Urban areas in developing countries are becoming more and more the ideal setting for the emergence of "new" infectious diseases and the resurgence of "old" epidemics, including TB [287]. For example, a study investigating the interaction between leprosy, tuberculosis, urbanisation and climate in 16 African countries found that an increment of 10% in the level of urbanisation corresponded in each country to 34 additional cases in the incidence rate of TB [196]. Overcrowding, unhealthy behaviours (i.e. smoking, alcohol abuse, unsafe sex), inefficient health care systems and poor physical environment are all urbanisation features justifying higher rates of TB in the urban areas. The rapid urbanisation in Europe in the 19th century and the consequent increase in TB incidence rates is often cited as an example to explain the stable, if not increasing, trends of TB in urban areas of developing countries despite the generally well-performing TB programs in these places [63, 288]. However, two aspects concur to make urbanisation in developing countries different, and somewhat worst, than the one involving the 19th industrialised nations: first, urbanisation in developing countries has not been accompanied by significant economic or industrial growth; secondly, whereas in developed countries 'soon the health of the poor was itself felt to be a burden on the whole society' [289], in developing countries there is little political attention to the problem and the ill health of the poor is often tolerated and perceived as inevitable [289]. Zambia, in particular, is a typical example of rapidly urbanising countries in Sub-Saharan Africa with approximately 40% of the population living in urban areas. Several studies have indicated a progressive deterioration of health indicators associated with the urbanisation rate of Zambia and it is possible that similar trends could be applied also to TB [275].

Compositional factors – Higher TB prevalence rates in relatively wealthier community could be due to the composition of the population living in these areas with regard to characteristics that influence TB at individual level. In Zambia, HIV is more common in urban areas [12]; Furthermore, looking at the age dependency ratio of the census population

(**Table 5.2** and **Table 5.3**), urban population are younger than rural populations. Given the association between HIV and TB and the fact that in Zambia most of the cases of TB occur among young adults [211], it follows that a certain degree of urban/rural disparity in the distribution of TB can be expected.

TB cases segregation – Urban areas may mask worse TB inequalities than rural settings. My ecological analysis has shown that, within the urban CSA, there was some evidence for a socioeconomic gradient in the spatial distribution of TB cases. The trend was not statistically significant, but it showed evidence of correlation between higher TB prevalence rates and CSA poor living conditions. This association could be due to the selective settlement of TB cases in the poorest CSA. This segregation of TB cases in poor urban settlements has been documented in at least two studies: TB rates among slum dwellers of the Philippines and TB incidence rates in a Peruvian shantytown far exceeded the national rates [131, 134]. A certain degree of concentration was observed also in this study with 'hyperendemic' CSA characterised by TB rates much higher than the overall rate for the urban/rural of residence. Unfortunately, because of the limited number of CSA, the study did not have enough power to detect a statistically significant difference between communities.

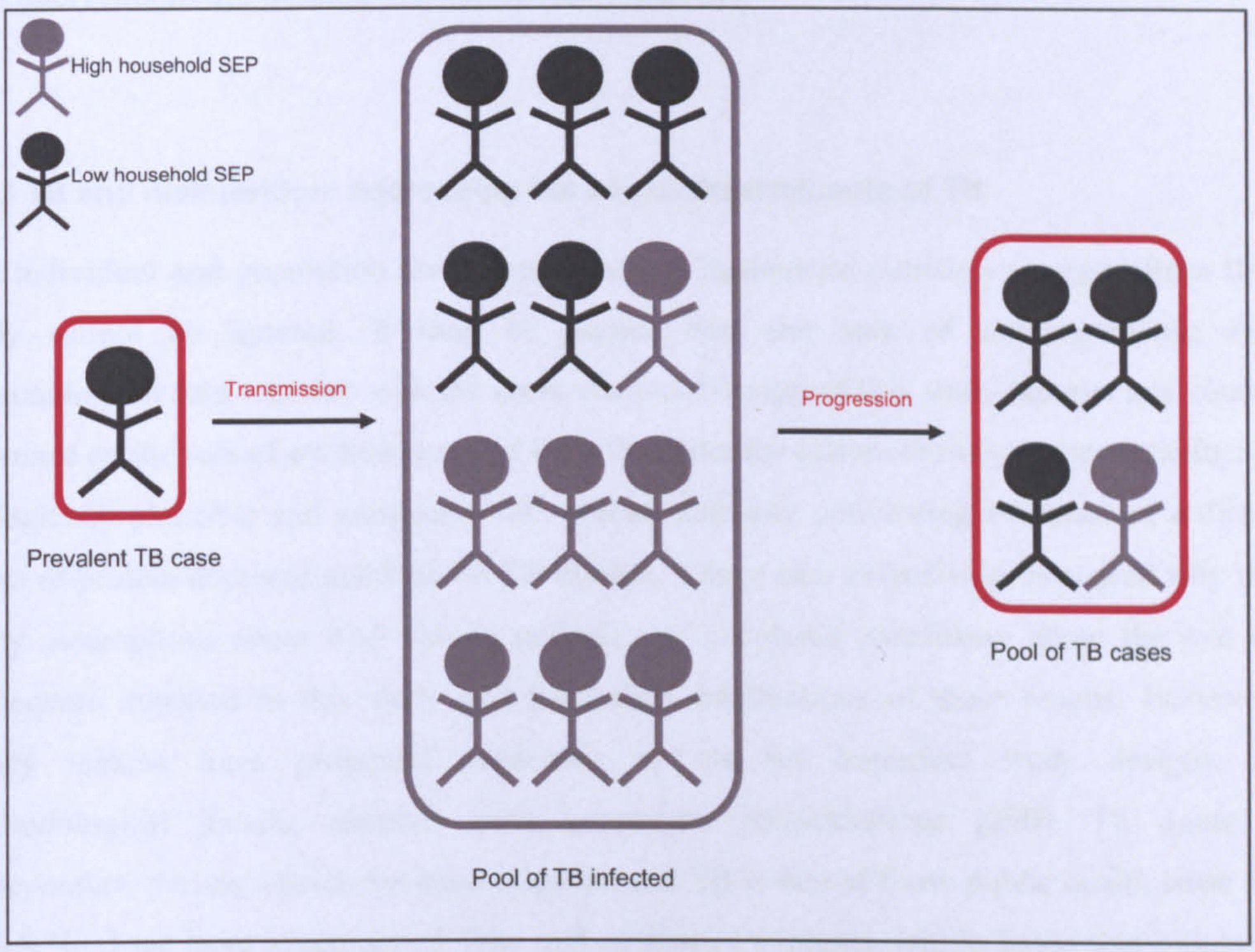
2. *Household SEP has an apparent opposite effect on the risk of TB infection and TB disease: despite an apparent higher frequency of TB infection among the wealthiest individuals of these communities, TB disease was substantially more common among the poor.*

As for this second paradox, it can be hypothesised that in a community characterised by socioeconomic inequalities, each case of prevalent TB will infect a mixed pool of people with different socioeconomic status. Among them, people with higher household SEP may be those more likely to be infected because of the frequency and dynamics of interaction conferred by their position. In the pool of TB infected, however, the poor will be those more likely to progress to TB disease. A visual representation of this hypothesis is outlined in **Figure 9.1.** The graph is just a rough representation of the hypothesis formulated and the number of subjects in each pool does not respect the actual number of TB infected and TB cases theoretically expected from each infectious case of TB.

In some way, this point is linked to the first one: it is unclear whether higher SEP individuals are more at risk of TB infection because wealth renders these people more vulnerable or whether their higher SEP just reflects the way that household SEP has been measured. In this second case, it would be more correct to consider these people as 'differently poor' rather

than wealthier. This second hypothesis seems to be more likely considering that TB infected people tended to be concentrated in the poorest CSA.

Figure 9.1: The opposite effect of household SEP on TB infection and TB disease



9.4.2 Policy implications

The majority of TB cases in high incidence settings reflect recent transmission, not reactivation [267]. TB prevalent cases are those mainly responsible for ongoing transmission in a community, therefore the first priority of TB control strategies in endemic settings should be to reduce the proportion of infected cases becoming ill and the time between onset of symptoms and appropriate diagnosis and treatment initiation. Despite the several limitations extensively discussed in this chapter, in this study few evidence emerged that could suggest further research if not policy changes in the control of TB.

My study suggests that interventions to address household SEP may be an effective way of reducing TB prevalence (and therefore TB transmission) by reducing the proportion of those infected who go on to develop disease and would complement currently control strategies mainly targeted to HIV-positive people, such as TB screening, anti-tuberculosis preventive therapy and antiretroviral drugs [267].

To reduce the impact of household SEP, interventions can be targeted to the mediators of this association (i.e. the social determinant level) or directly to household SEP (i.e. the structural determinant level). I have identified two possible entry points for TB control: one at the individual level and one at structural level. The potential advantages and limitations of both interventions are discussed in the next two sections.

9.4.3 TB and malnutrition: addressing the social determinants of TB

The individual and population level importance of inadequate nutrition emerged from this study cannot be ignored. It may be argued that the lack of anthropometric and immunological data together with the cross-sectional design of this study hamper any causal inference on the role of malnutrition and limit the room for action. However, these results are biologically plausible and consistent with studies showing convincing evidence of a direct effect of protein deprived nutrition on TB disease. I have also extensively discussed why the many assumptions about PAF should prevent any simplistic conclusion about the role of inadequate nutrition in this study and the policy implications of these results. However, policy makers have prompted academics to not let imperfect study designs, or methodological details, obscure more pragmatic considerations [290]. To quote a policymaker, the association between nutrition and TB is one of those public health issue for which '[...] we have accumulated clear and consistent evidence within limits that points to certain definite impact'[290].

Recently, in a review on TB and nutrition USAID has taken this approach [16]: although the review results show that there are still too limited programmatic evidence to make recommendations for specific programs, experiences and considerations lead to the conclusion that nutrition support can be usefully incorporated within TB control programs. Most of the few experiences reported are food assistance programs aiming to enhance TB patients treatment adherence under DOTS. The results are encouraging: a report of the World Food Programme concludes that food assistance in TB patients improves treatment adherence and consequently cure rates, it also improves the nutritional status, reduces TB morbidity and mortality, and encourages TB patients to participate in other programs such as HIV counselling and testing [291].

In this study the PAF and the pathway analysis results seem to suggest that food aid programs could be used to refer for TB screening, and possibly chemoprophylaxis and food support, all the TB-free individuals that because of their average food intake or actual nutritional status are particularly at risk of TB. The principle would be the same that today

recommends active TB screening in HIV positive patients and would have the potential to prevent TB cases, but also enhance case finding and improve early case detection.

A similar recommendation was recently formulated also in a study from Indonesia [292], however, several aspects need to be addressed before this suggestion is turned into policy change: first of all, a careful analysis should be undertaken to understand how best realise the integration of national TB control programmes and food aid programs in a way that avoids tasks overlaps and the overstretch of their respective competence. Secondarily, it is still unclear what level of household food insecurity or individual malnutrition should be targeted, and therefore how many individuals should be screened for each case of TB found; In conclusion, the feasibility, efficiency and the cost effectiveness of such an approach is all to be evaluated.

Another important consideration to make is the generalisation of these results. Given the equal importance of HIV and inadequate food observed in this study, such preventive approach would be epidemiologically sound in a setting like this one, but not necessarily in others where different social determinants may have a more prominent role (see, for example, diabetes in India [23]. For this reason, country-specific PAF data about nutrition and other risk factors are required to evaluate the appropriateness of using food aid programs for TB control purposes.

9.4.4 TB and household SEP: addressing the structural determinants of TB

It is relatively easy to suggest that TB control would benefit from interventions improving household SEP or, more broadly, by improving living conditions. Much more challenging is to understand which intervention should be prioritised and what would be the best way to implement it. Having in mind these challenges, this study was designed around a conceptual framework containing levels and constructs suitable for realistic TB preventive strategies. It is important to note that the suggestions provided here represent only two examples of the many ways through which household SEP improvement can be achieved. These two suggestions follow from the conceptual framework used in this research and from the evidence generated from this study. More importantly the solutions suggested come from a public health perspective and are finalised to TB prevention, more than global poverty alleviation or eradication.

(a) Food aid and support

The exploratory analysis of the household SEP suggested that TB control interventions do not have to address poverty as a whole, but they may selectively focus on the improvement of household nutrition standards to see a potential impact on prevalent TB.

If the results of this thesis are generalisable and correct, it is thinkable that areas affected by the double burden of TB and HIV, may benefit from food aid interventions aiming to reduce food insecurity and ensuring a better quality diet. Such interventions would have a critical role in directly improving strength and resistance to TB of the household SEP members. Furthermore, the mitigation of at least one aspect of what contributes to their low household SEP could have an indirect effect by reducing health risky behaviours increasing TB vulnerability including HIV, alcohol abuse and migration.

There is very little evidence on the effect of food aid programs for the prevention of TB. The only study is a very old controlled trial among the families of black TB patients in Harlem, New York city in 1949: 194 families were allocated alternately to receive vitamin and mineral supplement versus no supplements along with the health standard's health education programs. After 5 years of follow up, the risk of TB in the control group was 2.8 times the risk of TB in the vitamin groups. The authors further suggest that because of the several limitations of the trial, the impact observed was actually likely to be an underestimate of the true effect [38].

Provided a similar intervention would obtain ethical approval today, there seems to be no reason why it should not be replicated and evaluated in settings like Zambia these days. It may not be even necessary to implement a food aid program completely from scratch, but simply link up with existing initiatives currently operating in Zambia. According to a recent World Bank report [12], a variety of food aid programs, not specifically aimed to prevent TB, are already in place in Zambia. Recent reviews have concluded that most of the governmental initiatives (mainly played by the National Food and Nutrition Commission) are having minimal, if no impact, in increasing and stabilising household food production and security, for a series of reasons, including the awkward institutional set up for nutrition, lack of donor support and lack of operationalising policy into nutrition programs on the ground [12]. Along these unsuccessful examples, there is also a list of local NGOs that are actually implementing cost-effective interventions to prevent malnutrition at community level: the Program Against Malnutrition (PAM), a local NGO set up to distribute food during the drought in the early 1990s, for example is now working with farmer families to increase household food security. Some of these initiatives could be supported and scaled up to larger areas, if not national level, to see whether they can possibly have an impact on TB prevention.

(b) Financial support

Household SEP could be improved either through interventions aimed to enhance individual social protection (like cash transfer and micro insurance) and through initiatives strengthening the household livelihood and helping people to develop human and financial capital (like microcredit and skill transfer).

In developing countries, the number of initiatives using different strategies of financial support has increased dramatically over the past 40 years. There are convincing evidence from a range of settings that microcredit, for example, can be an effective approach to reduce poverty and support the achievements of the Millennium Development Goals. A number of studies have also proved a significant impact of microcredit in improving household SEP and improving livelihood security among those living in extreme poverty [293]. Much less is known when it comes to the health impact of these initiatives. The few examples available are, however, encouraging: microfinance programmes have been associated with a significant impact on several health indicators, including nutrition, child mortality, immunisation and contraceptive use and HIV prevention [294-297]. Among them, the IMAGE study in South Africa [45]: combined group-based microfinance with a gender and HIV training programme. The study did not show a direct impact on HIV transmission, but only after two years the study participants showed improvement in the economic wellbeing and the nine indicators of women empowerment. Young women participating in the programme reported higher level of HIV-related communication, HIV testing and greater condom use with non-spousal partners. Most remarkably, after two years the risk of physical and sexual partner violence – one of the main risk factor for HIV in South Africa – was reduced by 55% [298].

Conditional cash transfer programs in Latin America provide further evidence of the impact of income supplementation for health [299]: under these programs low income families receive cash, but their receipt of additional cash is conditional on children's enrolment and attendance to school, their compliance with regular visits to health centres for preventive care check-ups, and their participation to information sessions on nutrition and hygiene. Available studies demonstrate that compared to control groups, the intervention groups have increased immunisation rates, decreased illness rates, and reduced level of child stunting [300] [301].

In the field of TB, I could only find one example of financial support explicitly used to support TB treatment: between 1994 and 2001, an NGO in Cambodia developed two home-based and health centre based programs focused on microfinance and food supplementation to enhance compliance. Results show that the program achieved case detection and cure rates among the highest in the world. There was also a significant reduction in the delay between

symptoms onset and start of treatment. According to the authors, the microcredit program, in particular, reduced poverty in the community at large and simultaneously increased the visibility of the health care workers and the interventions, thus presumably increasing awareness of TB symptoms within the community, reducing disease-associated stigma, and increased case detection and treatment adherence [302]. Noticeably, the loan repayment approached 100% and the food supplementation scheme is now a nationwide component of the Cambodian National TB Control Program [302]. Despite the remarkable success of this intervention, the aim of the intervention was again not TB prevention in strict terms, but the support of case finding and treatment.

To my knowledge, the only intervention with a strong preventing focus is one currently implemented in Peru to evaluate the impact of microcredit (in association with vocational training and human health rights empowerment) to reduce intra-household TB transmission rates. Unfortunately, the impact evaluation is on process and no preliminary results are available as yet.

In summary, despite the increasing popularity of financial programs for health impact interventions, their use in TB is still incredibly rare to draw conclusions about their potential utility to support existing TB control paradigm and, even less, as preventive strategies. It is imaginable, though, that social protection and livelihood strengthening can impact the risk of TB through at least three different ways: by improving the household access to food (both in terms of quantity and quality), by affecting the health behaviours and the health seeking behaviours of its members; and by helping the household members to develop human and financial capital.

It is worth noting that socioeconomic interventions may be useful to prevent TB disease, but not necessarily TB infection. This study has provided some evidence that interventions improving the economic power of the household may not be sufficient to protect individuals, at least from infection. Actually, it may become a risk factor. Clearly, the policy message should not be that poverty reduction strategies may increase the risk of TB infection, but simply that higher household SEP may be not enough if not accompanied by interventions improving case finding and treatment and those community. Most importantly, interventions aiming to improve household SEP, may have less impact if not run in combination of interventions addressing the community physical and the environmental characteristics that increase the rates of infection (such as population density, poor urban planning, air pollution etc). Several studies have documented that most of TB transmission occurs at community level, with undiagnosed smear-positive patients as the main source of secondary infections [206-208, 303]. In this view, TB is considered an 'environmental' problem where TB bacilli, spread by undiagnosed contagious cases, represent the pollutant able to spoil the quality of

the air and make it unhealthy for all the community members, no matter how wealthy. It is my opinion, that due to its air-borne nature, even the best financial support program directed to households may not have the expected impact on TB transmission in a 'TB bacilli-polluted' environment. This is the main difference between socioeconomic interventions for HIV and TB: while the first ones can achieve important results through the economic empowerment of the individuals (and the consequent change in unhealthy sexual behaviours), in the case of TB, financial support (either at the individuals or the household level) needs to be accompanied by radical changes in the physical and environmental characteristics of the place where these people live.

A consequence of this is that interventions addressing 'places', rather than individuals or households, may have a bigger and longer-term impact. This recommendation does not want to undermine the importance of addressing TB disease as a way to reduce also TB transmission, I just made the point that in order to enhance TB control, TB infection and TB disease could be addressed from different directions, using different entry points: the household and the individual SEP for TB disease and the community SEP for TB infection.

Because this recommendation follows from a small study with several limitations, further studies are urgently needed to assess the many knowledge gaps still existing in our understanding of the community effect on TB epidemiology. In more than one point, this study has provided suggestions that tackling the adverse effect of living in urban settings may be a core component of addressing the structural determinants of TB; however, the study design does not allow any conclusive inference about the effect of urbanisation or contextual effect in general. In the next section I will provide concrete examples on how to address these knowledge gaps.

Another possible implication of these study results is that in future we may use household SEP as an entry point for active case finding for TB disease or – alternatively – to target people eligible for isoniazid preventive therapy. Given the strong and consistent association between SEP and TB this would be plausible, however, for the moment the implementation of this policy is practically hampered by the lack of an objective indicator of SEP. In other words, while malnutrition, food security, may be relatively easily defined and objectively measured – and therefore used for case finding targeting purposes - in the case of household SEP how could we achieve that? My study showed that TB inequalities are robust to the use of different indicators, but how can we assume that this is true in another context? And how can this can become a policy if the definition of household SEP is so time and setting specific and the measurement methods are all more or less inadequate? As discussed later,

further research on the conceptualisation and measurement of household SEP in developing countries is required before the above suggestion can become a policy recommendation.

9.4.5 Important notes on policy recommendations

Beside the specific recommendations, there are few more general issues that need to be taken into account when considering to prevention strategies for TB control:

1. Admittedly, at this stage, it seems to be clearer what we don't know about the implementation and impact of interventions addressing the social and structural determinants of TB than how to practically move forward; however, it is my view that a number of important lessons can be drawn from existing interventions even if targeting different health outcomes from TB.

2. It is important to stress again the concept that improving living conditions for TB control purposes does not coincide with embarking on complex structural adjustment programmes, but can be achieved by disaggregating the root causes of TB, like poverty and socioeconomic inequalities, into addressable pieces. In the case of malnutrition, for example, a study from the World Bank has shown that in some countries consumption inequalities do not always match with malnutrition inequalities. The authors concluded that 'there must be some form of mechanism in these countries that breaks the link between poverty and malnutrition' [304]. If we understand what this mechanism is we might be able to address a fundamental determinant of TB without having to eradicate poverty.

3. This study suggests that interventions improving household SEP, especially in the form of improving nutrition, should be included as core element of the Zambian National TB Program (NTP) if TB transmission is to be interrupted. It is clear that the Zambian NTP, as any other TB control program, cannot have the responsibility of such policy shift. However, it is recommended that Zambian NTP is more actively involved in any national initiative to fight poverty and improve living conditions, such as the development of poverty reduction strategies papers, and similar processes. Should the resources be available, the NTP should also try to understand better what social and structural determinants, beyond HIV, may be the driving forces of the TB epidemic in this country and identify any collaboration, within or outside the health care sector, potentially useful to tackle these forces.

4. It is evident that TB prevention strategies cannot be addressed by the health care system alone, but require a multisectoral effort, involving experts of different disciplines, and the political commitment to create a solid bridge between health services and the broad welfare and development sector.

9.5 Unanswered questions and future research

This PhD dissertation has provided some additional evidence on the social epidemiology of TB, but it also helped to identify at least three different areas of research, still poorly explored, that would enormously improve our knowledge on the social and structural determinants of TB. This section, that could be also entitled '*what I would do or do differently if I could do all over again*', will discuss how the research agenda on social determinants of TB should be moved forward and will attempt to provide some practical suggestions on how these research questions could be addressed in the future.

9.5.1 The contextual effect in TB epidemiology

The overall picture of this study suggests that the full understanding of the association between household SEP and prevalent TB may be hampered by the apparently complex interaction between area level living conditions and household SEP and, in particular, by the effect of urbanisation.

It is not clear, for example, whether prevalence rates are higher in the urban areas because of some intrinsic 'penalty' of the urban setting [305] or because in Zambia HIV prevalence tends to be higher in this type of settings. The study also clearly showed the existence of TB communities characterised by extremely high rates of prevalent TB, but because of the limited study power it was not possible to detect any significant difference in these communities compared to those with average rates of prevalent TB. It would be interesting to know, for example, whether these 'hot-spots' are due to ongoing local outbreaks or because of the physical and environmental characteristics of these communities independent from the people living in these areas.

This study was not designed to address these research questions, but even in the literature evidence of this type is scanty. To date, only two studies have tried to address the role of contextual variables in TB epidemiology in developing countries through a multilevel design: [178, 279]: in the study from South Africa the importance of the community effect

on self-reported TB changed according to the construct considered: level of employment and poverty rate at community level did not appear to have independent effect on the risk of TB once individual and household variables were taken into account. By contrast, individuals living in areas characterised by higher income inequalities were at twice as much the risk of TB independently from their individual and household level characteristics [178]. More recently, a study from Brazil showed a marked independent association between socioeconomic economic variables identified at individual level and at community level and notified cases of TB: when both types of variables were included in a multivariable model only a modest attenuation of the association between area level variables and the TB outcome was observed. In this study individual level variables were able to explain almost twice as many cases as the area level variables [279].

The limitations of these studies have been discussed by the authors and recently also in a commentary [306]. One additional limitation coming to light from the work presented here is that they do not address the issue separately for TB infection and TB disease. My hypothesis – suggested by the findings of this study – is in fact that area-level characteristics are important, but they may be much more important for TB infection rather than for TB disease.

These studies are often limited by the lack of good quality area-level data. As a result the choice of area-level indicators is often driven by the accessibility of information included in census and living conditions surveys rather than theory [149]. The use of Census or LSMS surveys is a possibility, but in my experience the access to these databases is often a complicate and time consuming process. Some authors have suggested the establishment of ongoing neighbourhood surveys, similar to census, but more in depth, covering not only structural dimensions, but also the environmental and psychosocial factors [94]. Such efforts are almost unthinkable for a developing country, but given the increasing interest in the structural determinants of health, in particular urbanisation [287], and the enormous opportunity for universal interventions the establishing of such a survey in one or two urban areas would justify the investment. In alternative, communities profile could be characterised through a participatory data collection process [287].

The paucity of studies addressing the role of contextual factors in TB epidemiology may also reflect the fact that investigating structural determinants, acting above and beyond the individual level, is often perceived as useless if the change of these constructs lies outside the public health domain [95]. This is understandable, but it may not be completely true: the Millennium Development Goals have set a specific target (Goal 7, target 11) to significantly improve the lives of at least 100 million in slum dwellers by 2020 [287]. Given the public health importance of TB and the close intersection between urbanisation and TB, it should be

relatively easy to convince funding agencies and governmental institutions working in development (for example, The World Bank or UN-Habitat) to support surveys trying to understand better the burden of TB in urban informal settlements and what social dynamics and physical characteristics drive the epidemic of TB in these settings.

With almost 1 billion people living in urban slums in developing countries and an annual population growth in Asia and Africa projected to be 2.4% it is likely that urbanisation will become one of the largest obstacle to the full implementation of current TB control strategies [31, 267] and therefore TB elimination. To respond to this phenomenon knowing whether area-level factors have an effect on the risk of TB independently from individual characteristics is a priority.

9.5.2 The collection of socioeconomic data in TB prevalence surveys

A better understanding of the social epidemiology of TB can only be achieved through a more systematic collection of reliable and comparable data on socioeconomic determinants of TB. Prevalence surveys have been indicated as ideal platforms for the collection of these data [23], however, more operational research is required to understand how to make the collection of these data more valuable.

One aspect relates to the best collection strategy. Nesting an analytical study, like this one, within these prevalence surveys seems to be the most sensible way to collect this information, but economic resources may not be available. The alternative option is the collection of these data through the questionnaires used in the prevalence surveys. It is still not clear, though, if and how this would affect the quality of the prevalence survey. A feasibility study conducted in the Philippines, for example, showed that the inclusion of socioeconomic status data and other risk factors to the questionnaire used for their prevalence survey in 2007 did not have any clear negative impact on the collection of the basic data of the prevalence survey; most importantly, there was no evidence that this affected TB cases detection by increasing the non-participation rate. Nonetheless, additional measures needed to be taken to ensure quality data collection (including additional training, supervision and staffing), and also for the data entry and the data analysis [23].

A second, probably more cumbersome issue, is how to best assess socioeconomic living conditions for the scope of a TB prevalence survey. This study did not solve the question on whether it is best to use an asset-based index or consumption expenditure data: if an asset-based index normally requires the collection of several assets to have a good discriminatory

power, on the other hand, the collection of consumption expenditure can extraordinarily complicate the implementation of a prevalence survey. In reality, Morris and colleagues have proposed a method through which the collection of consumption expenditure data can be replaced by the collection of a short list of key expenditure items though the analysis of existing survey data [254]. The method reflects the principle used in this thesis to predict consumption expenditure, with the difference that, while I have regressed assets type of items to predict consumption expenditures, Morris and colleagues have regressed expenditure data to identify a set of expenditure data that, when summed, best correlate with the overall total of expenditure. There is no guarantee that such relationship will be identified in each country, plus researchers are dependent on the analysis of complex databases, but this method could largely simplify the collection of expenditure data in the context of a prevalence survey and make the results interpretation much more straightforward.

Future prevalence surveys should be designed and appropriately budgeted to assess the advantages and disadvantages of both the expenditure and asset-based approach. Ideally each team should choose the most appropriate depending on what is the purpose of the socioeconomic data collection in the prevalence survey. If the aim is to outline the distribution of TB cases across different socioeconomic groups then an asset-based index or an econometric method based on income and expenditure may be equally valid. This reflects one of the main findings of this thesis suggesting that TB inequalities are robust to the choice of SEP indicators. However, investigators should remember that the magnitude of the inequalities observed may vary depending on the approach used. If, on the other side, the survey goes beyond the simple quantification of TB inequalities and aims to understand the determinants of these inequalities, then the use of one single indicator, whatever it is, it is not appropriate because it may mask the role of specific domains and aspects of household SEP driving the observed inequalities.

One often undervalued aspect is the potential usefulness of preliminary qualitative research to understand the meaning and perception of SEP by the local population. This approach can be particularly useful in identifying the list of items that best reflect the local concept of SEP to avoid the collection of redundant data. In this study, I underestimated the importance of such formative research. This resulted in the development of an extensive and relatively time consuming data collection tool, including items most of which were found to be not useful. Participatory wealth ranking could be a valid alternative to the asset index for studies being carried out in relatively small areas: this technique follows from the assumption that "the poor are the biggest expert of poverty". It involves the qualitative collection of data on concepts and definitions of SEP directly from the community members with the scope to

better characterise the distinctive features of households belonging to different SEP groups according to those that are most likely to know who is rich and who is poor in their community. The disadvantage is that this approach is possible only within a relatively small geographic area. It is also considerably more labour-intensive than an asset-based approach [192].

More pragmatic and more accurate strategies need to be tested in future; however, it seems unlikely we will ever find an easy way to measure SEP: living conditions are time and settings specific and, depending on that, may require different measurement strategies. Each method has advantages and disadvantages, the relative importance of which should be weighted against the level of accuracy required and the resources availability.

9.5.3 Waiting for structural interventions for TB control

I found a tension during my PhD: I was concerned about adding obvious evidence to the knowledge of TB epidemiology and – at the same time - I was also frustrated in seeing how current public health policies for TB often appear to be developed and implemented as if such evidence were not yet available. The dramatic lack of experiences on how to design, implement and evaluate interventions addressing the social and structural determinants of TB identified in this thesis limited the ability to provide useful and concrete policy recommendations, but more importantly represents today one of the biggest knowledge gap in the response to TB.

The reasons for the small attention to socioeconomic interventions go beyond the fear of diverting attention and resources away from the still urgently needed curative approach and include [307-308]:

1. The conceptual resistance in moving from individual focused interventions towards context level interventions;

2. The lack of tools to conceptualise and design social and economic interventions in public health and the lack of research tools to measure, monitor and evaluate them;

3. The erroneous identification of social and economic interventions with broad, national, economic reforms out of the specific competence and expertise of public health;

4. The lack of effective communication and synergy among different institutions and disciplines with different background;

5. The existence of a disease-focused funding mechanism;

Thanks to the Commission on Social Determinants of Health [309] and the increasing interest of the WHO Stop-TB department in health inequalities things are likely to invest more energy and funding in interventions addressing the socioeconomic determinants of TB. In the meantime, such dearth of evidence on the impact of structural interventions for TB can be at least partially compensated by creative and relatively cheap studies based on the analysis of secondary data. In particular, the research agenda could include the following initiatives:

1. Natural experiment studies to assess the unintended health impact of socioeconomic interventions implemented in developing countries. One concrete example particularly useful for the purpose of TB control would be the assessment of TB indicators '*before and after*' the implementation of slum upgrading projects, urban regeneration and community development and housing interventions.

2. The systematic review of studies reporting socioeconomic interventions, aimed to affect the socioeconomic and physical living conditions of a target population or community in order to produce an improvement in a given health outcome somehow related to TB (like HIV, malnutrition, respiratory diseases). The evidence so generated can be then appraised in terms of feasibility, sustainability, transferability and cost-effectiveness to verify their applicability to TB control.

3. Modelling studies inferring the impact of different socioeconomic scenario changes on the burden of TB.

Although imperfect and challenged by data availability, these types of studies could still help to improve our understanding of the social epidemiology of TB, the possible future path of the TB epidemic, the relative importance of different type of interventions addressing both the social and the structural determinants of TB, and how they can complement curative interventions.

10. Conclusions

I hate those who are indifferent

Gramsci, A[11], 1917

The evidence presented in this thesis suggests that today, in a developing country, socioeconomic factors are as important in TB epidemiology as they used to be 100 years ago in Europe and North America and that interventions addressing the structural and the social determinants of TB may be effective in preventing a significant proportion of cases of TB. The emergence of HIV has undoubtedly posed unprecedented challenges to the fight against TB and as a result a large proportion of TB research over the past two decades has been quite rightly invested to address this dual-epidemic. However, HIV seems to have not obscured the role of socioeconomic position. Actually, this study shows how these two factors – at least in this setting – seem to act independently from each other and to be at least equally important. Although no definitive conclusions can be drawn at this stage, this study has also suggested the importance of urbanisation and the living conditions in urban settings in the epidemiology of TB; despite the known limitations of the ecological study, it shown that the apparent advantage of living in an urban site can mask severe health inequalities within the site. It has also shown how – in these communities - relatively wealthier households may be more at risk of TB infection because of housing characteristics and proximity to highly densely populated areas. These observations need to be confirmed by larger and better designed studies, however, evidence point in the direction that TB interventions at context-level may have a strong impact on TB epidemiology. Nowadays, this is accepted for most of the diseases (even chronic) and it should be even more true for an infectious and air-borne disease like TB.

As discussed earlier, the results achieved are complicated and need to be confirmed; however, it is of great importance that this study coincides with the increasing acknowledgment from international organisations that poverty reduction is essential to fight TB [4, 30, 119].

Such renovated focus in the social determinant of TB should not undermine the importance of the DOTS programmes: millions of people have been successfully treated worldwide

because of DOTS and TB control programs should operate and be supported with increased intensity. A greater emphasis on role of poverty and inequalities in the epidemiology of TB should not lead to what has been called a "Luddite trap" [58], under which biomedical advances should be not pursued or welcomed because they do not address the root causes of TB: we should not wait for global poverty to be eradicated to fight TB and adequate investment for better diagnostics and treatment for TB remain a priority.

However, one should admit that TB control is unlikely to happen in countries still affected by poverty and socioeconomic inequalities as severe as they appear today.

What is urgently needed is a shift in the current paradigm for TB control towards a more integrated approach. There is no more time for debate on what should be the real focus of epidemiological studies and what level should be targeted in public health interventions: the proximate risk factors or the upstream determinants? Individuals or societies? Biomedical or socioeconomic interventions? Pills or poverty reduction strategies? Such endless discussion is becoming sterile, the choice between addressing the biomedical aspects and the societal causes of TB is a false dilemma. The real issue here is how we can best use all the knowledge we have about TB, from molecular biology to the role of structural determinants, to develop effective control programs based on strong interdisciplinary approaches. The question is not whether poverty eradication can become a public health programme for TB control [310], but how can we target structural determinants to effectively reduce the burden of TB in affected populations? If we are not prepared to address this question, calling TB 'a plague of poverty' is rhetorical and scientifically dishonest. If we keep ignoring the evidence on social determinants of TB, programs focused only on individual-level risk factors are unlikely to be useful and justifiable from an ethical point of view.

This research project was one of the many studies attempting to document TB inequalities and one of the few trying to explain the inequalities observed, but now it's time for the 'third generation' studies [311]. This stage involves the implementation and the evaluation of interventions in different settings and the use of the interventions results to promote policy changes. To provide concrete solutions, these studies should lie on two pillars: 1) community empowerment: socioeconomic interventions are unlikely to work and be sustainable if communities members are not actively involved in the interventions and if they are not convinced to take responsibility of their livelihood through emancipation and empowerment [312-313] ; and 2) "transdisciplinary" research: a process where "representatives of different disciplines are encouraged to transcend their separate conceptual, theoretical, and methodological orientations in order to develop a shared approach to the research, building on a common conceptual framework' [314].

In the field of HIV/AIDS this idea had taken hold under the definition of structural interventions, that is "interventions that work by altering the context within which health is produced or reproduced. Structural interventions locate the source of public-health problems in the social, economic and political environments that shape and constrain individual, community and societal health outcomes" [315].

In terms of TB control, this does not mean to eradicate poverty, but to break up the scientific questions on the social and structural determinants of TB into actionable pieces. This thesis was an attempt to do exactly this: it suggested, for example, that improving household SEP has the potential to reduce TB transmission in these communities and that among the possible approaches the one addressing household food insecurity and inadequate nutrition is the one more likely to have an impact. As Leonard Syme taught, when a problem is too difficult it may have to be approached by several directions [316]: in this respect, good quality multi-level studies addressing at the same time individual level risk factors for TB, household poverty and social and physical characteristics of the community can help us to identify new entry points of intervention and understand how to maximise our chance of success in fighting TB.

Unfortunately, structural interventions face often the ambivalent scepticism of those who consider the benefit of these interventions far too obvious to be worth proving it and those who claim there is no evidence on the effectiveness of any socioeconomic intervention to support their adoption in the control of TB. A further paradox is represented by the scarce availability of donors investing in this type of interventions: the lack of funding hampers the possibility to evaluate interventions to provide preliminary evidence and the lack of preliminary evidence limits the availability of required funding, creating the conditions for an "evidence lack- funding allocation" vicious cycle.

Despite this, many things suggest that times may be mature to move forward the agenda for structural interventions also for TB. The final report from the WHO Commission on Social Determinants of Health concludes that action to tackle health inequalities must start now and have to include: 1) improving daily living conditions; 2) addressing the inequitable distribution of power, money and resources; 3) measuring and understanding the problem and assessing the impact of action [309]. Following from the Commission recommendations, in January 2009, the Executive board of the World Health Assembly has urged the member states "to generate new or make use of existing methods and evidence tailored to national contexts in order to address the social determinants and the social gradients of health" and has requested the WHO Director general "to support research on effective policies and interventions to improve health by addressing the social determinants of health and to reduce

health inequalities" [317]. Consistently with the general climate, the WHO STOP TB department stated that to speed up the decline of TB incidence enough to get close to elimination by year 2050 it is essential to conjugate the current curative strategy with an intensified effort to reduce the impact of factors that increase vulnerability to TB (i.e. the social determinants of TB) and the underlying structural determinants, including material and social deprivation. [30]. These are all signs of the advent of a new promising era open to new approaches to TB control aiming to integration rather than exclusion.

It is my opinion that we should take advantage of this change to finally move from the sterile 'either/or' debate to a 'both/and' logic [53]: if we want to fight TB and give back hope to the people suffering from this unnecessary disease, we need to become 'competent facilitators' of multisectoral actions and capitalise upon rather than disperse the medical, technical and social notions we have of TB. Between the bohemian, romantic view of TB [318] and medicine, between passionate activism and arid cost-effectiveness thinking, there is a place for solid research resulting in ambitious, but realistic interventions addressing the root causes of TB.

This is where the social epidemiology of TB stands.

References

1. Amrit, S., *Plague of poverty? The World Health Organization, Tuberculosis and International Development. 1945-1980.* . 2002, Cambridge: King's college, University of Cambridge.

2. Waaler, H.T., *Tuberculosis and poverty.* Int J Tuberc Lung Dis, 2002. **6**(9): p. 745-6.

3. Gandy, M., *Life without germs: contested episodes in the history of tuberculosis*, in *The return of the white plague: global poverty and the 'new' tuberculosis*, M. Gandy and A. Zumla, Editors. 2003, Verso: London.

4. WHO, *Global Tuberculosis Control:surveillance, planning, financing: WHO Report.* 2008, The World health Organization: Geneva. p. 21.

5. WHO, *Addressing Poverty in TB Control: options for national TB control programmes.* 2005, WHO.

6. Gwatkin, D.G., M, *The Burden of Disease among the Global Poor*, in *Health, Nutrition and Population*. 2000, The World Bank,: Washington, DC.

7. Glynn, J.R., et al., *Tuberculosis: associations with HIV and socioeconomic status in rural Malawi.* Trans R Soc Trop Med Hyg, 2000. **94**(5): p. 500-3.

8. Schoeman, J.H., M.S. Westaway, and A. Neethling, *The relationship between socioeconomic factors and pulmonary tuberculosis.* Int J Epidemiol, 1991. **20**(2): p. 435-40.

9. Narayanan, D., *Voices of the poor (Volume 1): can anyone ear us*, ed. The World Bank. Vol. 1. 2000, Washington, DC: Oxford University Press, Inc.

10. Krieger, N., *Why epidemiologists cannot afford to ignore poverty.* Epidemiology, 2007. **18**(6): p. 658-63.

11. Milimo, J.T., T. Shilito, and K. Brock, *The poor of Zambia speak. Who would ever listen to the poor.* 2002, Zambia Social Investment Fund: Lusaka.

12. The World Bank, *Zambia: poverty and vulnerability assessment*, A.R. Human Development, Editor. 2005, World Bank: Washington D.C.

13. UNDP, *Millennium Development Goals: A Compact Among Nations to End Human Poverty*. 2003, Oxford University Press: New York.

14. Central Statistical Office - Republic of Zambia, *Living Conditions Monitoring Survey Report 2002-2003.* 2004, Central Statistical Office: Lusaka.

15. Central Statistical Office - Republic of Zambia, *Zambia Demographic and Health Survey 2001-2002*, Central Board of Health and ORC Macro, Editor. 2003, Central Statistical Office: Lusaka.

16. UNAIDS, *Report on the global HIV/AIDS epidemic 2008*. 2008: Geneva.

17. WHO, *Global Tuberculosis Control:surveillance, planning, financing: WHO Report.* 2009, World Health Organization, Geneva.

18. Mwaba, P., et al., *The relentless spread of tuberculosis in Zambia--trends over the past 37 years (1964-2000).* S Afr Med J, 2003. **93**(2): p. 149-52.

19. Bond, V., et al., *The converging impact of tuberculosis, HIV/AIDS, and Food Insecurity in Zambia and South Africa*, in *Final Working Report, February 2009, RENEWAL Programme*. 2009.

20. Needham, D.M., P. Godfrey-Faussett, and S.D. Foster, *Barriers to tuberculosis control in urban Zambia: the economic impact and burden on patients prior to diagnosis.* Int J Tuberc Lung Dis, 1998. **2**(10): p. 811-7.

21. Aspler, A., et al., *Cost of tuberculosis diagnosis and treatment from the patient perspective in Lusaka, Zambia.* Int J Tuberc Lung Dis, 2008. **12**(8): p. 928-35.

22. Nhlema Simwaka, B., et al., *Developing a socio-economic measure to monitor access to tuberculosis services in urban Lilongwe, Malawi.* Int J Tuberc Lung Dis, 2007. **11**(1): p. 65-71.

23. Lonnroth, K., et al., *Inclusion of information on risk factors, socio-economic status and health seeking in a tuberculosis prevalence survey.* Int J Tuberc Lung Dis, 2009. **13**(2): p. 171-6.

24. WHO, *Assessing tuberculosis prevalence through population-based surveys*. 2007, World Health Organization: Geneva.

25. Corbett, E.L., et al., *Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners.* Am J Respir Crit Care Med, 2004. **170**(6): p. 673-9.

26. Rieder, H.L., *Epidemiologic basis for tuberculosis control.* 1999, Paris: International Union Against Tuberculosis and Lung Disease.

27. Comstock, G.W., *Frost revisited: the modern epidemiology of tuberculosis.* Am J Epidemiol, 1975. **101**(5): p. 363-82.

28. Lienhardt, C., *From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis.* Epidemiol Rev, 2001. **23**(2): p. 288-301.

29. Comstock, G.W., *Epidemiology of tuberculosis.* Am Rev Respir Dis, 1982. **125**(3 Pt 2): p. 8-15.

30. Dye, C., et al., *Trends in tuberculosis incidence and their determinants in 134 countries.* Bull World Health Organ, 2009. **87**(9): p. 683-91.

31. Lonnroth, K. and M. Raviglione, *Global epidemiology of tuberculosis: prospects for control.* Semin Respir Crit Care Med, 2008. **29**(5): p. 481-91.

32. Baker, M., et al., *Tuberculosis associated with household crowding in a developed country.* J Epidemiol Comm Health, 2008. **62**(8): p. 715-21.

33. Wanyeki, I., et al., *Dwellings, crowding, and tuberculosis in Montreal.* Soc Sci Med, 2006. **63**(2): p. 501-11.

34. Escombe, A.R., et al., *Natural ventilation for the prevention of airborne contagion.* PLoS Med, 2007. 4(2): p. e68.

35. Canadian Tuberculosis Committee, *Housing conditions that serve as risk factors for tuberculosis infection and disease. An Advisory Committee Statement (ACS).* Can Commun Dis Rep, 2007. **33**(ACS-9): p. 1-13.

36. Clark, M., P. Riben, and E. Nowgesic, *The association of housing density, isolation and tuberculosis in Canadian First Nations communities.* Int J Epidemiol, 2002. **31**(5): p. 940-5.

37. Corbett, E.L., et al., *The growing burden of tuberculosis: global trends and interactions with the HIV epidemic.* Arch Intern Med, 2003. **163**(9): p. 1009-21.

38. Cegielski, J.P. and D.N. McMurray, *The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals.* Int J Tuberc Lung Dis, 2004. 8(3): p. 286-98.

39. Lin, H.H., M. Ezzati, and M. Murray, *Tobacco smoke, indoor air pollution and tuberculosis: a systematic review and meta-analysis.* PLoS Med, 2007. 4(1): p. e20.

40. Slama, K., et al., *Tobacco and tuberculosis: a qualitative systematic review and meta-analysis.* Int J Tuberc Lung Dis, 2007. **11**(10): p. 1049-61.

41. Rehfuess, E., C. Corvalan, and M. Neira, *Indoor air pollution: 4000 deaths a day must no longer be ignored.* Bull World Health Organ, 2006. **84**(7): p. 508.

42. Rehfuess, E., S. Mehta, and A. Pruss-Ustun, *Assessing household solid fuel use: multiple implications for the Millennium Development Goals.* Environ Health Perspect, 2006. **114**(3): p. 373-8.

43. Lonnroth, K., et al., *Alcohol use as a risk factor for tuberculosis - a systematic review.* BMC Public Health, 2008. **8**: p. 289.

44. Stevenson, C.R., et al., *Diabetes and tuberculosis: the impact of the diabetes epidemic on tuberculosis incidence.* BMC Public Health, 2007. **7**: p. 234.

45. Pronyk, P.M., et al., *Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial.* Lancet, 2006. **368**(9551): p. 1973-83.

46. Davies, P.D., *The world-wide increase in tuberculosis: how demographic changes, HIV infection and increasing numbers in poverty are increasing tuberculosis.* Ann Med, 2003. **35**(4): p. 235-43.

47. WHO, *Fory-fourth World Health Assembly*. 1991, World Health Organization: Geneva.

48. Raviglione, M.C. and M.W. Uplekar, *WHO's new Stop TB Strategy*. Lancet, 2006. **367**(9514): p. 952-5.

49. Snider, D.E., Jr. and W.L. Roper, *The new tuberculosis.* N Engl J Med, 1992. **326**(10): p. 703-5.

50. Dye, C. and B.G. Williams, *Eliminating human tuberculosis in the twenty-first century.* J R Soc Interface, 2008. **5**(23): p. 653-62.

51. Dubos, R.J., *The White Plague: Tuberculosis, Man, and Society*. Reissue edition (1 Jan 1987) ed. 1987: Rutgers University Press.

52. McKeown, T., *Medical issues in historical demography.* Int J Epidemiol, 2005. **34**(3): p. 515-20.

53. Krieger, N., *Theories for social epidemiology in the 21st century: an ecosocial perspective.* Int J Epidemiol, 2001. **30**(4): p. 668-77.

54. Grundy, E., *Commentary: The McKeown debate: time for burial.* Int J Epidemiol, 2005. **34**(3): p. 529-33.

55. Szreter, S., *Rethinking McKeown: the relationship between public health and social change.* Am J Public Health, 2002. **92**(5): p. 722-5.

56. Davies, R.P., et al., *Historical declines in tuberculosis in England and Wales: improving social conditions or natural selection?* Int J Tuberc Lung Dis, 1999. **3**(12): p. 1051-4.

57. Davies, P.D., *Medical classics. La Boheme and tuberculosis.* BMJ, 2008. **337**: p. a2587.

58. Grange, J.M., et al., *Historical declines in tuberculosis: nature, nurture and the biosocial model.* Int J Tuberc Lung Dis, 2001. **5**(3): p. 208-12.

59. Obermeyer, Z., J. Abbott-Klafter, and C.J. Murray, *Has the DOTS strategy improved case finding or treatment success? An empirical assessment.* PLoS ONE, 2008. 3(3): p. e1721.

60. Soemantri, S., et al., *Three-fold reduction in the prevalence of tuberculosis over 25 years in Indonesia.* Int J Tuberc Lung Dis, 2007. 11(4): p. 398-404.

61. Al-Maniri, A., et al., *Towards the elimination of tuberculosis in a developing country: 25 years of tuberculosis control in Oman.* Int J Tuberc Lung Dis, 2007. 11(2): p. 175-80.

62. Huong, N.T., et al., *Tuberculosis epidemiology in six provinces of Vietnam after the introduction of the DOTS strategy.* Int J Tuberc Lung Dis, 2006. 10(9): p. 963-9.

63. Dye, C., et al., *The decline of tuberculosis epidemics under chemotherapy: a case study in Morocco.* Int J Tuberc Lung Dis, 2007. 11(11): p. 1225-31.

64. Oakes, J.M. and J.S. Kaufman, *Methods in social epidemiology*. 2006, San Francisco: Jossey-Bass.

65. Krieger, N., *Embodiment: a conceptual glossary for epidemiology.* J Epidemiol Community Health, 2005. 59(5): p. 350-5.

66. Marmot, M., *Social determinants of health inequalities.* Lancet, 2005. 365(9464): p. 1099-104.

67. Regidor, E., *Social determinants of health: a veil that hides socioeconomic position and its relation with health.* J Epidemiol Community Health, 2006. 60(10): p. 896-901.

68. United Nations. *World Summit for Social Development.* 1995 [cited 2009; Available from: http://www.un.org/esa/socdev/wssd/.

69. Falkingham, J. and C. Namazie, *Measuring Health and Poverty: a review of approaches to identifying the poor*. 2002, Department for International Development (DFID), Health Systems Resouce Centre.

70. Organisation for economic co-operation and development (OECD). *DAC Guidelines on poverty reduction.* 2003 05/05/2009]; Available from: http://www.oecd.org/document/1/0,3343,en_2649_34621_1885953_1_1_1_1,00.html.

71. Krieger, N., *A glossary for social epidemiology.* J Epidemiol Community Health, 2001. 55(10): p. 693-700.

72. Commission on Social Determinants of Health. *A conceptual framework for action on the social determinants of health.* 2007 February 16, 2010]; Available from: http://www.who.int/social_determinants/resources/csdh_framework_action_05_07.pdf.

73. Berkman, L.F. and I. Kawachi, *Social Epidemiology*. 2000, New York: Oxford University Press.

74. Bartley, M., *Commentary: Relating social structure and health.* Int J Epidemiol, 2003. **32**(6): p. 958-60.

75. Blane, D., *Commentary: Socioeconomic health differentials.* Int J Epidemiol, 2001. **30**(2): p. 292-3.

76. Krieger, N., et al., *Racism, sexism, and social class: implications for studies of health, disease, and well-being.* Am J Prev Med, 1993. 9(6 Suppl): p. 82-122.

77. Bartley, M., *Health inequality. An introduction to theories, concepts and methods.* 2004, Cambridge: Polity Press.

78. Krieger, N., D.R. Williams, and N.E. Moss, *Measuring social class in US public health research: concepts, methodologies, and guidelines.* Annu Rev Public Health, 1997. **18**: p. 341-78.

79. Lynch, J. and G.A. Kaplan, *Socioeconomic position*, in *Social Epidemiology*, L.F. Berkman and I. Kawachi, Editors. 2000, Oxford University Press: New York. p. 13-35.

80. Graham, H., *Social determinants and their unequal distribution: clarifying policy understandings.* Milbank Q, 2004. **82**(1): p. 101-24.

81. Diez Roux, A.V., *Investigating neighborhood and area effects on health.* Am J Public Health, 2001. **91**(11): p. 1783-9.

82. Sampson, R.J., J.D. Morenoff, and T. Gannon-Rowley, *Aassessing "neighborhood effect": social processes and new directions in research.* Annual Review of Sociolohy, 2002. **28**: p. 443-478.

83. Deaton, A. and M. Grosh, *Consumption*, in *Designing household survey questionnaires for developing countries: Lessons from ten years of LSMS experience*, M. Grosh and P. Glewwe, Editors. 2000, The World Bank: Washington D.C. p. 91-133.

84. Gwatkin, D.R., et al., *Socio-economic differences in health, nutrition and population within developing countries. An overview*, in *HNP*, World Bank, Editor. 2007, World Bank.

85. Filmer, D. and L.H. Pritchett, *Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India.* Demography, 2001. 38(1): p. 115-32.

86. Rutstein, S.O. and K. Johnson, *The wealth index*, in *DHS Comparative Report N.6*, MEASURE DHS+, Editor. 2004, ORC Macro: Calverton, Maryland USA.

87. Montgomery, M.R., et al., *Measuring living standards with proxy variables.* Demography, 2000. **37**(2): p. 155-74.

88. Bollen, K.A., D.K. Guilkey, and T.A. Mroz, *Binary outcomes and endogenous explanatory variables: tests and solutions with an application to the demand for contraceptive use in Tunisia.* Demography, 1995. **32**(1): p. 111-31.

89. Ewer, K., et al., *Comparison of T-cell-based assay with tuberculin skin test for diagnosis of Mycobacterium tuberculosis infection in a school tuberculosis outbreak.* Lancet, 2003. **361**(9364): p. 1168-73.

90. Sahn, D.E. and D. Stifel, *Exploring alternative measures of welfare in the absence of expenditure data.* The review of income and wealth, 2003. **49**(4): p. 463-489.

91. Hentschel, J., et al., *Combining census and survey data to trace the spatial dimensions of poverty: a case study of Ecuador.* The World Bank Economic Review, 2000. **14**(1): p. 147-65.

92. Alderman, H., et al., *Unitary versus collective models of the household: is it time to shift the burden of proof?* . The World Bank Research Observer, 1995. **10**: p. 1-19.

93. Chandola, T., et al., *Social inequalities in health by individual and household measures of social position in a cohort of healthy people.* J Epidemiol Community Health, 2003. **57**(1): p. 56-62.

94. Rajaratnam, J.K., J.G. Burke, and P. O'Campo, *Maternal and child health and neighborhood context: the selection and construction of area-level variables.* Health Place, 2006. **12**(4): p. 547-56.

95. Pickett, K.E. and M. Pearl, *Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review.* J Epidemiol Community Health, 2001. **55**(2): p. 111-22.

96. O'Campo, P. and M. O'Brien Caughy, *Measures of residential community contexts*, in *Methods in social epidemiology*, J.M. Oakes and J.S. Kaufman, Editors. 2006, Jossey-Bass: San Francisco. p. 193-208.

97. Macintyre, S. and A. Ellaway, *Neighborhoods and health: an overview*, in *Neighborohood and health*, I. Kawachi and L.F. Berkman, Editors. 2003, Oxford University Press: New York. p. 20-44.

98. Macintyre, S., *The Black Report and beyond: what are the issues?* Soc Sci Med, 1997. **44**(6): p. 723-45.

99. Bartley, M., *Models of aetiological pathways, III: the materialistic model*, in *Health inequality: an introduction to theories, concepts and methods* M. Bartley, Editor. 2004, Polity Press: Cambridge. p. 90-102.

100. Blane, D.B., M. Davey Smith, G., *Disease aetiology and materialist explanations of socioeconomic mortality differentials.* European Journal of Public Health, 1997. 7(4): p. 385-391.

101. Lynch, J.W., et al., *Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions.* Bmj, 2000. **320**(7243): p. 1200-4.

102. Bartley, M., *Social ecology*, in *Health inequality: an introduction to theories, concepts and methods* M. Bartley, Editor. 2004, Polity Press: Cambridge. p. 116-134.

103. Daly, M.C., et al., *Macro to micro links in the relationship between income inequality and mortality.* Milbank Quarterly, 1998. **76**(3): p. 315-39.

104. Marmot, M., *Multilevel approaches to understanding social determinants*, in *Social Epidemiology*, L.F. Berkman and I. Kawachi, Editors. 2000, Oxford University Press: New York. p. 349-367.

105. Wilkinson, R.G., *Heakth inequalities within societies*, in *Unhealthy societies: the affliction of inequality*, R.G. Wilkinson, Editor. 1996, Routledge: London. p. 53-112.

106. Marmot, M. and R.G. Wilkinson, *Psychosocial and material pathways in the relation between income and health: a response to Lynch et al.* BMJ, 2001. **322**(7296): p. 1233-6.

107. Bartley, M., *Models of aetiological pathways, II: the psycho-social model*, in *Health inequality: an introduction to theories, concepts and methods* M. Bartley, Editor. 2004, Polity Press: Cambridge. p. 78-89.

108. Bosma, H., et al., *Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study.* BMJ, 1997. **314**(7080): p. 558-65.

109. Marmot, M.G., et al., *Contribution of job control and other risk factors to social variations in coronary heart disease incidence.* Lancet, 1997. **350**(9073): p. 235-9.

110. Lynch, J., et al., *Income inequality, the psychosocial environment, and health: comparisons of wealthy nations.* Lancet, 2001. **358**(9277): p. 194-200.

111. Krieger, N. and G. Davey Smith, *"Bodies count," and body counts: social epidemiology and embodying inequality.* Epidemiol Rev, 2004. **26**: p. 92-103.

112. Robert, S.A., *Neighborhood socioeconomic context and adult health. The mediating role of individual health behaviors and psychosocial factors.* Ann N Y Acad Sci, 1999. **896**: p. 465-8.

113. Montgomery, M.R. and P.C. Hewett, *Urban poverty and health in developing countries: household and neighborhood effects.* Demography, 2005. **42**(3): p. 397-425.

114. Stafford, M. and M. Marmot, *Neighbourhood deprivation and health: does it affect us all equally?* Int J Epidemiol, 2003. **32**(3): p. 357-66.

115. Johnston, W., *The modern epidemic: a history of tuberculosis in Japan*, ed. Council on East Asian Studies. 1995, Cambridge, Massachusset, US: Harvard University.

116. Fairchild, A.L. and G.M. Oppenheimer, *Public health nihilism vs pragmatism: history, politics, and the control of tuberculosis.* Am J Public Health, 1998. **88**(7): p. 1105-17.

117. Wilson, L.G., *Commentary: Medicine, population, and tuberculosis.* Int J Epidemiol, 2005. **34**(3): p. 521-4.

118. Hanson, C., *Tuberculosis, poverty and inequity: a review of the literature and discussion of issues*. 2002, World Bank WB.

119. Janssens, J.P. and H.L. Rieder, *An ecological analysis of incidence of tuberculosis and per capita gross domestic product.* Eur Respir J, 2008. **32**(5): p. 1415-6.

120. Beckhurst, C., et al., *Factors influencing the distribution of tuberculosis cases in an inner London borough.* Commun Dis Public Health, 2000. **3**(1): p. 28-31.

121. Kistemann, T., A. Munzinger, and F. Dangendorf, *Spatial patterns of tuberculosis incidence in Cologne (Germany).* Soc Sci Med, 2002. **55**(1): p. 7-19.

122. Tanrikulu, A.C., et al., *Tuberculosis in Turkey: high altitude and other socio-economic risk factors.* Public Health, 2008. **122**(6): p. 613-9.

123. Hadjichristodoulou, C., P. Christie, and S. O'Brien, *Pulmonary tuberculosis and deprivation in hospitalised patients in Scotland.* Eur J Epidemiol, 2001. **17**(1): p. 85-7.

124. de Pontual, L., et al., *Childhood tuberculosis in a low-income Paris suburb: lessons from a resurgence brought under control.* Int J Tuberc Lung Dis, 2004. **8**(8): p. 976-81.

125. Muniyandi, M. and R. Ramachandran, *Socioeconomic inequalities of tuberculosis in India.* Expert Opin Pharmacother, 2008. **9**(10): p. 1623-8.

126. Vendramini, S.H., et al., *Tuberculosis risks and socio-economic level: a case study of a city in the Brazilian south-east, 1998-2004.* Int J Tuberc Lung Dis, 2006. **10**(11): p. 1231-5.

127. Souza, W.V., et al., *The use of socioeconomic factors in mapping tuberculosis risk areas in a city of northeastern Brazil.* Rev Panam Salud Publica, 2000. **8**(6): p. 403-10.

128. Geraldes Santos Mde, L., et al., *Poverty: socioeconomic characterization at tuberculosis.* Rev Lat Am Enfermagem, 2007. **15 Spec No**: p. 762-7.

129. Vieira Rda, C., et al., *[Spatial distribution of new tuberculosis cases in Vitoria, State of Espirito Santo, between 2000 and 2005].* Rev Soc Bras Med Trop, 2008. **41**(1): p. 82-6.

130. Mota, F.F., et al., *[Spatial distribution of tuberculosis mortality in Salvador, Bahia, Brazil].* Cad Saude Publica, 2003. **19**(4): p. 915-22.

131. Sanghavi, D.M., et al., *Hyperendemic pulmonary tuberculosis in a Peruvian shantytown.* Am J Epidemiol, 1998. **148**(4): p. 384-9.

132. van Rie, A., et al., *Childhood tuberculosis in an urban population in South Africa: burden and risk factor.* Arch Dis Child, 1999. **80**(5): p. 433-7.

133. Munch, Z., et al., *Tuberculosis transmission patterns in a high-incidence area: a spatial analysis.* Int J Tuberc Lung Dis, 2003. 7(3): p. 271-7.

134. Tupasi, T.E., et al., *Tuberculosis in the urban poor settlements in the Philippines.* Int J Tuberc Lung Dis, 2000. 4(1): p. 4-11.

135. Akhtar, S., et al., *Hyperendemic pulmonary tuberculosis in peri-urban areas of Karachi, Pakistan.* BMC Public Health, 2007. 7: p. 70.

136. Barr, R.G., et al., *Neighborhood poverty and the resurgence of tuberculosis in New York City, 1984-1992.* Am J Public Health, 2001. **91**(9): p. 1487-93.

137. Mangtani, P., et al., *Socioeconomic deprivation and notification rates for tuberculosis in London during 1982-91.* BMJ, 1995. **310**(6985): p. 963-6.

138. Toungoussova, O.S., G. Bjune, and D.A. Caugant, *Epidemic of tuberculosis in the former Soviet Union: social and biological reasons.* Tuberculosis (Edinb), 2006. 86(1): p. 1-10.

139. Dye, C., *The science of social diseases.* Science, 2005. **307**(5707): p. 181.

140. Ilic, M., et al., *Trends and patterns of tuberculosis: interpretations and prospects of tuberculosis control in Vojvodina, Yugoslavia, 1987-2000.* Int J Tuberc Lung Dis, 2003. 7(11): p. 1064-9.

141. McKee, M., *The health consequences of the collapse of the Soviet Union*, in *Poverty, inequality, and health*, D.A.W. Leon, G., Editor. 2000, Oxford University Press: London. p. 17-36.

142. Shilova, M.V. and C. Dye, *The resurgence of tuberculosis in Russia.* Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1411): p. 1069-75.

143. Coker, R.J., R.A. Atun, and M. McKee, *Health-care system frailties and public health control of communicable disease on the European Union's new eastern border.* Lancet, 2004. **363**(9418): p. 1389-92.

144. Walberg, P., et al., *Economic change, crime, and mortality crisis in Russia: regional analysis.* BMJ, 1998. **317**(7154): p. 312-8.

145. Leon, D.A., et al., *Huge variation in Russian mortality rates 1984-94: artefact, alcohol, or what?* Lancet, 1997. **350**(9075): p. 383-8.

146. Diez-Roux, A.V., et al., *Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies.* Ann Epidemiol, 2001. **11**(6): p. 395-405.

147. Macintyre, S. and A. Ellaway, *Ecological approaches: rediscovering the role of the physical and social environment.*, in *Social epidemiology*, L.F. Berkman and I. Kawachi, Editors. 2000, Oxford University press: New York. p. 332-348.

148. Schwartz, S., *The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences.* Am J Public Health, 1994. **84**(5): p. 819-24.

149. Diez Roux, A.V., *The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches.* Epidemiol Rev, 2004. **26**: p. 104-11.

150. Diez Roux, A.V., *A glossary for multilevel analysis.* J Epidemiol Community Health, 2002. **56**(8): p. 588-94.

151. Muniyandi, M., et al., *The prevalence of tuberculosis in different economic strata: a community survey from South India.* Int J Tuberc Lung Dis, 2007. **11**(9): p. 1042-5.

152. Singh, M., et al., *Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis.* Arch Dis Child, 2005. **90**(6): p. 624-8.

153. Shetty, N., et al., *An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study.* Int J Tuberc Lung Dis, 2006. **10**(1): p. 80-6.

154. Kaulagekar, A. and A. Radkar, *Social status makes a difference: tuberculosis scenario during National Family Health Survey-2.* Indian J Tuberc, 2007. **54**(1): p. 17-23.

155. Rathi, S.K., et al., *Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulnionary tuberculosis cases in Umerkot, Pakistan.* Int J Tuberc Lung Dis, 2002. **6**(10): p. 851-7.

156. Karyadi, E., et al., *Social aspects of patients with pulmonary tuberculosis in Indonesia.* Southeast Asian J Trop Med Public Health, 2002. **33**(2): p. 338-45.

157. Tornee, S., et al., *Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand.* Southeast Asian J Trop Med Public Health, 2004. **35**(2): p. 375-83.

158. Jackson, S., et al., *Poverty and the economic effects of TB in rural China.* Int J Tuberc Lung Dis, 2006. **10**(10): p. 1104-10.

159. Pelly, T.F., et al., *Tuberculosis skin testing, anergy and protein malnutrition in Peru.* Int J Tuberc Lung Dis, 2005. 9(9): p. 977-84.

160. Olender, S., et al., *Low prevalence and increased household clustering of Mycobacterium tuberculosis infection in high altitude villages in Peru.* Am J Trop Med Hyg, 2003. **68**(6): p. 721-7.

161. Sanchez-Perez, H., et al., *Pulmonary tuberculosis and associated factors in areas of high levels of poverty in Chiapas, Mexico.* Int J Epidemiol, 2001. **30**(2): p. 386-93.

162. Perez-Padilla, R., et al., *Cooking with biomass stoves and tuberculosis: a case control study.* Int J Tuberc Lung Dis, 2001. 5(5): p. 441-7.

163. Buskin, S.E., et al., *Tuberculosis risk factors in adults in King County, Washington, 1988 through 1990.* Am J Public Health, 1994. **84**(11): p. 1750-6.

164. Saiman, L., et al., *Risk factors for latent tuberculosis infection among children in New York City.* Pediatrics, 2001. **107**(5): p. 999-1003.

165. Tocque, K., et al., *A case-control study of lifestyle risk factors associated with tuberculosis in Liverpool, North-West England.* Eur Respir J, 2001. **18**(6): p. 959-64.

166. Goldman, J.M., et al., *Childhood tuberculosis in Leeds, 1982-90: social and ethnic factors and the role of the contact clinic in diagnosis.* Thorax, 1994. 49(2): p. 184-5.

167. Ponticiello, A., et al., *Demographic risk factors and lymphocyte populations in patients with tuberculosis and their healthy contacts.* Int J Tuberc Lung Dis, 2001. 5(12): p. 1148-55.

168. Tekkel, M., et al., *Risk factors for pulmonary tuberculosis in Estonia.* Int J Tuberc Lung Dis, 2002. 6(10): p. 887-94.

169. Coker, R., et al., *Risk factors for pulmonary tuberculosis in Russia: case-control study.* Bmj, 2006. **232**: p. 85-7.

170. Crampin, A.C., et al., *Tuberculosis and gender: exploring the patterns in a case control study in Malawi.* Int J Tuberc Lung Dis, 2004. 8(2): p. 194-203.

171. Sinfield, R., et al., *Risk factors for TB infection and disease in young childhood contacts in Malawi.* Ann Trop Paediatr, 2006. **26**(3): p. 205-13.

172. Gustafson, P., et al., *Tuberculosis in Bissau: incidence and risk factors in an urban community in sub-Saharan Africa.* Int J Epidemiol, 2004. 33(1): p. 163-72.

173. Lienhardt, C., et al., *Investigation of the risk factors for tuberculosis: a case-control study in three countries in West Africa.* Int J Epidemiol, 2005. **34**(4): p. 914-23.

174. Gustafson, P., et al., *Risk factors for positive tuberculin skin test in Guinea-Bissau.* Epidemiology, 2007. **18**(3): p. 340-7.

175. Hill, P.C., et al., *Risk factors for pulmonary tuberculosis: a clinic-based case control study in The Gambia.* BMC Public Health, 2006. **6**: p. 156.

176. Lienhardt, C., et al., *Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia.* Am J Respir Crit Care Med, 2003. **168**(4): p. 448-55.

177. Lienhardt, C., et al., *Risk factors for tuberculosis infection in children in contact with infectious tuberculosis cases in the Gambia, West Africa.* Pediatrics, 2003. **111**(5 Pt 1): p. e608-14.

178. Harling, G., R. Ehrlich, and L. Myer, *The social epidemiology of tuberculosis in South Africa: a multilevel analysis.* Soc Sci Med, 2008. **66**(2): p. 492-505.

179. Singh, V., et al., *TB control, poverty, and vulnerability in Delhi, India.* Trop Med Int Health, 2002. **7**(8): p. 693-700.

180. Crampin, A.C., et al., *Field-based random sampling without a sampling frame: control selection for a case-control study in rural Africa.* Trans R Soc Trop Med Hyg, 2001. **95**(5): p. 481-3.

181. Robins, J.M. and S. Greenland, *Identifiability and exchangeability for direct and indirect effects.* Epidemiology, 1992. **3**(2): p. 143-55.

182. Victora, C.G., et al., *The role of conceptual frameworks in epidemiological analysis: a hierarchical approach.* Int J Epidemiol, 1997. **26**(1): p. 224-7.

183. Nhlema, B.J., K; Steenbergen, G; Theobald, S; Tang, S; Bertel Squire, S, *A Systematic Analysis of TB and Poverty.* 2003.

184. Killewo, J., *Poverty, TB, and HIV infection: a vicious cycle.* J Health Popul Nutr, 2002. **20**(4): p. 281-4.

185. Kuemmerer, J.M. and G.W. Comstock, *Sociologic concomitants of tuberculin sensitivity.* Am Rev Respir Dis, 1967. **96**(5): p. 885-92.

186. Engel, A. and J. Roberts, *Tuberculin skin test reaction among adults 25-74 years, United States, 1971-72.* Vital Health Stat 11, 1977(204): p. 1-40.

187. Reichman, L.B. and R. O'Day, *Tuberculous infection in a large urban population.* Am Rev Respir Dis, 1978. **117**(4): p. 705-12.

188. Bartley, M., *Models of aetiological pathways, I: Behavioural and 'Cultural" explanations*, in *Health inequality: an introduction to theories, concepts and methods* M. Bartley, Editor. 2004, Polity Press: Cambridge. p. 64-77.

189. Galobardes, B., et al., *Indicators of socioeconomic position*, in *Methods in social epidemiology*, J.M. Oakes and J.S. Kaufman, Editors. 2006, Jossey-Bass: San Francisco. p. 47-85.

190. Food and Agricultural Organization, *Rome Declaration on World Food Security and World Food Summit Plan of Action*, in *World Food Summit , 13-17 November*, FAO, Editor. 1996: Rome.

191. Food and Agricultural Organization, *Trade reforms and food security. Conceptualising the linkages*, Economic and Social Development Department, Editor. 2003, FAO,: Rome.

192. Howe, L.D., *The wealth index as a measure of socio-economic position*, in *Thesis submitted for the degree of Doctor of Philosophy*. 2008, University of London, Faculty of Medicine: London. p. 404.

193. Coudouel, A.H., JS; Wodon, QT, *Poverty measurement and analysis*, in *PRSP Book*, W. Bank, Editor. 2004, World Bank: Washington, D.C.

194. Løvendal, C.R., M. Knowles, and N. Horii, *Understanding vulnerability to food insecurity. Lessons from vulnerable groups profiling*, Agricultural and Development Economics Division, Editor. 2004, Food and Agricultural Organization, : Rome.

195. Hennekens, C.H. and J.E. Buring, *Analysis of epidemiologic studies: evaluating the role of confounding*, in *Epidemiology in medicine*, C.H. Hennekens and J.E. Buring, Editors. 1987, Lippincott Williams & Wilkins: Philadelphia. p. 287-304.

196. Hunter, J.M. and M.O. Thomas, *Hypothesis of leprosy, tuberculosis and urbanization in Africa.* Soc Sci Med, 1984. **19**(1): p. 27-57.

197. Cord, L., *Rural Poverty*, in *A source book for poverty reduction strategy*, J. Klugman, Editor. 2002, World Bank: Washington D.C. p. 67-115.

198. Holmes, C.B., H. Hausler, and P. Nunn, *A review of sex differences in the epidemiology of tuberculosis.* Int J Tuberc Lung Dis, 1998. **2**(2): p. 96-104.

199. Hudelson, P., *Gender differentials in tuberculosis: the role of socio-economic and cultural factors.* Tuber Lung Dis, 1996. **77**(5): p. 391-400.

200. Hafeman, D.M. and S. Schwartz, *Opening the Black Box: a motivation for the assessment of mediation.* Int J Epidemiol, 2009.

201. Perronne, C., *Tuberculosis, HIV infection, and malnutrition: an infernal trio in central Africa.* Nutrition, 1999. **15**(4): p. 321-2.

202. USAID, *Nutrition and Tuberculosis: A review of the literature and considerations for TB control programs*. 2008, USAID: Washington.

203. Andersen, P. and T.M. Doherty, *The success and failure of BCG - implications for a novel tuberculosis vaccine.* Nat Rev Microbiol, 2005. **3**(8): p. 656-62.

204. Brewer, T.F., *Preventing tuberculosis with bacillus Calmette-Guerin vaccine: a meta-analysis of the literature.* Clin Infect Dis, 2000. **31 Suppl 3**: p. S64-7.

205. Claessens, N.J., et al., *High frequency of tuberculosis in households of index TB patients.* Int J Tuberc Lung Dis, 2002. **6**(3): p. 266-9.

206. Garcia-Garcia, M., et al., *The role of core groups in transmitting Mycobacterium tuberculosis in a high prevalence community in Southern Mexico.* Int J Tuberc Lung Dis, 2000. **4**(1): p. 12-7.

207. Klovdahl, A.S., et al., *Networks and tuberculosis: an undetected community outbreak involving public places.* Soc Sci Med, 2001. **52**(5): p. 681-94.

208. Mangura, B.T., et al., *Mycobacterium tuberculosis miniepidemic in a church gospel choir.* Chest, 1998. **113**(1): p. 234-7.

209. Murray, E.J., et al., *A multidisciplinary method to map potential tuberculosis transmission 'hot spots' in high-burden communities.* Int J Tuberc Lung Dis, 2009. **13**(6): p. 767-74.

210. Chin, D.P., et al., *Spread of Mycobacterium tuberculosis in a community implementing recommended elements of tuberculosis control.* JAMA, 2000. **283**(22): p. 2968-74.

211. Ayles, H., et al., *Prevalence of tuberculosis, HIV and respiratory symptoms in two Zambian communities: implications for tuberculosis control in the era of HIV.* PLoS ONE, 2009. **4**(5): p. e5602.

212. Central Statistical Office - Republic of Zambia, *2000 Census of Population and Housing, Lusaka Province, Volume 5*. 2004, Central Statisical Office,: Lusaka.

213. Metchock, B.G., F.S. Nolte, and J.R. Wallace, *Mycobacterium*, in *Manual of clinical microbiology*, P.R. Murray, et al., Editors. 1999, American Society for Microbiology: Washington.

214. Kamerbeek, J., et al., *Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology.* J Clin Microbiol, 1997. **35**(4): p. 907-14.

215. Achikbache, B., et al., *Annex E - Strenghtening statistical systems: technical notes and case studies*, in *A source book for poverty reduction strategy*, J. Klugman, Editor. 2002, World Bank: Washington D.C. p. 67-115.

216. Krieger, N. *Geocoding and Monitoring US Socioeconomic Inequalities in Health: An introduction to using area-based socioeconomic measures.* 1998 [cited 2009 24/06/2009]; Available from: http://www.hsph.harvard.edu/thegeocodingproject/.

217. Krieger, N., et al., *Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project.* Am J Epidemiol, 2002. 156(5): p. 471-82.

218. Messer, L.C., et al., *Socioeconomic domains and associations with preterm birth.* Soc Sci Med, 2008.

219. Messer, L.C., et al., *The development of a standardized neighborhood deprivation index.* J Urban Health, 2006. 83(6): p. 1041-62.

220. Bedi, T., A. Coudouel, and K. Simler, *More than a pretty picture - Using poverty maps to design better policies and interventions*, The World Bank, Editor. 2007: Washington D.C.

221. Howe, L.D., J.R. Hargreaves, and S.R. Huttly, *Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries.* Emerg Themes Epidemiol, 2008. 5(1): p. 3.

222. Gwatkin, D.R., S; Johnson, K; Pande, R; Wagstaff, A, *Socio Economic Differences in Health Nutrition & Population in Zambia*. 2000, WB.

223. Vyas, S. and L. Kumaranayake, *Constructing socio-economic status indices: how to use principal components analysis.* Health Policy Plan, 2006. 21(6): p. 459-68.

224. Henry, C.S., M; Lapenu, C; Zeller, M, *Assessing the Relative Poverty of Microfinance Clients*. 2000, International Food Policy Research Institute.

225. McKenzie, D.J., *BREAD Working Paper N. 42: Measuring inequalities with asset indicators*. 2003, Bureau of Research in Economic Analysis of Development.

226. Pai, M., S. Kalantri, and K. Dheda, *New tools and emerging technologies for the diagnosis of tuberculosis: part II. Active tuberculosis and drug resistance.* Expert Rev Mol Diagn, 2006. 6(3): p. 423-32.

227. Houweling, T.A., A.E. Kunst, and J.P. Mackenbach, *Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter?* Int J Equity Health, 2003. 2(1): p. 8.

228. Grosch, M. and J. Baker, *Proxy means tests for social programs*. 1995, World Bank,: Washington.

229. Ward, P., T. Owens, and G. Kahyara, *Developing proxy predictors for household expenditure and income poverty*. 2002: Oxford Policy Press.

230. Bruzzi, P., et al., *Estimating the population attributable risk for multiple risk factors using case-control data.* Am J Epidemiol, 1985. 122(5): p. 904-14.

231. Miettinen, O.S., *Proportion of disease caused or prevented by a given exposure, trait or intervention.* Am J Epidemiol, 1974. 99(5): p. 325-32.

232. von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.* Lancet, 2007. 370(9596): p. 1453-7.

233. Farhat, M., et al., *False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?* Int J Tuberc Lung Dis, 2006. 10(11): p. 1192-204.

234. Villate, J.I., et al., *Analysis of latent tuberculosis and mycobacterium avium infection data using mixture models.* BMC Public Health, 2006. 6: p. 240.

235. Kerr-Pontes, L.R., et al., *Socioeconomic, environmental, and behavioural risk factors for leprosy in North-east Brazil: results of a case-control study.* Int J Epidemiol, 2006. 35(4): p. 994-1000.

236. Kerr-Pontes, L.R., et al., *Inequality and leprosy in Northeast Brazil: an ecological study.* Int J Epidemiol, 2004. 33(2): p. 262-9.

237. Raghunathan, P.L., et al., *Risk factors for Buruli ulcer disease (Mycobacterium ulcerans Infection): results from a case-control study in Ghana.* Clin Infect Dis, 2005. 40(10): p. 1445-53.

238. Pai, M., S. Kalantri, and K. Dheda, *New tools and emerging technologies for the diagnosis of tuberculosis: part I. Latent tuberculosis.* Expert Rev Mol Diagn, 2006. 6(3): p. 413-22.

239. Lalvani, A., *Diagnosing tuberculosis infection in the 21st century: new tools to tackle an old enemy.* Chest, 2007. 131(6): p. 1898-906.

240. Filmer, D. and K. Scott, *Assessing asset indices*, Development Research Group, Editor. 2008, The World Bank,: Washington, D.C.

241. Bingenheimer, J.B., *Wealth, wealth indices and HIV risk in East Africa.* Int Fam Plan Perspect, 2007. 33(2): p. 83-4.

242. McKenzie, D.J., *Measuring inequality with asset indicators.* Journal of population and economics, 2005. 18(2): p. 229-260.

243. Lindelow, M., *Sometimes more equal than others: how health inequalities depend on the choice of welfare indicator.* Health Econ, 2006. 15(3): p. 263-79.

244. Stifel, D. and L. Christiansen, *Tracking poverty over time in the abscence of comparable consumption data.* The World Bank Economic Review, 2007. 21(2): p. 317-41.

245. Bland, J.M. and D.G. Altman, *Measuring agreement in method comparison studies.* Stat Methods Med Res, 1999. 8(2): p. 135-60.

246. Huber, J.B., *Civilization and tuberculosis.* British Journal of Tuberculosis, 1907. **1**: p. 156-159.

247. Ayles, H.M., et al., *ZAMSTAR, The Zambia South Africa TB and HIV Reduction study: Design of a 2 x 2 factorial community randomized trial.* Trials, 2008. **9**: p. 63.

248. van der Werf, M.J., D.A. Enarson, and M.W. Borgdorff, *How to identify tuberculosis cases in a prevalence survey.* Int J Tuberc Lung Dis, 2008. **12**(11): p. 1255-60.

249. Schulz, K.F. and D.A. Grimes, *Sample size calculations in randomised trials: mandatory and mystical.* Lancet, 2005. **365**(9467): p. 1348-53.

250. Bollen, K.G., JL; Stecklov, G, *Socioeconomic status and class in studies of fertility in developing countries*, in *MEASURE Evaluation*. 1999, USAID: Washinghton.

251. Bills, D.B., D.S. Godfrey, and A.O. Haller, *Measure the socioeconomic status of occupations in Brazil.* Rural Sociology, 1985. **50**(225-50).

252. Bartholomew DJ, S.F., Moustaki I, Galbraith JI, *Principal Component Analysis*, in *The analysis and interpretation of multivariate data for social scientists*. 2002, Chapman & Hall/CRC. p. 115-142.

253. Ferguson, B., Tandon A., Gakidou E., Murray CJL. , *Estimating permanent income using indicator variables. Evidence and information for policy cluster*. 2002, World Health Organization, Geneva, WHO: Geneva.

254. Morris, S.S., et al., *Validity of rapid estimates of household wealth and income for health surveys in rural Africa.* J Epidemiol Community Health, 2000. **54**(5): p. 381-7.

255. Howe, L.D., et al., *Is the wealth index a proxy for consumption expenditure? A systematic review.* J Epidemiol Community Health, 2009.

256. Raby, E., et al., *The effects of HIV on the sensitivity of a whole blood IFN-gamma release assay in Zambian adults with active tuberculosis.* PLoS One, 2008. **3**(6): p. e2489.

257. van Lettow, M., et al., *Low plasma selenium concentrations, high plasma human immunodeficiency virus load and high interleukin-6 concentrations are risk factors associated with anemia in adults presenting with pulmonary tuberculosis in Zomba district, Malawi.* Eur J Clin Nutr, 2005. **59**(4): p. 526-32.

258. Macallan, D.C., *Malnutrition in tuberculosis.* Diagn Microbiol Infect Dis, 1999. **34**(2): p. 153-7.

259. Kenyon, T.A., et al., *Risk factors for transmission of Mycobacterium tuberculosis from HIV-infected tuberculosis patients, Botswana.* Int J Tuberc Lung Dis, 2002. **6**(10): p. 843-50.

260. Banerjee, A., A.D. Harries, and F.M. Salaniponi, *Differences in tuberculosis incidence rates in township and in rural populations in Ntcheu District, Malawi.* Trans R Soc Trop Med Hyg, 1999. **93**(4): p. 392-3.

261. Enarson, D.A., J.S. Wang, and J.M. Dirks, *The incidence of active tuberculosis in a large urban area.* Am J Epidemiol, 1989. **129**(6): p. 1268-76.

262. van Hooijdonk, C., et al., *Exceptions to the rule: healthy deprived areas and unhealthy wealthy areas.* Soc Sci Med, 2007. **64**(6): p. 1326-42.

263. Mweemba, A.J. and E. Webb, *Residential area as proxy for socio-economic status, paediatric mortality and birth weight in Lusaka, Zambia.* J Trop Pediatr, 2008. **54**(6): p. 406-9.

264. Krieger, N., et al., *Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US).* J Epidemiol Community Health, 2003. **57**(3): p. 186-99.

265. Krieger, N., et al., *Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures--the public health disparities geocoding project (US).* Public Health Rep, 2003. **118**(3): p. 240-60.

266. Bollen, K.A., J.L. Glanville, and G. Stecklov, *Economic status proxies in studies of fertility in developing countries: does the measure matter?* Population Studies - A Journal of Demography, 2002. **56**: p. 81-96.

267. Corbett, E.L., et al., *Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control.* International Journal of Tuberculosis and Lung Disease, 2009. **13**(10): p. 1231-1237.

268. Dai, G. and D.N. McMurray, *Altered cytokine production and impaired antimycobacterial immunity in protein-malnourished guinea pigs.* Infect Immun, 1998. **66**(8): p. 3562-8.

269. Cegielski, J.P., L. Kohlmier, and J. Cornoni-Huntley. *Malnutrition and Tuberculosis in a nationally representative cohort of adults in the United States, 1971-1987.* in *American Society of Tropical Medicine and Hygiene, 44th Annual Meeting.* 1995. Texas, 17-21 November 1995.

270. Corbett, E.L., et al., *Epidemiology of Tuberculosis in a High HIV Prevalence Population Provided with Enhanced Diagnosis of Symptomatic Disease.* PLoS Med, 2007. **4**(1): p. e22.

271. Wood, R., et al., *Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control.* Am J Respir Crit Care Med, 2007. **175**(1): p. 87-93.

272. Fylkesnes, K., et al., *The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women.* AIDS, 1997. **11**(3): p. 339-45.

273. Fylkesnes, K., et al., *Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys.* AIDS, 2001. **15**(7): p. 907-16.

274. Crampin, A.C., J.R. Glynn, and P.E. Fine, *What has Karonga taught us? Tuberculosis studied over three decades.* Int J Tuberc Lung Dis, 2009. **13**(2): p. 153-64.

275. Fotso, J.C., et al., *Progress towards the child mortality millennium development goal in urban sub-Saharan Africa: the dynamics of population growth, immunization, and access to clean water.* BMC Public Health, 2007. **7**: p. 218.

276. Simatele, S.M., *Kwacha ngwee: a snapshot at Zambia's contemporary migration patterns*, in *African Migration Workshop, 18-21 September 2007*. 2007: Accra, Ghana.

277. Bhugra, D., *Migration and mental health.* Acta Psychiatr Scand, 2004. **109**(4): p. 243-58.

278. McElroy, P.D., et al., *A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations.* Int J Tuberc Lung Dis, 2003. **7**(12 Suppl 3): p. S486-93.

279. Arraes de Alencar Ximenes, R., et al., *Is it better to be rich in a poor area or poor in a rich area? A multilevel analysis of a case–control study of social determinants of tuberculosis.* International Journal of Epidemiology, 2009. **38**: p. 1285-96.

280. Greenland, S. and J.M. Robins, *Conceptual problems in the definition and interpretation of attributable fractions.* Am J Epidemiol, 1988. **128**(6): p. 1185-97.

281. Kass, E.H., *Infectious diseases and social change.* J Infect Dis, 1971. **123**(1): p. 110-4.

282. Springett, V.H., *An interpretation of statistical trends in tuberculosis.* Lancet, 1952. **1**(6708): p. 575-80.

283. Drucker, E., et al., *Childhood tuberculosis in the Bronx, New York.* Lancet, 1994. **343**(8911): p. 1482-5.

284. Michelo, C., et al., *Steep HIV prevalence declines among young people in selected Zambian communities: population-based observations (1995-2003).* BMC Public Health, 2006. **6**: p. 279.

285. Barker, R.D., et al., *Traditional healers, treatment delay, performance status and death from TB in rural South Africa.* Int J Tuberc Lung Dis, 2006. **10**(6): p. 670-5.

286. Elender, F., G. Bentham, and I. Langford, *Tuberculosis mortality in England and Wales during 1982-1992: its association with poverty, ethnicity and AIDS.* Soc Sci Med, 1998. **46**(6): p. 673-81.

287. Beard, J.R., et al., *Neighborhood characteristics and change in depressive symptoms among older residents of New York City.* Am J Public Health, 2009. **99**(7): p. 1308-14.

288. Dye, C. and B. Williams, *Eliminating human tuberculosis in the twenty-first century.* Journal of the Royal Society Interface, 2007. **10.1098**: p. 1-10.

289. Konteh, F.H., *Urban sanitation and health in the developing world: reminiscing the nineteenth century industrial nations.* Health Place, 2009. **15**(1): p. 69-78.

290. Thomson, H., et al., *Do urban regeneration programmes improve public health and reduce health inequalities? A synthesis of the evidence from UK policy and practice (1980-2004).* J Epidemiol Community Health, 2006. **60**(2): p. 108-15.

291. World Food Programme, *Getting started: WFP food assistance in the context of tuberculosis care and treatment.* 2007, World Food Programme,: Rome.

292. Pakasi, T.A., et al., *Malnutrition and socio-demographic factors associated with pulmonary tuberculosis in Timor and Rote Islands, Indonesia.* Int J Tuberc Lung Dis, 2009. **13**(6): p. 755-9.

293. Kim, J., et al., *Exploring the role of economic empowerment in HIV prevention.* AIDS, 2008. **22 Suppl 4**: p. S57-71.

294. Khandker, S.R., *Fighting poverty with microcredit: experience in Bangladesh.* 1998, New York: Oxforf University Press.

295. Mushtaque, A., R. Chowdury, and A. Bhuiya, *Do poverty alleviation programmes reduce inequities in health? The Bangladesh experience*, in *Poverty inequality and health*, D.A. Leon and G. Walt, Editors. 2001, Oxford University Press: New York.

296. Schuler, S.R. and S.M. Hashemi, *Credit programmes, women's empowerment and contraceptive use in rural Bangladesh.* Studies on Family Planning, 1994. **25**: p. 65-76.

297. Kim, J.C., et al., *Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa.* Am J Public Health, 2007. **97**(10): p. 1794-802.

298. Pronyk, P.M., et al., *A combined microfinance and training intervention can reduce HIV risk behaviour in young female participants.* AIDS, 2008. **22**(13): p. 1659-65.

299. Rawlings, L.B., *Evaluating the impact of conditional cash transfer programs: lessons from Latin America.* World Bank Res Obs, 2005. **20**(1): p. 29-55.

300. Fernald, L.C., P.J. Gertler, and L.M. Neufeld, *Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades.* Lancet, 2008. **371**(9615): p. 828-37.

301. Paxson, C. and N. Schady, *Does money matter? Role of cash in conditional cash transfer and development in rural Ecuador.* 2007, The World Bank: Washington DC. p. 1-43.

302. Thim, S., et al., *A community-based tuberculosis program in Cambodia.* JAMA, 2004. **292**(5): p. 566-8.

303. Verver, S., et al., *Transmission of tuberculosis in a high incidence urban community in South Africa.* Int J Epidemiol, 2004. **33**(2): p. 351-7.

304. Wagstaff, A., E. Doorslaer van, and N. Watanabe, *On Decomposing the Causes of Health Sector Inequalities with an Applicatin to Malnutrution Inequalities in Vietnam 1993-98*, in *Research Policy Working Paper 2714.* 2001, The World Bank: Washington DC.

305. Harpham, T., S. Burton, and I. Blue, *Healthy city projects in developing countries: the first evaluation.* Health Promot Int, 2001. **16**(2): p. 111-25.

306. Marks, S., *Commentary: Socio-economic determinants of tuberculosis in Recife, Brazil.* Int J Epidemiol, 2009. **38**(5): p. 1295-6.

307. Porter, J., J.A. Ogden, and P.M. Pronyk, *The wat forward: an integrated approach to tuberculosis control*, in *Tuberculosis: and interdisciplinary perspective.*, J. Grange and J. Porter, Editors. 1999, Imperial College Press: London, UK. p. 359-378.

308. Pronyk, P., *Development finance, social capital and HIV/AIDS. The intervention with microfinance for AIDS & Gender Equity. The IMAGE study*, in *Department of Infectious and Tropical Diseases.* 2006, London School of Hygiene and Tropical Medicine: London.

309. Commission on Social Determinants of Health. *Commission on social determinants of health - Final report.* 2008 [cited February 16, 2010; Available from: http://www.who.int/social_determinants/thecommission/finalreport/en/index.html.

310. Rothman, K.J., H.O. Adami, and D. Trichopoulos, *Should the mission of epidemiology include the eradication of poverty?* Lancet, 1998. **352**(9130): p. 810-3.

311. Kilbourne, A.M., et al., *Advancing health disparities research within the health care system: a conceptual framework.* Am J Public Health, 2006. 96(12): p. 2113-21.

312. Sheuya, S.A., *Improving the health and lives of people living in slums.* Ann N Y Acad Sci, 2008. 1136: p. 298-306.

313. Thomas, S.B. and S.C. Quinn, *Poverty and elimination of urban health disparities: challenge and opportunity.* Ann N Y Acad Sci, 2008. 1136: p. 111-25.

314. Rosenfield, P.L., *The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences.* Soc Sci Med, 1992. 35(11): p. 1343-57.

315. Blankenship, K.M., et al., *Structural interventions: concepts, challenges and opportunities for research.* J Urban Health, 2006. 83(1): p. 59-72.

316. Syme, S.L. and L.F. Berkman, *Social class, susceptibility and sickness.* Am J Epidemiol, 1976. 104(1): p. 1-8.

317. World Health Assembly. *Resolution on Reducing Health Inequalities through action on the Social Determinants of Health.* 2009 September 2009]; Available from: http://apps.who.int/gb/ebwha/pdf_files/EB124/B124_R6-en.pdf.

318. Davies, P.D., *Commentary: Tuberculosis down the generations--a comment on 'Continued studies of Tuberculosis as a generation illness' by Kr F Andvord.* Int J Epidemiol, 2008. 37(5): p. 934-6.

Appendices

APPENDIX A

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

ETHICS COMMITTEE

APPROVAL FORM
Application number: 3008

Name of Principal Investigator	**Peter Godfrey-Faussett**
Department	**Infectious and Tropical Diseases**
Head of Department	**Professor Hazel Dockrell**

Title: Zamstar: Zambia and South Africa Tuberculosis and AIDS reduction study

Renewal of ethical clearance for 2006-2007

Approval of this study is granted by the Committee.

Chair T.W. Meade .....
Professor Tom Meade

Date 16 May 2006 .....

Approval is dependent on local ethical approval having been received.

Any subsequent changes to the consent form must be re-submitted to the Committee.

APPENDIX B

THE UNIVERSITY OF ZAMBIA

RESEARCH ETHICS COMMITTEE

Telephone: 260-1-256067
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: +260-1-250753
E-mail: unzarec@zamtel.zm

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia

Assurance No. FWA00000338
IRB00001131 of IORG0000774

21 June, 2006
Ref.: 007-10-04

Dr Helen Ayles
Zambart Project
Ridgeway Campus
University of Zambia
P.O. Box 50697
LUSAKA

Dear Dr Ayles,

RE: **ZAMSTAR STUDY**

Thank you for your letter of 30 May, 2006 and all the enclosures. I am pleased to note that all our concerns have been addressed satisfactorily.

Approval is hereby given for Phase II of the study to proceed. This approval expires on 31 May, 2007.

With best wishes.

Yours sincerely,

Prof. J. T. Karashani, MB, ChB, PhD
CHAIRMAN

APPENDIX C

The relationship between tuberculosis and poverty: a case control study in Zambia 1 o f12

Questionnaire for CASES

INDIVIDUAL BARCODE

Section A – General Information

Q01_IntCode **Interviewer's Code**

Q02_IntDate **Date of Interview**

D	D	M	M	Y	Y	Y	Y

Q03_Consent **Consent**

No	Yes	Absent	Excluded
0	1	2	3

Q04_Sex **Sex**

M	F
1	2

Q05_DOB **Date of Birth** (01/01/1800 if unknown)

D	D	M	M	Y	Y	Y	Y

Q05_1_Age **If not known, what was your age in years on your last birthday?** (999 if unknown)

Q06_Ethnic **Which ethnic group do you belong to?**

Black Zambian	Black other	White	Asian	Other	Unknown
0	1	2	3	4	9

Q07_Marital **What is your marital status?**

Married	0	Living as married	1
Single	2	Divorced	3
Widowed	4	Unknown	9

ONLY IF AGED BETWEEN 15-20, IF OLDER THAN 20 GO TO NEXT SECTION

		No	Yes	Unknown
Q08_Mother	**Is your biological mother still alive?**	0	1	9
Q09_Father	**Is your biological father still alive?**	0	1	9

Section B – Educational status of the case

SAY: I would now like to ask you some questions about your educational status

		No	Yes	Unknown
Q10_Write	Can you write and read?	0	1	9
Q11_Schnow	Are you currently attending school?	0	1	9
Q12_Schever	Have you ever attended school?	0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q14

Q12_1_Grade **What is the highest grade you have completed?**

None	1-4	5-7	8-9	10-12	College	Unknown
0	1	2	3	4	5	9

Q14_Neversch **Why have you never attended school? (Tick the most important option only)**

Couldn't get a place	0	Too expensive/couldn't find support	1
Too far away	2	Illness or injury	3
Unsafe to travel to school	4	School not important	5
Quality of school bad	6	Need to help family	7
Other, please specify	8	Unknown	9

Section C – Occupational status

SAY: I would now like to ask you some questions about the most important activity you were involved into at the time we first visited you in terms of earning money or goods for yourself or for the household

Q15_Activity **What was your main activity?**

Unemployed	0	Seasonal/ piece worker	1
Student	2	Employed	3
Retired	4	Self employed	5
Housewife	6	Unknown	9

IF UNEMPLOYED GO TO Q19

Q16_Job	**If employed/self-employed or a seasonal/piece worker, what type of job did you do?** *Ngati munali kusebenza, kodi munali kusebenzela muntu, kapena kuzisebenzela, olo kuchita maganyu?*	**Agriculture/fisheries**	**0**	**Transport**	**1**
		Construction/electricity/ gas/water	**2**	**Manufacturing**	**3**

INDIVIDUAL BARCODE

Nga mwalebomba/ukuibombela nagula bucibombebombe, ni ncito nshi mwalebomba?	**Community social services (health, education)**	**4**	**Household employee/ housekeeper**	**5**
	Mining/Quarrying	**6**	**Trader**	**7**
	Clerical and related	**8**	**Professional (finance, law, academic)**	**9**
	General worker	**10**	**Other**	**11**
			Unknown	**99**

		No	Yes	Unknown
Q16_1_Samejob	**Was this the main job you were you doing the 12 months before we first visited you?**	**0**	**1**	**9**

IF YES SKIP TO THE NEXT SECTION, IF NO GO TO Q18

Q18_Jobthen	**What was your job then?** *Kodi munali kusebenza inchito ya bwanji?* Bushe ni incito nshi mwalebomba?	**Agriculture/fisheries**	**0**	**Transport**	**1**
		Construction/electricity/ gas/water	**2**	**Manufacturing**	**3**
		Community social services (health, education)	**4**	**Household employee/ housekeeper**	**5**
		Mining/Quarrying	**6**	**Trader**	**7**
		Clerical and related	**8**	**Professional (finance, law, academic)**	**9**
		General worker	**10**	**Other**	**11**
				Unknown	**99**

19_Nojob	**What was the main reason you were not working at that time?**	**Low wage/salary/no profit**	**0**	**Fired**	**1**

Nicifukwa chiani chikulu cimene simunali kusebenzela?	**Enterprise closed/liquidated/ privatised**	**2**	**Sick because of TB**	**3**
Bushe cinshi icikalamba icalengele ukti mwilabomba pali ilya nshita?	**Sick because of disease other than TB**	**4**	**Made redundant**	**5**
	Invalid	**6**	**Student/Retired/ Housewife**	**7**
	Other	**8**	**Unknown**	**9**

Section D – Dietary habits and food consumption

SAY: I would now like to ask you some questions about your dietary habits and food consumption at the time we first visited you

Q20_Meals **How many meals did you normally use to eat per day?**

Munali kudya kangati pa siku?
Bushe miku inga mwalelya pa bushiku bumo?

0	1	2	3	>3	Unknown
0	1	2	3	4	9

Q21_Luxfood **How many times in a week did you normally use eat meat, fish, chicken or pork?**

Kodi ni kangati musabata limozi kamene munali kudya za kudya zopala nyama, nsomba, nkhuku kapena nyama ya nkhumba?

Bushe miku inga mu mulungu ilyo mwalelyapo inama, isabi, inkoko nangula inkumba?

0	1	2	3	4	5	>5	Unknown
0	1	2	3	4	5	6	9

Section E – Health status

SAY: I would now like to ask you some questions about your health, specifically about TB and HIV

Q22_TBTreat **Are you currently on TB Treatment?**

Kodi mukumwa mankhwala a chifuba cha TB pali ino nthawi?
Bushe pali nomba mulenwa umuti wa bulwele bwa TB?
IF YES CONTINUE, IF NO GO TO Q23

No	Yes	Unknown
0	1	9

Q22_1_Start **When did you start your treatment?**

Nicifukwa cani inu simukumwa mankwala ya TB?
Bushe cinshi mushinwena umuti wa TB?

M	M	Y	Y

Q23_Notreat **Why you are NOT on TB treatment?**

Nicifukwa cani inu simukumwa mankwala ya TB?

The relationship between tuberculosis and poverty: a case control study in Zambia 5 o f12

Bushe cinshi mushinwena umuti wa TB?

Treatment completed	**0**	**No drugs available**	**1**
Treatment unsuccessful	**2**	**Treatment not needed**	**3**
Treatment too expensive	**4**	**Clinic too far**	**5**
Feeling better	**6**	**Other, specify**	**7**
Unknown	**9**		

INDIVIDUAL BARCODE

Q24_Stigma **Since you fell sick have you ever experienced any of the following:**

Kuchokela pamene munadwalila TB, kodi zinakuchitikilaniponi izi?
Apo mwalwaalila TB, mwalitala amucitikilwapo nefili fyonse pali ifi fyakonkapo

	No	**Yes**
Been excluded from a social gathering *Ku patulidwa ku misonkhano* Uwo batamfyapo apo abantu bakolongene	**0**	**1**
Abandoned by spouse/partner *Kapena kusiiwa ku okondedwa awo.* Uwo babutuka ku mwina mwakwe/umutemwikwa	**0**	**1**
Isolated by your household *Ku musungila payekha mukati mwa banja* Uwo baapatulula ku ba pang'anda	**0**	**1**
Lost housing/been unable to rent a property *Kuchotsedwa munyumba kapena kukanizidwa nyumba yosonkhela* Uwo batamfya pa ng'anda nangu uwo bakanina ukupela ing'anda ya kusonkela	**0**	**1**
Lost respect or standing in the community *Anthu analeka kumu lemekeza mukomboni.* Uwalushilapo no mucinshi ukufuma ku Bantu mu ncende ekala	**0**	**1**
Been teased, insulted or sworn at *Anthu kumuseka,kumunyoza kapena kumu tukwana* Uusekwa, uutukwa nangula uwo balapisha	**0**	**1**
Been gossiped about *Ku munena kumbali* Uwo baamba	**0**	**1**

Your children or family have been isolated/shunned

0	1

Kapena ana anu ndi banja lanu si bvomelezedwa kucheza ndi anthu ena
Abana nangula balupwa benu balibapatulula/balabataluka

Been treated worse than other patients by health staff

0	1

Kapena simu mathandizidwa bwino ndi aja ogwila nchito muchipatala kulingana ndi aja odwala matenda ena
Tabamitangata ukufikapoku babomfi ba muchipatala ukulingana na balwele ba malwele yambi

Q25_BCG **Have you ever been vaccinated for TB**

No	Yes	Unknown
0	1	9

Q26_Scar **Can I see if you have any BCG scar on your arm?**

Examination not allowed	0
No visible scar on both arms	1
BCG confirmed by scar examination	2

Do you suffer from any of these clinical conditions?
Kodi munadwalapo matenda monga?
Bushe mwalilwalapo amalwele ayaba nga?

	No	Yes	Unknown
Q27_Sugar **Sugar**	0	1	9
Cancer	0	1	9

Q28_HospWalk **How long in minutes does it take to walk to the nearest health clinic/hospital from here?**
Kodi cimakutengelani mpindi zingati kuchokela kunyumbla kwanu kukafika kuchipatala chapa fupi?
Bushe cimusendela insa shinga pakufika ku cipatala icili mupepi?

0-15	16-30	31-45	45-60	> 60
0	1	2	3	4

Q29_HIVTest **Have you ever been tested for HIV?**
Kodi muna pimisa pa za kadoyo ka HIV?
Bushe mwalipimwapo palwa kashishi ka HIV?

No	Yes	Unknown
0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

Q29_1_HIVRes **Would you mind telling me what the result was?**
Kodi mungakondwele kuniuza zimene anapezamo?
Bushe kuti mwanjebako fyatumbwikemo?

Neg	Pos	Unknown
0	1	9

IF POSITIVE CONTINUE. IF NEGATIVE OR UNKNOWN GO TO NEXT SECTION

The relationship between tuberculosis and poverty: a case control study in Zambia 7 o f12

		No	Yes	Unknown
Q30_ARVever	**Have you ever taken ARV treatment?** *Kodi munamwapo mankhwala ya ARVs?* Bushe mwalitala amunwapo ama ARVs?	0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

		No	Yes	Unknown
Q31_ARVnow	**Are you currently taking ARV treatment?** *Kodi mukumwa mankhwala ya ARVs pali pano?* Bushe mulenwa ama ARVs pali ino nshita?	0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

		D	D	M	M	Y	Y	Y	Y
Q31_1_ARVdate	**When did you start ARV treatment** *Kodi munayamba liti kumwa mankhwala ya ARVs?* Nilisa mwayambile ukunwa ama ARVs?								

INDIVIDUAL BARCODE

Q31_2_ARVwhere	**Which health centre do you receive treatment from?** *Kodi mumalandilila ku chipatala chiti mankhwala aya?* Chipatala ninshi musendelako uyu umuti?	

Section F – Hazardous behaviours of the case

SAY: Now I would now like to ask you some questions about your behaviours and habits

		Never/ Rarely	1-4 times/month	Many times/week	Every day	Unknown
Q32_Alcohol	**How often do you drink any type of alcohol?** *Kodi kamwedwe ka kanu ka moba kali motani?* Bushe iminwene ya bwalwa yenu yaba shani?	0	1	2	3	9

Q33_Drinks **When you drink, how many drinks containing alcohol might you have on a typical day?**

Kodi mumamwa mabotolo angati pa siku lanu lo kumwa moba?
Bushe ngamulenwa munwa amabotola yanga pa bushiku bumo?

1-3	4-6	7-9	>= 10	Unknown
0	1	2	3	9

No	Yes	Unknown

Q34_Drunk1

During the last year have you ever drunk so much that you were unable to remember what you were doing?
Mu chaka chatha ichi kodi munalezelapo kufikila kuti simuna kumbukile zimene zinachitika kamba kaka mwedwe kanu?
Muli uyu mwaka wapwile mwaliikolelwepo icakuti mwalaba nefyo mwalecita?

0	1	9

Q35_Drunk2

During the last year have you been unable to remember what happened the night before because of your drinking?
Mu chaka chatha ichi simunakumbukilepo zamene zinacitika usiku wamene munalezela kwambili cifukwa ca kamwedwe kanu?
Bushe muli uyu mwaka wapwile mwalikolwapo icakuti mwalaba ne fyacitike ubushiku bwafumineko?

No	Yes	Unknown
0	1	9

Q36_Drunk3

During the last year have you been criticised by a relative, friend or a doctor because of the amount of alcohol you drink?
Kodi caka chatha anakuzuzulaniponi abanja kapena anzanu olo adotolo cifukwa ca kamwedwe kanu?
Muli uyu mwaka wapwile mwallikalilwapo na balupwa, banenu nangula ba shing'anga ba kucipatala pa mulandu wa minwene yenu?

No	Yes	Unknown
0	1	9

Q37_Drunk4

During the last year, have you or anybody else been injured because of your drinking?
Kodi chaka chatha, imwe kapena munthu wina anacitiwapo ngozi cifukwa cha kamwedwe kamoba kanu?
Muli uyu mwaka wapwile, bushe mwaliicenapo nangu umuntu umbi ukucenekwa pa mulandu wa kunwa kwenu?

No	Yes	Unknown
0	1	9

Q38_Drunk5

During the last year, have you been arrested, been held at a police station, or had any legal problems because of your drinking?
Kodi chaka chatha, imwe munamangiwapo ndi a kapokola kapena kuimbiwa mulandu cifukwa ca kamwedwe kamoba kanu?
Muli uyu mwaka wapwile, balimwiketepo kuli bakapokola nangula ukulubululapo imilandu pa mulandu ne minwene yenu?

No	Yes	Unknown
0	1	9

Q38_Drunk5

During the last year, have you been arrested, been held at a police station, or had any legal problems because of your drinking?
Kodi chaka chatha, imwe munamangiwapo ndi a kapokola kapena kuimbiwa mulandu cifukwa ca kamwedwe kamoba kanu?
Muli uyu mwaka wapwile, balimwiketepo kuli bakapokola nangula ukulubululapo imilandu pa mulandu ne minwene yenu?

No	Yes	Unknown
0	1	9

Q39_Smoke

How would you classify your smoking habits?
Kodi munga ziike mugulu la bwanji pakakokedwe kafwaka?
Kuti mwabikapisa imipeepele ya fwaka yenu?

Never smoked *Mukalibe kukokapo* Tamwapeepapo	0	**Daily smoker** *Masiku onse* Mupeepa cilabushiku	1
Occasional smoker *Kukoka panthawi* Mupeepa pa kashita	2	**Ex-smoker** *Munaleka* Mwalileka ukupeepa	3

IF SMOKER OR EX-SMOKER CONTINUE, IF NEVER SMOKED GO TO Q40

Q39_1_Longsmoke

How long have you/did you smoke for?(years)
(if less than 1 year: put 00 for less than 6 months and 01 for 6months-1 year. 999 if Unknown
Mwakoka fwaka pa zaka zingati /muna koka fwaqka zaka zingati?
Myaka inga iyapitapo apo mwayambila ukupeepa fwaka/myaka inga mwapeepelepo fwaka?

Q39_2_Cigarettes

How many cigarettes do/did you smoke per day? (999 if Unkwown)
Mukoka mishanga ingati pa siku / munali kukoka mishanga ingati pa siku?
Mupeepa/mwalepeepa imishanga inga pa bushiku bumo?

INDIVIDUAL BARCODE

		No	Yes	Unknown
Q40_Dagga	**Do you ever smoke marijuana/dagga?** *Kodi mumakoka chamba kapena zina zotele?* Bushe mulapeepa ibange/icamba?	0	1	9

		No	Yes	Unknown
Q41_Drugs	**Do you take any other recreational drugs?** *Mumakoka fwaka kapena kusebenzesa zilizonse zamene zimapangisa munthu kukhala osangalala kapena kuzunguza bongo?* Bushe mwalibomfyapo ifimiti (ama drugs) fimbi ukucila pa camba ifyakumusansamusha?	0	1	9

		Never/ rarely	Yes	Unknown
Q42_Cooking	**Are you exposed to cooking fire indoor?** *Kodi munankhalapo pafupi na kumoto munyumba?* Bushe mwalisangwapo ku mulilo wa kwipikila mukati kang'anda?	0	1	9

Q43_OthSmoke **Are you exposed to any other type of smoke on a regular basis?**
Kodi mumankhala pafupi ndi chusi kambili kambili?
Bushe mulekala mupepi ne chushi inshita iitali?

No	Yes	Unknown
0	1	9

Q44_TBcontact **Had you been in contact with anyone with TB before we first visited you?**
Kodi munankhalapo ndi munthu odwala chifuba ca TB pamene tikalibe kukutandalilani?
Bushe mwalisangwapo no mulwele wa TB lintu tatulamutandalila?

No	Yes	Unknown
0	1	9

Q45_TBprof **Have you ever been professionally exposed to TB cases as a health care worker?**
Kapena muna gwilapo nchito yosamala odwala TB?
Limbi mwalibombapo incito yakundapa ubulwele bwa TB?

No	Yes	Unknown
0	1	9

Q46_Prison **Have you ever been in prison?**
Kodi munankalapo mundende?
Limbi mwalikalapo muchifungo?

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q47

Q46_1_Prisonwhen **When?**

< 1 year ago	1-5 years ago	> 5 years ago	Unknown
0	1	2	9

Q46_2_Prisonlong **For how long?**

< 1 year	1-5 years	> 5 years	Unknown
0	1	2	9

Q47_Video **Do you ever go to VIDEO CLUBS?**
IF YES CONTINUE, IF NO SKIP TO => Q48

No	Yes	Unknown
0	1	9

Q47_1_OftenVideo **How often?**

< 1 month	1/month	> 1/month	Unknown
0	1	2	9

≤ 1 hour	1-2 hours	> 2 hours	Unknown

Q47_2_LongVideo	How long do you typically remain there?	0	1	2	9

		No	Yes	Unknown
Q48_Bars	Do you ever go to BARS? IF YES CONTINUE, if NO SKIP TO => Q49	0	1	9

		≤ 1 month	1/month	> 1/month	Unknown
Q48_1OftenBars	How often?	0	1	2	9

		≤ 1 hour	1-2 hours	> 2 hours	Unknown
Q48_2_LongBars	How long do you typically remain there?	0	1	2	9

		No	Yes	Unknown
Q49_Hair	Do you ever go to HAIR SALOONS/BARBER SHOP? IF YES CONTINUE, IF NO SKIP TO => Q50	0	1	9

		≤ 1 week	1/week	> 1/week	Unknown
Q49_1_OftenHair	How often?	0	1	2	9

		≤ 1 hour	1-2 hours	> 2 hours	Unknown
Q49_2_LongHair	How long do you typically remain there?	0	1	2	9

INDIVIDUAL BARCODE

		No	Yes	Unknown
Q50_Church	Do you ever go to CHURCH? IF YES CONTINUE, IF NO SKIP TO => Q51	0	1	9

		≤ 1 week	1/week	> 1/week	Unknown
Q50_1_OftenChurch	How often?	0	1	2	9

		≤ 1 hour	1-2 hours	> 2 hours	Unknown
Q50_2_LongChurch	How long do you typically remain there?	0	1	2	9

No	Yes	Unknown

Q51_Migr	**In the 5 years before we first visited you, have you lived anywhere other than where you live today for a period of at least 6 months?** *Pa zaka zisano zapita pamene ife tikalibe kukutandalilani munakhalapo kwina kwach mu miyezi isami ndi umodzi posacedwa apa?* Bushe pamyaka yapita isano, mwalikalapo kunchende imbi panshita iilingene ne myeshi mutanda?	**0**	**1**	**9**

IF YES CONTINUE, IF NO SKIP TO NEXT SECTION

		Somewhere else in Zambia	Outside Zambia	Outside Africa	Unknown
Q51_1_Wheremigr	**Where?**	**0**	**1**	**2**	**9**

THANK THE RESPONDENT FOR PARTICIPATION AND ASK HIM/HER WHAT HE/SHE THINKS POVERTY IS – FILL IN THE BOX WITH THE DEFINITION PROVIDED

		Date								
	Interviewer's Code	**d**	**d**	**m**	**M**	**y**	**y**	**y**	**Y**	**Signature**
Interviewer										
Field Manager										
1st data entry										
2nd data entry										

Questionnaire for CONTROLS

INDIVIDUAL BARCODE

Section A – General Information

Q01_Intcode **Interviewer's Code**

Q02_Intdate **Date of Interview** D D M M Y Y Y Y

Q03_Consent **Consent**

No	Yes	Absent	Excluded
0	1	2	3

Q04_Sex **Sex**

M	F
1	2

Q05_DOB **Date of Birth** (01/01/1800 if unknown) D D M M Y Y Y Y

Q05_1_Age **If not known, what was your age in years on your last birthday?** (999 if.unknown)

Q06_Ethnic **Which ethnic group do you belong to?**

Black Zambian	Black other	White	Asian	Other	Unknown
0	1	2	3	4	9

Q07_Marital **What is your marital status?**

Married	0	Living as married	1
Single	2	Divorced	3
Widowed	4	Unknown	9

ONLY IF AGES BETWEEN 15-20, IF OLDER THAN 20 GO TO NEXT SECTION

		No	Yes	Unknown
Q08_Mother	**Is your biological mother still alive?**	0	1	9
Q09_Father	**Is your biological father still alive?**	0	1	9

Section B – Educational status of the control

SAY: I would now like to ask you some questions about your educational status

		No	Yes	Unknown
Q10_Write	**Can you write and read?**	0	1	9
Q11_Schnow	**Are you currently attending school?**	0	1	9
Q12_Schever	**Have you ever attended school?**	0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q14

Q12_1_Grade **What is the highest grade you have completed?**

None	1-4	5-7	8-9	10-12	College	Unknown
0	1	2	3	4	5	9

Q14_Neversch **Why have you never attended school? (Tick the most important option only)**

Couldn't get a place	0	Too expensive/couldn't find support	1
Too far away	2	Illness or injury	3
Unsafe to travel to school	4	School not important	5
Quality of school bad	6	Need to help family	7
Other, please specify	8	Unknown	9

Section C – Occupational status

SAY: I would now like to ask you some questions about the most important activity you were involved into at the time we first visited you in terms of earning money or goods for yourself or for the household

Q15_Activity **What was your main activity?**

Unemployed	0	Seasonal/ piece worker	1
Student	2	Employed	3
Retired	4	Self employed	5
Housewife	6	Unknown	9

IF UNEMPLOYED GO TO Q19

Q16_Job	**If employed/self-employed or a seasonal/piece worker, what type of job did you do?** *Ngati munali kusebenza, kodi munali kusebenzela muntu, kapena kuzisebenzela, olo kuchita maganyu?*	**Agriculture/fisheries**	**0**	**Transport**	**1**
		Construction/electricity/ gas/water	**2**	**Manufacturing**	**3**

INDIVIDUAL BARCODE

Ngamwalebomba/ukuibombela nagula bucibombebombe, ni ncito nshi mwalebomba?	**Community social services (health, education)**	**4**	**Household employee/ housekeeper**	**5**
	Mining/Quarrying	**6**	**Trader**	**7**
	Clerical and related	**8**	**Professional (finance, law, academic)**	**9**
	General worker	**10**	**Other**	**11**
			Unknown	**99**

		No	**Yes**	**Unknown**
Q16_1_SameJob	**Was this the main job you were you doing the 12 months before we first visited you?**	**0**	**1**	**9**

IF YES SKIP TO THE NEXT SECTION, IF NO CONTINUE

Q18_Jobthen	**What was your job then?** *Kodi Munali kusebenza inchito ya bwanji?* Bushe ni incito nshi mwalebomba?	**Agriculture/fisheries**	**0**	**Transport**	**1**
		Construction/electricity/ gas/water	**2**	**Manufacturing**	**3**
		Community social services (health, education)	**4**	**Household employee/ housekeeper**	**5**
		Mining/Quarrying	**6**	**Trader**	**7**
		Clerical and related	**8**	**Professional (finance, law, academic)**	**9**
		General worker	**10**	**Other**	**11**
				Unknown	**99**

Q19_Nojob	**What was the main reason you were not working at that time?**	**Low wage/salary/no profit**	**0**	**Fired**	**1**

Nicifukwa chiani chikulu cimene simunali kusebenzela?	**Enterprise closed/liquidated/ privatised**	2	**Sick because of TB**	3
Bushe cinshi icikalamba icalengele ukti mwilabomba pali ilya nshita?	**Sick because of disease *other than TB***	4	**Made redundant**	5
	Invalid	6	**Student/Retired/ Housewife**	7
	Other	8	**Unknown**	9

Section D – Dietary habits and food consumption

SAY: I would now like to ask you some questions about your dietary habits and food consumption at the time we first visited you

Q20_Meals **How many meals did you normally use to eat per day?**
Munali kudya kangati pa siku?
Bushe miku inga mwalelya pa bushiku bumo?

0	1	2	3	>3	Unknown
0	1	2	3	4	9

Q21_Luxfood **How many times in a week did you normally use eat meat, fish, chicken or pork?**
Kodi ni kangati musabata limozi kamene munali kudya za kudya zopala nyama, nsomba, nkhuku kapena nyama ya nkhumba?
Bushe miku inga mu mulungu ilyo mwalelyapo inama, isabi, inkoko nangula inkumba?

0	1	2	3	4	5	>5	Unknown
0	1	2	3	4	5	6	9

Section E – Health status

SAY: I would now like to ask you some questions about your health, specifically about TB and HIV

Q22_ETB_C0 **Has a doctor ever told you had TB**

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO GO TO Q25_BCG

In which year it was treated?

Y	Y	Y	Y

Q23_1_CO **First time treated**

Q23_2_CO **Second time treated**

Q23_3_CO **Third time treated**

Unknown

Q23_4 **Cannot remember the year** | 9 |

Q24_CATT_CO **The last time you had TB, did you finish the course of treatment?**
Kodi pamene munadwala chifuba chaTB munasiliza kumwa mankwala a chifuba chaTB?
Bushe ilyo mwalwele TB kunuma uku, Mwalipwishishi ukunwa umuti wa TB?

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO GO TO Q26_Notreat_CO

INDIVIDUAL BARCODE

Q25_SET_CO **Did you have a sputum examined at the end of the treatment**
Kodi munapimwa vikola pamene munasiliza kumwa munkwala wachifuba cha TB?
Bushe mwalipimishepo ifkola ilyo mwapwishishi ukunwa umuti wa TB?

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO GO TO Q24_Stigma

Q25_SRET_CO **What was the result?**

Negative for TB	Positive for TB	Awaiting results	Unknown
0	1	2	9

Q26_Notreat_CO **Why you HAVE NOT completed your treatment?**
Nichifukwa chiani chimene simunasilize kumwa munkwala?
Bushi cinshi calengele ukuti mwipwisha ukunwa umuti?

Still on treatment	0	**No drugs available**	1
Treatment unsuccessful	2	**Treatment not needed**	3
Treatment too expensive	4	**Clinic too far**	5
Feeling better	6	**Other, specify**	7
Unknown	9		

Q24_Stigma **Since you fell sick have you ever experienced any of the following:**
Kuchokela pamene munadwalila TB, kodi zinakuchitikilaniponi izi?
Apo mwalwaalila TB, mwalitala amucitikilwapo nefili fyonse pali ifi fyakonkapo

	No	Yes
Been excluded from a social gathering *Ku patulidwa ku misonkhano* Uwo batamfyapo apo abantu bakolongene	0	1
Abandoned by spouse/partner *Kapena kusiiwa ku okondedwa awo* Uwo babutuka ku mwina mwakwe/umutemwikwa	0	1
Isolated by your household *Ku musungila payekha mukati mwa banja* Uwo baapatulula ku ba pang'anda	0	1

	No	Yes
Lost housing/been unable to rent a property *Kuchotsedwa munyumba kapena kukanizidwa nyumba yosonkhela* Uwo batamfya pa ng'anda nangu uwo bakanina ukupela ing'anda ya kusonkela	0	1
Lost respect or standing in the community *Anthu analeka kumu lemekeza mukomboni* Uwalushilapo no mucinshi ukufuma ku Bantu mu ncende ekala	0	1
Been teased, insulted or sworn at *Anthu kumuseka,kumunyoza kapena kumu tukwana* Uusekwa, uutukwa nangula uwo balapisha	0	1
Been gossiped about *Ku munena kumbali* Uwo baamba	0	1
Your children or family have been isolated/shunned *Kapena ana anu ndi banja lanu si bvomelezedwa kucheza ndi anthu ena* Abana nangula balupwa benu balibapatulula/balabataluka	0	1
Been treated worse than other patients by health staff *Kapena simu mathandizidwa bwino ndi aja ogwila nchito muchipatala kulingana ndi aja odwala matenda ena* Tabamitangata ukufikapoku babomfi ba muchipatala ukulingana na balwele ba malwele yambi	0	1

		No	Yes	Unknown
Q25_BCG	**Have you ever been vaccinated for TB**	0	1	9

Q26_Scar	**Can I see if you have any BCG scar on your arm?**	**Examination not allowed**	0
		No visible scar on both arms	1
		BCG confirmed by scar examination	2

Q27_Sugar **Do you suffer from any of these clinical conditions?**

Kodi munadwalapo matenda monga?
Bushe mwalilwalapo amalwele ayaba nga?

	No	Yes	Unknown
Sugar	0	1	9
Cancer	0	1	9

Q28_Hosp Walk **How long in minutes does it take to walk to the nearest health clinic/hospital from here?**
Kodi cimakutengelani mpindi zingati kuchokela kunyumbla kwanu kukafika kuchipatala chapa fupi?
Bushe cimusendela insa shinga pakufika ku cipatala icili mupepi?

0-15	16-30	31-45	45-60	> 60
0	1	2	3	4

INDIVIDUAL BARCODE

	No	Yes	Unknown
Q29_HIVTest **Have you ever been tested for HIV?** *Kodi muna pimisa pa za kadoyo ka HIV?* Bushe mwalipimwapo palwa kashishi ka HIV?	0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

	Neg	Pos	Unknown
Q29_1_HIVRes **Would you mind telling me what was the result?** *Kodi mungakondwele kuniuza zimene anapezamo?* Bushe kuti mwanjebako fyatumbwikemo?	0	1	9

IF POSITIVE CONTINUE. IF NEGATIVE OR UNKNOWN GO TO NEXT SECTION

	No	Yes	Unknown
Q30_ARVever **Have you ever taken ARV treatment?** *Kodi munamwapo mankhwala ya ARVs*? Bushe mwalitala amunwapo ama ARVs?	0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

	No	Yes	Unknown
Q31_ARVnow **Are you currently taking ARV treatment?** *Kodi mukumwa mankhwala ya ARVs pali pano?* Bushe mulenwa ama ARVs pali ino nshita?	0	1	9

IF YES CONTINUE. IF NO OR UNKNOWN GO TO NEXT SECTION

Q31_1_ARVdate **When did you start ARV treatment**
Kodi munayamba liti kumwa mankhwala ya ARVs?
Nilisa mwayambile ukunwa ama ARVs?

D	D	M	M	Y	Y	Y	Y

Q31_2ARVwhere	**Which health centre do you receive treatment from?** *Kodi mumalandilila ku chipatala chiti mankhwala aya?* Chipatala ninshi musendelako uyu umuti?	

Section F – Hazardous behaviours of the case

SAY: Now I would now like to ask you some questions about your behaviours and habits

		Never/ Rarely	1-4 times/month	Many times/week	Every day	Unknown
Q32_Alcohol	**How often do you drink any type of alcohol?** *Kodi kamwedwe ka moba kanu kali muntani?* Bushe iminwene ya bwalwa yenu yaba shani?	0	1	2	3	9

Q33_Drinks **When you drink, how many drinks containing alcohol might you have on a typical day?**

Kodi mumamwa mabotolo angati pa siku lanu lo kumwa mooba?
Bushe ngamulenwa munwa amabotola yanga pa bushiku bumo?

1-3	4-6	7-9	>= 10	Unknown
0	1	2	3	9

		No	Yes	Unknown
Q34_Drunk1	**During the last year have you ever drunk so much that you were unable to remember what you were doing?** *Mu chaka chatha ichi kodi munalezelapo kufikila kuti simuna kumbukile zimene zinachitika kamba kaka mwedwe kanu?* Muli uyu mwaka wapwile mwaliikolelwepo icakuti mwalaba nefyo mwalecita?	0	1	9

		No	Yes	Unknown
Q35_Drunk2	**During the last year have you been unable to remember what happened the night before because of your drinking?** *Mu chaka chatha ichi simunakumbukilepo zamene zinacitika usiku wamene munalezela kwambili cifukwa ca kamwedwe kanu?* Bushe muli uyu mwaka wapwile mwalikolwapo icakuti mwalaba ne fyacitike ubushiku bwafumineko?	0	1	9

Q36_Drunk3

During the last year have you been criticised by a relative, friend or a doctor because of the amount of alcohol you drink?
Kodi caka chatha anakuzuzulaniponi abanja kapena anzanu olo adotolo cifukwa ca kamwedwe kanu?
Muli uyu mwaka wapwile mwallikalilwapo na balupwa, banenu nangula ba shing'anga ba kucipatala pa mulandu wa minwene yenu?

No	Yes	Unknown
0	1	9

Q37_Drunk4

During the last year, have you or anybody else been injured because of your drinking?
Kodi chaka chatha, imwe kapena munthu wina anacitiwapo ngozi cifukwa cha kamwedwe kamoba kanu?
Muli uyu mwaka wapwile, bushe mwaliicenapo nangu umuntu umbi ukucenekwa pa mulandu wa kunwa kwenu?

No	Yes	Unknown
0	1	9

INDIVIDUAL BARCODE

Q38_Drunk5

During the last year, have you been arrested, been held at a police station, or had any legal problems because of your drinking?
Kodi chaka chatha, imwe munamangiwapo ndi a kapokola kapena kuimbiwa mulandu cifukwa ca kamwedwe kamoba kanu?
Muli uyu mwaka wapwile, balimwiketepo kuli bakapokola nangula ukulubululapo imilandu pa mulandu ne minwene yenu?

No	Yes	Unknown
0	1	9

Q39_Smoke

How would you classify your smoking habits?
Kodi munga ziike mugulu la bwanji pakakokedwe kafwaka?
Kuti mwabikapisa imipeepele ya fwaka yenu?

Never smoked *Mukalibe kukokapo* Tamwapeepapo	0	**Daily smoker** *Masiku onse* Mupeepa cilabushiku	1
Occasional smoker *Kukoka panthawi* Mupeepa pa kashita	2	**Ex-smoker** *Munaleka* Mwalileka ukupeepa	3

(IF SMOKER OR EX-SMOKER CONTINUE, IF NEVER SMOKED GO TO Q40)

Q39_1_Longsmoke

How long have you/did you smoke for?(years)
(if less than 1 year: put 00 for less than 6 months and 01 for 6months-1 year. 999 if Unknown)
Mwakoka fwaka pa zaka zingati /muna koka fwaqka zaka zingati?
Myaka inga iyapitapo apo mwayambila ukupeepa fwaka/myaka inga mwapeepelepo fwaka?

Q39_2_Cigarettes **How many cigarettes do/did you smoke per day? (999 if Unkwown)**
Mukoka mishanga ingati pa siku / munali kukoka mishanga ingati pa siku?
Mupeepa/mwalepeepa imishanga inga pa bushiku bumo?

Q40_Dagga **Do you ever smoke marijuana/dagga?**
Kodi mumakoka chamba kapena zina zotele?
Bushe mulapeepa ibange/icamba?

No	Yes	Unknown
0	1	9

Q41_Drugs **Do you take any other recreational drugs?**
Mumakoka fwaka kapena kusebenzesa zilizonse zamene zimapangisa munthu kukhala osangalala kapena kuzunguza bongo?
Bushe mwalibomfyapo ifimiti (ama drugs) fimbi ukucila pa camba ifyakumusansamusha?

No	Yes	Unknown
0	1	9

Q42_Cooking **Are you ever exposed to cooking fire indoor?**
Kodi munankhalapo pafupi na kumoto munyumba?
Bushe mwalisangwapo ku mulilo wa kwipikila mukati kang'anda?

Never/ rarely	Yes	Unknown
0	1	9

Q43_OthSmoke **Are you exposed to any other type of smoke on a regular basis?**
Kodi mumankhala pafupi ndi chusi kambili kambili?
Bushe mulekala mupepi ne chushi inshita iitali?

No	Yes	Unknown
0	1	9

Q44_TBcontact **Had you been in contact with anyone with TB before we first visited you?**
Kodi munankhalapo ndi munthu odwala chifuba ca TB pamene tikalibe kukutandalilani?
Bushe mwalisangwapo no mulwele wa TB lintu tatulamutandalila?

No	Yes	Unknown
0	1	9

Q45_TBprof **Have you ever been professionally exposed to TB cases as a health care worker?**
Kapena muna gwilapo nchito yosamala odwala TB?
Limbi mwalibombapo incito yakundapa ubulwele bwa TB?

No	Yes	Unknown
0	1	9

Q46_Prison **Have you ever been in prison?**

No	Yes	Unknown
0	1	9

Kodi munankalapo mundende?
Limbi mwalikalapo muchifungo?

IF YES CONTINUE, if NO SKIP TO => Q47

Q46_1_Prisonwhen **When?**

< 1 year ago	1-5 years ago	> 5 years ago	Unknown
0	1	2	9

Q46_2_Prisonlong **For how long?**

< 1 year	1-5 years	> 5 years	Unknown
0	1	2	9

Q47_Video **Do you ever go to VIDEO CLUBS?**

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q48

Q47_1_OftenVideo **How often?**

< 1 month	1/month	> 1/month	Unknown
0	1	2	9

INDIVIDUAL BARCODE

Q47_2_LongVideo **How long do you typically remain there?**

≤ 1 hour	1-2 hours	> 2 hours	Unknown
0	1	2	9

Q48_Bars **Do you ever go to BARS?**

No	Yes	Unknown
0	1	9

IF YES CONTINUE, if NO SKIP TO => Q49

Q48_1_OftenBars **How often?**

≤ 1 month	1/month	> 1/month	Unknown
0	1	2	9

Q48_2_LongBars **How long do you typically remain there?**

≤ 1 hour	1-2 hours	> 2 hours	Unknown
0	1	2	9

Q49_Hair **Do you ever go to HAIR SALOONS/BARBER SHOP?**

No	Yes	Unknown
0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q50

		≤ 1 week	1/week	> 1/week	Unknown
Q49_1_OftenHair	**How often?**	0	1	2	9

		≤ 1 hour	1-2 hours	> 2 hours	Unknown
Q49_2_LongHair	**How long do you typically remain there?**	0	1	2	9

		No	Yes	Unknown
Q50_Church	**Do you ever go to CHURCH?**	0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q51

		≤ 1 week	1/week	> 1/week	Unknown
Q50_1_OftenChurch	**How often?**	0	1	2	9

		≤ 1 hour	1-2 hours	> 2 hours	Unknown
Q50_2_LongChurch	**How long do you typically remain there?**	0	1	2	9

		No	Yes	Unknown
Q51_Migr	**In the 5 years before we first visited you, have you lived anywhere other than where you live today for a period of at least 6 months?** *Pa zaka zisano zapita pamene ife tikalibe kukutandalilani munakhalapo kwina kwach mu miyezi isami ndi umodzi posacedwa apa?* Bushe pamyaka yapita isano, mwalikalapo kunchende imbi panshita iilingene ne myeshi mutanda?	0	1	9

IF YES CONTINUE, IF NO SKIP TO NEXT SECTION

		Somewhere else in Zambia	Outside Zambia	Outside Africa	Unknown
Q51_1_Wheremigr	**Where?**	0	1	2	9

THANK THE RESPONDENT FOR THEIR PARTICIPATION AND ASK HIM/HER WHAT HE/SHE THINKS POVERTY IS – FILL IN THE BOX WITH THE DEFINITION PROVIDED

	Interviewer's Code	Date								Signature
		d	d	m	m	y	y	y	Y	
Interviewer										
Field Manager										
1st data entry										
2nd data entry										

HOUSEHOLD QUESTIONNAIRE

INDIVIDUAL BARCODE

Section A – General info on the head of the household

Q01_HH_Intcode **Interviewer's Code**

Q02_HH_Intdate **Date of Interview**

D	D	M	M	Y	Y	Y	Y

Q03_HH_Consent **Consent**

No	Yes	Absent	Excluded
0	1	2	3

Q04_HH_Sex **Sex**

M	F
1	2

Q05_HH_DOB **Date of Birth** (01/01/1800 if unknown)

D	D	M	M	Y	Y	Y	Y

Q05_1_HH_Age **If not known, what was your age in years on your last birthday?** (999 if unknown)

Q06_HH_Ethnic **Which ethnic group do you belong to?**

Black Zambian	Black other	White	Asian	Other	Unknown
0	1	2	3	4	9

Q07_HH_Marital **What is your marital status?**

Married	0	Living as married	1
Single	2	Divorced	3
Widowed	4	Unknown	9

Q08_HH_Bread **Are you the one who most contribute the economic resources of the household? IF NO go to NEXT SECTION, if YES skip to SECTION C**

No	Yes	Unknown
0	1	9

Section B – General info on the main bread winner

TO BE FILLED OUT ONLY IF THE HEAD OF THE HOUSEHOLS IS NOT THE MAIN BREAD WINNER

Q09_HH_Breadcode **Bread winner's code**

Q10_HH_Breadsex **Sex of the bread winner**

M	F
1	2

Q11_HH_BreadDOB **Date of Birth** (01/01/1800 if unknown)

D	D	M	M	Y	Y	Y	Y

Q11_1_HH_BreadAge **If not known, what was your age in years on your last birthday?** (999 if unknown)

Q12_HH_BreadEthnic **To which ethnic group does the breadwinner belong**

Black Zambian	0	Black other	1
White	2	Asian	3
Other	4	Unknown	9

Q13_HH_BreadMarital **What is his/her marital status?**

Married	0	Living as married	1
Single	2	Divorced	3
Widowed	4	Unknown	9

Section C – Household Roster

SAY: Now I am going to ask you some questions about the people who live in this household. That is the people who live together, usually pool their income and eat at least one meal together when they are at home. This does not include people who have permanently migrated or are considered visitors.

Q14_HH_Number **How many people live in your household, including yourself, the main bread winner (if different from you) and the person we have just interviewed**

Can you please tell me about each person:

	Q14_HH_Sex	Q14_HH_Age	Q14_HH_ Occupied	Q14_HH_ Unemployed/ retired	Q14_HH_ Infant/ student	Q14_HH_ Physically*/ mentally ill

ID serial number	M	F	Y	M	No	Yes	No	Yes	No	Yes	No	Yes
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1

INDIVIDUAL BARCODE

	Q14_HH_Sex		Q14_HH_Age		Q14_HH_ Occupied		Q14_HH_ Unemployed/ retired		Q14_HH_ Infant/ student		Q14_HH_ Physically*/ mentally ill	
ID serial number	M	F	Y	M	No	Yes	No	Yes	No	Yes	No	Yes
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1
	0	1			0	1	0	1	0	1	0	1

***SAY: Does he or she have any permanent health problem, either physical or mental, that stops him/her performing normal daily activities?**

Section D – Educational status of the main breadwinner

If the head of the household is also the main breadwinner
SAY: I would now like to ask you some questions on your educational status.

If the head of the household is not the main bread winner
SAY: I would like to ask you some questions on the educational status of the person who most contributes to economic resources of the household. If this person is at home we can ask this information to him/her if you want so, otherwise you can answer on his/her behalf.

		No	Yes	Unknown
Q15_HH_Write	Can you write and read?	0	1	9
Q16_HH_Schever	Have you ever attended school?	0	1	9

IF YES CONTINUE, IF NO SKIP TO => Q17

Q16_1_HH_Grade **What is the highest grade completed**

None	1-4	5-7	8-9	10-12	College	Unknown
0	1	2	3	4	5	9

Q17_HH_Schnever **Why have you never attended school? (Tick the most important option only)**

Couldn't get a place	0	Too expensive/couldn't find support	1
Too far away	2	Illness or injury	3
Unsafe to travel to school	4	School not important	5
Quality of school bad	6	Need to help family	7
Other, please specify	8	Unknown	9

Section E – Occupation status of the main bread winner

If the head of the household is also the main breadwinner
SAY: I would now like to ask you some questions about the most important activity you have been involved into at the time we first visited your household in terms of earning money or goods for yourself or for the household.

If the head of the household is not the main bread winner
SAY: I would like to ask you some questions about the most important activity the main breadwinner have been involved into in terms of earning money or goods for himself/herself or for the household.
If this person is at home we can ask this information to him/her if you want so, otherwise you can answer on his/her behalf.

Q18_HH_Activity **What was your main activity?**

Unemployed	0	Seasonal/ piece worker	1
Student	2	Employed	3
Retired	4	Self employed	5
Housewife	6	Unknown	9

IF UNEMPLOYED GO TO Q19

Q18_1_HH_Job	**If employed/self-employed or a seasonal or piece worker, what type of job did you do?** *Ngati munali kusebenza, kodi munali kusebenzela muntu, kapena kuzisebenzela, olo kuchita maganyu?* Nga mwalebomba/ukuibombela nagula bucibombebombe, ni ncito nshi mwalebomba?	**Agriculture/fisheries**	**0**	**Transport**	**1**
		Construction/electricity/ gas/water	**2**	**Manufacturing**	**3**
		Community social services (health, education)	**4**	**Household employee/ housekeeper**	**5**

INDIVIDUAL BARCODE

		Mining/Quarrying	**6**	**Trader**	**7**
		Clerical and related	**8**	**Professional (finance, law, academic**	**9**
		General worker	**10**	**Other**	**11**
				Unknown	**99**

Q18_2_HH_Income

Beside you, how many other member of the household were involved in any income generating activities?
Kuchoselako imwe, nibangati mubanja lanu bamene benzekutandidzira kubwetsa ndalama kapene chakudya?
Ukufumyako imwe, nibanga pabo mwikala nabo pa ng'anda abaleletako ulupiya lwa kubomfya pa ng'anda?

0	1	2	3	>3	Unknown
0	1	2	3	4	9

Q19_HH_NoJob	**What was the main reason you were not working at that time?** *Nicifukwa chiani chikulu cimene simunali kusebenzela?* Bushe cinshi icikalamba icalengele ukti mwilabomba pali ilya nshita?	**Low wage/salary/ no profit**	**0**	**Fired**	**1**
		Enterprise closed/liquidated/ privatised	**2**	**Sick because of TB**	**3**
		Sick because of disease other than TB	**4**	**Made redundant**	**5**
		Invalid	**6**	**Student/Retired/ Housewife**	**7**
		Other	**8**	**Unknown**	**9**

Section F – Diet habits and food consumption of the household

SAY: I would now like to ask you some questions about your dietary habits and food consumption in your household at the time we first visited you

Q20_HH_Meals

How many meals did you normally eat per day in your family?
Munali kudya kangati pa siku pa nyumba panu?
Bushe miku inga mwalelya ifyakulya pabushiku bumo pang'anda pamyenu?

0	1	2	3	>3	Unknown
0	1	2	3	4	9

Q21_HH_Luxfood

How many times in a week did you normally eat meat, fish, chicken, or pork in your household?
Kodi ni kangati musabata limozi kamene munali kudya za kudya zopala nyama, nsomba, nkhuku kapena nyama ya nkhumba?
Bushe miku inga mu mulungu ilyo mwalelyapo inama, isabi, inkoko nangula inkumba?

0	1	2	3	4	5	>5	Unknown
0	1	2	3	4	5	6	9

Q22_HH_Nofood

During the last 12 months, for how many months did your household had without enough to eat
Pa mwezi kumi ndi iwili, ndi myezi ingati imene banja lanu linankala kopanda za kudwa zo kwanila?
Bushe pa myeshi ikumi na ibili iyapitile, myeshi inga iyo ulupwa lwenu lwaikele ukwabula ukukwata ifyakulya ifingi pa ng'anda?

0	1	2	3	4	5	6	>6	Unknown
0	1	2	3	4	5	6	7	9

Once they answered probe answers by asking: "so in the last 12 months you did have enough to eat for – NUMBER ANSWERED- months? Is it correct?

Section G – Assets and quality of housing

SAY: Now I would like to ask you some questions about the place where you live and the assets you own

Q23_HH_Sleep

How many sleeping room are in your dwelling?
Kodi nyumba yanu ili ndi vipinda vingati vogonamo?
Bushe ing'anda yenu ya kwa imiputule ya kulalamo inga?

Q24_HH_Windows

Overall how many windows/vents you have in your sleeping rooms
Pamozi niyangati mawindo yamene mulinayo kuvipinda vogonamo?
Bushe mwaba amaawindo yanga mu miputule ya kulalamo?

Windows	
Vents	

Q25_HH_Rent **Is the dwelling?**

Your own	**1**	**Other**	**3**
Rented	**2**	**Unknown**	**9**

Q26_HH_Assets **In your household is there**
Kodi muli na zinthu izi panyumba panu?
Pa N'ganda pa mwenu mwalikwata ifintu ifif ifyakonkapo
READ OUT EVERY ITEM - MARK EVERY ITEM

	No	**Yes**
A radio *Wailesi* Ichilimba	**0**	**1**
A television *Wailesi ya kanema* TV	**0**	**1**

INDIVIDUAL BARCODE

	No	**Yes**
A refrigerator *Filigi*	**0**	**1**
A bicycle *Njinga* Inchinga	**0**	**1**
A motorcycle *Honda* Mpumpu	**0**	**1**
A car *Galimoto* Motoka	**0**	**1**
A domestic worker not related to household head *Wanchito wapanyumba* Umubomfi wapa n'ganda ushili lupwa wabene ba N'ganda	**0**	**1**

Q27_HH_Light **What is the main source of ENERGY used for lighting in your household?**
Kodi mu nyumba yanu chenicheni chamene musanikila niciani?
Bushe finshi mubomfya sana pa kusanika mu ng'anda yenu?
READ OUT EVERY ITEM - MARK ONE OPTION ONLY

Kerosine/parafine Parafini	**1**	**Electricity** *Malaiti* Amalaiti	**2**
Candles *Kandulo* Kandulo	**3**	**Open Fire** *Moto* Umulilo	**4**
Torch *Toci* Toci	**5**	**Solar panels** *Batili ya magesi ya zuba* Amabatili ayafumya amaka ukufuma ku kasuba	**6**

Other
Zinangu
Fimbi

7

Unknown

9

Q28_HH_Water

What is the main source of DRINKING WATER for this household?
Kodi manzi akumwa pano panyumba a chokela kuti?
Bushe amenshi munwa pa N'ganda ilingiline yafuma kwi?
DO NOT READ OUT SOURCES. MARK ONE OPTION ONLY

Piped water in the residence
Kupompi ya pabanja
Ku mipipi (ku Tap) ya pa N'ganda

1

Piped water from a public tap
Ku pompi ya gulu
Ku mipipi (ku Tap) ya cintu bwingi

2

Inside well
Ku chisime cha mukati
Mu chishima cenu pa ng'anda

3

Shallow well
Chisime cha pa nyumba
Ichishima

4

Traditional well
Chisime cha mwambo
Mu chishima ica cintu bwingi

5

Bore hole
Ku borehole

6

River, stream, lake etc
Ku msinje, mumana, nyanja
Mumumana,muchishiba

7

Other
Kwina
Kumbi

8

Unknown

9

Q29_HH_Toilet

What is the main type of TOILET facility for this household?
Nanga ni chimbuzi chabwanji chimene mu sebenzesa pano panyumba?
Cimbusu ca musango shani caba pano pa N'ganda?
DO NOT READ OUT FACILITY - MARK ONE OPTION ONLY

Private flush toilet
Chogujumula Chamunyumba
Icimbusu caku kumpa ico mubomfyafye pa N'ganda mweka

1

Shared flush toilet
Chogujumula cha anthu ambil
Icimbusu caku kumpa ico mubomfya na mayanda yambi ayali mupepi

2

Pit Latrine
Cha mugodi
Icimbusu

3

VIP Latrine
Cha mugodi chimene chilli ndi ka paipi?
Icimbusu (ico mwabikako umupaipi uwakupisha umwela.

4

None- use bush/field
Kapena kulibe, olo musebenzesa kusanga kapena kuminda
Tapaba; tuya mu mpanga

5

Other
Kwina
Kumbi

6

Unknown

9

Q30_HH_Floor

What is the main type of FLOORING for this household?
Nanga pansi pa mukati ya nyumba yanu pana mangidwa ndi chiani?
Bushingulo (floor) nshi bwaba muli ino N'ganda?
DO NOT READ OUT OPTIONS - MARK ONE ONLY

Dirt/earth
Dothi
Yaku shingula

1

Wood, plank
Mapulanga
Ya mapulanga

2

Parquet, lino Matailo	3	**Cement** Simenti	4
Tile flooring *Amatawelo*	5	**Other** *Zinangu* Cimbi	6
Unknown	9		

Q31_HH_Walls **What is the main type of WALLS in your dwelling?**

Kodi nyumba yanu ili navipupa vo tani?
Bushe ing'anda yenu ya pangwa ne cibumba ca shani?

DO NOT READ OUT OPTIONS - MARK ONE ONLY

Concrete brick *Mabloko ya sementi* Amabuloko	1	**Mud brick** *Mabloko ya doti* Njelwa	2

INDIVIDUAL BARCODE

Mud, burnt brick *Mabloko ya doti yoshyoka* Njelwa shakoca	3	**Pole & Dagga** *Mitengo* Fimiti	4
Mud *Madoti* Yakumasa	5	**Other** *Zinangu* Fimbi	6
Unknown	9		

Q32_HH_Roof **What is the type of ROOFING in your dwelling?**

Nanga mutenge wanyumba yanu niyabwanji?
Bushe mutenge nshi inganda yenu yakwata?

DO NOT READ OUT OPTIONS - MARK ONE ONLY

Asbestos *Malata ya masibesita* Amalata ya asibesita	1	**Iron sheet** *Malata ya nsimbi* Amalata	2
Grass/Straw *Mauzu* Ifyani	3	**Other** *Zinangu* Fimbi	4
Unknown	9		

Q33_HH_Cook **What is the main type of COOKING energy?**

Muma pika pa ciani chenicheni?

Bushe finshi mubomfya sana sana mukwipikila?

READ OUT EVERY ITEM - MARK ONE OPTION ONLY

Collected fire wood *Nkuni* Inkuni	1	**Charcoal purchased** *Malasha* Amalasha	2
Electricity *Magesi* Amalaiti	3	**Other** *Zinangu* Fimbi	4
Unknown	9		

Q34_HH_Land **Do any members of your household own, rent or borrow land for agricultural work?**

Kodi mubanja lanu, alimo amene alinawo munda , kapena wolipila olo obweleka wamene amalimamo?

Bushe pa nganda pamwenu paliba abakwata, ablipila nangula abashima impanga apakulima?

No	Yes
0	1

IF YES CONTINUE, IF NO GO TO => Q35_Cattle

Q34_1_HH_Landrent **Is the biggest plot you use**

Owned	Rented	Borrowed
1	2	3

Q34_2_HH_Lima **How much of this plot is used for agricultural purposes**

Ndimbali ikulu bwanji yamene isebenzeselwa ku ulimi?

Bushe impanga mubomfya kubulimi yaba shani ubukalamba?

Lima	
Acre	
Hectares	

Q34_3_HH_Unknown **Don't know how much cultivable land is owned**

Unknown
999

Q35_HH_Cattle **How many heads of CATTLE do you own?**

Kodi muli na ng'ombe zingati?

Bushe mwakwata ingombe shinga?

Q36_HH_Goats **How many heads of GOATS do you own?**

Kodi muli na mbuzi zingati?

Bushe mwakwata imbushi shinga?

Q37_HH_Pigs **How many heads of PIGS do you own?**

Kodi muli na nkumba zingati?

Bushe mwakwata inkumba shinga?

Q38_HH_Chicken **How many heads of CHICKEN do you own?**

Kodi muli na nkhuku zingati?

Bushe mwakwata inkoko shinga?

Q39_HH_Ducks **How many heads of DUCKS/GEESE do you own?**
Kodi muli na vibakha zingati?
Bushe mwakwata ingombe shinga?

Q40_HH_Rabbits **How many heads of RABBITS do you own?**
Kodi muli na akalulu angati?
Bushe mwakwata tu kalulu tunga?

Q41_HH_Horses **How many heads of HORSES/DONKEYS do you own?**
Kodi muli na mahosi yangati?
Bushe mwakwata amahosi yanga?

Q42_HH_Walk **How long in minutes does it take from here to walk to the nearest...**
Kodi cimamutengerani nthawi bwanji kuyenda ku chokela a kuno kufika....pafupi
Bushe mupwisha insa shinga ku fika.....mupepi

	0-15	16-30	31-45	46-60	> 60
Supply of drinking water *Uko mutapa amenshi ya kunwa* Kumene mutapa manzi yo khumwa					
Food market *Ku maliketi kwamene baguriza vokundya* Ukukobashitisha ifya kulya					
Public transportation *Komwe mumakwerera mabasi* Apakuninina abamotoka					
Primary school *Kuprimari sukulu* Kuprimari sukulu					
Secondary school *Kusekondari sukulu* Kusekondari sukulu					
Health clinic *Kukiliniki* Kukiliniki					

INDIVIDUAL BARCODE

Q43_HH_Poor **Do you consider your household to be**
Kodi muona kuti banja lanu muli otani?
Bushe panganda pamwenu mwaliba ?

Not poor	Moderately poor	Very poor	Don't know
0	1	2	9

Q44_HH_Today **Compared to last year do you consider your household to be**
Kulinganiza ndi chaka chata muona monga mubanja lanu vinthu vankhala motani?
Ukulinganya no mwaka wapiti bushe mulemona shani imikalile yenu pang'anda lelo?

Better off	The same	Worse off	Don't know
0	1	2	9

SAY: Finally, there are a number of ways people can cope in time of hardship. Did your household have to rely on any of the following during the last 12 months?

		No	Yes	Unknown
Q45_HH_Cope	**Piecework on farms belonging to other households** *Kodi munali kuchita ma pisi work kapena maguyu miminda ya anthu ena.* Uku bombela mumabala ya bantu	0	1	9
	Working on "work for food" or "food for assets" *Kapena munali kusebenzela chakudya kapena katundu* Ukubombela ifyakulya nangula ifintu fimbi ifyakubomfya	0	1	9
	Relief food, free food from government and other bodies *Kapena chakudya chopasiwa kapena chaulele cho chokela kuboma* Nalimo ifyakulya ukufuma kubuteko ifyakula ukulipila	0	1	9
	Eating wild food only *Kapena kudya zo kudya zamusanga* Nangula ukulya ifi nsabwa nsabwa fya mumpanga	0	1	9
	Substituting ordinary meals with mangos, pumpkin, sweet potatoes, etc... *Mumalo yakudya nsima kapena munali kukdya mango,matanga,kandolo kapena zina zace* Ukulya ifyakulya ifyapala mango,ifipushi,ifyumbu nefyashala	0	1	9
	Reducing number of meals or food in-take *Kucepesa nambala kapena vakudya* Ukuchefya imiku iyakulya nangula ubukulu bwa chakulya	0	1	9
	Reducing other households items (i.e. soap, tissues) *Kuchefya zofunikila za panyumba monga, sopo, tissue ndi zina cane* Ukuchefyako ifintu fyapa nganda nge sopo nefyashala	0	1	9
	Informal borrowing (e.g. kaloba, borrowing from friends) *Kubweleka kocekela kwa anzanu, kapena kutenga kaloba* Ukukongola(kaloba nangula ifintu ukufuma kubanenu)	0	1	9
	Formal borrowing in cash or kind (e.g. borrowing from bank,) *Kubweleka kololedwa kwa ndalama kapena zinthu kuchokela monga kuma banki* Uku kongola indalama ukufuma ku basunga ifipao kwati kuma banki	0	1	9

Rely on church charity or other NGOs

0	1	9

Kuchetela thandizo kuchokela kuma chaliti kapena kuma bungwe ena othandizila.
Uku chetekela ku machalichi nangula utubungwe tumbi kubwafwilisho

Sale of assets

0	1	9

Kapena kugulisa katundu
Na ngula ukushitisha ifipe

Petty vending

0	1	9

Kuchita tu malonda
Nangula ukushitisha ifya kushitisha shitisha

Pulling children out of school

0	1	9

Kulesa bana kuyenda kusukulu
Kulesha abana ukulaya kusukulu

Migrated to find work

0	1	9

Kapena munapita kwina ku kafuna nchito
Ukukila kumbi mukufwaya inchito

Begging from the streets

0	1	9

Kuphempa mimiseu
Kulombalomba mumisebo

INDIVIDUAL BARCODE

THANK THE RESPONDENT FOR PARTICIPATION AND ASK HIM/HER WHAT HE/SHE THINKS POVERTY IS – FILL IN THE BOX WITH THE DEFINITION PROVIDED

	Interviewer's Code	Date								Signature
		d	d	m	m	y	y	y	Y	
Interviewer										
Field Manager										
1st data entry										
2nd data entry										

APPENDIX F

Am J. Trop. Med Hyg., 80(6), 2009, pp. 1004–1011

Tuberculosis Infection in Zambia: The Association with Relative Wealth

Delia Boccia,* James Hargreaves, Helen Ayles, Katherine Fielding, Musonda Simwinga, and Peter Godfrey-Faussett

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom; Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom; ZAMBART Project, Lusaka, Zambia

Abstract. This study aimed to assess the association between household socioeconomic position and tuberculosis (TB) infection in two communities of Zambia. For this purpose we implemented a cross-sectional investigation, nested within a larger case control study. Infection was assessed using Quantiferon-TB Gold. A socioeconomic position index was constructed through principal component analysis combining data on human resources, food availability, housing quality, and access to services and infrastructures. In this study, higher socioeconomic position, rather than lower, was associated with significantly higher risk of TB infection. None of the traditional risk factors for TB infection mediated this association, suggesting that in these two communities TB transmission may occur through exposure to as yet undefined risk factors that are associated with higher socioeconomic position. Although further studies are needed, these results suggest emerging new patterns of TB transmission and a role of socioeconomic position on the risk of TB infection opposite to that expected.

INTRODUCTION

Tuberculosis (TB) is considered to be a disease of poverty.[1] Its association with low socioeconomic position (SEP) is well established at the ecologic level: 17 of the 22 highest burden countries accounting for 80% of the world's TB cases are classified as low income.[2] The World Health Organization (WHO) estimates that 98% of the 2 million annual TB deaths and 95% of the 8.4 million new TB cases occur in developing countries.[3] Furthermore, recent data from the United States suggest that socioeconomic factors act independently from the human immunodeficiency virus (HIV) epidemic.[4,5]

In contrast, the association between TB and low SEP at the individual level is less well characterized and studies provide more conflicting results.[5,6] This is probably because living conditions are time- and setting-specific and because of the inconsistency of the measurement strategies adopted. In TB studies, the most frequently used SEP indicators are median household income, expenditure, crowding, level of education, and housing quality.[4,7–17] Composite indicators have also been used, such as the Townsend deprivation index[8] and the Jarman index.[9]

Results interpretation is also made difficult by the two-stage nature of TB, characterized by an infection and a disease stage. Often studies do not clearly differentiate between TB infection and TB disease, and it is not yet clear how SEP is associated with the risk of becoming infected, the risk of developing the disease, or both.

Understanding the association between SEP and risk of TB infection (rather than disease) is further complicated by the fact that TB infection has traditionally been assessed by the tuberculin skin test, a tool in which TB extracts are injected and skin induration 2 days later is considered a sign of TB infection. Tuberculin skin test is prone to false positive results as a consequence of bacilli Calmette-Guérin (BCG) vaccination and exposure to environmental bacteria,[18,19] both of which are associated with SEP.[20–22]

These problems in assessing SEP and TB infection may explain the conflicting results of the few published studies. Research in North America and Europe showed that tuberculin skin test positivity was least frequent in households with higher educational level, income, skilled occupations, and room size.[7,23,24] In contrast, studies in the Gambia,[10] Malawi,[11] and Peru[25] found that the risk of tuberculin skin test positivity was not associated with socioeconomic indicators.

Recently, an easier and more standardized approach in the assessment of SEP has been proposed by Filmer and Pritchet,[26] whereby households are ranked according to the ownership of assets. In this approach, the relative weight of each asset is computed through principal component analysis, a data reduction strategy used to reduce a number of exposures to a single proxy measure.[26,27] Principal component analysis produces a set of linear combinations of the original variables and typically the first combination is the composite index extracted, having the largest amount of information common to all the variables. The creation of this composite index results in the computation of a SEP score.[27,28]

The diagnosis of TB infection has also recently been enhanced with an *in-vitro* interferon-gamma (IFN)-γ release assay that is unaffected by BCG vaccination and environmental mycobacteria, allowing more accurate assessment of TB infection.[29,30]

The aim of this study was to use these improved methods of measuring SEP and TB infection to investigate the association between SEP and risk of TB infection in Zambia, which has one of the highest tuberculosis incidences in the world.[31]

METHODS

Study design. Between June 2005 and March 2006 a population-based HIV-tuberculosis prevalence survey was conducted among all residents over 15 years of age from two Zambian communities: one rural (~13.000 inhabitants) and one urban (~11.000 inhabitants). Both areas are located in the Lusaka province, where ~40% of the population live in extreme poverty.[32] The prevalence of TB was estimated to be 650/100,000 (95% confidence interval [CI] 360–940/100,000) in the rural and 1200/100,000 (95% CI 750–1640/100,000) in the urban area (Ayles H and others, unpublished data).

*Address correspondence to Delia Boccia, Infectious Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom. E-mail: delia.boccia@lshtm.ac.uk

We conducted a case control study nested in the previous population-based prevalence survey, recruiting 106 prevalent cases and 318 controls aimed to assess the effect of household SEP on the risk of prevalent tuberculosis. A case was defined as any person enrolled in the prevalence survey from which *Mycobacterium tuberculosis* was grown in at least one sputum culture. A control was defined as any person whose sputum culture was negative for *M. tuberculosis*. Controls were frequency matched to the cases by age group and area of residence (i.e., urban versus rural). For the selection of these general-population controls, all individuals enumerated in the prevalence survey have been first stratified by area and age groups and then randomly selected from each category according to the age group and area distribution of the cases.

In order to assess the effect of household SEP on the risk of TB infection we adopted an unusual, but *opportunistic* design: we restricted the analysis only to these 318 controls and considered them as an age and rural/urban stratified random cross-sectional sample of the population. Each subject was classified as tuberculosis infected or non-tuberculosis infected and these groups were compared in terms of their socioeconomic characteristics.

Because this analysis was a secondary analysis, restricted only to these controls available from the case control study, formal *a priori* power calculations were not made in respect of the exploratory hypothesis we investigate here.

Informed written consent was requested for study participation. Ethical approval was obtained both from the University of Zambia and the London School of Hygiene and Tropical Medicine research ethics committee.

Infection status assessment. Infection status was assessed using Quantiferon®—TB Gold (In Tube), (Cellestis; Carnagie, Australia), an *in vitro* laboratory test using a whole blood specimen for the diagnosis *M. tuberculosis* complex infection. The test is based on the measurement of INF-γ released by sensitized T cells after stimulation with tuberculosis antigens.[30] Blood samples were collected, stored, tested, and results interpreted according to the instructions (Cellestis; Carnegie, Australia).

Because of health and safety considerations, blood samples were collected in two clinics in the urban and rural area. To better monitor participation, consenting people were given an appointment for blood testing. Each participant giving consent, but not coming to the clinic on the agreed date, was visited three times before being excluded from blood collection.

Prevalence of infection was defined as the number of Quantiferon positives divided by the total number of participants with interpretable results. Individuals having indeterminate Quantiferon results were excluded from the analysis.

Definition of socioeconomic position. SEP was defined only at household level. For the construction of the SEP index we took into account four different dimensions or aspects relevant to the definition of SEP in this context: 1) human resources, 2) food availability and vulnerability, 3) assets ownership and housing quality, 4) access to community infrastructures and facilities (Figure 1). This latter domain describes features of the area and neighborhood more than household characteristics.

Analysis was conducted combining households from both areas. Variables from across all four domains were included in a principal component analysis[26] (PCA) and screened out accord-

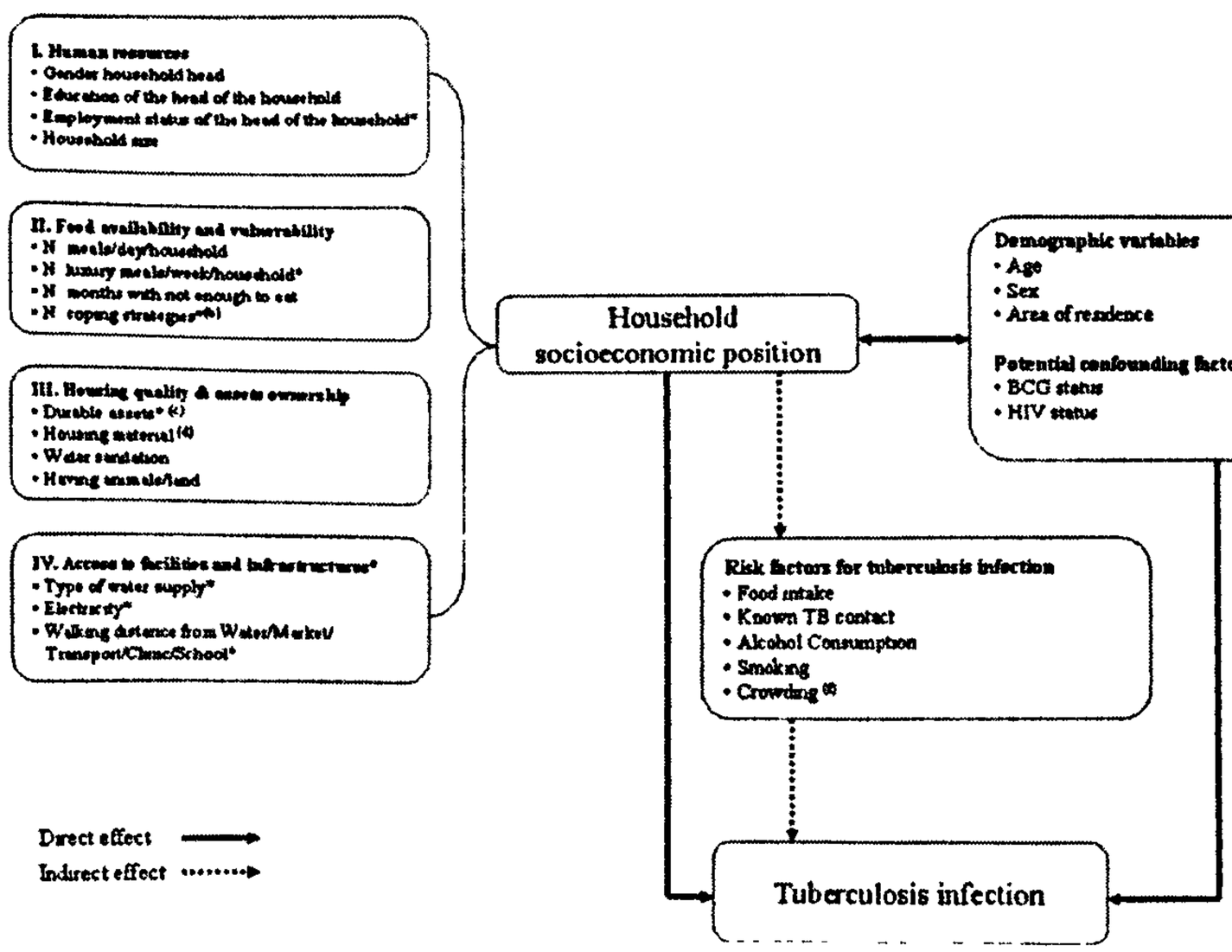


FIGURE 1. Conceptual framework showing the hypothesized relationship between SEP[a] and tuberculosis infection. * Variables included in the first principal component extracted (i.e., the SEP composite index). [a] For each domain of SEP considered, only some of the SEP proxies characterized in the questionnaire are included in the figure. [b] Coping strategies were defined as the number of activities implemented in time of hardship during the 12 months prior to the study. [c] Durable assets included radio, TV, bicycle, fridge, car, or motorbike. [d] Housing material included roof, floor, and wall quality. [e] Variables in this domain are more area/neighborhood features rather than household characteristics. [f] Crowding was defined according to the number of people sharing the same sleeping room. Three levels of crowding were identified: less crowded (< 3 people sharing the same sleeping room), crowded (3–4 people sharing the same sleeping room), more crowded (> 4 people sharing the same sleeping room). The cutoffs were decided according to the average number of people per sleeping room in the sample (i.e., ~3 people per sleeping room).

ing to their loading factor (i.e., strength of correlation with each other). A total of 21 variables were considered for inclusion, of which 11 were included into PCA. The principal component so generated was used to assign to each household a SEP score in relation to all other households in the sample. Finally, households were grouped into SEP groups: very poor, poor, and less poor. To convert the SEP scoring into categories made the results more easily interpretable, because a unit increase of the SEP score does not correspond to any meaningful level of SEP.

Data collection and data analysis. Data was collected during visits to respondents' homes and used two different structured questionnaires: one assessed individual socio-demographic characteristics (i.e., age, gender, education, occupation) and known risk factors for TB infection and disease (i.e., HIV status, contact with TB cases, smoking, alcohol, migration, past disease history); the second questionnaire assessed the household SEP, including the variables discussed previously.

Data were collected over 12 months (March 2006–March 2007), double entered and checked using Epi-Info Software, and analyzed with Stata Software (Version 9; Stata Corporation, College Station, TX). Odds ratios (OR) with 95% Confidence Intervals (CI) for determinants of infection were assessed through univariable and multivariable logistic regression analysis. The likelihood ratio test was used to assess the overall significance of risk factors, test for trend, and test for interaction.

Figure 1 outlines the conceptual framework we used to guide our investigation of the possible association between SEP (and each composing domain) and TB infection. Two pathways have been hypothesized: one affecting the risk of TB infection directly; and the other one indirectly through the effect of mediating factors (i.e., more proximal risk factors) that are on the causal pathway leading from SEP to TB infection.[33] In researching the mediating pathway, we seek to explore the extent to which variation of TB infection risk across SEP terciles is explained by differential exposure to these risk factors across SEP groups. The list of mediating factors explored is illustrated in Figure 1.

The association between SEP and TB infection was explored looking both at the composite index and the individual SEP proxies from which the composite index was derived. This second approach allowed us to explore whether any of the domains were more important in explaining the association between SEP and TB infection.

The association between SEP and TB infection was first adjusted for BCG vaccination and HIV status, considered to be potential confounding factors because both are associated with SEP and the likelihood of TB infection and Quantiferon positivity. The mediation effect was assessed through the construction of a multivariable model containing SEP, the outcome of interest and, one by one, each of the risk factors explored. A reduction in the OR for SEP associated with TB infection upon inclusion of any of these variables was considered suggestive of mediation. Each model was always adjusted for age, gender, and area of residence. The same approach was used for each of the SEP proxies investigated.

RESULTS

Population sampled. Consent to blood testing was given by 265 out of the 318 participants (83.3%) originally enrolled as controls in the case control study. Blood samples were collected from 188 of these subjects (Figure 2) and Quantiferon analysis performed on 174 of the 188 (92.4%) available samples. The remaining samples ($N = 14$) were not tested because of insufficient blood collection or because they were not incubated within the required time.

Eighty-three out of 174 participants (47.7%) whose samples were available were Quantiferon positive and therefore classified as TB infected. Seven participants were classified as indeterminate (4.0%) and excluded from further analysis. The difference in the estimated proportion of infection between areas (45.9% and 51.9%, respectively, for the rural and urban area) was not statistically significant ($P = 0.5$).

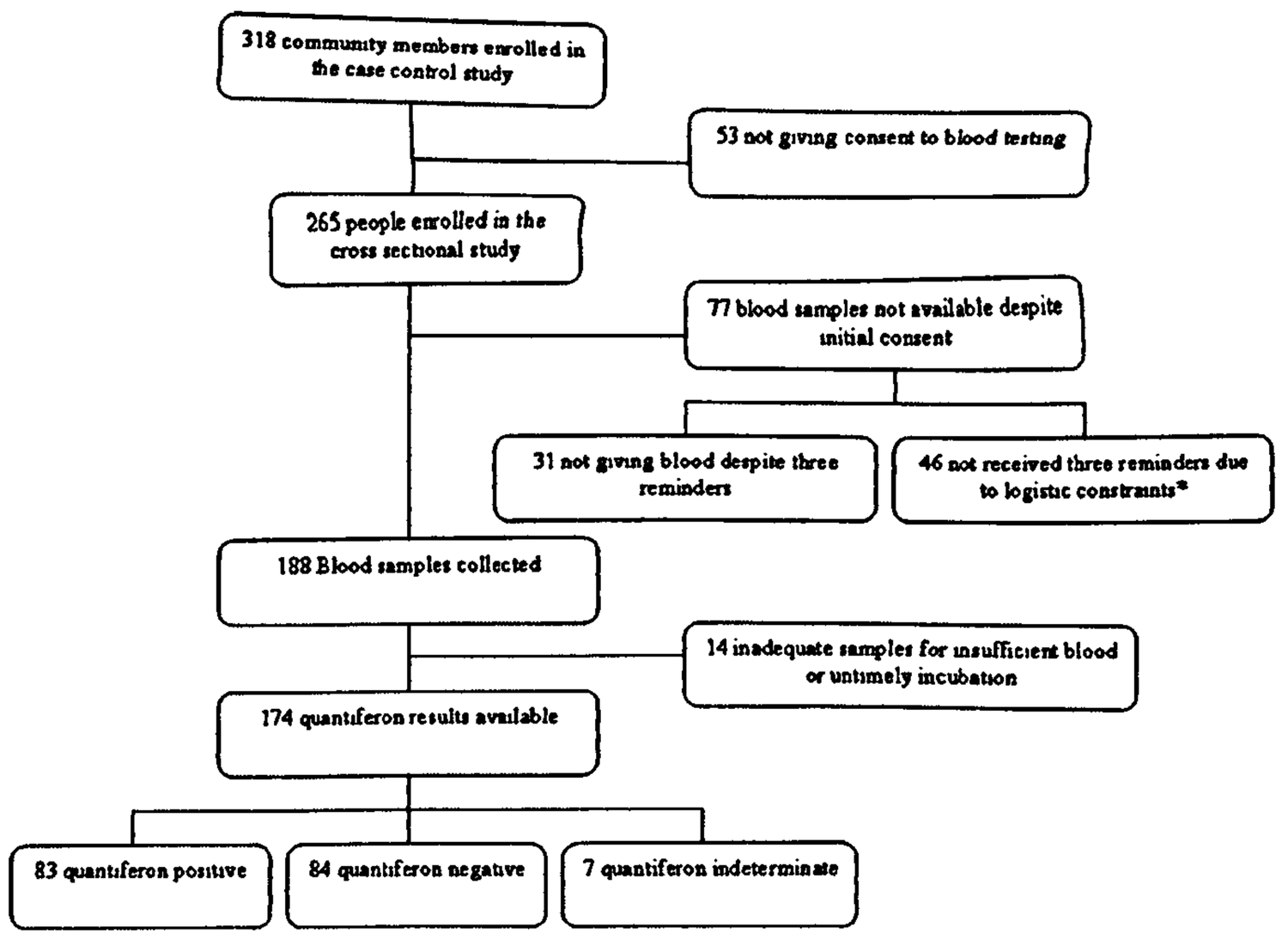


FIGURE 2. Flow diagram of study participation. * Bad weather conditions and poor road access.

Compared with individuals for whom blood samples were unavailable (N = 144), subjects included in the final analysis (N = 174) were more likely to be literate (OR = 2.9, 95% CI: 1.5–5.6, $P < 0.001$), to come from the urban area (OR = 5.6, 95% CI: 3.4–9.2, $P < 0.001$), and to be less poor (OR = 2.0, 95% CI: 1.2–3.6 and OR = 4.4, 95% CI: 2.5–7.9, $P < 0.001$, respectively, for the poor and less poor compared with those classified as more poor); However, after these variables were included in a logistic regression model together with gender and age, only coming from the urban area was independently associated with an increase odds of having given blood (OR = 3.8, 95% CI: 2.2–6.7, $P < 0.001$).

Risk factors for TB infection. In the univariable analysis relatively wealthier household SEP was associated with higher prevalence of TB infection (Figure 3A, Table 1).

Infection status was not associated with any other sociodemographic variables considered, including gender (P = 0.6) and age (P = 0.9). Among the known risk factors explored, only people having more than two meals containing proteins per week were more likely to be TB infected (P = 0.04) (Table 1), and there was a weak association of TB infection with increased crowding (P = 0.09).

There was little evidence that the association between TB infection and SEP was confounded by any of the confounding factors considered (Model 2, Table 2). There was also little evidence that the hypothesized mediating variables explained any part of the association between TB infection and SEP (as suggested by the unchanged value of the OR of SEP in the multivariable models, including SEP, age, gender, area of residence, and each of the potential mediators considered) (Model 3, Table 2). Crowding showed no mediation effect, but when included in the same model, became significantly associated with TB infection independently from SEP (OR = 1.8, 95% CI: 0.9–3.9 and OR = 3.0, 95% CI: 1.2–7.4 for crowded and very crowded households, P = 0.01, compared with less crowded ones) (Table 2). When the relationship between crowding and SEP was explored further, data showed that prevalence of TB infection increased with SEP across all levels of crowding (Figure 3B).

Association between socioeconomic position proxies and TB infection. We also explored the association between each of the SEP proxies and TB infection after adjustment for age, gender, and area of residence. Crowding was also included in each model as the only other variable significantly associated with TB infection. Because of the high number of variables considered in the analysis, Table 3 presents the results only for SEP variables associated with TB infection at a value of $P \leq 0.2$ in the univariable analysis. Of the four domains considered, variables from the domains "food availability and vulnerability" (e.g., weekly number of meals containing proteins and number of coping strategies) and "access to infrastructures" (e.g., having electricity, walking distance from the market) were significantly associated with TB infection. By contrast, variables considered as indicators of "human resources" and "housing quality" were not generally associated with TB infection. Consistent with the analysis based on the SEP composite index, prevalence of TB infection tended to be higher in wealthier categories for each of the SEP proxy analyzed, and for several individual SEP proxies this association between relative wealth and TB infection reached statistical significance.

DISCUSSION

This study describes the association between SEP and TB infection using a quantitative, validated method for measuring SEP and a more specific test for TB infection (Quantiferon) in a general population sample.

Despite TB being historically associated with deprivation, in this study TB infection was more frequent among individuals from households with relatively higher SEP. This observation was confirmed when looking at single SEP covariates: TB infection was consistently associated with markers of higher SEP.

In this study the risk factors traditionally researched in tuberculosis epidemiology were not significantly associated with TB infection. It was therefore not surprising that we found little evidence that these variables mediated the association between SEP and TB infection. It is unclear why these factors were not associated with risk of TB in our study. A reason could be that in this study we were able to detect that part of TB transmission occurring among relatively wealthier people and therefore explained by risk factors associated with higher SEP.

Although these risk factors have not been yet identified in this study, a possible interpretation of our results comes from the analysis of indicators of SEP grouped into different conceptual domains: we found that the strongest association was observed for SEP proxies linked to "food security and vulnerability" and "access to infrastructures and facilities," such as

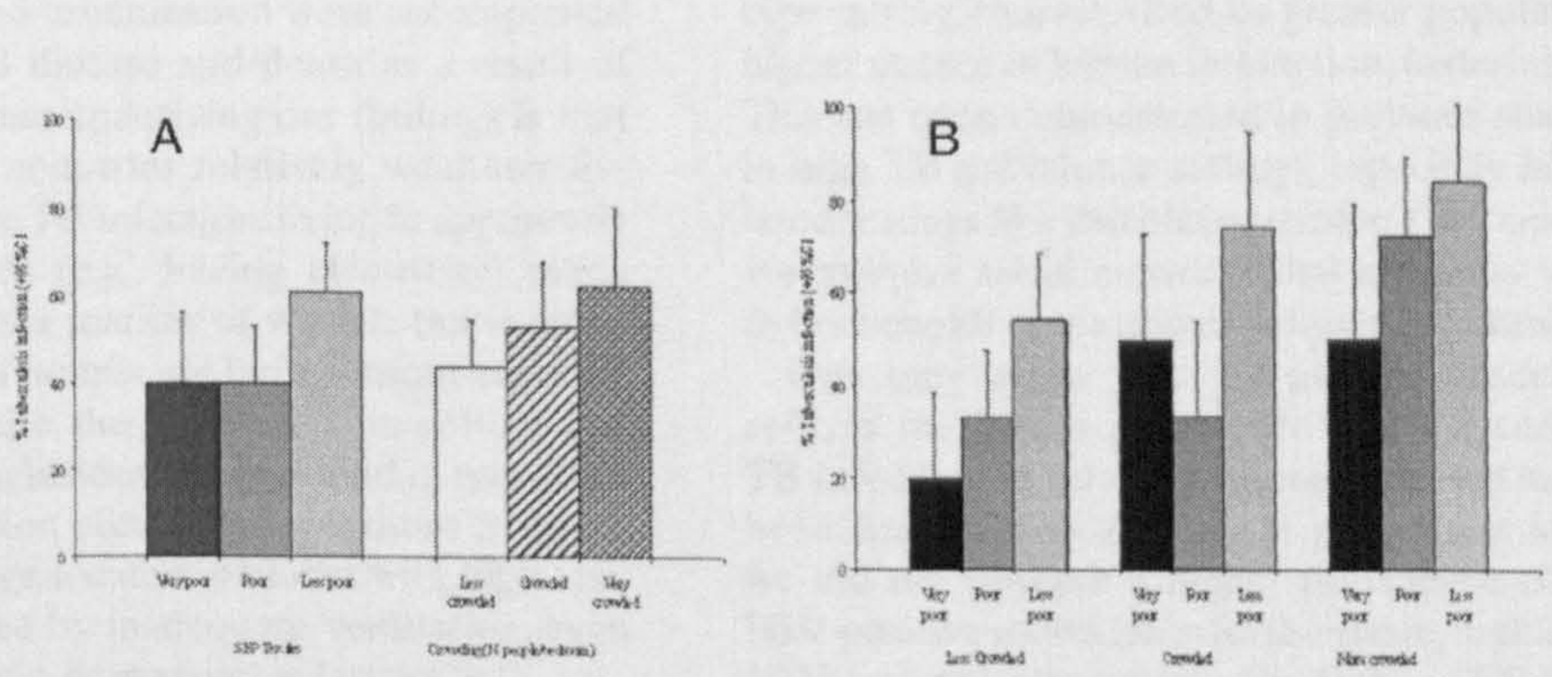


FIGURE 3. Associations between tuberculosis (TB) infection, socioeconomic position (SEP), and crowding. (**A**) Associations shown separately. (**B**) Association between TB (and SEP by level of crowding).

APPENDIX F

TABLE 1

Risk factors for tuberculosis (TB) infection: univariable analysis (N. 84 Quantiferon negative; N. 83 Quantiferon positive)*

Exposures	Quantiferon positives n/N (%)	Unadjusted odds ratio (95% CI)	P value
SEP index			
Less poor	47/77 (61.0)	1.0	0.02
Poor	21/53 (39.6)	0.4 (0.2–0.8)	
Very poor	15/37 (40.5)	0.4 (0.2–0.9)	
Gender			
Male	36/76 (47.4)	1.0	0.6
Female	47/91 (50.0)	1.2 (0.6–2.2)	
Age group (years)			
15–29	29/58 (50.0)	1.0	0.9
30–44	37/76 (48.7)	0.9 (0.5–1.9)	
≥ 45	17/33 (51.5)	1.1 (0.4–2.5)	
Area of residence			
Rural	38/61 (45.9)	1.0	0.5
Urban	55/106 (51.9)	1.3 (0.7–2.4)	
BCG vaccination			
Yes	78/160 (48.7)	1.0	0.6
No	3/5 (60.0)	1.6 (0.2–9.7)	
HIV status			
Positive	24/49 (49.0)	1.0	0.9
Negative	59/118 (50.0)	1.0 (0.5–1.9)	
Alcohol consumption†			
No	58/118 (49.1)	1.0	0.8
Yes	25/49 (51.0)	1.1 (0.6–2.1)	
Smoking‡			
No	72/144 (50.0)	1.0	0.8
Yes	11/23 (47.8)	0.9 (0.4–2.2)	
Known contact with TB			
No	59/120 (49.2)	1.0	0.9
Yes	16/32 (50.0)	1.0 (0.5–2.2)	
Weekly number of meals containing proteins§			
0	14/40 (35.0)	1.0	0.04
1	25/46 (54.3)	2.2 (0.9–5.3)	
2	17/37 (46.0)	1.6 (0.6–3.9)	
> 2	27/44 (61.4)	3.0 (1.2–7.2)	
Crowding			
Less crowded	37/85 (45.5)	1.0	0.09¶
Crowded	35/47 (53.2)	1.5 (0.7–3.0)	
More crowded	21/35 (60.0)	1.9 (0.9–4.3)	

* CI = confidence interval; SEP = socioeconomic position; BCG = bacilli Calmette-Guérin; HIV = human immunodeficiency virus.
† Drinking more than 3 drinks containing alcohol per day.
‡ Including current or past smoking.
§ Weekly consumption of meals containing protein as reported by the respondent.
¶ Test for trend.

TABLE 2

Testing the conceptual framework: Association between tuberculosis infection and socioeconomic position in multivariable models accounting for confounding and mediating effect of risk factors for tuberculosis infection (N. 84 Quantiferon negative and N. 83 Quantiferon positive)

		Adjusted odds ratio (95% CI)	P value
Model 1*			
SEP index	Less poor	0.4 (0.1–1.1)	0.01
	Poor	0.4 (0.2–0.8)	
	Very poor	1.0	
Model 2†			
SEP index	Less poor	0.3 (0.1–1.0)	0.01
	Poor	0.4 (0.2–0.8)	
	Very poor	1.0	
Model 3‡			
SEP index	Less poor	0.4 (0.1–1.0)	0.02
	Poor	0.4 (0.2–0.8)	
	Very poor	1.0	
Alcohol consumption	No	1.0	0.6
	Yes	1.2 (0.6–2.4)	
SEP index	Less poor	0.4 (0.1–1.0)	0.02
	Poor	0.4 (0.2–0.8)	
	Very poor	1.0	
Smoking	No	1.0	0.7
	Yes	1.2 (0.4–3.4)	
SEP index	Less poor	0.2 (0.08–0.9)	0.01
	Poor	0.4 (0.2–0.9)	
	Very poor	1.0	
Known contact with TB	No	1.0	0.5
	Yes	1.4 (05–3.3)	
SEP index	Less poor	0.5 (0.2–1.5)	0.09
	Poor	0.4 (0.2–0.9)	
	Very poor	1.0	
Weekly number of meals containing proteins	0	1.0	0.2
	1	2.0 (0.8–4.9)	
	2	1.2 (0.4–3.3)	
	> 2	2.1 (0.8–5.9)	
SEP index	Less poor	0.3 (0.09–0.8)	0.005
	Poor	0.4 (0.2–0.8)	
	Very poor	1.0	
Crowding	Less crowded	1.0	0.01§
	Crowded	1.9 (0.9–4.0)	
	More crowded	3.0 (1.2–7.4)	

* Model 1: Socioeconomic position (SEP) adjusted for demographic variables (i.e., gender, age, and area of residence)
† Model 2 SEP adjusted for potential confounding factors (i.e., bacilli Calmette-Guérin (BCG) vaccination status, human immunodeficiency virus (HIV) infection status)
‡ Model 3 SEP adjusted demographic variables and each potential mediator. A reduction in the odds ratio (OR) of SEP upon inclusion of each previous variable is suggestive of mediation.
§ Test for trend

having electricity and living closer to infrastructures. These last two variables describe characteristics of the local context rather than of households or individuals.

Historically, the industrial revolution in the 19th Century in Europe and the associated urbanization were accompanied by an initial increase in TB disease and death as a result of TB.[34–38] One possible dynamic underlying our findings is that in modern-day developing countries relatively wealthier living conditions could increase TB infection: living in apparently better equipped households (e.g., having electricity) made with more solid materials is a marker of wealth, but it could also mean poor air quality if houses are built without ensuring adequate ventilation. Despite the differences in setting and study design, analogous conclusions can be found in two other studies where TB transmission occurrence was more frequent in newer apartment buildings, located in blocks with high residential density, characterized by inadequate ventilation, even after adjustment for key socio-demographic factors.[39,40]

Similar interpretation can be given to the findings on the distance from main infrastructures: households closer to facilities are likely to be wealthier, but also likely to reflect a more urban-type setting, characterized by greater population density and a higher chance of human interaction, fostering TB transmission. This has been demonstrated in previous studies, showing that in high TB prevalence settings, especially high densely populated settings like this one, extensive TB transmission can occur via complex social networks that are likely to be as important as households contact in maintaining transmission.[41,42]

One may argue that the pattern observed in this study reflects the higher prevalence of HIV, and thus perhaps of TB infection, in relatively more educated individuals that has been described in Zambia in previous studies.[43–45] However, we did not observe a higher prevalence of infection among HIV-positive individuals; furthermore, in this study population HIV was not associated with higher SEP (data not shown). Our study also confirmed the importance of crowding in the epidemiology of tuberculosis.

TABLE 3

Association between tuberculosis (TB) infection and proxies* of socioeconomic position (N. 84 Quantiferon negative and N. 83 Quantiferon positive)

Exposures		Quantiferon positives n/N (%)	Unadjusted Odds ratio (95% CI)	*P* value	Adjusted Odds ratio (95% CI)†	*P* value
Human resources						
Occupation of the head of the household	Self used	33/62 (53.2)	2.3 (0.6–8.3)	0.2	4.1 (0.9–18.7)	0.07
	Used	46/93 (49.5)	1.9 (0.6–6.9)		3.9 (0.6–17.0)	
	Unemployed	4/12 (3.3)	1.0		1.0	
Food availability and vulnerability						
Weekly number of meals containing proteins†	0	12/34 (35.3)	1.0	0.2	1.0	0.04§
	1	23/47 (48.9)	1.7 (0.7–4.3)		1.9 (0.7–5.1)	
	2	23/40 (57.5)	2.5 (1.0–6.3)		3.2 (1.1–9.0)	
	3	25/46 (54.3)	2.2 (0.9–5.4)		2.9 (1.0–8.1)	
Daily number of meals in the household	≤ 2	21/49 (42.9)	1.0	0.2	1.0	0.2
	> 2	62/118 (52.5)	1.5 (0.7–2.9)		1.5 (0.7 3.1)	
Number of coping strategies†	> 3	21/51 (41.2)	1.0	0.07	1.0	0.03§
	1–3	25/55 (45.4)	1.2 (0.5–2.6)		1.3 (0.6–3.0)	
	0	37/61 (60.7)	2.2 (1.0–4.7)		2.5 (1.1–5.9)	
Assets ownership and housing quality						
Having animals	Yes	19/48 (39.6)	1.0	0.09	1.0	0.1
	No	64/119(53.8)	1.8 (0.9–3.5)		0.5 (0.2–1.2)	
Floor material†	Dirt/earth	28/64 (43.7)	1.0	0.2	1.0	0.2
	Cement	55/103 (53.4)	1.5 (0.8–2.7)		1.7 (0.8–3.5)	
Roof material†	Grass	10/20 (50.0)	1.0	0.08	1.0	0.1
	Iron sheet	30/74 (40.5)	0.7 (0.2–1.8)		0.7 (0.2–2.2)	
	Asbestos	43/73 (58.9)	1.4 (0.5–3.9)		1.6 (0.5–5.3)	
Type of water supply†	Piped outside house	52/109 (47.7)	1.0	0.1	1.0	0.7
	Piped inside house	30/57 (52.6)	1.2 (0.6–2.3)		1.1 (0.6–2.4)	
Type of water sanitation	Latrines	67/143 (46.8)	1.0	0.07	1.0	0.08
	Flush toilets/other	16/24 (66.7)	2.3 (0.9–5.6)		2.3 (0.9–5.9)	
Access to infrastructures and facilities						
Having electricity†	No	49/112 (43.7)	1.0	0.03	1.0	0.03
	Yes	34/55 (61.8)	2.1 (1.1–4.0)		2.2 (1.1–4.6)	
Walking distance from the market (minutes)†	> 15	31/80 (38.7)	1.0	0.07	1.0	< 0.001
	0–15	52/87 (60.0)	2.3 (1.2–4.6)		2.7 (1.3–5.6)	

*SEP proxies are grouped by dimension considered in the conceptual framework.
†Variables included in the principal component analysis.
‡Multivariable analysis including gender, age group, area of residence, and crowding.
§Test for trend.

As in other studies,[46–48] our data suggest a strong independent effect of both crowding and SEP on the risk of TB infection. Crowding and SEP perhaps represent two key forces, acting in different contexts and explaining different cases of TB infection: 1) those infected at household level through overcrowding in poor households, and 2) those infected at community level because of dynamics reflecting higher SEP described previously. Thus, there may be aspects of SEP that act over and above the role played by household overcrowding in fostering TB transmission.

The study had potential limitations. The small sample size, imposed by the nested design in the case control study, limited the power of the study. We therefore advise caution in the interpretation of our findings. However, because low study power most seriously increases the potential for false negative conclusions in studies, it does not provide a highly plausible explanation for the unusual and interesting pattern of infection we report here.

The people who participated in the study were more likely to come from the urban area compared with the rural one. This raises issues on the generalizability of these findings and, most importantly, about the possibility of selection biases (as people living in urban areas are generally wealthier compared with the residents in urban ones). Nonetheless, once the effect of the area of residence was taken into account, we found no significant difference in terms of demographics, SEP, any risk factors increasing the likelihood of TB infection (i.e., known contact with tuberculosis cases), or immune system impairment and therefore Quantiferon response (i.e., HIV status, food intake, alcohol consumption) between people who gave blood and those who did not.

It has been suggested that malnutrition may reduce the expression of INF-γ and other mycobactericidal substances.[49,50] If this were to be the case, our study may have missed infected individuals among the poor (i.e., false negatives) and may have been more likely to detect TB infection among better nourished, wealthier individuals potentially biasing our results. However, a better immunologic response probably can only help in providing an interpretable result, but does not necessarily translate into a positive Quantiferon result. Most importantly, we have shown that almost all markers of high SEP were associated with higher risk for TB infection, suggesting that the false negative bias is unlikely to be relevant.

Finally, the importance of previous contact with a TB case in the household (considered to be one of the most important risk factor for TB infection),[51–56] could have been underestimated as our data were based only on respondent recall.

In conclusion, our findings suggest that unexpectedly, higher SEP, rather than lower, was associated with TB infection in two Zambian communities.

None of the traditional risk factor variables that we explored appeared to mediate this association, suggesting that in this setting TB transmission may occur through the exposure to risk factors for TB other than those traditionally explored and associated with poverty, such as alcohol consumption, inadequate nutrition, smoking, and contact with a TB case. It is possible that, especially in urban settings, higher SEP is associated with housing characteristics that reduce ventilation and life-styles that increase social mixing and therefore the likelihood of contact between cases and susceptible people.

Further studies are needed to test these hypotheses and to verify whether even increasing the household SEP may not be sufficient to reduce the risk of TB infection, if it is not accompanied by the improvement of community living conditions (including better housing quality, crowding reduction, and access to public services) and a strong TB control program.

Received October 3, 2008. Accepted for publication February 3, 2009.

Authors' addresses: Delia Boccia, Infectious Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1 E 7HT, United Kingdom, Tel: +44(0)20-7958-8145, Fax: +44(0)20-7612-7860, E-mail: delia.boccia@lshtm.ac.uk. James Hargreaves, Infectious Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, Tel: +44(0)20-7927-2955, Fax: +44(0)20-7612-7860, E-mail: james.hargreaves@lshtm.ac.uk. Helen Ayles, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, ZAMBART Project, Ridgeway Campus, University of Zambia, Lusaka, Zambia, Tel: +260-1-254710, Fax: +260-1-254710, E-mail: helen@zambart.org.zm. Katherine Fielding, Infectious Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, Tel: +44(0)20-7927-2889, Fax: +44(0)20-7636-8739, E-mail: katherine.fielding@lshtm.ac.uk. Musonda Simwinga, ZAMBART Project, Ridgeway Campus, University of Zambia, Lusaka, Zambia, Tel: +260-1-254710, Fax: +260-1-254710, E-mail: musonda@zambart.org.zm. Peter Godfrey-Faussett, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, Tel: +44(0)20-7958-8127, Fax: +44(0)20-7612-7860, E-mail: peter.godfrey-faussett@lshtm.ac.uk.

REFERENCES

1. Dubos RJ, 1987. *The White Plague: Tuberculosis, Man, and Society.* Piscataway, NJ: Rutgers University Press.
2. Bank W, 2000. *World Development Indicators.* Washington, DC: The World Bank.
3. WHO, 2005. *Addressing Poverty in TB Control: Options for National TB Control Programmes.* Geneva: WHO.
4. Cantwell MF, McKenna MT, McCray E, Onorato IM, 1998. Tuberculosis and race/ethnicity in the United States: impact of socioeconomic status. *Am J Respir Crit Care Med 157:* 1016–1020.
5. CDC, 1992. Prevention and control of tuberculosis in U.S. communities with at-risk minority populations. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR Recomm Rep 41:* 1–11.
6. Hanson C, 2002. *Tuberculosis, Poverty and Inequity: A Review of the Literature and Discussion of Issues.* Washington, DC: World Bank.
7. Kuemmerer JM, Comstock GW, 1967. Sociologic concomitants of tuberculin sensitivity. *Am Rev Respir Dis 96:* 885–892.
8. Spence DP, Hotchkiss J, Williams CS, Davies PD, 1993. Tuberculosis and poverty. *BMJ 307:* 759–761.
9. Bhatti N, Law MR, Morris JK, Halliday R, Moore-Gillon J, 1995. Increasing incidence of tuberculosis in England and Wales: a study of the likely causes. *BMJ 310:* 967–969.
10. Lienhardt C, Fielding K, Sillah J, Tunkara A, Donkor S, Manneh K, Warndorff D, McAdam KP, Bennett S, 2003. Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia. *Am J Respir Crit Care Med 168:* 448–455.
11. Claessens NJ, Gausi FF, Meijnen S, Weismuller MM, Salaniponi FM, Harries AD, 2002. High frequency of tuberculosis in households of index TB patients. *Int J Tuberc Lung Dis 6:* 266–269.
12. Schoeman JH, Westaway MS, Neethling A, 1991. The relationship between socioeconomic factors and pulmonary tuberculosis. *Int J Epidemiol 20:* 435–440.
13. Sanchez-Perez H, Flores-Hernandez J, Jansa J, Cayla J, Martin-Mateo M, 2001. Pulmonary tuberculosis and associated factors in areas of high levels of poverty in Chiapas, Mexico. *Int J Epidemiol 30:* 386–393.
14. Glynn JR, Warndorff DK, Malema SS, Mwinuka V, Ponnighaus JM, Crampin AC, Fine PE, 2000. Tuberculosis: associations with HIV and socioeconomic status in rural Malawi. *Trans R Soc Trop Med Hyg 94:* 500–503.
15. Tekkel M, Rahu M, Loit HM, Baburin A, 2002. Risk factors for pulmonary tuberculosis in Estonia. *Int J Tuberc Lung Dis 6:* 887–894.
16. Coker R, McKee M, Atun R, Dimitrova B, Dodonova E, Kuznetsov S, Drobniewski F, 2006. Risk factors for pulmonary tuberculosis in Russia: case-control study. *BMJ 232:* 85–87.
17. Mangtani P, Jolley DJ, Watson JM, Rodrigues LC, 1995. Socioeconomic deprivation and notification rates for tuberculosis in London during 1982–91. *BMJ 310:* 963–966.
18. Farhat M, Greenaway C, Pai M, Menzies D, 2006. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? *Int J Tuberc Lung Dis 10:* 1192–1204.
19. Villate JI, Ibanez B, Cabriada V, Pijoan JI, Taboada J, Urkaregi A, 2006. Analysis of latent tuberculosis and *Mycobacterium avium* infection data using mixture models. *BMC Public Health 6:* 240.
20. Kerr-Pontes LR, Barreto ML, Evangelista CM, Rodrigues LC, Heukelbach J, Feldmeier H, 2006. Socioeconomic, environmental, and behavioural risk factors for leprosy in North-east Brazil: results of a case-control study. *Int J Epidemiol 35:* 994–1000.
21. Kerr-Pontes LR, Montenegro AC, Barreto ML, Werneck GL, Feldmeier H, 2004. Inequality and leprosy in Northeast Brazil: an ecological study. *Int J Epidemiol 33:* 262–269.
22. Raghunathan PL, Whitney EA, Asamoa K, Stienstra Y, Taylor TH Jr, Amofah GK, Ofori-Adjei D, Dobos K, Guarner J, Martin S, Pathak S, Klutse E, Etuaful S, van der Graaf WT, van der Werf TS, King CH, Tappero JW, Ashford DA, 2005. Risk factors for *Buruli* ulcer disease (*Mycobacterium ulcerans* infection): results from a case-control study in Ghana. *Clin Infect Dis 40:* 1445–1453.
23. Engel A, Roberts J, 1977. Tuberculin skin test reaction among adults 25–74 years, United States, 1971–72. *Vital Health Stat 11 204:* 1–40.
24. Reichman LB, O'Day R, 1978. Tuberculous infection in a large urban population. *Am Rev Respir Dis 117:* 705–712.
25. Pelly TF, Santillan CF, Gilman RH, Cabrera LZ, Garcia E, Vidal C, Zimic MJ, Moore DA, Evans CA, 2005. Tuberculosis skin testing, anergy and protein malnutrition in Peru. *Int J Tuberc Lung Dis 9:* 977–984.
26. Filmer D, Pritchett LH, 2001. Estimating wealth effects without expenditure data–or tears: an application to educational enrollments in states of India. *Demography 38:* 115–132.
27. Bartholomew DJSF, Moustaki I, Galbraith JI, 2002. *Principal Component Analysis: The Analysis and Interpretation of Multivariate Data for Social Scientists.* Chapman & Hall/CRC, 115–142.
28. CGAP, 2002. *Assessing the Relative Poverty of Microfinance Clients: A CGAP Operational Tool.* Washington, DC: CGAP.
29. Lalvani A, 2007. Diagnosing tuberculosis infection in the 21st century: new tools to tackle an old enemy. *Chest 131:* 1898–1906.

30. Pai M, Kalantri S, Dheda K, 2006. New tools and emerging technologies for the diagnosis of tuberculosis: part I. Latent tuberculosis. *Expert Rev Mol Diagn 6:* 413–422.
31. World Health Organization, 2008. Global tuberculosis control: surveillance, planning, financing. WHO Report 2008. Geneva, Switzerland.
32. LCMS, November 2004. Living Conditions Monitoring Survey Report, 2002–2003. Lusaka, Zambia: Central Statistical Office, Republic of Zambia.
33. Victora CG, Huttly SR, Fuchs SC, Olinto MT, 1997. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol 26:* 224–227.
34. Kass EH, 1971. Infectious diseases and social change. *J Infect Dis 123:* 110–114.
35. Springett VH, 1952. An interpretation of statistical trends in tuberculosis. *Lancet 1:* 575–580.
36. Hunter JM, Thomas MO, 1984. Hypothesis of leprosy, tuberculosis and urbanization in Africa. *Soc Sci Med 19:* 27–57.
37. Enarson DA, Wang JS, Dirks JM, 1989. The incidence of active tuberculosis in a large urban area. *Am J Epidemiol 129:* 1268–1276.
38. Drucker E, Alcabes P, Bosworth W, Sckell B, 1994. Childhood tuberculosis in the Bronx, New York. *Lancet 343:* 1482–1485.
39. Acevedo-Garcia D, 2001. Zip code-level risk factors for tuberculosis: neighborhood environment and residential segregation in New Jersey, 1985–1992. *Am J Public Health 91:* 734–741.
40. Wanyeki I, Olson S, Brassard P, Menzies D, Ross N, Behr M, Schwartzman K, 2006. Dwellings, crowding, and tuberculosis in Montreal. *Soc Sci Med 63:* 501–511.
41. Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, Musser JM, 2001. Networks and tuberculosis: an undetected community outbreak involving public places. *Soc Sci Med 52:* 681–694.
42. McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, Lambert LA, Ridzon R, 2003. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. *Int J Tuberc Lung Dis 7:* S486–S493.
43. Michelo C, Sandoy IF, Dzekedzeke K, Siziya S, Fylkesnes K, 2006. Steep HIV prevalence declines among young people in selected Zambian communities: population-based observations (1995–2003). *BMC Public Health 6:* 279.
44. Fylkesnes K, Musonda RM, Kasumba K, Ndhlovu Z, Mluanda F, Kaetano L, Chipaila CC, 1997. The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women. *AIDS 11:* 339–345.
45. Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F, Monze M, 2001. Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. *AIDS 15:* 907–916.
46. Clark M, Riben P, Nowgesic E, 2002. The association of housing density, isolation and tuberculosis in Canadian First Nations communities. *Int J Epidemiol 31:* 940–945.
47. Barker RD, Millard FJ, Malatsi J, Mkoana L, Ngoatwana T, Agarawal S, de Valliere S, 2006. Traditional healers, treatment delay, performance status and death from TB in rural South Africa. *Int J Tuberc Lung Dis 10:* 670–675.
48. Elender F, Bentham G, Langford I, 1998. Tuberculosis mortality in England and Wales during 1982–1992: its association with poverty, ethnicity and AIDS. *Soc Sci Med 46:* 673–681.
49. van Lettow M, West CE, van der Meer JW, Wieringa FT, Semba RD, 2005. Low plasma selenium concentrations, high plasma human immunodeficiency virus load and high interleukin-6 concentrations are risk factors associated with anemia in adults presenting with pulmonary tuberculosis in Zomba district, Malawi. *Eur J Clin Nutr 59:* 526–532.
50. Macallan DC, 1999. Malnutrition in tuberculosis. *Diagn Microbiol Infect Dis 34:* 153–157.
51. Rathi SK, Akhtar S, Rahbar MH, Azam SI, 2002. Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulnionary tuberculosis cases in Umerkot, Pakistan. *Int J Tuberc Lung Dis 6:* 851–857.
52. Lienhardt C, 2001. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiol Rev 23:* 288–301.
53. Kenyon TA, Creek T, Laserson K, Makhoa M, Chimidza N, Mwasekaga M, Tappero J, Lockman S, Moeti T, Binkin N, 2002. Risk factors for transmission of *Mycobacterium tuberculosis* from HIV-infected tuberculosis patients, Botswana. *Int J Tuberc Lung Dis 6:* 843–850.
54. Tornee S, Kaewkungwal J, Fungladda W, Silachamroon U, Akarasewi P, Sunakorn P, 2004. Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. *Southeast Asian J Trop Med Public Health 35:* 375–383.
55. Singh M, Mynak ML, Kumar L, Mathew JL, Jindal SK, 2005. Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. *Arch Dis Child 90:* 624–628.
56. Sinfield R, Nyirenda M, Haves S, Molyneux EM, Graham SM, 2006. Risk factors for TB infection and disease in young childhood contacts in Malawi. *Ann Trop Paediatr 26:* 205–213.

APPENDIX G

Quotes References

1. Pasteur L, quoted in many websites.

2. Berrigan D, 1971. See Infection and Inequalities. The modern plague. Farmer P. 1999. University of California Press. p. 18.

3. Heinstein A. See http://en.wikiquote.org/wiki/Albert_Einstein (Last accessed September 2009).

4. Nietzsche F. See Introducing Nietzsche. Gane L, Chan K. 1999. Cambridge: Icon books.

5. Morris, JN. Uses of Epidemiology. Read in opening a discussion at the Section of Preventive Medicine and Infectious Diseases at the Annual Meeting of the British Medical Association, Glasgow, 1954, and since expanded. First published BMJ 1955; 2: 395–401. Reprinted with permission in International Journal of Epidemiology, 2007; 36: 1165-72.

6. Harrington, M. The other America. Macmillan publishing company. New York; 1962.

7. Lynch J, Kaplan GA. Socioeconomic position. In: Berkman LF, Kawachi I, editors. Social Epidemiology. New York: Oxford University Press; 2000. p. 29.

8. Dubos RJ. The White Plague: Tuberculosis, Man, and Society. Reissue edition (1 Jan 1987) ed: Rutgers University Press; 1987. p.202.

9. Narayanan D. Voices of the poor (Volume 1): can anyone ear us. Washington: Oxford University Press, Inc; 2000.

10. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol. 2001 Aug; 30(4). p. 674.

11. Gramsci, A. Prisons notebook. Columbia University Press. New York; 2007.