Determinants of child mortality in Angola:

An econometric analysis

by

Ingrid Hoem Sjursen

Master thesis

Master thesis to complete the degree

Master in Economics

(Professional Studies in Economics)

University of Bergen, Department of Economics

December 2011

UNIVERSITY OF BERGEN

Acknowledgements

First and foremost, I want to thank my supervisor, Gaute Torsvik, for invaluable guidance in the process of writing this thesis, for useful comments and advice, and for always being available to answer my endless series of questions. I also thank my second supervisor, Arild Aakvik, for helping me with the econometric methodology and analysis. Without the two of you, I could never have finished this thesis.

To the Christian Michelsen Institute (CMI): thank you for accepting me as an affiliated master student and giving me the opportunity to be a part of an inspiring research environment. In particular, I would like to thank Ottar Mæstad and Mona Frøystad for letting me take part in the Angola health project, for providing useful STATA tips and for taking the time to answer my many questions. Thanks also to Magnus Hatlebakk and Ida Lindkvist for good advice and to the rest of the health cluster for giving me the opportunity to present my project, and for enlightening meetings. I also thank my fellow students at the CMI and at the Department of Economics for providing good company and entertaining breaks.

Furthermore, I would like to thank mamma, Stein Jenssen and Ingeborg Forthun for useful comments on my drafts, and to Eirik Høyer Leivestad, Ragne Rommetveit and Johannes Oldervoll for proofreading. A special thank you goes to Kristine Husøy Onarheim for introducing me to the topic, for enduring countless discussions and questions, and for general support.

Finally, I would like to thank Roman Linneberg Eliassen for his patience and support, both moral and academic.

The data used in the analysis are collected by the Christian Michelsen Institute (CMI) and Centro de Estudos e Investigação Científica (CEIC). All interpretations and errors are my own responsibility.

Ingrid Hoem Sjursen

Ingrid Hoem Sjursen, Bergen 30. November 2011

Abstract

Determinants of child mortality in Angola: an econometric analysis

by

Ingrid Hoem Sjursen

Master in Economics (Professional Studies in Economics)

University of Bergen, 2011

Supervisor: Gaute Torsvik

Second supervisor: Arild Aakvik

The aim of this thesis is to investigate determinants of child mortality in the regions of Luanda and Uíge in Angola. The country has one of the highest child mortality rates in the world. The literature on the subject has found extensive variation in causes of death between and within countries, and knowledge on local conditions is a prerequisite for shaping sound and efficient policies addressing the problem.

The analysis is conducted using data collected jointly by the Christian Michelsen Institute (CMI) and Centro de Estudos e Investigação Científica (CEIC) in 2010. In the descriptive analysis, I find large differences in infant and under-five mortality between wealth quintiles, education levels and households belonging to different public health facilities. In the econometric analysis, OLS, Poisson and binary logit models are estimated using both number of infant and under-five children dead as the dependent variable. The analysis puts particular emphasis on the effect of wealth, education, use and access to health services. While I find no effect of wealth, education, use and access to delivery services are found to be significant determinants of child mortality. All calculations and estimations are conducted using STATA version 11.1.

Contents

Tables

Figures

1 Introduction

> *Improving the health of women and children contributes extensively to economic development, which in turn contributes to better conditions for women and children*
>
> Jens Stoltenberg[1]

The right to health is a fundamental part of the Human Rights declaration of 1948 (United Nations High Commissioner for Human Rights, 2008). Striving for improvements in health is therefore a moral obligation for policy-makers around the world, at the international as well as the national and local level. In addition to being a goal in itself, researchers and the international community are starting to recognize that health plays an important role in other aspects of development. This represents a shift in development strategy; earlier, health has been viewed as an *end* of development, but now the general tenet is that that improvement of health standards is a *means* to achieve other aspects of development.

Millions of people die every year from diseases that could have been easily prevented and treated. The world is in the possession of the medical knowledge and the resources to reduce mortality, and improve on the quality of life of the world population. Failing to do so would be an unacceptable moral failure of the international community and mean opportunities lost for economic and social development. However, health is not only affected by direct biological factors, but is influenced by the economic, the social, the cultural and the institutional environment. More research is needed to acquire knowledge and understanding about how these factors influence health in local settings.

Premature deaths are unequally distributed worldwide, and life expectancy at birth depends heavily on which continent, country and social class one is born into. While a child born in Japan can expect to live as long as 83 years, the life expectancy is just the half of this (44 years) in Afghanistan, the country with the lowest life expectancy in the world (World Bank,

[1] The Global Campaign for the Health Millenium Development Goals 2011. Innovating for Every Woman, Every Child. Thematic Report. Oslo.

2011). Large disparities in life expectancy are also observed across income groups, both between and within countries (Commission on Social Determinants of Health, 2008). According to the World Health Organization (2011c), the probability of a child born in a low-income country dying before the age of five is nearly 18 times higher than for children born in high income countries.

The Commission on Macroeconomics and Health was established by the World Health Organization in year 2000 and given the task of assessing the role of health in global development. They find evidence that improvements in health are essential for economic development and growth. Other researchers have made similar findings. There are many explanations for this. At the macro-level, healthiness of the workforce is a determinant of its productivity and thus GDP. The health condition of the population also affects the GDP through people's ability to work, receive an income and pay taxes. At the micro-level, poor health could be a large financial burden to the households; income is lost if one is unable to work, and out-of-pocket payments must be made for utilization of health services and medicines. Low life expectancy also shifts the allocation of resources from investments with long-term benefits to investments with short-term benefits because it is less attractive to invest in the future if you don't expect to live to enjoy the fruits from them. Failing to make long-term investments could have severe economic consequences for the households.

The fact that health is a part of economic and social development does not imply that economic and social development is not important in improving health. At the macro-level, richer and more developed countries have better means to provide their countries with the necessary health promoting goods and services to keep their populations healthy. At the micro-level, the richer households are more likely to be educated, live in areas with well-developed infrastructure, to have access to high quality health and social services and have the means to buy health promoting goods and services.

The relationship between economic development and improvements in health run in both directions. A special feature of this bidirectional relationship is that it can be mutually reinforcing and form vicious or virtuous circles. In a vicious circle, poverty can contribute to ill health through for example lack of nutrition and access to health services. When health conditions worsen, poverty increases through loss of income. This results in a further

deterioration of health, which reinforces poverty and thus a downwards spiral is set into motion. Oppositely, good health breeds more income and increased possibilities for employment. The higher income leads to further improvements in health, which in turn brings prospects of additional income (Wagstaff, 2002, Bloom et al., 2004). This makes health issues very hard to ignore when shaping sound economic policies: health matters to poverty and economic development and poverty matters to health.

In this thesis, I will investigate determinants of health in the two regions Luanda and Uíge in Angola, and pay particular attention to the effect of wealth. Angola was chosen on the basis that it has some of the worst health outcomes in the world. I focus on two health indicators, infant and under-five mortality, because children's, as well as and women's, health is increasingly recognized as being of special importance in both human and economic development (Ki-moon, 2010). The analysis is based on data consisting of a household survey, a health facility survey and a health worker survey conducted jointly by the Christian Michelsen Institute (CMI) in Bergen and Centro de Estudos e Investigação Cientifica (CEIC) in Luanda, Angola, in 2010.

The rest of the thesis is organized as follows. Chapter 2 provides basic information about the specifics of child health and mortality. Chapter 3 gives a short background on the Angolan context including geography, demography, the economy, health and health system. In chapter 4, a theoretical backdrop is developed. Here, an economic model of fertility, an analytical framework and a short review of the empirical literature will be presented. Chapter 5 describes the data and the variables in the analysis. Econometric models are discussed in chapter 6. A descriptive analysis of some key variables is conducted in chapter 7. Results and analysis of the regression results are presented in chapter 8. Finally, chapter 9 gives a short discussion, and sketches some policy implications and ideas for further research on child mortality.

2 Child health

Children are more vulnerable to all kinds of hazards than adults in the sense that they are dependent on parents or other care takers to provide what they need in order to survive. A newborn's chances of survival are dependent on whether she gets the right and enough nutrition, on whether she is immunized and on the hygienic environment she is born into (Skolnik, 2008). As children are physically and mentally unable to take care of themselves, child health is, as it should be, a special priority to the international community. That children are in fact an international priority is reflected by The Millennium Development Goals (MDG), where three of the goals can be directly[2], and all eight can be indirectly, tied to child health (UN, 2011, Skolnik, 2008). Child health, and child mortality in particular, is not only an important issue in itself, but is commonly regarded as an indicator of the overall health status in a county or a region (Avogo and Agadjanian, 2010). Child mortality is also often used as an indicator of social development (Hill, 1991). Figure 2.1 gives some key definitions of health indicators that will be discussed in this and the following chapters.

Adult mortality rate: probability that a 15 year old person will die before reaching his/her 60th birthday.

Neonatal mortality rate (per 1000 live births): probability of death during the first 28 completed days of life

Postneonatal (per 1000 live births): probability of death before age 1 year conditional on surviving to age1 month

Infant mortality rate (per 1000 live births): probability of a child born in a specific year or period dying before reaching the age of one, if subject to age-specific mortality rates of that period.

Under- five mortality rate (per 1000 live births): probability of a child born in a specific year or period dying before reaching the age of five, if subject to age-specific mortality rates of that period.

Life expectancy at birth: Average number of years that a newborn is expected to live if current mortality rates continue to apply.

Maternal mortality rate (per 100 000 live birth): Annual number of female deaths from any cause related to or aggravated by pregnancy or its management (excluding accidental or incidental causes) during pregnancy and childbirth or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, per 100,000 live births, for a specified year.

Figure 2.1: Definitions, important health indicators (World Health Organization, 2010b, Rajaratnam et al., 2010)

[2] Goal 4: Reduce child mortality rate by two thirds between 1990 and 2015, Goal 5: Reduce maternal mortality by three quarters between 1990 and 2015 and Goal 6: Combat HIV/AIDS, Malaria and other diseases.

2.1 Where do children die?

In 2011, it is estimated that 7.2 million children under the age of five will die worldwide and that 99% of these deaths will occur in developing countries. Almost 50% of these under-five deaths can be accounted for by sub-Saharan Africa and more than 30% by South Asia (Lozano et al., 2011). Comparing these percentages to the total population in the two areas, namely 839 and 1567 millions[3] (World Bank, 2011), it is evident that the African region is not only the most heavily affected in absolute, but also in relative, terms. Some countries within the sub-Saharan and South Asian regions are particularly affected. According to estimates done by Lozano et al. (2011), 50% of child deaths will occur in five countries; India, Nigeria, Democratic Republic of the Congo, Pakistan and Ethiopia, in 2011. Compared to the fraction of world population these countries account for, 24%[4], their proportion of child deaths is substantial. Well above a third of under-five deaths will befall India and Nigeria alone in 2011. The world map below, where the size of each country corresponds to their proportion of worldwide deaths in the age group one to four, clearly illustrates that the African and South-Asian regions are the most severly affected by child deaths.

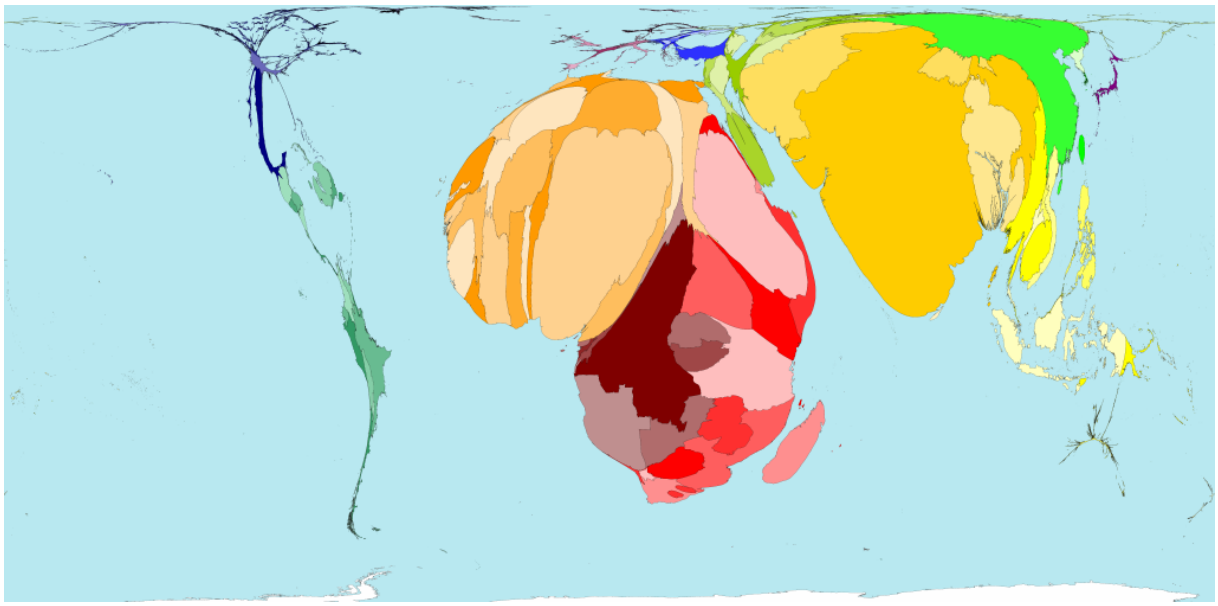


Figure 2.2: World map with territory size proportional to the countries fraction of world total child deaths[5].

[3] 2009 estimates
[4] This number is calculated on the basis of population estimates for 2009 presented in WDI 2011
[5] The map is downloaded from the webpage worldmapper.org (http://www.worldmapper.org/images/largepng/263.png) accessed 23th September 2011 and is based on data from 2002.

2.2 Why do children die?

According to the World Health Organization, more than two thirds of under-five deaths are caused by conditions that could be treated with access to simple low-cost preventive measures and treatments. The major cause of under-five deaths is infectious diseases[6] (accounted for 68% in 2008) and health problems in the neonatal period. The most important neonatal death causes are preterm birth complications and birth asphyxia[7], while diarrhoea and pneumonia were the major killers of children aged 1-59 months. Malnutrition is estimated to be the underlying cause of one third of all under-five deaths (Black et al., 2010, World Health Organization, 2011c). Figure 2.1 gives a graphic representation of child deaths by cause.

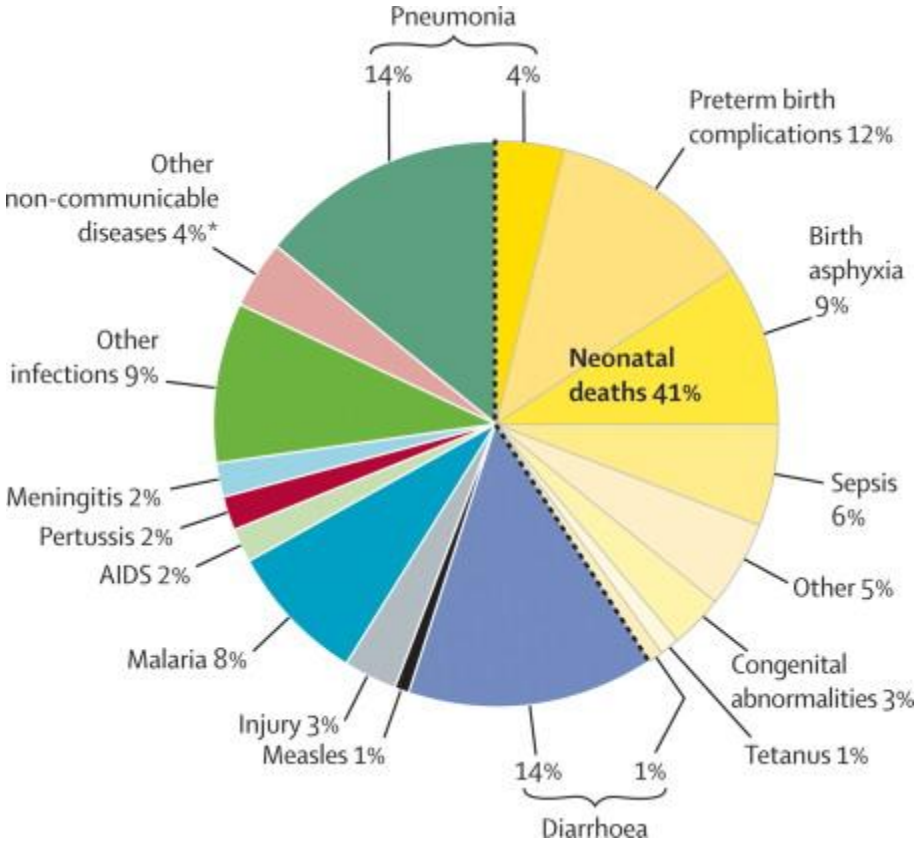


Figure 2.3: Distribution of child deaths by cause (Black et al., 2008)

[6] WHO definition: "Infectious diseases are caused by pathogenic microorganisms, such as bacteria, viruses, parasites or fungi; the diseases can be spread, directly or indirectly, from one person to another." (World Health Organization 2011c)

[7] Defined as the failure to establish breathing at birth (Spector & Daga 2008)

2.3 When do children die?

Of the 7.2 million estimated under-five deaths in 2011, 2.2 millions are predicted to be early neonatal deaths, 0.7 million to be late neonatal, 2.1 postneonatal and 2.2 in the ages between one and four years old. Thus, 40% of the child deaths will occur during the first month of life (Lozano et al., 2011). The World Health Organization estimates that approximately half of neonatal deaths occur within 24 hours after birth, and three quarters within one week. In a study from 2010, Rajaratam et al (2010) find that child mortality in all age groups has been substantially declining in the 40-year period between 1970 and 2010, but that the decline has been smaller for neonates than for the other two groups. Thus, neonatal deaths are not only the largest in number, but they are also rising in importance as their share of total deaths is becoming larger relative to postneonatal and childhood deaths.

2.4 Regional and Country differences

Although overall statistics provide valuable information about world trends in child mortality, it is important to recognize the large differences in disease burden between regions and countries. For example, 92% of all child deaths caused by malaria and 90% of deaths caused by AIDS worldwide occurred in the African region in 2008. Age of death also differed immensely between the two regions: in Southeast Asia, 54% of children died in the neonatal period whereas this number was significantly lower, 29%, in Africa (Black et al., 2010). Because of these large differences, country specific knowledge about social and epidemiological conditions is crucial when aiming to promote child health. This point is highlighted in Black et al.'s "Where and why are 10 million children dying every year?" The authors estimate the distribution of death causes for children under-five in 42 countries. These countries were chosen on the basis that they together accounted for 90% of child deaths worldwide in 2002. Black et al. group the countries into five profiles distinguished by variations in the distribution of child death causes. Whereas the proportion of deaths attributed to pneumonia and diarrhoea reveal similar patterns across the countries, differences are large in proportion attributed to malaria, AIDS and neonatal causes. The following five profiles were constructed to capture these differences in death causes:

- *Profile 1*: Malaria and AIDS each account for less than 10% of deaths and neonatal causes for less than 40%.

- *Profile 2*: Malaria accounts for at least 10%, but AIDS accounts for less than 10%.
- *Profile 3*: Malaria and AIDS each account for less than 10% and neonatal causes for at least 40%.
- *Profile 4*: Both malaria and AIDS account for at least 10%.
- *Profile 5*: Malaria accounts for less than 10% and AIDS for at least 10% of deaths.

This paper contains an important message: even between a selection of countries in sub-Saharan Africa, which is often thought of, and internationally treated, as one homogenous region, there are large differences in disease topology. In fact, the sub-Saharan countries were grouped into four out of the five different profiles, namely profile 1, profile 2, profile 4 and profile 5 (Black et al., 2003). The large inter-country differences in disease topology make it essential for policy makers to use available evidence and country specific information when evaluating how they should address the particular challenges they face.

2.5 Development and trends in child mortality

After the introduction of the Millennium Development Goals in 2000, child mortality and other indicators have been carefully monitored. Worldwide, child mortality has been continuously declining over the last 20 years, but the decline has not been uniform across age groups. Neonatal mortality decline has been slower than that of infants and children in the age group of one to four years (Lozano et al., 2011). This can be linked to slow progress in

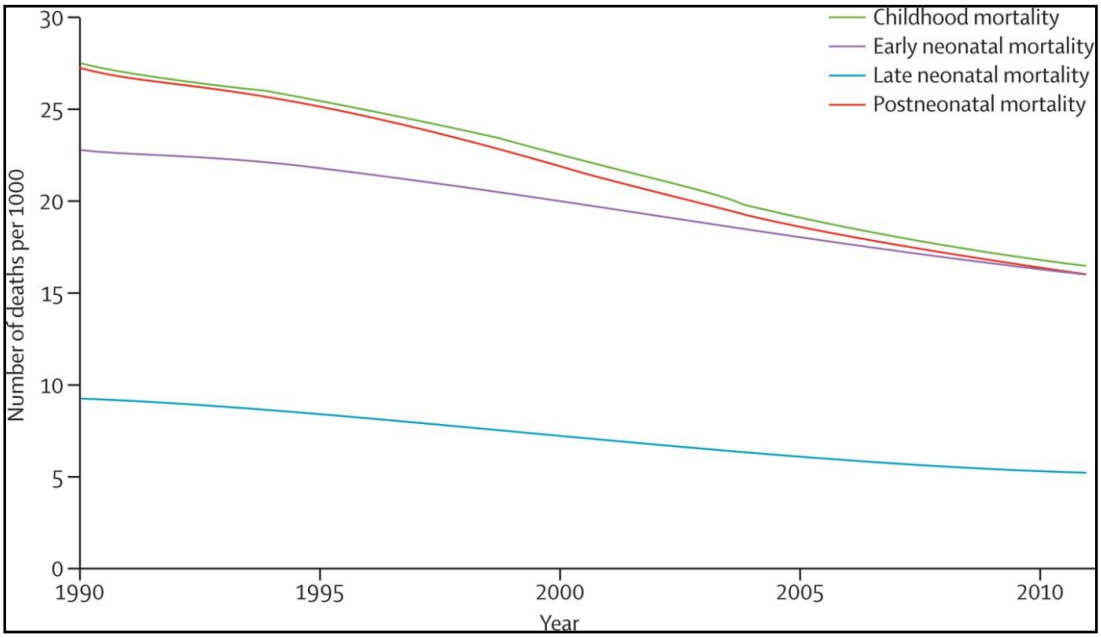


Figure 2.4: Worldwide neonatal, late neonatal, postneonatal and childhood mortality, 1990-2011 (Lozano et al., 2011)

improvement of maternal health (Bhutta et al., 2010). Figure 2.3 illustrates worldwide reduction in overall childhood mortality as well as in the different age groups; early neonatal mortality, late neonatal mortality and postneonatal mortality, in the 20-year period between 1990 and 2010. The figure is from the article "Progress towards millennium Development Goals 4 and 5 on maternal and child mortality: an updated systemic analysis", where Lozano et al. investigate and update estimates on the progress made towards achieving the Millennium Development Goals 4 and 5. They estimate the average annual decreases in mortality the last two decades to be 2.2% for under-fives, 1.7% for early neonatal, 2.7% for late neonatal, 2.5% for postneonatal and 2.4% for children aged one to four years old.

Lozano et al. also estimate progress towards the millennium development goal for child mortality reduction in the individual countries. The good news of the report is that 31 developing countries (accounting for 27% of world deaths) are estimated to reach the targeted child mortality reduction by 2015, and eleven more by 2020. The prospect for countries in sub-Saharan Africa is however quite gloomy. Only one country, Madagascar, is doing well enough to reach MDG 4 by 2015. The eight countries next in line (Eritrea, Ethiopia, Ghana, Liberia, Malawi, Rwanda, São Tome and Príncipe and Sierra Leone) are lagging 10 years behind and are not estimated to achieve the goal before 2025. Yet 23 sub-Saharan countries are not going to reach the targeted reduction until 2040. Nevertheless, compared to the period between 1990 and 2000, 39 out of 48 countries in the region have experienced an accelerating decline in child mortality between 2000 and 2011.

2.6 Socioeconomic determinants

The predominant causes of child deaths worldwide are diarrhoea, pneumonia and malaria. Cause of death is defined as "disease or injury which initiated the train of morbid events leading directly to death" (International Classification of Diseases 9, 2000). According to this definition, statistics on causes of children under-five deaths do not take into account factors that have indirect impacts on child health, and that constitute the underlying causes of death. Examples of such factors are living environment and demographic characteristics of the household. In understanding, addressing and ultimately mitigating the problem of excess child mortality, it is important to investigate the impact and causeway of these non-medical factors. This has lately been recognized by both medical and social scientists and large international

organizations such as the World Health Organization, The World Bank and the International Monetary Fund. The renewed focus on underlying causes of death has lead to the emergence of the buzz-phrase *social determinants of health*. Social determinants of health are the conditions in which people live, and that affect their opportunities to lead healthy lives (Labonté and Schrecker, 2007). Examples of social determinants of health are education, income, social status, physical environment and housing conditions, social support networks, genetics and gender, but also health systems and services (World Health Organization, 2011e). The point of introducing social determinants of health is therefore not to exclude factors traditionally focused upon in health research, but rather to *include* the root causes of health outcomes. A key aspect of this widening of focus is prevention; social determinants like clean water, sanitation and education does not only improve chances of survival when people have fallen ill, but could prevent people from falling ill in the first place. In chapter 4, I will present some theoretical models for the mechanisms through which social determinants affect child health.

3 The Angolan context

The aim of this chapter is to provide insights into the Angolan context to give a better understanding of the country and the reasons for why I want to study it in relation to child mortality.

Since the liberation from the Portuguese in 1975 and up until the peace agreement in April 2002, Angola was torn by civil wars between different nationalist movements that left the country's infrastructure and social services in ruins (Hodges, 2004). This contributed to rendering it one of the least developed countries in the World in terms of socio-economic indicators (UNDP, 2010). In 2010, Angola was ranked as far down as number 146 out of 169 on the Human Development Index (HDI)[8] by the United Nations Development Programme (UNDP, 2010). Despite having a slightly higher public per capita expenditure on health than the average in sub-Saharan Africa, Angola's infant and under-five mortality rate is substantially higher than both the regional average (World Bank, 2010) and the average in the group of countries classified as "nations with low human development" by UNDP (UNDP, 2010). Some key indicators for Angola are presented in table 2.1.

Table 3.1: Key indicators, Angola (2009)[9]

Indicator	*Value*
Population	18 million
Population growth (average annual growth 1990-2009)	2.9%
Surface area	1 247 000 km^2
GNI per capita (PPP-adjusted)	$5190
Life expectancy at birth	48 years
Population below $2 per day	70.2%
Access to an improved water source (2006)	50%
Urban population	58%

[8] HDI is a composite measure of human development consisting of indicators of health, education and income.
[9] Source: World Bank 2011. *World Development Indicators 2011,* Washington, D.C., The World Bank.

3.1 Geography

Angola is situated on the west coast of Africa and is, with its more than 1.2 million square kilometres, the 7th largest country on the continent and the 23rd largest in the world. The country is rich on natural resources including petroleum, diamonds, iron ores, phosphates, feldspar, copper, bauxite, uranium and gold (CIA, 2011). Geographically, Angola can be divided into three longitudinal zones, stretching from the border to the Democratic Republic of the Congo in the North to the Namibian border in the south. The coastal zone is up to 200 kilometres wide and consists of dry lowland and desert in the south. A belt of highland with fertile land and mountains as high as 2500 meters traverses the middle of the country. The east is dominated by a high plateau with dry vegetation that take up two thirds of the country width (Kapuscinski, 1976).

3.2 Demography

The country population is estimated to be about 18-18.5 million (World Health Organization and UNICEF, 2010, The World Bank, 2010), but estimates vary largely between organizations and methods due to a lack of birth and death registers. More than half of the population lives in urban areas, many of them in the country capital Luanda, which has an estimated size of 4.8 million people (Frøystad et al., Forthcoming 2011)

3.3 People

The main ethno-linguistic groups in Angola are Ovimbundu, Mbundu and Bakongo speaking Umbundu, Kimbundu and Kikongo, respectively. Many Angolans also speak Portuguese. This is partly due to the fact that rapid urbanization has lead to extensive interaction between people of different ethnic origins. Portuguese is the sole official language in Angola, and is used by teachers in schools as well as in the military (Hodges, 2004). Christianity is the largest religion in the country, with the Catholic Church as the main denomination, as well as a variety of smaller Christian communities. Many Angolans also practice animism [10] (Norwegian Embassy in Angola, 2011).

[10] "Animism is generally the doctrine that some vital principle, or some kind of soul, produces the living phenomena of organized bodies and yet exists apart from those bodies" (Barfield, 1997).

3.4 The Economy

In the last decade, Angola experienced a large increase in economic growth. Between 2000 and 2008, the annual average growth rate of GDP was 13.7%, a number only beaten by Azerbaijan and Turkmenistan during the same period (World Bank, 2009). Most of this growth can be ascribed to increases in oil production, which accounted for more than 80% of the country's GDP and 90% of export revenues in 2009 (USAID, 2009). The main importers of Angolan oil are the United States and China, together accounting for 60% of the oil exports. The remaining 10% of revenue comes from diamond and mineral export and the Angolan economy is heavily dependent on natural resources, making it susceptible to price fluctuations in the oil and mineral market (Hodges, 2004).

The World Bank estimates the Angolan PPP-adjusted gross national income per capita of $5190 to be more than twice the size of the regional average in Sub Saharan Africa (World Bank, 2011). According to British Petroleum's *Statistical review of World energy* from June 2010, Angola is the second largest oil producer in Africa and the 15th largest in the world. In 2009, the Angolan oil production amounted to 2.3% of the world total production (British Petroleum, 2010). The oil sector is controlled by the state owned conglomerate *Sonangol*. In contrast to other Angolan institutions and companies, Sonangol is regarded as well functioning. The company has a good international reputation and is known for hard bargaining and technical competence. According to de Oliveira, a top oil executive of a major European oil company referred to Sonangol as "the Angolan miracle" in an interview in 2005 (de Oliveira, 2007). The fact that Angola's oil industry is well managed does not, however, mean that oil revenue is trickling down to the larger masses of the population. Since the formation of Sonangol in June 1976, the company has "been a private tool for the interests of the president clique known as the Futungo de Belas" (de Oliveira, 2007). The Futungo is a group of unelected officials and businessmen who are functioning as a part of the power structure around President José Eduardo dos Santos. This group sprung out in the 1980's, when dos Santos took over the position as President after the death of the Agosthino Neto in 1979 (Hodges, 2004).

Global Witness, an NGO working on resource-related conflict and corruption in Angola, claims that up to one third, $1.4 billion, of the Angolan state budget was unaccounted for in

2001 (Global Witness, 2002). Investigations made by the same organization showed that huge sums disappear into offshore money laundering, but they underline that lack of governmental transparency makes it impossible to know how exactly much money is evaded, and where it ends up (Global Witness, 2004). Money does not only accrue to the elite families through illegitimate channels, but also through privileged access to state scholarships for education and support from the national health board to get medical treatment in hospitals abroad (Hodges, 2004). Against this background, it comes as no surprise that Angola is listed as the 10th most corrupt country in the world by Transparency International (2010).

The large resource export revenues and relatively high income per capita combined with a strong and wealthy elite, makes the distribution of income in Angola extremely uneven. In 2000, the richest 10 percent of the population accounted for almost 45 percent of the total household expenditure (World Bank, 2010, Hodges, 2004), and in 2011, 70% of the population is estimated to live on less than $2 a day. The high degree of inequality in income is also reflected by an estimated GINI coefficient[11] of 58.6, which is twice as high as the one for Norway (World Bank, 2011). The following quote provides a picture on how these inequalities play out in daily life in Angola

> (...) the contrast between very rich and very poor is evident in images encountered daily throughout the country: the sight of malnourished children begging next to brand new Land Rover Discoveries, or destitute amputees leaning against an advertisement for cellular telephones.
>
> (Hodges, 2004, pp. 41-42)

At first glance, the situation in Angola does not look too bad compared to other countries in the Sub-Saharan region. The country is endowed with abundant resources and GNI per capita is relatively high. A closer look at the socio-economic indicators turns the picture up-side-down: hardly any country has worse education and health outcomes, and nowhere does public expenditure yield lower returns.

[11] The GINI coefficient is a measure of to what extent the income is unequally distributed in a population. A GINI coefficient of 0 represents perfect equality in distribution and a coefficient of 100 represents perfect inequality in distribution of income.

3.5 Health

According to the Angola Health System Assessment carried out by USAID in 2010, the availability of health data in the country is limited. In 2010, a Demographic Health Survey was conducted, but the data has not yet been released (USAID, 2011). As the source data are limited, estimates of health indicators for Angola should be treated with caution. Nevertheless, indicators, though imprecise, leave no doubt that the country is facing large health challenges. As seen in table 3.2, Angola is doing worse than the regional[12] average in terms of infant and under-five mortality rate, but has a lower maternal mortality. The fertility is also higher than the average in the Sub-Saharan region. Health expenditure[13] per capita is somewhat higher than the regional average. However, knowing that a lot of this money is spent on treatment of elite society members in foreign hospitals, the average number should regarded with suspicion. It is unlikely to give a reliable reflection of government spending on public health services. In addition, health expenditures as percentage of total GDP in Angola is among the lowest in the world. This can be interpreted as reluctance to commit to the improvement of health from the government's side. Also note that the HIV-prevalence, i.e. percentage of total population infected with HIV, is quite low; less than half of the regional average. Hodges (2004) ties this together with the fact that Angola has been poorly integrated with the rest of the region in terms of migration and transportation. The low prevalence is probably not a result of successful health polices, but rather the special political circumstances in the country in the previous decades. Because of the civil war the regional integration of the country has been small in terms of transport and migration, which are two important transmission vectors. HIV could therefore emerge as a public health problem in the future, following more interaction with neighbouring countries and improved transport and infrastructure.

[12] Sub-Saharan Africa as defined by The World Bank
[13] Total expenditures (governmental + private)

Table 3.2: Health indicators in Angola compared to regional average

Indicators	*Angola*	*Regional average*
Infant mortality rate (per 1000 in 2011)	90**	65**
Under-five mortality rate (per 1000 in 2011)	134*	101**
Maternal mortality ratio (per 100 000 live births 2011)	335*	380**
Health expenditure per capita (both private and governmental, 2008)	$183***	$132***
Physicians per 1000 people (2004-2009)	0.1***	0.2***
Nurses and midwives per 1000 people (2004-2009)	1.4***	1.0***
HIV prevalence (2009)	2.0%***	5.4%***
Fertility rate (births per woman in 2009)	5.6***	5.1***

*Source: from Lozano et al. (2011), **Calculated on the basis of the estimates in Lozano et al. (2011),
***Source: World Development Indicators (World Bank, 2011)

3.6 Health system

The colonial administrative system (including the health system) that Angola inherited from its Portuguese colonizers, was extremely centralized and exclusively suited to cater the needs of the colonizers, while neglecting the local population. Plunging straight into civil war after the country gained independence, little was done to change the system in keeping with the Angolan population's needs until 2001, when the first official moves towards decentralization of the health system were made as part of the broader administrative reform, *Reforma Administrativa do Estado*.

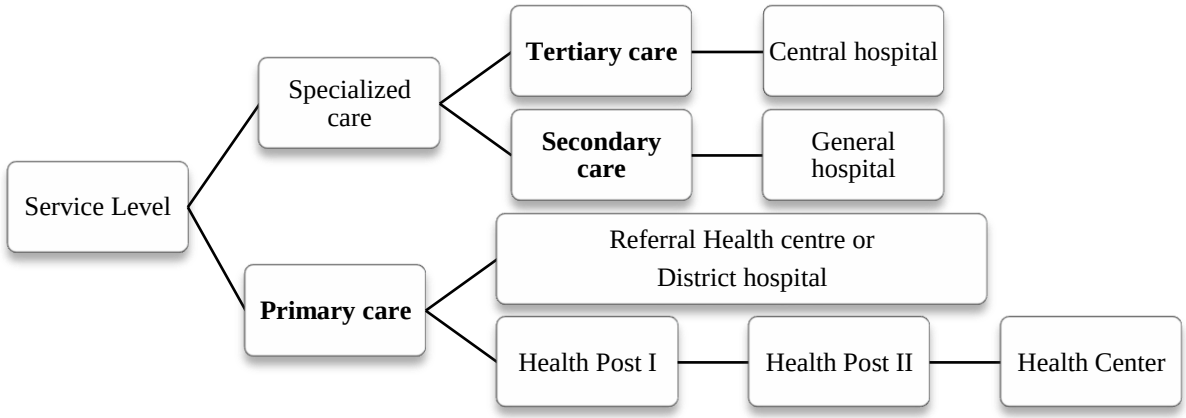


Figure 3.1: Organizational structure, health service system in Angola (Connor, 2010)

In the health sector, this reform has led to an ongoing decentralization process where the responsibility and decision making in primary health care is being transferred from provincial to district level. Primary care is the lowest level of care followed by secondary and tertiary care. Figure 3.4 above shows how the health service delivery in Angola is organized according to level of care. The lowest unit of service delivery is "Health Post" (Connor, 2010).

3.7 Chapter summary

There is no doubt that the Angolan population is desperately in need for improvements in quality of life. In the introduction I argued that health, and child health in particular, is an important aspect of social development. To my knowledge, not much research has been done on health issues in Angola, partly because of data shortage. A study of child mortality in relation to health service delivery and household characteristics in Angola is useful, as country and region specific data is crucial in developing efficient policies that address the enormous challenge the country faces in improving population health.

4 Theoretical backdrop and hypotheses

This chapter provides a theoretical basis for the analysis of child mortality. The first section gives a brief account of the historical decline in mortality that has taken place in the developed world, and presents some causes behind this development. Thereafter, an economic model of the household decision of fertility is presented. The third section introduces an analytical framework of determinants of health that allows us to tie household characteristics such as education and wealth to more direct causes of child mortality. The fourth section gives a short review of evidence from empirical research, aiming at relating our theory to the real world and determining the variables to be included in the analysis. Before concluding the chapter, I formulate some theory inspired hypotheses that will guide my empirical work. The hypotheses are formulated on the basis of the theory, analytical framework and empirical evidence presented.

4.1 A very brief history of mortality

The following section is based on Cutler, Deaton and Lleras-Muney's description of the historical decline in mortality laid out in their article "The determinants of mortality" (2006).

Ten thousand years ago, life expectancy at birth was probably as low as 25 years. In 1700, this expectancy had only increased by 12 years to 37 years in the richest countries at that time, England and the Netherlands. Since then, mortality rates have been dramatically reduced worldwide. In England the decline started around 1750, and by 1820 life expectancy had risen from 37 to 41 years, a significant improvement considering the short time interval. The period between 1820 and 1870 was characterized by the industrial revolution bringing extensive changes to society. Manual labour was widely substituted by mechanical production, infrastructure was developed, and people moved from the countryside to large cities to work in factories. It was a period of economic growth and prosperity. Still, housing and sanitary conditions in the rapidly expanding cities were poor. People lived in congested spaces and disease and plagues spread fast. Not until large improvements had been made in public health, did the mortality in England start to decline (around 1870). During the first part of the twentieth century, life expectancy underwent a spectacular 50 years rise, to 77 years, which is the life expectancy in England today. According to Cutler et al. much of the decline in

mortality can be attributed to a decline in infectious diseases, but there has been extensive debate in the literature on why this decline has taken place.

The debate can be summarized by dividing the historical decline in mortality into three phases: The first phase took place between 1750 and 1850, and in this period, higher living standards and improved nutrition occasioned by economic growth were the main contributors to the reduction of mortality. In the second phase, between 1850 and the early 1900's, mortality was further reduced because of public health service improvements such as the delivery of clean water, collection of waste, providing sanitation system and encouraging better personal health practises. The third and last phase dating from the 1930s and onwards, has been characterized by medical inventions, both preventing (e.g. immunization) and treating (e.g. antibiotics) illnesses and thus reducing mortality.

Even though the overall mortality has declined and life expectancy increased, worldwide, there are large differences between countries and between countries grouped by income; while the average life expectancy at birth in low income countries was 57 in 2009, this number was 80 for countries in the high-income group (World Bank, 2011). The child and infant mortalities in the two groups differ accordingly. Most inhabitants in developed countries have the means and knowledge to acquire sufficient nutrition, as well as access to important public health services and opportunity to utilize modern medical inventions. In developing countries, on the other hand, one might lack all or some of these important factors in order to achieve increased life expectancies and lower mortality rates. While developed countries have gone through the phases in the historical succession described above, this is not necessarily the case for developing countries. They could benefit from utilizing the knowledge and experience that the developed countries have acquired to achieve more rapid reductions in mortality. In principle, this can be done by focusing on all of the abovementioned factors at once, and these should be kept in mind when choosing variables for, and conducting, the analysis.

In industrialized countries, the decline in mortality has been coupled with a decrease in fertility. This process is known in the literature as *demographic transition*, and in most countries, the decline in fertility occurred some time after the reduction in mortality. A possible interpretation is that people choose to have more children when the chances of

children dying young, are high. It is also likely that fertility affects household child mortality due to the fact that more mouths to feed reduces the amount of resources available to secure the survival of each individual child. Thus, fertility is of significance to the study of child mortality, not only because the number of children born into a household or community sets an upper boundary to the number of children dying, but also because it interacts with mortality.

4.2 An economic model for fertility

In this section, I introduce an economic model of fertility decisions in a utility maximizing household. The model stems from work done by Gary Becker and Gregg Lewis in the 1960s and 70s and the version presented here is based on that of Bardhan and Urdy (1999, pp. 22-31). A main result in the model is that households face a trade-off between the number of children and the level of human capital of these children often referred to as "quality" of children in the literature.

Many factors come into play when parents make a decision about fertility: biological factors such as the need to reproduce or to enjoy the company of own children, economic factors like costs and benefits of bringing up a child and cultural factors like norms and expectations about the number of children in a family. In the model, only economic factors are included explicitly. Other factors are implicitly modelled through a vector of household preferences. In poor societies, economic considerations are expected to be relatively more important in fertility decisions than in richer ones. People struggle to get by, and having children can be motivated by the expectation that they will yield economic returns in the future. This economic return could be divided into two categories. Firstly, the children could contribute to the household by engaging in income generating activities. Secondly, in the absence of pensions and other social security systems, having children is a way for parents to secure financial support in their old age. Economic motives like these are less relevant in more developed countries where child labour is prohibited and there are pension systems that enable people to make financial arrangements for their old age (Bardhan and Udry, 1999, Todaro and Smith, 2006).

4.2.1 The model

If we assume that the household's fertility is a rational choice made to maximize utility, and that household utility is dependent on parental consumption of market goods, *x*, the number of surviving children, *n*, and the human capital level of these children, *z*, the utility function can be written as

$$U(x, n, z; \boldsymbol{\alpha}), \qquad \frac{\partial U}{\partial x}, \frac{\partial U}{\partial z} > 0 \tag{4.1}$$

where $\boldsymbol{\alpha}$ is a vector of exogenous factors affecting the preferences of the household. These might be cultural norms, biological needs etc. The utility of the household is, in accordance with the convention in economic literature, assumed to be increasing in parental consumption. Utility is also assumed to increase in level of child human capital, *z*, because more human capital increases expected wage and thus the amount of money contributed to the household income. The utility parents get from the human capital level of their children can also be attributed to altruistic feelings (Becker, 1992). Parents wish the best for their children and value their well-being. Investments in child human capital affect the children's well-being positively and the parents therefore want to make these investments if they are altruistic towards their children. An increase in the number of children is also assumed to lead to an increase in utility, at least for small values of *n* (Bardhan and Udry, 1999).

For the sake of simplicity, it is assumed that all children receive the same treatment (i.e. that an equal amount of resources is spent on each child), so that the human capital level is identical for all children in the household. Human capital includes health status and education and is dependent on consumption, *c*, and time and effort laid down in the children by parents, *t*. In the model, child consumption is not only consumption of goods such as food and clothes, but also services important to other human capital aspects like education and medical care. The human capital level of a child can be written as:

$$z = \frac{Z(c, t; \boldsymbol{\beta})}{n} \tag{4.2}$$

where $\boldsymbol{\beta}$ is a vector of exogenous factors affecting the possibilities for accumulating child human capital in the household, for example education policy. The amount of time and effort that parents put down in their children depends on the opportunity cost, namely the wage

earned if they had devoted their time to work instead. Let *w* be the household wage if all available time is used for work, p_x the price of adult consumer goods and p_c the price of child consumer goods. The household budget constraint is then given by

$$w(1-t) = p_x x + p_c c \tag{4.3}$$

The household wants to maximize its utility subject to the constraints given by (4.2) and (4.3), leading to the following maximization problem:

$$\max_{x,n,c,t} U(x, n, z; \boldsymbol{\alpha}) \tag{4.4}$$

4.2.1.1 Trade-offs

Household utility depends positively on both number of children, *n*, and the human capital level of these children, *z*, and that the human capital level is negatively related to the number of children. In maximizing the utility in (4.4), parents therefore face a direct trade-off between *n* and *z*. A higher fertility will give an increase in utility through *n*, but will also reduce it through a decrease in *z*; the more children in a household, the lower is the human capital level of each child for any given investment level *c* and *t*. Similarly, parents face a trade-off between their own and children's consumption. For a given household budget and constant prices, an increase in parental consumption, *x*, will lead to a decrease in child consumption, *c*, and consequently a reduction in, *z*. The allocation of time between work and child rearing represents yet another trade-off. Time spent caring for children, *t*, has a positive direct effect on child human capital, but also has a negative indirect effect through reducing time available to income generating activities and thereby possibilities of consumption. An increase in *t* will reduce the budget available for consumption, $w(1-t)$. More time spent on children will thus have a negative effect on *z* through a reduction of *c*, but a positive effect through an increase in *t*, in (4.2).

So far, the model has revealed that, for any given amount of resources, the fertility decision of the household must be made with regard to the fact that more children leads to less resources available for each of them, which in turn leads to lower human capital per child. Why is this helpful in dealing with mortality?

An important aspect of human capital is health. By showing that there is a relationship between investments in human capital and number of children, I have pointed to a decisive mechanism in child health. From a health point of view, if resources are scarce, it is preferable that parents have fewer children with more human capital, particularly better health, rather than many children with worse health. In developing countries, however, high fertility rates prevail regardless of limited resources. I now proceed to seek explanations for this by analysing how changes in the different factors affect the choice of fertility in the model.

4.2.1.2 Income

According to economic theory, an increase in income will generally increase the consumption of goods. In "An Economic Analysis of Fertility" published in 1960, Becker classifies children as normal consumer goods; they are a source of emotional satisfaction and can provide the family with extra income. We could therefore also expect to see a higher "demand" for children when the wage rate is increased. This higher demand does not necessarily manifest itself in a higher number of children; it could also take the form of a higher desired level of child human capital ("quality"). According to Becker, the income elasticity of the quantity of normal goods is usually small compared to the corresponding elasticity for the quality. If the classification of children as normal consumer goods is adequate, a household will respond to an increase in income by increasing expenditures on child human capital relatively more than on an increasing number of children. If the income elasticity of child human capital is sufficiently much larger than that of quantity, an increase in income will not bring about any increase in the number of children, only more investments in child human capital. In the model, an increase in income will, through this mechanism, improve child health and thus reduce mortality.

Wage affects child human capital and health in the model through changing the *opportunity cost* of spending time on child rearing and the *purchasing power* of the household. While the increase in opportunity cost will induce parents to work more and thus have fewer children, the higher purchasing power makes spending time on child-rearing more affordable to the household. The former is referred to as a *substitution* effect, and the latter as an *income* effect, in microeconomic theory, and our model does not provide an answer to which effect is dominating the other. In the case that a wage increase induces the parents to work more, time spent with children, t, will decrease and the household budget $w(1-t)$ will increase. The

reduction of *t* will, all other things being equal, lead to a lower level of child human capital. However, since parental utility is dependent on child human capital, it is likely that parents will make sure that that this decrease is (at least) compensated for. This compensation can be made either by spending some of the increased income on child consumption, or by reducing the number of children. Parents are likely to choose the latter if the expected return (increase in utility) from spending a given amount on one child excels the expected return of spending the same amount on several children. Choosing to have fewer children is also the most likely option if the wage raise reduces the need for financial security provided by children or if the costs of investing in child health are very high.

4.2.1.3 Mortality

Mortality and fertility affect each other in many ways – both directly and indirectly. It is reasonable to assume that parents care about the number of surviving children and not how many that are born. Child mortality in a household, i.e. the probability of a child passing away, will therefore affect the number of births required to reach the desired number of surviving children. This can be said to be an indirect effect of mortality on fertility; the probability of a child dying affects the household "demand" for children (Becker, 1992). A more direct effect of mortality on fertility comes through the influence on the cost of bringing up a surviving child. Carrying out a pregnancy and giving birth requires both time and money. Mortality affects the average number of births needed to get a surviving child, and a reduction in mortality will therefore reduce the cost of having a survivor. Hence, one should expect to see an increase in fertility when mortality falls. This is contrary to empirical evidence. An explanation could be that the reduction in mortality concurred with economic growth and the relative increase in the value of time and returns to investment in child human capital. Child mortality also has a direct negative impact on the health of mother and children if it leads to more children being born. Frequent pregnancies wear the woman's body out and increase the chance of complications, which have adverse consequences for both mother and child. High mortality leads to more pregnancies that in turn affect the initial child human capital negatively. In our model, the mortality rate would manifest its effect on child human capital through a higher *n* which gives a reduction of child human capital in (3.2).

4.2.1.4 Culture and norms

The household decision about fertility is affected by the social environment that the parents reside in. Fertility choices made by neighbours and family influence preferences concerning

the number of children in the household. If the social convention is to have large families, parents will probably prefer having more children. Norms about the number of children also affect relative prices through their influence on the demographic structure of the society and demand. Relative prices in turn affect fertility decisions. Social norms about gender and work are also likely to affect preferences about fertility. In the framework described above, social norms and conventions can be modelled through $\boldsymbol{\beta}$.

A considerable drawback of the model is that it is based on the assumption that the choice of having a child is a rational and conscious one. However, pregnancy is often not planned for by the household, especially not in poor developing countries with low education levels and lack of contraceptives and family planning. Furthermore, many factors important to child health are not directly included in the model. In the following section I expand the theoretical framework for the analysis of child health and incorporate a wider spectrum of factors than the economic model took account of.

4.3 Determinants of child mortality

4.3.1 Mosley and Chen's analytical framework for the study of child survival

> *A child's death is the ultimate consequence of a cumulative series of biological insults rather than the outcome of one single biological event.*
>
> (Mosley and Chen, 1984, pp. 29)

In the mid 1980's, Henry Mosley and Lincoln Chen (1984) published the article "An analytical framework for the study of child survival in developing countries", where they provided a framework for analysing causes of child mortality that addressed both biological (direct causes) and social and economic factors (underlying or distal causes that affect biological mechanisms). They claimed that all socioeconomic variables have to work through a limited set of proximate variables that directly affect mortality. Proximate determinants are variables with both behavioural and biological aspects that link other underlying variables such as income and education to the biological process causing illness. An example of a proximate variable is the use of contraceptives. Use of contraceptives is affected by

underlying variables such as education, place of residence and wealth, but is also affecting the biological variable fertility (Lewis et al., 2007). Moseley and Chen grouped proximate variables into the following five categories:

1. *Maternal factors*: age, parity and birth interval
2. *Environmental contamination*: air, food/water/fingers, skin/soil/inanimate objects, insect vectors
3. *Nutrient deficiency*: calories, protein, micronutrients
4. *Injury*: accidental, intentional
5. *Personal illness control*: personal preventive measures, medical treatment

Figure 4.1 illustrates how Mosley and Chen thought the socioeconomic determinants affected health and, ultimately, death (mortality) through these proximate determinants. While determinants in the first four groups (maternal factors, environmental contamination, nutrient deficiency and injury) affect the rate at which children is moving between healthy and sick, the fifth category (personal illness control) is also affecting the rate at which children recover from illness (Hill, 2003).

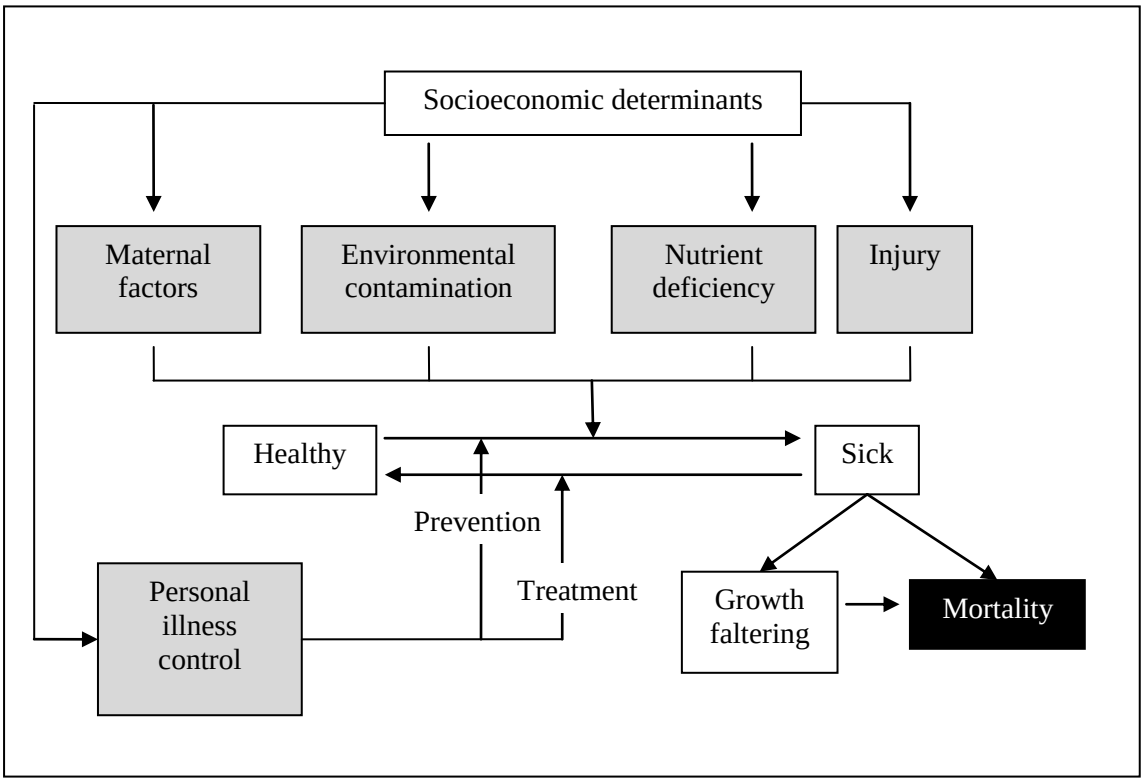


Figure 4.1: Proximate determinants and the ways they work on health

An illustrative example of how socioeconomic factors affect child health through proximate factors in a developing country is the *education of women*. Education is a social determinant that affects health through several of the proximate determinants. Educated women tend to marry and give birth later in life than uneducated women. Delaying the first childbirth beyond the teenage years decreases the chance of complications during delivery, and thus increases the chance of survival of both mother and child. In this way, the education level of a child's mother as a social determinant affects a proximate determinant of child health, namely the mother's age at birth. Because they start making family later in life and work outside the household to a larger extent, educated women have less time to both procreating and fostering children and consequently tend to have fewer children than women with less education. The number of children affects the individual child's health through the health of the mother (which is likely to deteriorating in step with number and frequency of childbirths) and through the amount of resources (such as food) available to the fostering of the child. The mother's age at birth and the number of children fall into the "maternal factors" category in figure 4.1 (Hobcraft, 1993). The mother's education also affects child health through knowledge. Both her ability to prevent herself and the children from getting ill and to take the appropriate measures if they do, is influenced by her level of education. This knowledge is part of the proximate determinant group named "personal illness control" in the figure. This example focussed on the conditions in a developing country. Of course, women's education matter to child health in more developed countries too. However, the mechanism through which it works and the relative importance of the effects differs between the two settings, and the example was chosen because it is relevant to the subsequent analysis.

In an article from 2011, Macassa, Hallquist and Lynch acknowledge the importance of Mosley's and Chen's work, but also criticize it for failing to incorporate more indirect factors affecting child health, such as national health, and health related, polices, institutions and macroeconomic variables. According to the authors, a consequence of this has been an over-emphasis on individual-level decision-making, while neglecting factors like the political, the geographical and the cultural environment which all affect child health (Macassa et al., 2011). The authors also propose a conceptual framework that includes a number of social components. This will not be presented here. Instead, I take a closer look at a more comprehensive and, to me, clearer framework for analyzing child health that takes into account the social context that Macassa, Hallquist and Lynch are calling for. The addition of a

macro-level category of determinants to the framework provides a more complete overall picture of the many factors and mechanisms that affect health, but comes at the price of increased complexity.

4.3.2 The Huynen, Martens & Hildernik framework

In "The health impacts of globalisation: a conceptual framework", Huynen, Martens and Hildernik (2005b) present a framework for analysing the impacts of globalization on population health. The core concepts in their model are similar to those of Mosley and Chen, but a group of determinants at the level above the distal determinants, namely "contextual determinants", is also incorporated. The framework is, contrary to that of Mosley and Chen, not meant for child health in particular, but rather for population health in general. However, most of the mechanisms described below will also be applicable to children.

Figure 4.2 depicts the succession of hierarchal causality levels in the framework. Proximate factors work directly on health, while distal factors work through proximate factors that in turn affect health. Contextual factors are macro-level conditions shaping the environment in which factors at the distal and proximate level develop, and precede the distal determinants in the chain of causality. The figure also shows how contextual determinants work on proximate determinants both directly and indirectly through distal determinants.

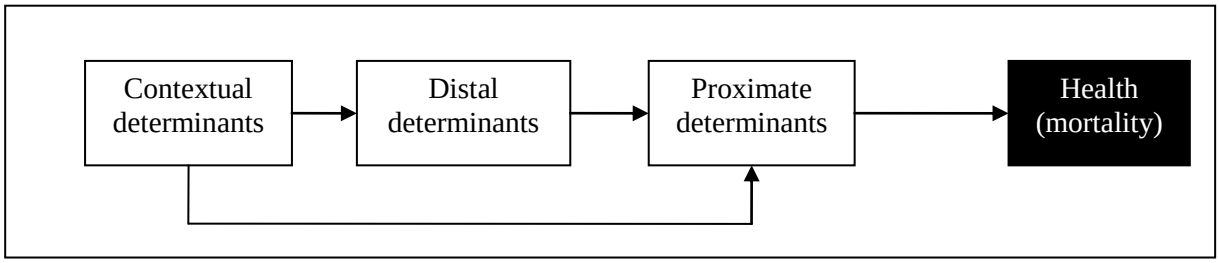


Figure 4.2: Health determinants: different hierarchical levels of causality (Huynen et al., 2005a)

Though the framework indicates a unidirectional causality running from a set of different determinants to health, health could be, and is, also affecting these determinants. In addition, the various determinants interact and influence each other (Huynen et al., 2005a).

Huynen et al. divide all determinants of health into four different categories; socio-cultural, economic, environmental and institutional factors. Determinants in each of these categories

then work at different levels in a chain of causality for health outcome. By analyzing other health models and empirical literature, the authors have chosen a number of general factors appropriate for a graphical representation of both the hierarchal levels of causality and the four different categories of determinants. These are described below.

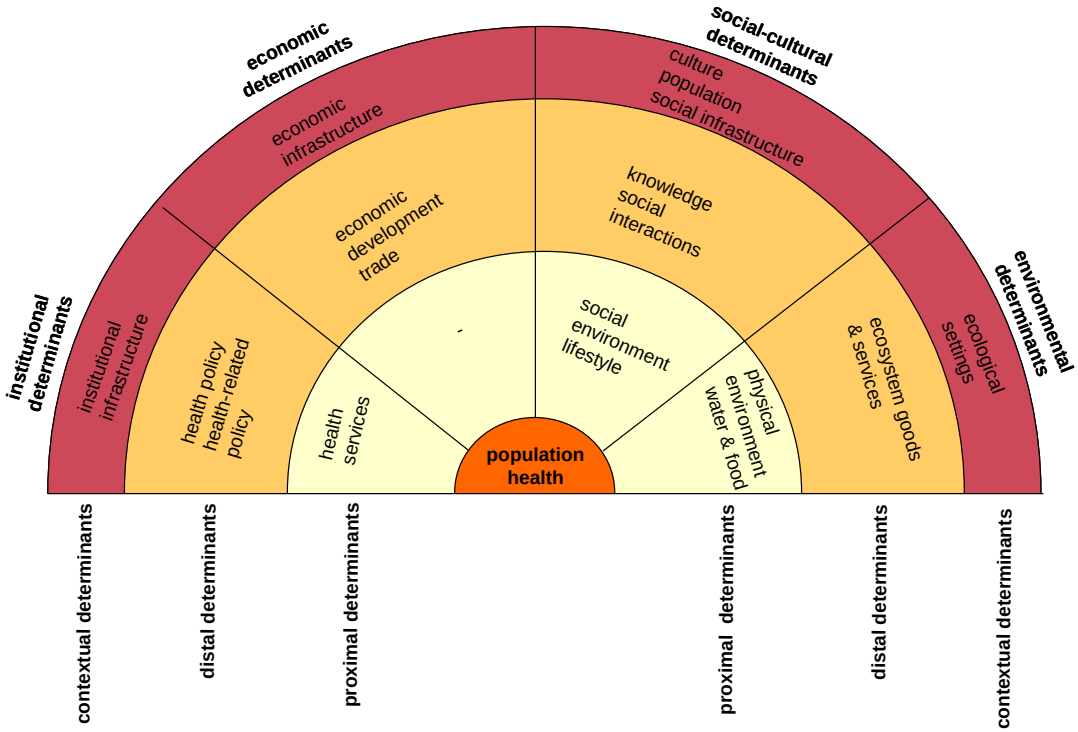


Figure 4.3: Multi-nature and multi-level framework for population health (Huynen et al., 2005a)

4.3.2.1 Proximate determinants

As portrayed in the figure 4.3, "health services" are considered to be a factor at the proximate level in the group of institutional determinants. The provision of, and access to, health services is thus thought to have a direct impact on population health. Amongst economic determinants, none are believed to have a direct impact on health, as illustrated by the empty area representing proximate determinants in the figure. "Social environment" includes informal care and social support, as well as factors having a direct negative impact on health such as abuse and violence in the household or community. "Lifestyle" also affects health directly through diet, use of unhealthy substances like alcohol, tobacco and drugs, sexual behaviour and amount of exercise. Social environment and lifestyle are both factors in the social-cultural group. Within the environmental determinants group, "physical environment" and "water and food" are proximate determinants. Physical environment includes quality of

housing and work (school) condition. Important factors in this category are chemical pollution, temperature and presence of bacteria and viruses and affect the rate at which people recover from illness. Availability of clean water and sufficient and nutritious food are also factors that influence health directly.

4.3.2.2 Distal determinants

Distal determinants affect health indirectly through proximate factors. "Health and health-related policy" are distal factors affecting quality and availability of health, and other public, services through budget allocations and priority settings. In the category of economic determinants, "economic policy" and "trade" are distal determinants affecting health through their influence on income levels. "Social interactions" are distal factors in the group of socio-cultural determinants and encompass migration patterns, conflicts, social equality and travelling. Conflicts affect health directly through psychical stress and damage of soldiers and civil population, but also indirectly through damages on infrastructure that affect the supply of food and water. Migration and travelling to new places might affect health by exposure to new diseases and unfamiliar bacteria floras. In the group of environmental determinants, "ecosystems goods and services" are factors at the distal level. The state of the ecosystem where people live sets the conditions of production, and affects quality and quantity of food and water consumed as well as the epidemiological environment. While proximate factors are relatively easy to identify and measure and effects may be predicted, this is much harder in the case of distal determinants.

4.3.2.3 Contextual determinants

The contextual determinants are factors forming the macro-level conditions in which distal and proximate factors are shaped. These are "institutional infrastructure" (governance structure, political environment, system of law, regulation), "economic infrastructure" (occupational structure, tax system, markets), "culture" (religion, ideology, customs), "population" (population size, age structure, geographical distribution), "social infrastructure" (social organization, knowledge development, social security, insurance system, mobility and communication) and "ecological settings" (ecosystems, climate). The complexity of the interaction between contextual determinants and health makes effects and causalities difficult to identify. For instance, occupational structure might be affected by population health and population health might affect the occupational structure (Huynen et al., 2005a).

In the analysis, I will investigate determinants of child mortality in two regions in Angola. Contextual determinants like tax systems, systems of law and governance structures do not, or only to a very small degree, vary within a country. I will therefore not include such determinants explicitly in the analysis. However, I do recognize their importance in shaping distal and proximate determinants also in Angola, and that my results must be interpreted with certain reservations as to the macro-level conditions.

4.4 Empirical work on child mortality

In this section, I present a brief overview of some of the empirical work done on child mortality. I've focused my attention to the factors in the economic model and the analytical framework and to those the data contain information about. The overview is not intended to be exhaustive.

4.4.1 Income/wealth

The empirical evidence shows a significant negative relationship between child mortality and wealth. The evidence is strong and the association is found at both the macro level and the micro level. Since this thesis deals with household characteristics and health service delivery, I focus on the micro evidence. In a review of international evidence on child mortality in low and middle income countries, Houweling and Kunst (2010) use a household asset index to measure wealth. They find that the child mortality is significantly higher in the poorest compared to the richest bracket of society in 55 developing countries. There were not only differences between the poorest group and the rest, but also across other income groups. Anyamele (2011) also finds a significant negative association between wealth quintile and infant mortality in 20 countries in Sub-Saharan Africa and Sswanyana et al. (2002) find a negative effect of wealth measured by household assets in Uganda. Explanations for this relationship are plentiful. In addition to its direct impact on purchasing power and the ability to buy health promoting goods and services, wealth is closely related to proximate determinants of child mortality; most of these factors show worse levels for the poorest (Houweling and Kunst, 2010). Wealthier households are likely to dwell in conditions with better sanitation, and display a higher level of education. Hence, they will probably have more knowledge about disease prevention and treatment, and to have access to higher quality health service than the less wealthy.

4.4.2 Education

Education, and mother's education in particular, is one of the most frequently described social determinants of child mortality in developing countries, and empirical evidence strongly suggests that educated women have fewer and more healthy children than the less educated (Kiros and Hogan, 2001, Cleland and van Ginneken, 1988, Houweling and Kunst, 2010). Studying the relationship between maternal education and child mortality in 17 developing countries, Bicego and Boerma (1993) found a significantly higher mortality rate for children aged 0-23 months that had mothers with low levels of education. They also found neonatal mortality to be significantly less sensitive to maternal education than mortality among children aged 1-23 months. This is consistent with Hobcraft et al. (1984) who found an increasing impact of mother's (and father's) education on child mortality as children grow older. In the same study, as well as in Caldwell (1979), it is also found that even for very low levels of education, positive effects on child survival are observed. While some of the effects of education on child mortality found in the literature are likely to be associated with household wealth (Houweling and Kunst, 2010), there are many plausible explanations for why education has additional direct effects on child mortality. One is that educated women have better knowledge about prevention and treatment of illness. Hobcraft (1993) found that while the evidence is strong for higher prevalence of diseases among the children of uneducated mothers, the difference is much larger in treatment of diseases, when investigating data from 25 developing countries. In his study, educated mothers proved to have better knowledge about illnesses and were more likely to take their children to a health facility when falling ill. Moreover, he found evidence that educated women were more likely to make sure that their children were fully vaccinated, to receive prenatal care and be vaccinated against tetanus during pregnancy, and to give birth in the presence of skilled personnel. Hobcraft also found evidence that children of educated mothers were less likely to be undernourished in terms of stunting[14]. To what extent these differences translate into improved chances of survival for the children of educated mothers is not estimated in the article, but it is reasonable to assume that the two are positively related (Hobcraft, 1993).

[14] Stunting is a key indicator of chronic malnutrition and a child is defined by the World Health Organization (2011b) as stunted if her "height-for-age is less than -2 standard deviations of the WHO Child Growth Standards median".

4.4.3 Maternal factors

In addition to her education level, a number of other characteristics of a child's mother are found to have significant effects on survival. *Age at childbirth* is one of these. In a study of determinants of child mortality in Malawi, Manda (1999) found risk of infant mortality to be considerably higher for relatively young and relatively old mothers. This could be explained by biological factors. While young women in their teenage years have increased risks of complications during delivery because they are not fully developed, older women have a higher risk of complications because their bodies and reproductive systems are "worn". Fertility characteristics such as the number and frequency of child births and the number of the child in the succession of births are also found to have significant effects on child survival (Manda, 1999). These findings could also be accounted for by biological factors. If child births occur with brief intervals, this could drain the mother of nutritional and reproductive resources and give her weak children, more likely succumb to infections. *Under- and malnutrition* is one of the largest direct causes of child deaths worldwide (World Health Organization, 2009a) and is closely related to another important maternal factor, namely *breastfeeding*. Breastfeeding is found to have a significant bearing on child mortality (Manda, 1999), and this can be explained by the fact that breast milk is very nutritious, contains antibodies that help protect the infant from infection causing diseases like diarrhea and pneumonia (World Health Organization, 2011a) and is sterile (Palloni and Tienda, 1986). All of these properties make breastfeeding especially important in low income settings where good alternatives to breast milk are not readily available and households have limited access to clean water. Like education, the maternal factors are generally worse for the least wealthy, with one exception: breastfeeding. Poor women with little or no education are usually breastfeeding their children for a longer period than the wealthier women (Houweling and Kunst, 2010, Alemayehu et al., 2009). Last, but not least, *use of health services*, especially those directly related to pregnancy and delivery are important maternal determinants of child health. Closely related to this, is access to and quality of health services which are also important determinants of child health.

4.4.4 Access to and quality of health services

In the literature there seems to be a broad consensus on the significance of health services, especially antenatal[15], delivery, postpartum[16] and immunizations services, to child health, but empirical evidence on the subject is scarce. Some studies do, however, exist; Lavy et al. (1996) find a negative relationship between access to public child health services and child mortality in Ghana, and a recent study of neonatal mortality in China shows large positive effects of delivering in hospital compared to delivering at home. This study also reveals large differences in chances of survival between hospital deliveries in urban and rural regions, the rural mortality being much higher (Feng et al., 2011). The reasons for why access to and quality of health services matter are obvious. They play an important role in both prevention (for example immunization and health education of patients) and treatment of illness. Because most maternal and child deaths occur during or shortly after delivery, antenatal care, which serves to discover and treat micro-nutrient deficiencies and assess other risk factors, is of crucial importance to survival of both mother and child. So is the attendance of skilled personnel that have access to necessary equipment and medicines needed if complications occur during delivery. Follow-up services in the period after delivery are also central to detect and treat infections and other conditions that are likely to occur to mother and child postpartum (World Health Organization, 2009b).

4.4.5 Geographic variables

Many researchers have found significantly lower child mortality rates in urban compared to rural areas in developing countries (Pandey, 1998, Wang, 2003). This probably reflects differences in underlying socioeconomic factors (income, education level, household demography) that are often large between urban and rural areas. In a study of urban-rural differences in child mortality in Brazil, Naryan Sastry (1997) finds that in addition to underlying socioeconomic differences, rural-urban differences are also explained by community characteristics such as the quality of social services (including health service

[15] Antenatal care constitutes screening for health and socioeconomic conditions likely to increase the possibility of specific adverse pregnancy outcomes, providing therapeutic interventions known to be effective; and educating pregnant women about planning for safe birth, emergencies during pregnancy and how to deal with them. (World Health Organization, 2011b)

[16] The postpartum period starts about an hour after delivery of the placenta and last for six weeks. Postpartum care addresses the special needs of the mother and child during this phase and should include the prevention and early detection and treatment of complications and disease, and the provision of advice and services on breastfeeding, birth spacing, immunization and maternal nutrition (World Health Organization, 1998)

delivery), infrastructure and sanitation, education and geographic and epidemiological environment.

4.5 Research question and hypotheses

Both the theory presented and the limited empirical literature on child health and mortality tell us that this is a complex matter, involving a large spectrum of factors ranging from medical to geographic and education variables. Nonetheless, some factors of particular importance and interest have been singled out. In seeking an answer to my research question it would be appropriate to formulate some hypotheses about these factors. The research question that I want to investigate is:

What are the determinants of child mortality in Angola?

The first factor that I want to make a hypothesis about is wealth. I expect *wealth to be negatively related to child and infant mortality* in the sample. This will be referred to as hypothesis H1a from now on. I also expect to find differences across all levels of wealth, but not necessarily a linear relationship. According to Houwling and Kunst, the countries with the highest child mortality rates in Sub-Saharan Africa exhibit high mortality rates in all segments of the population except from in the elite. This, taken in conjunction with the income distribution in Angola, gives rise to expecting a *larger gap between the elite (for instance the richest quintile) and the rest of the population than between the other wealth groups* in the sample. This will be referred to as H1b.

The second factor that has emerged as essential to child health is education. I expect to find a *negative relationship between under-five and infant mortality and education* (H2a). Based on findings in the empirical literature, I also expect *the effect of education to be larger on under-five than infant mortality* (H2b).

Thirdly, I hypothesise that *under-five and infant mortality is significantly different between groups of households belonging to different health facilities* (H3a). Provided this is true, I expect that *health facility characteristics (reflecting access to and quality of services) to be negatively related to under-five and infant deaths* (H3b).

4.6 Chapter summary

This chapter commenced with a very brief examination of the history of mortality that extracted improvement in nutrition, economic growth, improvements in public health and medical inventions as the main sources of the historical decline in mortality during the last two centuries. Then, a formal economic model was developed. This gave valuable insights into the household decision about fertility and the trade-off between the number of children and their level of human capital. The subsequent section introduced two quite similar frameworks for the analysis of child health that both emphasized the role of non-medical determinants such as education and economic development and how they can be modelled to work on child health through proximate determinants. A short review of the empirical literature gave us evidence supporting the significance of the variables in the economic model, and some additional factors were pointed out. As illustrated in the analytical frameworks, child mortality is a complex affair, affected by a wide range of factors. This makes the study of child mortality laborious, but all the more important.

5 Data and variables

In this chapter data and variables for the analysis are described.

5.1 The data

The analysis is based on data collected in collaboration between Christian Michelsens Institute (CMI) for international development in Bergen and Centro de Estudos e Investigação Científica (CEIC) in Luanda. The aim of the collection was to provide information about health service quality and availability in Angola, particularly how health seeking behaviour is related to socio-economic status (Frøystad et al., Forthcoming 2011).

Figure 5.1: Angola and its regions

The data consists of three separate surveys: a health facility survey, a household survey and a health worker survey. These were conducted in two provinces, Luanda and Uíge, in April-May and September-October 2010 respectively. The Luanda province is situated along the coastal line to the Atlantic Ocean and holds the capital with the same name. While all households in this province are classified as urban, more than half of the households surveyed in the province of Uíge are rural. Uíge is an inland province and lies in the North of Angola, bordering the Democratic Republic of the Congo. The locations of the regions are illustrated in figure 5.1. Uíge is by far the largest of the two provinces in terms of geographical size

(58 000 vs. 2 418 km^2 (Frøystad et al., Forthcoming 2011)), but Luanda, which is the most densely populated province in Angola, has a population more than four times larger than Uíge. In each of the two regions, three municipalities were chosen for conducting the survey. The selection had the purpose of ensuring variation in the income level. In Uíge, the municipalities selected were geographically close to each other due to the rainy season making the more distant locations inaccessible. Table 5.1 lists the municipalities by region.

Table 5.1: Municipalities by region

Luanda	*Uíge*
Cazenga	Uíge
Kilamba Kiaxi	Quitexe
Ingombota	Puri

The *health facility survey* provides information about health facility "infrastructure" such as services offered, health service outputs, user fees, equipment, infection control, health workforce, drugs and commodities and laboratory. In total, 40 public health facilities were surveyed. Six (three in each region) of these were hospitals chosen in accordance with the purpose of the study. 19 health centres (twelve in Luanda and seven in Uíge) and 15 health posts (five in Luanda and ten in Uíge) were selected randomly. 25 households belonging to each facility's catchment area were then included in the *household survey*. About half of the households selected were situated in close vicinity of the facility; the rest had more distant locations of approximately five km away. This survey contains information about household characteristics such as education and asset ownership, access to health services, recent illness episodes, health seeking behaviour and maternal health issues. The total number of households in the sample is 999. In the *health worker survey*, patient case simulations (PCS) were used to map health workers skills and abilities to diagnose illness. One health worker, preferably the technical worker in charge of services, was surveyed at each facility and the sample thus contains 40 health worker assessments. The number of facilities and health workers are equally distributed in number between the two provinces, but the number of households in Uíge exceeds Luanda by one. Table 5.2 shows total number of surveys conducted and distribution between the regions.

Table 5.2: Number of households, facilities and health workers by region

Region	Households	Facilities	Workers
Luanda	499	20	20
Uíge	500	20	20
Total	999	40	40

The sample of households can be thought of as *stratified*. That is, the population that the sample is drawn from is divided into different subgroups where each unit can belong to one, and only one, group, and samples are drawn from the different subgroups to assemble a population sample. In the sample, subgroups of the households are classified by health facility catchment area. Connected to each health facility there are two subgroups: one consisting of households close to the facility and one consisting of households with more distant locations. Since the health facilities are randomly selected within the municipalities, the households can be considered as a random sample from the population within these. The municipalities themselves are not randomly selected *per se*, but are chosen, on the basis of knowledge about the local context, to reflect income differences within the regions. Thus, the municipalities might not be representative for their regions in a statistical sense, but because of the careful selection the results obtained here could probably be generalized to account for the two regions. Region results are however not representative for the whole Angolan population. Uíge was amongst the most heavily affected by the civil war and its health situation is probably worse than in the less affected regions. Luanda is the economic centre of the country and is likely to show higher levels than the majority of regions in terms of health. I will use the three surveys to investigate determinants of child health in the regions of Luanda and Uíge in Angola.

5.2 Variables

In this subsection the variables in the analysis are presented. I start with the dependent variable and then proceed to the independent variables grouped in three categories; household characteristics, health facility characteristics and health worker characteristics.

5.2.1 The dependent variable; child deaths

The conventional measure of under-five child mortality in medical literature is child mortality rate, that is, number of under-five deaths per 1000 live births. Such measure can be constructed from questions to the households about number of children born and number of

children under the age of five that have passed away in the course of the five years preceding the survey. Similarly, infant mortality, i.e. number of under-one deaths per 1000 live births, can be constructed. Using these measures as dependent variables in the analysis will, however, make regression coefficients hard to interpret because in reality, there will be two dependent variables on the left hand side; number of children dead and number of children born. For this reason, I will use the number of children dead as a dependent variable, and control for number of children born by including it among the explanatory variables. I will also estimate models with number of infants dead as the dependent variable. It should be noted that the number of under-five deaths includes number of infant deaths in addition to deaths occurring between the age of one and five years. The rationale for estimating infant death separately is that the determinants of mortality differ between the age groups. Ideally, I would also run separate regressions for children aged one to four years, but our data do not contain a sufficient number of observations in this group.

In the analysis, I only include households where at least one child was born the last five years. This reduces the sample from 999 to 946 households. 467 of these are situated in the Luanda region and 479 in Uíge. In the included households, a total of 1995 children were born and 184 passed away in the course of the five years preceding the survey. Of the children passing away, 109 died before the age of one. That the largest fraction of children died during their first year, and especially first month, of life is in line with findings in the empirical literature.

Table 5.3: Number and frequency of children born

Number of children born	Frequency (number of households)
1	304
2	369
3	189
4	51
5	23
6	6
7	2
8	1
9	1
Observations	946

Table 5.3 presents number and frequency of children born and table 5.4 reports number and frequency of children passing away in the households the last five years preceding the survey.

Table 5.4: Number and frequency of children under-five and infants dead

Number of children under-five dead	Frequency (number of households)	Number of infants dead	Frequency (number of households)
0	791	0	849
1	109	1	88
2	22	2	6
3	7	3	3
5	2	-	-
Total	931	Total	946

Both under-five and infant deaths are count variables, i.e. variables that can take on nonnegative and (often) relatively few integers, including zero. Taking a closer look at table 5.4, I observe that in a few households two or more children under-five died in the last five years. The same is the case for infant deaths, where an even smaller fraction of households experienced two or more deaths. Having only a few observations in the upper range of the number of deaths could be a potential problem when analysing the data. When a small number of observations deviate from the rest of the sample, they could have a disproportionate influence on the regression results compared to when the observations are excluded. Observations like these are often referred to as *outliers*. In some cases, outliers are given disproportionate weight in regressions because one or very few extreme observations alter the results and generate over- or under-estimated outcomes. The effects of the outliers on the regression results are most severe when the number of explanatory variables is low. A solution to the outlier problem is to merge the groups with few observations into one larger group. For example, the groups of two, three and five children under-five dead and the groups of two and three infants dead could be added together in one group. The dependent variables would then have three possible outcomes: zero children dead, one child dead and two or more children dead. This approach will not solve the problem completely, but reduce it.

Alternatively, the dependent variables can be converted to have only two possible outcomes, namely “zero” and “one or more” or “no deaths” and “death(s)”. Variables that only have two possible outcomes are commonly referred to as binary variables. Table 5.5 describes the frequency of households in the two groups when the dependent variables only distinguish between no deaths and one or more deaths.

Table 5.5: Number and frequency of under-five and infant deaths with binary outcome

Number of children under-five dead	Frequency (number of households)	Number of Infants dead	Frequency (number of households)
0	791	0	849
1 or more	140	1 or more	97
Total	931	Total	946

5.2.1.1 Limitations of the dependent variables

There are several potential problems with the dependent variables that I use here. One is related to the reporting of children born and deceased. For a start, children that die immediately after birth might not be counted in the sample due to difficulties distinguishing between live births and stillbirths[17]. Furthermore, the households are asked to report events as far back as five years before the survey. It might be difficult for the respondents to remember exactly when a child was born, and, in the case of death, the exact age of the child. This could give rise to measurement errors in the dependent variables. Ideally these errors occur randomly across households and are not related to any of the independent variables. In that case, the measurement error will not affect the unbiasedness of the estimates I obtain. However, it is not unlikely that misreporting is related to some of the explanatory variables. For example, it might be the case that the more educated households are better in keeping track of time, and therefore report child-births and deaths with more accuracy than their less educated counterparts. If our data are characterized by such relationships, our estimates will be biased.

Another important issue regarding the dependent variables arises from the fact that the survey did not follow all children until they reached the age of five. The sample contains some observations of young children that, unfortunately, are likely to die/to have died since the survey was conducted, but that are reported to be alive in the survey. The children born in the last five years that passed away before the age of five, but to whom death occurred after the survey was conducted are not reported as dead in the sample. Number of children dying before the age of five is therefore underreported in the sample. The same problem applies to infant deaths, but to a less extent.

[17] Definition varies across countries. For international comparisons WHO recommends using the following definition: stillbirths as pregnancy loss after 28 weeks of pregnancy or at birth weight of 1000 g

5.2.2 Independent variables

The independent variables in the analysis are chosen on the basis of theory and findings in empirical research presented in the previous chapter. The data consist of more than 1400 variables, most of which can be assumed to affect the probability of a child dying in a household. Selection of variables is therefore a difficult task and, though made to the best of knowledge, the choices here could always be criticized and argued against.

A challenge is that while the variables in the data are mostly reported on the household-level, empirical findings frequently point out maternal (individual) factors such as education of the mother to be the most important. Furthermore, the lack of individual information about the children is a clear limitation of the data. As a consequence of this, I might have a problem of *omitted variables*. A variable is *omitted* if it belongs in the true population model, but is not included in the empirical model estimated because it is unobservable or because of limited data. For instance, the intervals between the deliveries a mother goes through, birth spacing, and the birth order of a child has been found in the literature to have significant effects on survival. Unfortunately, our data do not contain information on these variables. Birth spacing and birth order are then variables entering the true population model of child deaths, but that are omitted from the estimated model. Omission of these variables will lead to biased estimates unless they are uncorrelated with all other independent variables in the estimated model.

As with the dependent variables, the independent variables might also be characterized by measurement errors. For example, I would like to include the age of the mother at the time the child was born, but in the data I only have information about the mother's age at the time that the survey was conducted. Because the time since the woman last gave birth differs across households and because some women had more than one child the last five years, it is difficult to correct for this. The measurement error in the variable is then the difference between the age reported in the data and the age at which the woman last gave birth. The *classical errors-in-variables assumption* assumes the measurement error to be uncorrelated with the unobserved variable (here: mother's age at the time she last gave birth). When this assumption

is satisfied, the OLS estimator will be biased and inconsistent[18]. It can be proved that, if the coefficient of the variable that we want to estimate is positive, the OLS estimator will have a downward bias, meaning that it will underestimate the effect of the variable. This is commonly referred to as *attenuation bias* in the theory of statistics (Wooldridge, 2006). On the other hand, if the classical errors-in-variables assumption does not hold and the measurement error is uncorrelated with the observed variable, the OLS estimator is consistent. A third possibility is that the measurement error is correlated with both the observed and unobserved variable, in which case the OLS estimator is inconsistent.

5.2.2.1 Household characteristics

This group of variables contains socioeconomic indicators and health seeking behaviour of the households.

Wealth

Finding a good measure of wealth or income is a challenge. Traditionally it has been measured using monetary variables, but collection of information on income, household expenditure and household consumption is a demanding task and requires a lot of resources, especially in developing countries. An alternative way of measuring wealth is collecting data on variables reflecting living standards that can be directly observed by a surveyor. Examples of such variables are durable assets like television and washing machine, infrastructure (water, sanitation etc.) and housing conditions. This information can be used to construct an asset-based wealth index. To account for the different items having different significances to living standard, the variables need to be weighted. An increasingly used method of deriving these weights is principal component analysis (PCA) (Vyas and Kumaranayake, 2006, Wang, 2003).

PCA is a statistical method that allows us to reduce the number of variables in the dataset and get a smaller set of summary indicators. Dunteman (1989: p.10) describes the method the following way: "Principal Component Analysis searches for a few uncorrelated linear combinations of the original variables that capture most of the information in the original variables". PCA has a number of applications ranging from biology, medicine, chemistry,

[18] An estimator is unbiased if its expected value is equal to the mean value in the population. An estimator is consistent if its expected value converges to the mean value in the population when the number of observations in the sample is increasing towards ∞ (Wooldridge 2006).

psychology and geology to social sciences. When using the method to derive a wealth measure in economics, the first step is to choose some variables that are believed to reflect aspects of wealth and to capture inequalities between households. The second step is to compute the covariance matrix and the corresponding eigenvalues and eigenvectors[19] with the help of a statistical software package (Smith, 2002). Principal components are then derived. They are weighted linear combinations of the included variables where the weights of each component is given by the eigenvectors of the correlation matrix. Each principal component is uncorrelated with all other principal components (Vyas and Kumaranayake, 2006). This method generates as many principal components as there are variables, but the aim of the procedure being to reduce the number of variables, only one or a few are usually kept in the further analysis. There exists a number of ways of deciding how many and which components to retain. One of them is the eigenvalue-criterion. It states that all principal components with an eigenvalue greater than one should be retained. The eigenvalue being the variance of a principal component, this means that only components that account for a large fraction of the variation in the variables are kept.

The data contain an income variable for the households, but this is reported as missing for as many as 695 of the 946 households. Hence, even if income were a good measure of socioeconomic status, it would not be useful in our analysis. I therefore construct a wealth index using Principal Component Analysis on dummies for roof, floor and wall material, for household assets, for whether or not the household owns land, and, in the case it does: how much. The principal component can either be used directly in the regression, or to construct wealth quintile dummies. According to hypothesis H1a, I expect our principal component(s) to be negatively related to under-five and infant mortality in the sample.

Education

In the dataset, education is divided into six categories, and the number of adults (>21 years old) with a specific education category as their highest attained level is reported. The categories are the following: university, medium, professional training, secondary school, primary school and not completed primary school. The data do not contain individual information about the education level of the mother or the father in the household. For the

[19] Definitions from Simon and Blume (1994). Eigenvalue: Let A be a square matrix. The scalar r is an eigenvalue of A if, when it is subtracted from the matrix converts it into an identity (singular) matrix, I. Eigenvector: Let A be a square matrix and r be an eigenvalue of A. The non-zero vector **v** is then an eigenvector of A if (A-rI)**v**=0.

analysis, I construct dummies for education levels that take the value 1 if at least one adult in the household have attained that level, but no adults have attained a higher level of education, and 0 otherwise. Because there are so few observations in the three lowest groups of education, I construct one common dummy for these three. Table 5.6 describes frequency of households in each category of education.

Table 5.6: Education level and frequency

Education level	Number of households
No education	29
Primary	60
Secondary	141
Professional training	300
Medium	303
University	113
Total	946

Number of children

Number of children born in the household the last five years is reported in the survey and constitutes an important control variable; it sets an upper limit to possible number of deaths in a household and effects resources available to each child. I expect to find a positive relationship between number of children born and infant and under-five deaths in the sample. Number of children aged 6-16 and 17-21 in the household is also reported in the survey. The effects of these variables on child mortality are more difficult to predict. While more children older than five years old in the household create more competition for scarce resources, it could also increase the production of the household if the children are working.

Maternal age

As explained above, the age of mother reported in the data might be biased by measurement errors, because what I am really interested in is the age of the woman at the time she gave birth to each of her children, not her age at the time of the survey. The interpretation of the coefficient of this variable should therefore be conducted with care. In addition, there are reasons to believe that the effect of mother's age is not linear, but U-shaped, with high mortalities for relatively young and relatively old mothers and lower mortalities in the intermediate age groups. I control for this by including a quadratic age term among the explanatory variables. If there is indeed a U-shaped relationship between mortality and

mother's age in the sample, the age and age squared should have a negative and positive sign respectively.

Food available

Nutrition is a central issue in child health. In the survey, households were asked if they always had enough food for the household members. Access to food in a household is an evident aspect of nutrition and I therefore expect the dummy for households always having sufficient food to be negatively related to number of infant and under-five deaths.

Use of health services: Antenatal, delivery and postpartum care

Questions about antenatal care attendance, place of-, and personnel attending-, delivery, and postpartum follow-up were asked to the woman in the household who most recently gave birth. I expect the households where the women attended antenatal care to experience fewer deaths than the ones that did not receive these services. I also expect to find that women who were attended by skilled medical personnel during their most recent delivery have lower expected numbers of infant and under-five deaths than those who were not, and that postpartum service is negatively associated to deaths. Women giving birth in health facilities are also expected to do better in terms of number of household infant and under-five deaths. The effect of antenatal, delivery and postpartum services should all be more significant for infant mortality than for under-five.

Breastfeeding

Whether or not the last childbearing woman in the household breastfed her child, and, if she did; how many months, is reported in the data. These variables might affect my dependent variables in two ways. The first way is through the direct nutrition- and immune system-effects acting on the children that are actually being breastfed. The other is through the indirect effects on the other children in the household. If the mother does not breastfeed, it might increase the chances of her getting pregnant sooner after giving birth. Shorter birth intervals and more children are both factors that could be thought to affect the survival chances of the children in a household negatively.

Urban/Rural geographic position and region

For all households, region of residence and whether the household is situated in an urban or a rural area is reported. All of the 467 households in Luanda and 190 in Uíge are defined as urban. The rest of the 289 households are rural households situated in Uíge. The following three dummies can then be used to capture regional and urban/rural differences.

- Rural
- Urban-Uíge
- Luanda

I expect both the urban-Uíge and the Luanda dummy to be negatively associated with the number of child and infant deaths in the households when rural is used as a reference dummy. I also expect the coefficient of the Luanda-dummy to be larger in size than the one of urban-Uíge.

Perception of quality of health facility

How the household perceive quality of health services is likely to affect to what degree its members seek medical help when falling ill. Household perceptions probably also reflect, at least partly, the real quality of the health services. Inclusion of a perception variable could therefore capture differences in mortality due to variations in the actual quality of health services, but could also reflect differences in use of health services. Households were asked to rank the quality of the nearest public health facility and choosing among the following alternatives: very low, low, medium quality, high quality and very high quality. On the basis of this information, I construct three dummies; one for households ranking quality of services provided at the closest public health facility as very low or low, one for medium and one for high or very high ranking. I expect household that rank quality as low or very low to have a larger probability of infant and under-five deaths than those that rank it high or very high.

5.2.2.2 Health facility characteristics

This group of variables hold information about health facility infrastructure, services offered, equipment and staff available. In the sample each household is connected to the closest public facility surveyed, but there is of course no guarantee that this is the facility they normally use. However, I can control for this, because questions about whether or not the household normally uses the closest public health facility were asked in the survey. Regardless of whether the household is using the closest facility or not, health facility characteristics are

likely to be related to child health outcomes, because they are indicators of quality of institutional infrastructure in the area in which they live.

Electricity

Availability of electricity in the health facilities could be directly associated with quality of health service. For example, electricity is required to keep medicines refrigerated and to use certain diagnostic tools. Electricity might also be looked upon as a more general indicator of quality and infrastructure. The survey collected information about whether or not electric power was continuously available in the facilities during all opening hours the past week before the survey. I expect the households in the catchment area of facilities that are reported to have electricity available during all opening hours to be negatively related to number of infant and under-five deaths.

Services available

What child and maternal services are available at the closest public health facility of a household could affect the number of children dead in the households directly, if it leads to conditions (not) being prevented or treated. Immunization, vitamin A supplementation, antenatal, delivery and postpartum, malaria, TB and HIV/AIDS services are of particular significance to child health. The health facility survey include dummies for whether the health facilities offer these services, and it is expected that these will all enter the analysis of child deaths with a negative sign.

Equipment

How well health workers are able to help their patients is, among several other things, determined by availability of equipment. Like water and electricity, equipment available at a facility could also be a more general indicator of its quality. The data reports availability of 16 equipment items. Of these, stethoscope and thermometer can be singled out in the analysis to separate the poorly equipped facilities from the rest, because these are basic tools in making diagnoses. Refrigerator, on the other hand, is a more advanced piece of equipment that might not affect diagnostics directly, but influences other aspects of quality such as storage of medicines. This item is also an indicator of infrastructure because it requires electricity and is therefore useful to separate the most equipped facilities from the rest. Assuming that availability of equipment affects possibilities for health workers to do a good job, I expect to

find a negative relationship between the availability of equipment in a facility and the number of child deaths in its catchment area.

Health workforce

The more training the workers at a health facility have, the more capable they should be to deal with complicated cases. A facility with more skilled personnel should therefore deliver better health services and provide better health outcomes than those with less trained workers. In the data, number of health workers employed at each facility is reported by staff categories. There are eleven different categories of staff, but I include only two in the analysis, namely *midwife* and *specialized doctor.* The availability of a midwife is particularly important to child health outcomes, whereas availability of a specialized doctor could be a rough indicator of quality of the facility. For the analysis, I construct dummies for the two staff categories, and I expect the number of child deaths in catchment areas of facilities that have a midwife and/or specialized doctor in its staff to be lower than in catchment areas of facilities that don't.

5.2.2.3 Health worker characteristics

In addition to equipment and medication availability in the facilities, the abilities of the health workers to reach correct diagnoses, prescribe the right treatments and to educate patients in order to prevent future illness incidents are important aspects of the quality of health services. This group of variables contains information about knowledge and skills, as well as personal characteristics like age and formal training, of the health workers.

Performance in patient case simulations

In the health worker survey, patient case simulations were used to measure performance. A member of the survey team played the part as a patient with a specific condition and the health worker was asked to do a consultation on him/her. Because the actor did not actually have the condition, he or she, the "patient", would tell the health worker about his symptoms in response to questions asked. The health worker stated what examinations he wanted to perform and the patient answered what he/she would find. Another member of the surveyor team noted down and reported what the health worker did during the consultation. The following five patient case simulations were presented to each health worker: child <5 years with malaria and anaemia (difficult), child <5 years with acute diarrhoea and dehydration (difficult), child <5 years with pneumonia (simple), woman with pelvic inflammatory disease (difficult) and man with pulmonary tuberculosis (simple).

The performance of the health worker was assessed according to a predetermined list of relevant items, divided into the following six categories: history taking, physical examination, tests, diagnosis, treatment and health education. For the analysis, variables for the total number of items performed in all of the five patient case simulations in each of the categories were constructed. In the history taking-category, 13 items were listed for the first patient case, eleven for the second and fifth, eight for the third and ten for the fourth. Hence, the highest score a health worker can get in the history taking category is 53. Three of the performance categories, history-taking, physical examination and correct diagnosis, are included in the analysis. These are the measures thought to best reflect knowledge and abilities of the health worker and to be of the highest relevance to child health.

I now proceed to describe the econometric models that will be used to estimate the relationship between the dependent and the explanatory variables in the analysis in chapter 8.

6 Empirical methodology

As pointed out, the dependent variables that I want to model in the analysis are count variables that take on a limited number of non-negative values. Using ordinary least squares (OLS) estimation on count variables is problematic because some of the assumptions of the model are violated. However, it is not obvious that another econometric model will do better. While most econometric textbooks argue that the linear regression is inappropriate for estimation of limited dependent variables such as binary and count data, others disagree and claim that linear regression models works just as well as others (Angrist and Pischke, 2009). According to Jeffrey Walker and Sean Maddan (2005), the debate as to whether to OLS is the most appropriate technique for count data has not yet been settled. For this reason, as well as for comparison, it is useful to estimate the OLS model in addition to other model(s). In this chapter, I present the three empirical models applied in the analysis in chapter 8, based on Verbeek (2008) and Wooldridge (2006). The choice of models is based on empirical literature on child survival and investigation of the data. For each model I describe the underlying assumptions and how the regression coefficients can be interpreted. The problem of clustered standard errors is discussed in section 6.4.

6.1 The ordinary least squares method of estimation

The ordinary least squares method of estimation is one of the most important in econometrics, and is used to estimate unknown parameters in a linear regression model. I want to model the expected number of infant and child deaths in a household conditional on a number of household, health facility and health worker characteristics.

Let *y* be number of children dead the in a household the last five years and $x_1, x_2, \dots, x_k$ be the household characteristics, health facility and health worker variables that we want to estimate the effects, $\beta_0, \beta_1, \dots, \beta_k$, of. A linear statistical model for the relationship between number of child deaths and the explanatory variables in the population can be specified as

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \qquad (6.1)$$

or, in matrix notation,

$$y = x_i'\beta + \varepsilon \quad (6.2)$$

where ε is an error term containing unobserved factors affecting y, and x_i is a vector of explanatory variables. Before deriving the OLS estimator for the β's, the assumptions of the model must be presented. The wording is taken from Wooldridge (2006: pp. 166-167)

OLS Assumption 1: Linearity in parameters
The model in the population can be written as $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$ *where* $\beta_0, \beta_1, \ldots, \beta_k$ *are unknown parameters (constants) of interest and ε is an unobserved random error or disturbance term.*

OLS Assumption 2: Random sampling
We have a random sample of n observations $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, \ldots, n\}$*, following the population model in Assumption 1.*

OLS Assumption 3: No perfect collinearity
In the sample (and therefore in the population), none of the independent variables are constant, and there is no exact linear relationship among the independent variables.

OLS Assumption 4: Zero conditional mean
The error ε has an expected value of zero given any values of the independent variables. In other words; $E(\varepsilon|x_1, x_2, \ldots, x_k) = 0$

OLS Assumption 5: Homoskedasticity
The error term ε has the same variance given any values of the explanatory variables. In other words; $Var(\varepsilon|x_1, x_2, \ldots, x_k) = \sigma^2$

OLS Assumption 6: Normality
The population error u is independent of the explanatory variables $x_1, x_2, \ldots, x_k$ *and is normally distributed with zero mean and variance* σ^2 *;* $\varepsilon \backsim N(0, \sigma^2)$

Under OLS Assumption 1 to 4, the OLS estimator is *unbiased*. An estimator is unbiased if its expected value is equal to the true, unknown, value of the parameter we want to estimate. Mathematically, the estimator W is an unbiased estimator of the parameter θ if:

$$E(W) = \theta$$

Adding OLS Assumption 5 to the four first, the OLS estimator is also what is referred to as the *Best Linear Unbiased Estimator* (BLUE). In this context, "best" has the significance "with the smallest variance". Thus, when Assumption 1 to 5 is satisfied, the OLS model obtains the unbiased estimator with the smallest variance of all unbiased estimators.

In the case that OLS Assumption 4 is violated, the OLS estimator will be biased. A violation of OLS Assumption five on the other hand, does not lead to unbiasedness, but to a situation where the OLS estimator is no longer the best estimator in the sense that it is not the one with the smallest variance. Together, OLS Assumptions 1 to 6 is called the *classical linear model* (CLM) assumptions. The addition of OLS Assumption 6 to the first five assumptions is necessary to draw sound, reliable conclusions from the sample (statistical inference). On the basis of the classical linear model assumptions I now derive the OLS-estimator.

6.1.1 Deriving the OLS-estimator

For simplicity, I use a linear model with only one explanatory variable, β_1, and a constant term β_0. The equation for an individual in a given population can be written as

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \tag{6.3}$$

Rearranging (6.3), the error term, ε_i, can be expressed as:

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_1 \tag{6.4}$$

The Ordinary Least Squares estimator is obtained by *minimizing squares*, more specifically, by minimizing the sum of the squared error term over all *n* observations in the sample. Because the true values of the parameters β_0 and $\beta_1$are unknown, these are substituted with $\tilde{\beta}_0$ and $\tilde{\beta}_1$ in (6.4) to indicate that these are estimated values for the parameters. The squared sum of error terms is then given by

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i) \tag{6.5}$$

And the minimization problem is

$$\min_{\tilde{\beta}_o, \tilde{\beta}_1} \sum_{i=1}^{n} \varepsilon_i^2 \tag{6.6}$$

To solve this, we partial derivate (6.6) with respect to $\tilde{\beta}_0$ and $\tilde{\beta}_1$ respectively, and equal the expressions to zero:

$$\frac{\partial}{\partial \tilde{\beta}_0} \sum_{i=1}^{n} (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2 = 0$$
$$\frac{\partial}{\partial \tilde{\beta}_1} \sum_{i=1}^{n} (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2 = 0 \quad (6.7)$$

Solving these one at the time, starting with $\tilde{\beta}_o$, the estimators, $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are obtained:

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x}$$
$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_1 - \bar{x}_1)\,(y_i - \bar{y})}{\sum_{i=1}^{n} (x_1 - \bar{x}_1)^2} \quad (6.8)$$

Where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \qquad \bar{x}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (6.9)$$

6.1.2 Interpretation of OLS regression coefficients

Coefficients obtained with OLS estimation are straightforward to interpret. They are marginal effects indicating the change in the dependent variable (conditional on x_i) resulting from a one unit change an explanatory variable, when holding all other variables fixed. When the independent variable is a dummy, its coefficient measures the average difference in the dependent variable between groups for which the dummy takes the value 1 and groups for which the dummy takes 0. For instance, *luanda* is a dummy that takes on the value 1 for household situated in the Luanda region and 0 for households situated in Uíge. Then the coefficient of the explanatory variable *luanda* is then measuring the average difference in number of under-five or infant deaths between households in the two regions, holding all other explanatory variables constant. In mathematical terms, the OLS coefficient can be expressed as $\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \tilde{\beta}_k$

6.1.3 The OLS estimator and count variables

Having presented the underlying assumptions that need to be fulfilled in order for the OLS-estimator to be BLUE, I can now make an argument for whether or not it is applicable to the

data. The dependent variables are both count variables for which a large fraction of the observations in the sample takes the value zero. While the assumption about normally distributed error terms (which also entails the dependent variable conditional on $x_1, x_2, \ldots, x_k$ being normally distributed) is straightforward to make when the dependent variable is continuous and takes on all values, this is problematic when dealing with count dependent variables. The explanation for this is twofold. Firstly, for most of the observations in the sample, the independent variables take zero. Secondly, the dependent variable outcome is limited to nonnegative values. Both of these properties contribute to the dependent variables not being normally distributed. When dependent variables are not normally distributed, the same must be true for the error terms. Graphically, the distribution of infant and child deaths in my sample have only one "tail", as opposed to the normal distribution that has two. The consequence of the normality assumption being violated is difficulties with making reliable statistical inference from the sample. However, when the sample is sufficiently large, the assumption can be dropped[20] (Wooldridge, 2006).

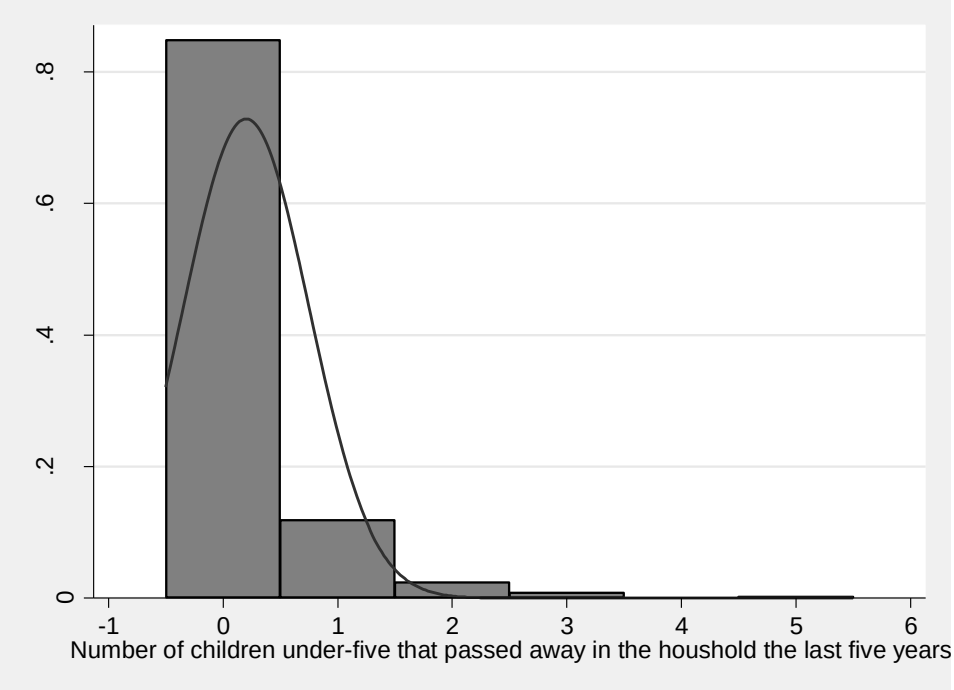


Figure 6.1: Distribution of number of under-five children dead and the normal distribution

[20] According to the *central limit theorem*, all sufficiently large samples drawn from a population with finite mean and variance will be approximately normally distributed, regardless of the shape of the distribution in the population. Distributions that converge towards the normal distribution as the number of observations is increasing towards ∞ are referred to as *asymptotic normal distributions* (Thomas, 2005).

Figure 6.2: Distribution of infants dead and the normal distribution

Figures 6.1 and 6.2 above give graphical representations of the distribution of the independent variables compared to calculated normal distributions for the sample. The black solid curves indicate normal distributions and the grey columns depict the distribution and density of under-five and infant deaths. The graphical illustrations is clearly showing that under-five and infant deaths are not normally distributed in our sample.

Another classical linear model assumption that is violated using count data dependent variables is Assumption 5: homoskedasticity. Because the dependent variable only takes on a limited number of values, the outcome of the error term, ε_i, conditional (on x_i), is also limited. Thus, the conditional variance of ε_i is not constant and independent of the explanatory variables. The consequence of this assumption being violated is that the OLS-estimator is not the unbiased estimator with the least variance. Hence, the estimator obtained when applying the OLS-method on count data is not BLUE.

Furthermore, when performing regressions on count data, what we really do is to estimate the probability of a given outcome of the dependent variable, i.e. $0 \cdot P(y_i = 0|x_i) + 1 \cdot P(y_i = 1|x_i) + 2 \cdot P(y_i = 2|x_i) + \cdots$. A basic characteristic of probabilities is that they

always lie between zero and one in value. This is not ensured in the linear OLS model where outcomes can be both negative and larger than one (Verbeek, 2008).

To sum up, when applying the OLS method of estimation to dependent variables that are counts, both Assumption 5 and Assumption 6 are violated. Because of this, the OLS-estimator will not be BLUE and, unless the sample is sufficiently large, not knowing the underlying distribution of the data makes statistical inference from regression results less exact. In addition, outcomes are not guaranteed to be between zero and one. However, it should at this point be emphasized once more that the application of the OLS method of estimation on count data will yield unbiased estimators. I now proceed to present the Poisson model which is especially designed to address the problems of count data estimation.

6.2 The Poisson model for count variables

To avoid negative predicted probabilities, the expected value of y_i is modeled as an exponential function (which is nonzero for all possible values of x_i) in the Poisson model:

Poisson Assumption 1
The expected value of y_i conditional on a given set of characteristics x_i, is given by $E(y_i \mid x_1, x_2, \dots, x_k) = exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$ or, in matrix notation

$$E(y_i|x_i) = exp\,(x_i'\beta) \tag{6.10}$$

where x_i is a vector of the explanatory variables, $x_1, x_2, \dots, x_k$, and β is a vector of the unknown parameters that we wish to estimate, $\beta_0, \beta_1, \dots, \beta_k$.

Taking the logarithm of (6.10) obtains:

$$ln\,E(y_i \mid x_i) = x_i'\beta \tag{6.11}$$

Thus, the logarithm of the expected value of y conditional on x_i is a linear function. In order to estimate the probability of a given outcome of y, for example $P(y_i = 1|x_i)$, an assumption about the distribution of the dependent variable and error terms must be made. The normal probability distribution fits badly with count data, and we therefore make the assumption that the dependent variables are distributed according to the *Poisson distribution*.

Poisson Assumption 2

For a given x_i, the count variable y_i, is distributed according to a Poisson distribution. Let y be the number of times the event that we want to model happens. Then the probability mass function of the Poisson distribution is given by

$$f(y) = \frac{ex\,p(-\lambda)\lambda^y}{y!}, \qquad y = 0,1,2,\ldots \tag{6.12}$$

Where y! ,y factorial, equals $y \cdot (y-1) \cdot (y-2) \cdot (y-3) \cdot \ldots \cdot 2 \cdot 1$ and λ is the expected value of the number of times the event occurs.

Following the definition of the expected value of *y* conditional on x_i in Poisson assumption 1, we can then write $E(y_i|x_i) = exp(x_i'\beta) \equiv \lambda_i$. The conditional probability of y_i being equal to the value *y* can be expressed as

$$P(y_i = y|x_i) = \frac{exp(-\lambda_i)\lambda_i^h}{y!} \tag{6.13}$$

This equation forms the basis of Poisson regression model. The most common method of estimation of models with limited dependent variables is Maximum Likelihood Estimation (MLE). In short, MLE is a method of estimation based on the assumption that we know the distribution of y_i conditional on x_i. The unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$, are then estimated in such a way that the probability of the observed sample fitting the assumed distribution is maximized (Verbeek, 2008). All estimation results for the Poisson model presented in chapter 8 and the appendixes are obtained using the standard Poisson procedure in STATA.

When Poisson Assumption 1 holds and our sample of y_i and x_i is random, the maximum likelihood estimator produces a consistent, asymptotically efficient[21] and asymptotically normal distributed estimator of β.

A problem with assuming that the dependent variables are distributed according to the Poisson distribution is that imposes another assumption; the conditional variance of *y* must be equal to the mean

[21] An estimator is efficient if it is unbiased and no other estimator has a smaller variance. An estimator is asymptotically efficient if it converges to the true population mean when the number of observations increases towards ∞ (consistent) and no other consistent estimator is approaching the true value of the population mean at a faster rate when the number of observations is increasing.

$$Var(y_i|x_i) = E(y_i|x_i) = \exp(x_i'\beta) \quad (6.14)$$

(6.14) is referred to as *equidispersion*, and the assumption is often violated. If the conditional variance in a sample is larger than the conditional mean for all x_i, the distribution is characterized by *overdispersion.* Opposite, if the conditional variance of y_i is smaller than the conditional mean for all x_i, the distribution is *underdispersed*. When a sample is characterized by overdispersion the application of regular ML on the NegBin II model (that assumes the sample to have a negative binomial distribution), is an alternative to the Poisson Model. This model is often referred to as the Negative Binomial regression model and is similar to the Poisson model, but assumes overdispersion and that the variance is increasing in the conditional mean (Verbeek, 2008).

6.2.1 Interpretation of regression coefficients

The Poisson model estimates the expectation of the dependent variable conditional on a vector of explanatory variables, x_i, as an exponential function; $E(y_i|x_i) = exp(x_i'\beta)$. When x_{ik} is a continuous explanatory variable, the marginal effect of a one unit change in its value on the expected value of y_i is given by the partial derivative:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \exp(x_i'\beta)\,\beta_k \quad (6.15)$$

Because $\exp(x_i'\beta)$ is always nonzero and nonnegative, the sign of the marginal effect is also always the same as the sign of the coefficient β_k. The marginal effects can be calculated at the sample averages of the independent variables using STATA.

Alternatively, (6.15) can be transformed into a semi-elasticity. Using the fact that $E(y_i|x_i) = exp(x_i'\beta)$, the semi-elasticity can be written as:

$$\beta_k = \frac{\partial E(y_i|x_i)}{\partial x_{ik}} \frac{1}{E(y_i|x_i)} \quad (6.16)$$

β_k is interpreted as the relative change in mean expectation of the dependent variable conditional on a one unit change in the k^{th} explanatory variable (Cameron and Trivedi, 1998).

In the case that x_{ik} is a dummy variable with only two outcomes; $x_{ik} = 0$ or $x_{ik} = 1$, we estimate the effect of the variable by comparing the expected conditional means of y_i. It can be proved that

$$\frac{E(y_i|x_{ik}=1, x_i^*)}{E(y_i|x_{ik}=0, x_i^*)} = \exp(\beta_k) \leftrightarrow \ln\left(\frac{E(y_i|x_{ik}=1, x_i^*)}{E(y_i|x_{ik}=0, x_i^*)}\right) = \beta_k \qquad (6.17)$$

Where x_i^* is the x_i-vector exluding the *k*th element. The interpretation is that the expected mean of y_i is $\exp(\beta_k)$ times higher in the case that $x_{ik} = 1$ compared to when $x_{ik} = 0$, regardless of the other explanatory variables. In the Poisson model, y is interpreted as the expected number of children dead.

6.3 The logit model for binary response

As pointed out in subsection 5.2.1, the fact we have very few observations of household that have experienced more than one infant or under-five death the last five years could lead to a overestimation of parameters because we have a outlier problem. I also discussed the possibility of converting the dependent variables from counts to binary responses. This will require yet another model and method of estimation than those described in 6.3 and 6.4., namely a *binary choice* model. In a binary choice model, the dependent variable has only two possible outcomes, one or zero. We want to describe the probability that $y_i = 1$ conditional on a vector of explanatory variables, x_i: $\mathrm{P}(\mathrm{y_i} = 1|\mathrm{x_i})$. In order to estimate this probability, we need to make an assumption about the distribution of y. While the probit model rests on the assumption that the dependent variable is distributed according to a standard normal distribution function, the logit model assumes a standard logistic distribution. Both distributions have an expectation of zero and gives estimated probabilities between 0 and 1. In applied empirical work the two models yield very similar results. Here, I choose the logit model specified by:

$$P(y_i = 1|x_i) = \frac{\exp(x_i'\beta_k)}{1 + \exp(x_i'\beta_k)} = \Lambda(x_i'\beta_k) \qquad (6.18)$$

(6.18) is estimated using maximum likelihood and all results for the logit model presented in chapter 8 and the appendixes are obtained by using the standard logit procedure in STATA.

6.3.1 Interpretation of logit coefficients

The estimated β-coefficients in the logit model are difficult to interpret directly and are usually only made use of to indicate the sign and significance of the estimated effects. When interpreting size of the effects, we look at the marginal change in probability. Marginal effect of an explanatory variable on the dependent variable in the logit model is defined as the partial derivative of the probability of $y_i = 1$. Taking the partial derivative of (6.18) with respect to x_i obtains:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ik}} = \frac{e^{x_i'\beta}}{\left(1 + e^{x_i'\beta}\right)^2}\beta_k$$

The marginal effect of a change in the variable x_{ik} is dependent on, and therefore varies with, values of x_i. It is common to calculate marginal effects in the logit model at the sample averages of the independent variables. This can be done in STATA. y is interpreted as the probability of a child passing away the last five years.

6.4 Clustered standard errors

Because observations on health facility and health worker characteristics are identical for all households that belong to the same catchment area, the standard errors of these observations will be correlated. If the correlation of variables with groups belonging to the same facility is not taken into account, the number of independent observations will be exaggerated and standards errors to small. While the OLS estimator of β will remain unbiased, the underestimation of standard errors will make statistical inferences from the regressions less exact. Clustered standard errors allow for non-independence of the error terms within specified groups by calculating variance on the basis of aggregated scores within each of these rather than the aggregated score of each individual in the sample. In all regressions presented in chapter 8 and the appendixes, I calculate clustered standard errors by applying the *cluster* command at facility level in STATA.

I now proceed to a short descriptive analysis in the next chapter before applying the three models presented here to the data in chapter 8.

7 Descriptive analysis

This chapter presents descriptive statistics on the distribution of infant and child deaths by region, urban-rural residence, wealth, education and closest public health facility. A complete listing of all variables in the analysis, description, mean, min and max values is presented in table A.1 in appendix A.

7.1 Number of children born, number of under-five and infants dead and mortality rate

The overall under-five mortality rate is calculated to be *92 per 1000* live births in the sample. This is substantially lower than both the World Bank estimate of 161 (2011), the Angolan Ministry of Health 2008 estimate of 193.5 (Instituto Nacional de Estatística, 2010) and the recently published 2011 estimate of 134 (Lozano et al., 2011). This is worrying with respect to the quality of our data, but might be explained by one or more of the following factors. As pointed out, the survey did not follow all children until they turned five years old. It is therefore not accounting for all under-five deaths in the households. The calculated under-five mortality rate underestimates the actual rate, and is not comparable with conventional measures presented in publications such as the World Development Indicators. In addition, observations in the sample are unevenly distributed between rural and urban locations. 70% of the households in the sample are urban, while according to the World Bank (2011) this number is 58% for the country as a whole. In the literature, many researchers have found substantially higher mortality rates in rural areas compared to urban ones (Sastry, 1997) and the excess number of urban households in the sample could also contribute to the under-five mortality being underestimated. Infant mortality rate is calculated to be *55 per 1000*. As with the under-five mortality rate, this is significantly lower than the World Bank estimate of 98 (2011) and the recently published 2011 estimate of 90 (Lozano et al., 2011). Factors explaining the large difference are the same for infant as for under-five mortality.

In table 7.1 the number of children born, children under-five dead, infants dead, under-five mortality rates and infant mortality rates are reported for the whole sample, the two regions, for urban and rural locations and for the different wealth quintiles. “Total” is equal to the overall number of children born, dead and overall rate, and “Per household” is the average

household value, in each category. The total numbers are interesting when comparing Angola to other countries because international statistics on child mortality are based on this kind of aggregated information. However, I want to analyze differences in mortality between *households* and therefore focus my attention on per household numbers.

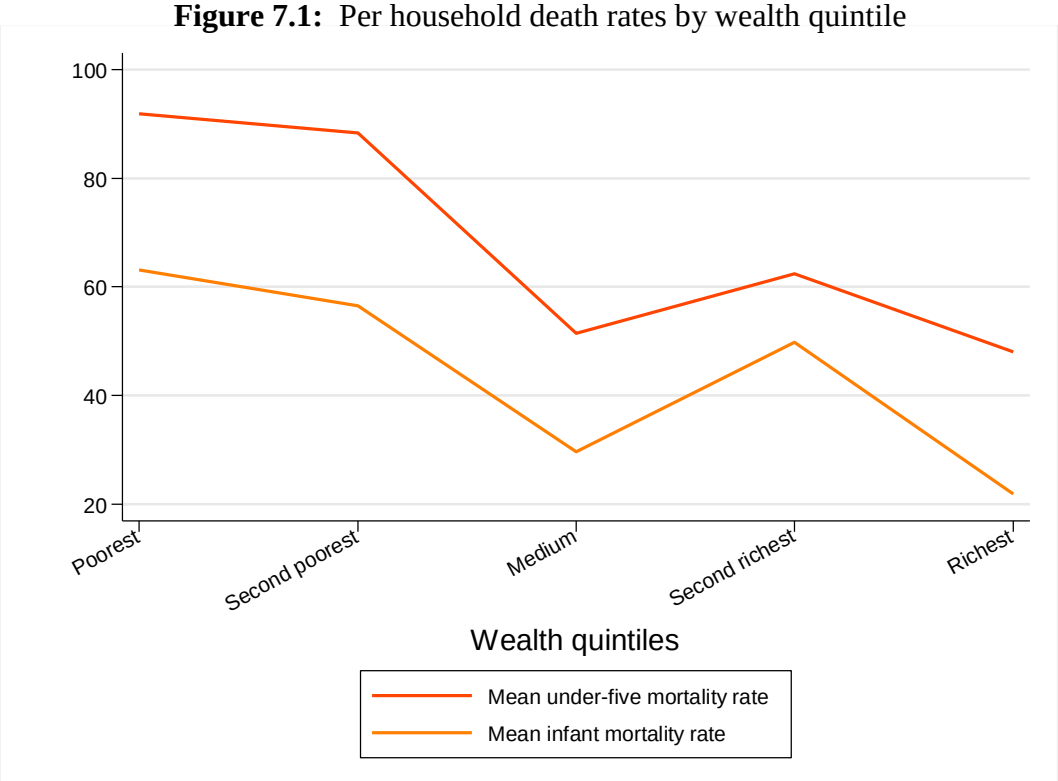
Table 7.1: Total number and per household values for children born, under-five and infants dead, and under-five and infant death rates

		Number of children born alive	Number of children under-five dead	Number of infants dead	Under-five deaths per 1000 born	Infant deaths per 1000 born
Overall						
	Total	1995	184	109	92	55
	Per household	2.11	0.20	0.11	69.77	44.27
Region						
Luanda						
	Total	915	59	38	64	42
	Per household	1.96	0.13	0.08	55.68	34.15
Uige						
	Total	1080	125	71	116	66
	Per household	2.25	0.26	0.15	83.07	54.13
Geographic area						
Urban						
	Total	1319	92	54	70	41
	Per household	2.01	0.14	0.08	57.01	33.41
Rural						
	Total	676	92	55	136	81
	Per household	2.34	0.32	0.19	98.13	68.96
Wealth quintile						
1 (poorest)						
	Total	444	53	32	119	72
	Per household	2.31	0.28	0.17	91.85	63.11
2 (second poorest)						
	Total	442	56	29	127	66
	Per Household	2.31	0.30	0.15	88.35	56.54
3 (middle)						
	Total	408	30	18	74	44
	Per household	2.16	0.16	0.09	51.50	29.60
4 (second richest)						
	Total	337	26	21	77	62
	Per household	1.80	0.14	0.11	62.45	49.82
5 (richest)						
	Total	360	18	9	50	25
	Per household	1.97	0.10	0.05	48.01	21.86

The table shows that the average household in Uíge had more under-five and infant deaths than the average household in Luanda. Per household under-five mortality is 27 higher, and infant mortality is 20 higher, per 1000 live birth in Uíge than in Luanda. Differences are even larger between urban and rural areas. In the average rural household, 41 more children under-five and 36 more infants died per 1000 live birth than in the average urban households. In no other group are under-five and infant mortality rates higher than in the rural areas.

Per household under-five mortality rate is actually higher in the second richest than in the middle quintile. However, there seem to be a clear division between the two poorest and three richest quintiles, the per household under-five rate being substantially higher in the poorest and second poorest quintiles. For per household infant mortality rates, the middle and richest quintiles exhibits significantly lower numbers than the others. The infant mortality in the second richest quintile is standing out even more than for under-five mortality and is close to the value in the second poorest group. For the overall infant mortality rates on the other hand, the richest quintile stands out and is almost half that of the next lowest rate in the middle quintile. I conclude that the data lends support to H1a, which hypothesized a negative relationship between child mortality and wealth. However, the relationship found is not linear in wealth quintiles. This is illustrated graphically in figure 7.1. Thus, the hypothesis is only found to hold on average, and not for all wealth quintiles.

Figure 7.1: Per household death rates by wealth quintile



In terms of total under-five and infant mortality, I also find evidence in support of hypothesis H1b; there is a gap in the total death rate between the richest and the other wealth quintiles. However, when number of households in each group is taken into account and I look at the household values, the hypothesis is rejected.

A last thing worth mentioning is the variation between the different groups in number of children born. In Uíge 0.29 more children were born per household than in Luanda. This could reflect differing norms on issues like family size and age of marriage in the two regions. Fertility difference is even larger when comparing urban and rural areas. The average urban household had 0.33 fewer children than the rural. Again, culture and norms could explain the difference, but it could also be the case that rural households need the labour from children more than urban ones because they live mainly from farming. This could also explain the fertility difference between the two regions; while Luanda has only urban households, the sample from Uíge consists of 60% rural households. Per household fertility is higher in the poorest and second poorest quintiles than in the middle, and in the middle compared to the second richest and richest. This is in line with Becker's hypothesis that the income elasticity of quantity of children is relatively smaller than that of child human capital ("quality"); when income increases people will spend relatively more of the increase on child human capital than on having more children. Another interpretation is that the opportunity cost of having children is higher for richer people because they earn higher wages. Only four of the households in the two poorest wealth quintiles are situated in Luanda and wealth differences could be an explanation of the regional differences in fertility discussed above.

Table 7.2 reports distribution of total number and per household values of children born, children under-five dead, infants dead, under-five and infant mortality rates across education levels. The per household under-five mortality rate is highest in the households where no adults have attained a higher level of education than secondary school. These findings support hypothesis H2a, which suggested a negative relationship between parents' education level and child mortality. Under-five mortality is higher in the group of households with university level education than amongst those with medium level education, but the difference is only 1 per 1000. The average infant mortality exhibits a similar pattern, but here the university level rate is smaller than the medium level by 7 per 1000.

Table 7.2: Number of children born, under-five and infants dead, and under-five and infant death rates by education level

		Number of children born alive	Number of children under-five dead	Number of infants dead	Under-five deaths per 1000 born	Infant deaths per 1000 born
No education/ Primary/ Secondary						
	Total	519	67	40	129	77
	Per household	2.26	0.29	0.17	88.62	60.41
Professional						
	Total	651	64	39	98	60
	Per household	2.17	0.22	0.13	77.81	50.28
Medium						
	Total	595	41	23	69	39
	Per household	1.96	0.14	0.08	53.03	32.95
University						
	Total	230	12	7	52	30
	Per household	2.04	0.11	0.06	54.10	25.81

For both under-fives and infants, per household mortality is in fact decreasing relatively more between the group with professional and medium level education than between the no education/primary/secondary and the professional level group. For the per household infant mortality rate, the percentage decrease is also larger between the medium and university level than between the no education/primary/secondary and professional categories. These findings could indicate that higher level education is very important to child survival, but might also reflect the importance other factors common to more educated groups of households, for instance income, access to health services, food available or region. These other factors will be controlled for in the analysis in chapter 8. I find no support for our hypothesis that the effect of education is larger on under-five than infant deaths (H2b). In fact, the percentage decrease in mortality between the lowest and second lowest, and between the second lowest and second highest education level, is larger for infants than under-fives.

Table A.2 (appendix A) reports per household number of children born, under-five dead, infants dead, under-five and infant mortality rate sorted by facility catchment areas. In the households in the catchment area of three facilities, *C.S. Paz* (Luanda), *Centro da Pedeira* (Uíge) and *Malangino* (Luanda), no children under-five were reported to have passed away

the last 5 years. The highest mean under-five rate amongst the facilities was found at *C.S. Vila da Mata* in Luanda. In its catchment area, households had an average under-five mortality rate as high as 245 per 1000 children born, which is equivalent to every fourth child passing away before turning five years old. *CS Siga* in Luanda is the facility with the highest per household infant mortality rate in its catchment area, and has a rate of 139 per 1000. These findings lend support to our hypothesis that child mortality depend on the closest public health facility (H3a).

7.2 Summary of descriptive statistics

The descriptive statistics show that there are large differences in number of children born, under-five dead and infant deaths and mortality rates across different groups in the sample. Amongst all the different groups, I find that households situated in rural areas have the highest average mortality rates for both under-fives and infants. In number of deaths, as many children passed away in rural as in urban areas, despite the urban group being more than twice the size of the rural.

I also find a negative relationship between child mortality and wealth. Mortality differs across wealth quintiles; both under-five and infant per household mortality is decreasing from the poorest to second poorest and from the second poorest to the medium quintile, then increasing substantially from the medium to the second richest quintile and decreasing again from the second richest to the richest quintile. Furthermore I find that mean mortality is highest in the households with least education and is decreasing in groups of higher education. While these findings are supporting some of the hypotheses put forward in section 4.5, it is not unlikely that differences described her are actually caused by other factors. I now proceed to the econometric analysis where such factors are controlled for.

8 Regression results

In this chapter, I present the regression results. The analysis is focused around three factors; income, wealth and quality of health services, following the three sets of hypotheses that were put forward on the basis of theory and empiric evidence in chapter 5. I start by testing hypothesis H3a, which stated that under-five and infant deaths in a household is dependent on the health facilities it belongs to. If I find support the hypothesis, I would like to include more specific health service measures in the further analysis. I then proceed to investigate the impact of wealth, education, health facility characteristics and health worker characteristics on infant and under-five mortality.

All calculations and estimates reported here are obtained using STATA. Regressions are based on the OLS, Poisson and logit models described in chapter 6[22]. For all of these, I tested regressions with different definitions of the dependent variables as discussed in subsection 5.2.1. The purpose of this was to check if the outliers of the dependent variables are a problem. On the basis of these regressions, I have chosen to report the following three models here:

1. OLS with original definitions of dependent variables
2. Poisson with original definitions of dependent variables
3. Binary logit

To test whether the equidispersion assumption holds in the Poisson model I run a negative binomial regression. In the this model, an overdispersion parameter, α, is estimated. When α is significantly larger than zero, the Poisson model is not appropriate. For under-five deaths, the estimated overdispersion is 3.88e-07≈0 and I conclude that the Poisson model is appropriate. For infant deaths, alpha is 0.065241 and the appropriateness of the Poisson model is more ambiguous. I therefore perform a goodness of fit test. The test statistics are not significant for either under-five nor infant deaths and I conclude that the equidispersion assumption cannot be rejected; the Poisson model is appropriate for our data (see table B.1 and B.2, appendix B). As discussed in chapter 6, there are good reasons to suspect that the OLS model is not appropriate here. While the coefficient estimates obtained in this model are

[22] I also ran ordered logit regressions, but Poisson and binary logit was found to be more appropriate

unbiased, the significances of the results are not trustworthy. The Poisson and logit models should give rise to more reliable results. Since the Poisson model utilizes more of the variation in the dataset than the binary logit model, it is also likely to yield better results. For the Poisson and binary logit model, both coefficients and marginal effects are reported.

8.1 Differences in mortality across health facilities

The first step of the analysis is to choose which variables to include. I started with running bivariate OLS regressions with all variables listed in section 5.2. On the basis of the regressions, variables that were significant at 0.1, 1 or 5% level for either under-five or infant deaths were kept. I also retained some variables that were not significant, but for which there are strong theoretical or empirical reasons to include in the analysis.

The next step is to investigate hypothesis H3a and see if there are significant differences in mortality between facilities. In the descriptive statistics I found large differences in average under-five mortality between facilities, and the mortality ranged from zero per 1000 in the catchment areas of some facilities to 245 per 1000 at the worst. I now want to see if these differences are statistically significant when controlling for household characteristics such as income, education, geographic area and place of delivery.

Compared to the reference facility (automatically chosen by STATA), I find that a large part of the facility dummies have a significantly different per household mortality even after controlling for important household characteristics such as education and wealth in both the OLS and binary logit model (see table C.1 and C.2, appendix C). Most of the coefficients are positive, implying that the average number of deaths in the households in the catchment areas of the facilities is higher relative to the reference catchment area. This is because the reference facility, *Cs. Paz* has a per household under-five and infant mortality rate of zero. In the Poisson model, marginal effects are estimated to be very large and for the most part not significant. The explanation is that in some facilities the dependent variable contains no variation. Evidence in the other models is however so strong, that I conclude that there are statistically significant differences in mortality between facilities. I therefore include more specific health facility and health worker variables to seek further explanations of why chances of survival are better for children born in some catchment areas than others. Though I

control for important factors here, it is still possible, and even likely, that the large difference in under-five and infant deaths between facilities is not only reflecting unequal availability and quality of health services but also other factors that I am not able to control for here.

8.2 Wealth

Results from the OLS, Poisson and the binary logit model are reported, with under-five and infant deaths as dependent variables in table 8.1 and 8.2, respectively. In the OLS model, the F-test tells me that the hypothesis that all coefficients are zero can be rejected at 1% significance level for both under-five and infant deaths. For this model, R^2, which tells us how much of the observed variance in the dependent variable that is explained by the variables included in the regression, is also reported. 24% of the variation in under-five and 13% of variation in infant, deaths is explained by the variables included in the OLS model. Normally, STATA displays loglikelihoods and loglikelihood-tests for the Poisson and logit models, but because I use clustered standard errors, pseudolikelihood and wald-chi test statistics are reported here. The Wald-chi statistic is used to tests the hypothesis that all coefficients are simultaneously zero. In both the under-five and infant model this is rejected at a 1% level. Pseudo R^2 is reported in the logit model, but it is not directly comparable to the regular goodness of fit the regular R^2 measure. While I started out with 931 observations for under-five deaths and 946 for infants, these numbers are 879 and 893 in the regressions, respectively. In STATA, observations are left out from the regressions when one or more of the explanatory variables are reported as missing. Observations are missing if the respondent did not answer one or more of the questions posed in the survey.

In none of the models, the wealth measure (land_pc1) is estimated to have a significant effect on number under-fives and infants dead when controlling for number of children born, education, use antenatal controls, place of delivery, region, urban residency and health facility and health worker characteristics. Hence, the large differences in child mortality between wealth quintiles found in the descriptive analysis is not explained by wealth differences themselves, but by other features distinctive to the different income groups.

Table 8.1: Results from OLS, Poisson and binary logit results with under-five deaths as dependent variable

		Poisson		Binary Logit	
	OLS (Std. Err)	Coefficients (Std. Err)	Marginal effects (Std. Err)	Coefficients (Std. Err)	Marginal effects (Std. Err)
children_born	0.225***	0.682***	0.077***	0.977***	0.086***
	(0.038)	(0.042)	(0.008)	(0.112)	(0.011)
land_pc1	0.010	0.015	0.002	0.067	0.006
	(0.011)	(0.048)	(0.005)	(0.062)	(0.005)
noeduc_primary_secondary (d)	0.134*	0.803***	0.114**	0.854*	0.090*
	(0.067)	(0.286)	(0.050)	(0.460)	(0.055)
professional (d)	0.083	0.615**	0.079*	0.722	0.071
	(0.054)	(0.303)	(0.045)	(0.458)	(0.049)
medium (d)	0.063	0.477**	0.059*	0.644*	0.063
	(0.038)	(0.238)	(0.033)	(0.377)	(0.039)
a_control4 (d)	0.014	0.005	0.001	0.040	0.003
	(0.044)	(0.216)	(0.024)	(0.267)	(0.023)
del_home_post_other (d)	0.133**	0.501**	0.060*	0.787**	0.075**
	(0.054)	(0.255)	(0.033)	(0.356)	(0.038)
del_centre_other_fac (d)	-0.004	0.020	0.002	-0.038	-0.003
	(0.045)	(0.323)	(0.037)	(0.443)	(0.038)
att_doctor (d)	0.102	0.045	0.005	0.310	0.026
	(0.074)	(0.199)	(0.022)	(0.303)	(0.025)
breastf (d)	-0.078	-0.526	-0.076	-0.855	-0.102
	(0.077)	(0.407)	(0.072)	(0.574)	(0.086)
urban_uige (d)	-0.106	-0.149	-0.016	-0.508	-0.040
	(0.081)	(0.404)	(0.042)	(0.522)	(0.037)
luanda (d)	-0.130	-0.505	-0.057	-1.067*	-0.095*
	(0.086)	(0.484)	(0.056)	(0.601)	(0.055)
power (d)	0.033	0.379	0.040	0.416	0.035
	(0.042)	(0.343)	(0.033)	(0.475)	(0.036)
im_facility_outreach (d)	0.145**	0.556**	0.051**	0.724*	0.051**
	(0.062)	(0.276)	(0.020)	(0.423)	(0.023)
antenatal_serv (d)	0.015	0.083	0.009	0.138	0.012
	(0.039)	(0.157)	(0.017)	(0.281)	(0.024)
del_serv (d)	-0.004	-0.570**	-0.064**	-0.564	-0.049
	(0.055)	(0.274)	(0.030)	(0.464)	(0.040)
malaria_serv (d)	0.015	0.182	0.019	0.362	0.029
	(0.037)	(0.184)	(0.019)	(0.306)	(0.022)
thermometer (d)	-0.062	0.000	0.000	-0.155	-0.014
	(0.046)	(0.207)	(0.023)	(0.305)	(0.030)
midwife (d)	0.006	0.386	0.046	0.392	0.036
	(0.060)	(0.323)	(0.043)	(0.472)	(0.047)
spes_doctor (d)	-0.017	-0.098	-0.011	-0.142	-0.012
	(0.041)	(0.285)	(0.031)	(0.429)	(0.036)
nquest_tot	-0.004	-0.032**	-0.004**	-0.028	-0.002

	(0.003)	(0.014)	(0.002)	(0.022)	(0.002)
nexam_tot	0.005	0.065**	0.007**	0.046	0.004
	(0.005)	(0.027)	(0.003)	(0.042)	(0.004)
diag_tot	-0.014	-0.220**	-0.025**	-0.138	-0.012
	(0.018)	(0.102)	(0.010)	(0.144)	(0.012)
F(23, 38)	*6.12*				
Prob > F	*0.0000*				
Log pseudolikelihood		*-365.43*		*-293.75*	
Wald chi2(23)		*1765.80*		*146.80*	
Prob > chi2		*0.0000*		*0.0000*	
R^2	*0.2375*				
pseudo R^2				*0.195*	
Number of observations	*879*	*879*		*879*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table 8.2: Results from OLS, Poisson and Binary Logit with infant deaths as dependent variable

	OLS (Std. Err)	Poisson		Binary Logit	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
children_born	0.107***	0.587***	0.042***	0.784***	0.048***
	(0.020)	(0.063)	(0.005)	(0.109)	(0.006)
land_pc1	0.005	-0.002	-0.000	0.024	0.001
	(0.008)	(0.067)	(0.005)	(0.089)	(0.005)
noeduc_primary_secondary (d)	0.093*	0.782*	0.070	0.762	0.056
	(0.047)	(0.401)	(0.045)	(0.527)	(0.046)
professional (d)	0.057	0.613	0.050	0.652	0.045
	(0.039)	(0.407)	(0.038)	(0.528)	(0.041)
medium (d)	0.027	0.279	0.021	0.405	0.026
	(0.026)	(0.349)	(0.028)	(0.454)	(0.032)
a_control4 (d)	-0.045	-0.361	-0.028	-0.191	-0.012
	(0.040)	(0.305)	(0.026)	(0.336)	(0.022)
del_home_post_other (d)	0.100**	0.559**	0.042*	0.811**	0.054*
	(0.040)	(0.281)	(0.024)	(0.359)	(0.028)
del_centre_other_fac (d)	-0.002	-0.034	-0.002	-0.280	-0.016
	(0.027)	(0.345)	(0.024)	(0.406)	(0.021)
att_doctor (d)	0.103**	0.301	0.020	0.663**	0.037**
	(0.044)	(0.263)	(0.018)	(0.336)	(0.019)
breastf (d)	-0.106	-0.933**	-0.105	-1.202**	-0.118
	(0.064)	(0.413)	(0.067)	(0.546)	(0.075)
urban_uige (d)	-0.064	-0.389	-0.025	-0.663	-0.035
	(0.045)	(0.473)	(0.028)	(0.533)	(0.024)
luanda (d)	-0.056	-0.459	-0.033	-0.927	-0.058
	(0.058)	(0.647)	(0.047)	(0.807)	(0.052)
power (d)	0.040	0.596*	0.039*	0.424	0.024

	(0.028)	(0.357)	(0.020)	(0.449)	(0.024)
im_facility_outreach (d)	0.043	0.206	0.013	0.005	0.000
	(0.035)	(0.377)	(0.023)	(0.454)	(0.028)
antenatal_serv (d)	-0.007	0.128	0.009	0.186	0.011
	(0.025)	(0.276)	(0.019)	(0.356)	(0.020)
del_serv (d)	-0.040	-0.807***	-0.057***	-0.768*	-0.047*
	(0.030)	(0.303)	(0.020)	(0.429)	(0.025)
malaria_serv (d)	0.019	0.229	0.015	0.421	0.023
	(0.025)	(0.200)	(0.012)	(0.311)	(0.015)
thermometer (d)	-0.042	0.102	0.007	-0.124	-0.008
	(0.030)	(0.201)	(0.013)	(0.263)	(0.018)
midwife (d)	0.042	0.687**	0.054**	0.744**	0.050*
	(0.031)	(0.296)	(0.027)	(0.378)	(0.029)
spes_doctor (d)	0.016	0.168	0.013	0.014	0.001
	(0.022)	(0.279)	(0.021)	(0.389)	(0.024)
nquest_tot	-0.002	-0.022	-0.002	-0.012	-0.001
	(0.002)	(0.022)	(0.002)	(0.026)	(0.002)
nexam_tot	0.002	0.045	0.003	0.027	0.002
	(0.003)	(0.037)	(0.003)	(0.047)	(0.003)
diag_tot	-0.011	-0.215**	-0.015**	-0.150	-0.009
	(0.012)	(0.108)	(0.007)	(0.141)	(0.008)
F(23, 38)	*6.53*				
Prob > F	*0.0000*				
Log pseudolikelihood		*-274.70*		*-241.15*	
Wald chi2(23)		*1113.84*		*205.99*	
Prob > chi2		*0.0000*		*0.0000*	
R^2	*0.131*				
pseudo R^2				*0.161*	
Number of observations	*893*	*893*		*893*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Because theory and empirical research so strongly suggest that wealth, or socioeconomic status, is a key determinant of health, I tried several measures of wealth. In appendix D, table D.1 and D.2, results from regressions where the continuous wealth index is replaced by quintile dummies are reported. Replacing the continuous variable should pick up nonlinearities in the health-wealth relationship. Apart from the logit coefficient of the second richest group being significant at a 10% level for under-fives, no evidence is found for wealth to have a significant impact on child mortality with this specification neither.

I also run separate regressions for the regions and construct separate wealth indexes for the two (see appendix D, table D.3-D.6). The rationale for doing so is that the PCA constructed wealth index measures relative wealth and it might be the case that importance of the various items in it is of different significance in the two regions. For example, having a gas cooker

might be a sign of a household having high socioeconomic status in Uíge, while this is a common thing to have in all households in Luanda, regardless of socioeconomic status. I include the same variables in both wealth measures apart from dummies for land ownership and number of plots owned, that I only include in the Uíge index. In this case as well, the wealth index remains insignificant in all three models, for both under-five and infant deaths in both regions. Though not generating any interesting results on account of the wealth index, the regressions done on the split sample gives some interesting results in terms of other variables. I comment on some of these in section 8.8.

Following Dunteman (1989), I also try substituting the wealth index with a small set of variables from which the index was constructed from, that exhibits high correlation with the index. Doing this did not give better and more meaningful results, and these are therefore not reported here.

After trying various ways of measuring wealth without finding any significant effect, I conclude that there is no evidence to support our hypothesis that wealth is a key determinant of child mortality in the regions of Luanda and Uíge in Angola, and H1a and H1b are rejected.

8.3 Education

Compared to households with a university degree, those in the lowest education category have a significantly higher likelihood (11.4% in the Poisson model) of having children that dies before the age of five. A similar effect is found for households in the two middle categories (7.9% and 5.9% for professional training and medium level in the Poisson model, respectively). These latter effects are, however, not very precisely estimated, as the difference between the middle and highest level is only significant at a 10% level. I conclude that the data lends support to H2a, which suggested a negative relationship between child mortality and household education level. No significant marginal effects of education on infant mortality are found in the Poisson and logit model. The findings lend support to the effect of education being stronger on under-five than infant mortality, as hypothesised in H2b.

8.4 Other household characteristics

Perhaps the most striking feature of the regression results is the estimated effects of number of children born in the households the last five years preceding the survey. The variable is statistically significant at 1% level in all three models for both under-five and infant deaths. In the Poisson and logit model, an increase of one child is associated with a 7.7% and 8.6% higher likelihood of the expected number of under-fives dead respectively. The effect on infant deaths is roughly half the size of the under-five estimates; 4.2% and 4.8%. These findings are however not surprising. It is a logical necessity that that the more children are born in a household, the larger is the expected number of children dying. Nevertheless, an interesting question to pose is whether the increase in expected number of children dead is only a consequence of the household having more children that could possibly die or if a child has a larger chance of dying if it is born in a household with many children than a household with fewer children. That is: is the probability of death for each individual child higher the more siblings it has, or is the probability increasing in number of children in the household when all other factors are equal? In order to answer this question I need detailed information about when the children die, how many siblings they have at time of death and how many children that were ever born in the household. The data do not contain this type of information and I cannot answer the question. Making a guess on the basis of theory and empirical research is also difficult: while the economic model presented in chapter 4 suggested that probability of dying is increasing in number of siblings (because human capital level is reduced), Magne Mogstad and Matthew Wishall (2010) find an inverse U-shaped relationship between family size and child outcome[23] when testing the model on Norwegian population data.

Compared to households where the last woman to give birth delivered in a public hospital, households where the last delivery took place in their own home, someone else's home, a public health post or "other" place, have a higher likelihood (6% in the Poisson, 7.5% in the logit model at 10 and 5% levels of significance respectively) of having children under-five die. I find similar effects on number of infant deaths; households in the home delivery category had a 4.2% higher expectancy of number of children dead compared to households in the hospital delivery category, in the Poisson model. These findings are in line with

[23] Measured by educational attainment

expectations. Most maternal and child deaths occur during or shortly after birth any many researchers are emphasising the importance of access to assisted delivery and emergency obstetric care in reducing these child and maternal deaths. Women giving birth in their own homes, other people's homes, health posts or "other" places are less likely to be attended by skilled health workers that can provide the needed help if complications occur during delivery than women giving birth in a hospital. This is probably part of the explanation of why I observe higher expected number of children dead in the former group.

A problem that could occur using this kind of variable is *selection*. Women who experience complications during pregnancy or delivery tend to seek medical care to a larger degree than others. These women have increased risks for negative health outcomes both with regards to themselves and their children. Women seeking medical care thus might have a higher probability of child deaths than others and this would then affect the size and, if the selection effect is very strong, the sign of regression coefficient. I can then mistakenly be lead to believe that the higher probability of child deaths is caused by some factors connected to the health facility, while the real cause could be that they have more ill patients with higher probabilities of losing their children than others. No statistically significant differences in child mortality is found between households where the last delivery took place in a public hospital and households where last the delivery took place at a public health centre or other (public or private) health facility.

Compared to households in Uíge, the likelihood of having a child below five years of age die is 9.5% lower for households in Luanda. However, the regional dummy is only found to be significant in the logit model and the estimate has a low level of precision (10%). Regional differences in likelihood of infant deaths are negative, but not significant.

Women who were attended by skilled personnel (medical doctor, nurse or midwife) during their last delivery have a higher likelihood (3.7% in the logit model at 5% significance level) of losing a child before the age of one, than those who were not attended by skilled personnel. This is surprising as I, from a theoretical point of view, would expect skilled birth attendance to improve chances of survival. The finding can, however, be explained by selection of women with higher probability of negative health outcomes to the group.

8.4.1 A note on maternal age

The maternal variables age and age^2 are not included in our regressions despite clear indications of these being important to child mortality in the empiric literature. The reason for this is two-fold. First, I suspect measurement errors in the variable. Secondly, when I included the age variables in the analysis, the coefficients were significant, but had signs opposite of the expectation. Instead of being U-shaped, the relationship between mother's age and child mortality is hyperbolic in the sample: the young and old ages are associated with lower mortalities than the age groups between. While a lower mortality rate among children of older women could be explained by their experience in child caring, the low child mortality for young women could be linked to the fact that the survey did not follow all children until they reached the age of five. The low mortality among children of young mothers is thus probably owing to the fact that they had not "had the time" to die yet when the survey was conducted, not the age of the mother itself.

8.5 Health facility characteristics

In the descriptive analysis significant differences in per household under-five and infants mortality were observed between catchment areas of the facilities in the sample, and in 8.1 I found that these differences persist when controlling for household characteristics. I now turn to investigate the effects of specific health facility characteristics.

A dummy variable for facilities that offer immunization services, either in the facility or as an outreach service, or both, is included in the regressions. In the sample, households in catchment areas of facilities that offer immunization have a 5.1% higher likelihood of number of under-five deaths than households residing close to facilities that does *not* offer any form of immunization services. This effect is significant at a 5% level. The finding is surprising form an epidemiological point of view; immunization is an important factor in disease prevention and households residing close to facilities where it is offered should have better health outcomes because of a lower rate of infection. A possible explanation is reversed causality; that immunization services are offered to a larger degree in the worst-off areas. However, when taking a closer look at the distribution of the variable, I find that only four facilities are not offering immunization and the result could reflect some other factor common to these four facilities. Availability of immunization services is not estimated to have any effect on number

of infant deaths in the households. Another result that is hard to explain is that likelihood of infant deaths is 5.4% higher in catchment areas of facilities that have midwife in its staff compared to those who don't.

The effect of availability of delivery services is more in line with expectations. In the Poisson model, households living in the catchment areas of facilities that offer delivery services are estimated to have a 6.4% lower likelihood of under-five, and 5.7% lower likelihood of infant deaths than other households at 5 and 1% significance level, respectively.

Compared to households in the catchment area of facilities that did not have power during all opening hours the last week before the survey, households belonging to facilities that had power available continuously have a 3.9% higher likelihood of infant deaths. This finding is difficult to explain as I expect that in general facilities with stabile access to power are of better quality than others. The precision of the estimate not being very great with significance at 10% level, I do however not put too much emphasis on it.

8.6 Health worker characteristics

In the data, I find the performance of health workers in terms of history taking, to be negatively related to child mortality. An increase in performance of one more question asked in the patient case simulation is associated with a 0.4% reduction in likelihood of under-five deaths in the Poisson model, significant at a 5% level. No significant effect of health workers history taking performance on infant deaths was found.

Health worker performance in terms of physical examination is positively related to under-five mortality. An increase in number of examinations done is associated with a 0.7% higher (in the Poisson model at significance level of 5%) likelihood of children under-five dead in the households in the catchment area of the facility that the health worker belongs to. This is contrary to expectations; assuming that the score the health workers obtain during patient case simulations reflects performance with real patients, a higher fraction of relevant physical examinations done should indicate better quality of health services provided. A possible explanation is that health workers in areas with high mortality are better in performing physical examinations because their patients are generally more ill. In this case that the

relationship observed is a matter of reversed causality. However, when looking at the distribution of health worker physical examination scores across the regions, urban/rural locations and wealth quintiles, I find the score to be generally higher for the groups that were found to have the lowest mortalities in the descriptive analysis.

Ability to make the correct diagnosis is the only health worker characteristic found significant to both under-five and infant mortality. An increase in right diagnoses made by the health worker gives a reduction of 2.5% in likelihood of under-five, and 1.5% of infant, deaths at a 5% significance level in the Poisson model.

It should be noted that the extent to which the scores obtained in the patient case simulations reflect their performance with real patients is not clear. For example, Lindkvist (2011) find that health workers in Tanzania perform better in patient simulations than with real patients.

8.7 Separated sample between regions

In table D.3-D.6 in appendix D regression results for separate regressions for the two regions with separate wealth indexes are reported. According to these results, determinants of child mortality differ between the regions.

8.7.1 Household characteristics

8.7.1.1 Under-five dead

Regressions with number of under-five dead for Luanda and Uíge are found in table D.3 and table D.5, respectively. No significant effect of being in the lowest education category on under-five mortality is found in either of the two regions. In fact, none of the education categories have any effect on child mortality in Uíge. In Luanda, households in the professional training category have a higher (7.7% at 5% level in the Poisson model) likelihood of under-five child deaths compared to households with a university degree. This is in line with Bicego and Boerma (1993) who find the effects of (maternal) education to be larger in urban than rural areas. They propose that the urban-rural difference in education is caused by urban women being constrained by access social and economic support rather than physical access to health services that might be a larger problem in rural areas. Because education is a factor in overcoming social and economic constraints, it is of larger importance

to urban than rural women. Another explanation put forward by Bicego and Boerma is that educated women residing in rural areas might be under more social pressure to follow traditional prevention, treatment and eating practices, even if they know them to be potentially harmful, than women living in a more modern urban context.

In Luanda, the women who gave their last birth in their own home, some else's home, a health post or "other" place to have a 10.8% higher expected number of children under-five dead than those who last delivered in an hospital, in the Poisson model (significant at a 10% level). No significant differences in mortality are found between women delivering in a health centre or other health facility and women delivering in a public hospital. Opposite, in the Uíge region women giving birth in a health centre or other health facility, have an 11.4% lower expectation of number of under-five deaths than those giving birth in a hospital, at a 1% level of significance. This finding is contrary expectations. Hospitals are the highest level of public health services delivered in Angola I expect that hospitals produce better health outcomes than facilities at lower levels of services because they are better equipped and have more skilled workers. In section 8.4 I mentioned the possibility of a selection effect on this variable. The large estimated negative effect of the dummy for health centres and other health facilities could be an implication of selection of the worst-off women with the most severe complications to the hospital in Uíge. That the selection of more ill women to hospital deliveries is stronger in Uíge than in Luanda, where I do not detect any selection effects, can possibly be explained both by differing social norms connected to childbirth or accessibility of hospital services. 60% of households in Uíge are classified as rural. For these households, the distance to a facility other than the closest public health facility is probably much larger than for the households in Luanda. This is a possible explanation for the Uíge households being more reluctant to use time and resources travelling to a hospital unless complications occur that cannot be dealt with at home or in a local facility. Another interpretation is that the quality of health services at the hospitals in Uíge is very poor.

8.7.1.2 Infants dead

Regressions with number of infant deaths as dependent variable in Luanda and Uíge are presented in table D.2 and table D.4, respectively. In Luanda, households with medium level education have a 2.5% higher likelihood of number of infants dead compared to households with university level education, though the estimate is significant only at a 10% level. As was

the case for under-five mortality, none of the education dummies are significant in the Uíge regressions.

In Luanda, women who attended four or more antenatal controls during their last pregnancy have a lower (7.6% in the Poisson model) likelihood of having children who die before the age of one than those who did not. Antenatal care serves to discover and deal with complications and risk factors prior to birth, and is also educating pregnant women about nutrition, breastfeeding, prevention of illnesses etc. It is therefore according to expectations that women getting this kind of follow-up have lower expectations of child infant deaths than those who do not. I do not find any effect of antenatal controls in the Uíge region. A possible explanation for this is that quality of antenatal services is poor in Uíge.

Regarding place of delivery, I observe the same differences between the regions as was the case with under-five mortality.

8.7.2 Health facility characteristics

8.7.2.1 Under-five dead

In Luanda, the only health facility characteristic that is estimated to have an effect on likelihood of under-five deaths is whether or not the closest public facility offers antenatal services, though this is only significant in the OLS and logit model.

In Uíge, several health facility variables have an effect on under-five deaths. Contrary to what I would expect, the facilities in Uíge reporting to have power available during all opening hours the last week has a 16.4% higher likelihood of under-five deaths in their catchment areas than those that were reported not to have power continuously available. The effect is significant at a 1% level in both the Poisson and logit model. Compared to the catchment areas of facilities where no immunization services are offered, households nearby facilities that offer immunization have a 13.8% higher likelihood of having children die under the age of five. Households in catchment areas of health facilities that offer malaria services is also found to have a 6.7% higher expected number of under-five deaths than households in the catchment areas of those who don't in the Poisson model. These findings are not easily explained, but could be caused by reversed causality.

Furthermore, while not significant in Luanda, whether the closest health facility is offering delivery services is affecting likelihood of under-five deaths in Uíge. In the catchment areas of facilities that are offering delivery services expected number of under-five deaths in the Poisson model is 12.3% lower than in the catchment areas of those who do not. The effect is significant at a 5% level. A similar, but a bit smaller, effect is estimated in the logit model. An interesting question to pose is why delivery services do not seem to have an effect in Luanda. A possible answer is that that for a household residing in an area where the closest public health facility does not offer delivery services, it is probably easier to find, and shorter way of travelling to, another facility that does offer these services in Luanda than in Uíge where population density is lower and infrastructure is in worse shape.

In the Poisson model, households in the catchment area of facilities that employ midwifes have a 10.4% lower likelihood of under-five deaths compared to households in the catchment areas of facilities without midwifes. The effect is significant a 1% level. As midwives are thought to be important in assisting women in labour this finding is not very surprising. More interesting is the fact that this variable is not estimated to be statistically significant in the Luanda region. Again, the explanation for this could be better access to other health facilities in Luanda.

8.7.2.2 Infants dead

The dummy variable for whether a facility had power during all opening hours the last week has as significant effect on number of infant deaths in Uíge. There, households in catchment areas of facilities with power continuously available during all opening hours had 11.4% higher likelihood of number infants dead than households nearby facilities that did not have power in Poisson model. This effect is significant at a 1%. In comparison, the Luanda estimate is 3.8%, significant at a 5% level.

Households in the catchment areas of facilities with a midwife have a 4.9% higher expected number of infants dead in the Poisson model for Luanda, but the effect is only significant at a 10% level. Whether the closest public health facility has a midwife in their staff is not estimated to have any effect on infant deaths in Uíge.

The effects of availability of delivery and malaria services on infant deaths are similar to those found for under-fives.

8.7.3 Health worker characteristics

The effects of health worker performance on under-five and infant deaths for the two regions are similar to those in the whole sample.

8.7.4 Goodness of fit:

Although R^2 and pseudo-R^2 is not directly comparable to each other, I can use these statistics to compare the goodness of fit of each model between the regions. While R^2 in the OLS model is 0.134 in the regression with under-five deaths as dependent variable in Luanda, the variables included in the regression are much more successful in explaining the variation in the Uíge region where the R^2 is 0.349. The same pattern is observed in the logit model. Pseudo-R^2 is 0.1886 in the Luanda model, and 0.2822 in Uíge, indicating a better fit for the latter. In the infant-regressions I find similar differences. R^2 for the OLS is 0.114 in Luanda and 0.180 in Uíge, and in the logit model pseudo-R^2 is 0.1817 and 0.2059, respectively. No goodness of fit measure is presented for the Poisson model, but it is safe to assume that is displays a similar pattern with greater power of explanation for the Uíge regressions. It is interesting to note that while the goodness of fit is generally better for regressions with under-five as the dependent variable than those with infant deaths, this is not the case in the logit model in Luanda.

8.8 Summary of findings

In the analysis I find significant differences in mortality across households in catchment areas of different public health facilities. This could be interpreted as evidence to support the notion that access to and quality of health services are important determinants of health, because the types of services offered, equipment available and performance of the health workers varies from facility to facility. However, I cannot exclude the possibility that the relationship is spurious, i.e. caused by a third variable not included in the regression. For instance, it could be the case that more children die in some catchment areas than others because the rate of disease transmission is higher, not because the health services are less accessible and of poorer quality.

While a strong socio-economic gradient in under-five and infant deaths was found descriptively in chapter 7, the wealth effect did not persist when other important factors such as number of children born, education, use of, access to and quality of health services were

controlled for. There are two possible interpretations of this finding. One is that wealth itself is not important to child survival, and that the large differences in per household mortality across wealth quintiles found in the descriptive analysis is caused by other factors common to the households in the wealth quintiles such as education and region of residence. The other possibility is that the wealth index is not a good measure of wealth in the sample.

The descriptive differences in mortality across education categories are large, but when controlling for other factors the effect is less pronounced. Nevertheless, I conclude that the findings lend some support to the hypothesis that education and mortality is negatively related. Due to the fact that the education variables are measured imprecisely (the categories are course and do not reflect mother's education directly), the estimated effects of education could be thought to have attenuation biases. In the Uíge sample, no significant effects of education were found.

Place of delivery has a significant effect in the Poisson and logit model on both under-five and infant mortality. Access to delivery services in the closest public health facility is also found to be significant, both in the whole sample and in the Uíge region.

The significance of the health worker performance measures differs between the models, but coefficients are generally small.

9 Discussion and policy implications

The aim of this thesis has been to investigate determinants of child mortality in Angola, with a particular focus on wealth, education and health service utilization and access. Reducing child mortality is an internationally recognized development goal, and both policy makers and researchers have been paying much attention to the subject in the recent years. Because causes of death differ substantially between and within countries, understanding the local conditions and causes is crucial for obtaining further reductions in child mortality. While plentiful studies have been conducted elsewhere, there has, to my knowledge, been done little research on this matter in Angola.

In the analysis, I find large descriptive differences in child mortality between wealth quintiles. However, no effect is found in the regressions when including important control variables. Thus, the differences in child mortality across wealth quintiles observed in the descriptive analysis are caused by some other factor(s) common to the households in the same wealth groups, and not wealth itself. Investigating the distribution of these other factors between wealth segments would provide the knowledge needed to understand and address the large inequalities in child mortality between wealth quintiles in Angola, and is a topic for further research.

Education is found to be important both in the descriptive analysis and in the regressions. For under-five mortality, all education categories are associated with a higher likelihood of death compared to the university category in the sample, even after controlling for wealth. Hence, investing in education should be considered as part of a policy for reducing child mortality.

Another main finding in the thesis is the importance of use and access to delivery services. While *use* of delivery services, indicated by place of delivery, is important in both the whole sample and in the separate regions, *access* to delivery services at the closest public facility is not found to have any effect on child mortality in Luanda. These findings suggest that increasing the proportion of deliveries in health facilities should be part of a policy for reducing child mortality in the regions of Luanda and Uíge in Angola. This could be done by launching information campaigns or by using demand side incentives such as paying women to deliver in health facilities. Further research on the reasons for women not giving birth in

health facilities is needed, and policies should be made according to the findings from such investigations. In Uíge, access should also be improved by increasing the number of health facilities offering delivery services.

Furthermore, I find a significant effect of the number of children born in a household on child mortality. The data do not contain enough information to determine whether this effect only reflects that more children die when there are more children in a household, or rather that the number of children increases the probability of each child dying. However, the estimated average number of births per woman in Angola is quite high, 5.6, (World Bank, 2011) and reduction of fertility could be a possible way of reducing child mortality in a context where many households have limited resources. Thus, promoting family planning and prevention could be part of a health policy aiming to reduce child mortality, but the relationship between number of children born and health outcomes should be investigated further.

I wanted to investigate the effect of quality of health services by including availability of electricity, equipment and performance of health workers. However, the effects of these are ambiguous, and I do not find evidence to support that the quality of health services affects child mortality in our sample. Finding better indicators for it should be a topic for further research, because it will enable better assessments of the impact of the quality of the services on health outcomes.

A shortcoming of the study is the lack of data on important maternal factors; age (at the time of delivery), education level and birth spacing and infrastructure: water, sanitation, waste management and electricity. In addition, our calculated under-five and infant mortality rates cannot be compared to other estimates because the children were not followed until they reached the age of five. Despite these drawbacks, the data contain information on central issues in improving the health of people in Uíge and Luanda in Angola. The results presented here are hopefully a small contribution to the scarce literature on causes of child mortality in Angola.

While this thesis has investigated proximate and distal determinants of child mortality, a crucial contextual determinant has been left uncommented, namely the willingness of the Angolan government to commit to improvements in health. This together with policies made

on the basis of knowledge about the local conditions is needed to support a healthier future for the children of Angola.

Literature

Akoto, E. & B. Tambashe 2002. Socioeconomc Inequalities in Infant and Child Mortality among Urban and Rural areas in Sub-Saharan Africa. *IUSSP Committee on Emerging Health Threats.* Max Planck Institute for Demographic Research, Rostock

Alemayehu, T., J. Haidar & D. Habte 2009. Determinats of exclusive brestfeeding practices in Ethiopia. *Ethiopian Journal of Health Development,* 23**,** 12-18.

Angrist, J. D. & J.-S. Pischke 2009. *Mostly Harmless Econometrics: An Empiricist's Companion,* New Jersey, Princton University Press.

Anyamele, O. D. 2011. The Role of Wealth in Infant Mortality in Sub-Saharan Africa Within Urban and Between Urban and Rural Areas. *International Journal of Economics and Reserach,* 2**,** 73-81.

Avogo, W. A. & V. Agadjanian 2010. Forced migration and child health and mortality in Angola. *Social Science & Medicine,* 70**,** 53-60.

Bardhan, P. & C. Udry 1999. *Development Microeconomics,* Oxford, Oxford University Press.

Barfield, T. J. 1997. *The Dictionary of anthropology,* Oxford, Blackwell.

Becker, G. S. 1992. Fertility and the economy. *Journal of Population Economics,* 5**,** 185-201.

Bhutta, Z. A., M. Chopra, H. Axelson, P. Berman, T. Boerma, J. Bryce, F. Bustreo, E. Cavagnero, G. Cometto, B. Daelmans, A. de Francisco, H. Fogstad, N. Gupta, L. Laski, J. Lawn, B. Maliqi, E. Mason, C. Pitt, J. Requejo, A. Starrs, C. G. Victora & T. Wardlaw 2010. Countdown to 2015 decade report (2000-10): taking stock of maternal, newborn, and child survival. *The Lancet,* 375**,** 2032-2044.

Bicego, G. T. & J. Ties Boerma 1993. Maternal education and child survival: A comparative study of survey data from 17 countries. *Social Science & Medicine,* 36**,** 1207-1227.

Black, R. E., L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. de Onis, M. Ezzati, C. Mathers & J. Rivera 2008. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet,* 371**,** 243-260.

Black, R. E., S. Cousens, H. L. Johnson, J. E. Lawn, I. Rudan, D. G. Bassani, P. Jha, H. Campbell, C. F. Walker, R. Cibulskis, T. Eisele, L. Liu & C. Mathers 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The Lancet,* 375**,** 1969-1987.

Black, R. E., S. S. Morris & J. Bryce 2003. Where and why are 10 million children dying every year? *The Lancet,* 361**,** 2226-2234.

Bloom, D. E., D. Canning & D. T. Jamison 2004. Health, Wealth and Development. *Finance & Development,* 41.

British Petroleum 2010. BP Statistical Review of World Energy.

Caldwell, J. C. 1979. Education as a Factor in Mortality Decline An Examination of Nigerian Data. *Population Studies,* 33**,** 395-413.

Cameron, A. C. & P. K. Trivedi 1998. *Regression analysis of count data,* New York, Cambridge University Press.

CIA. 2011. *The World Factbook* [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/geos/ao.html [Accessed 24th March 2011].

Cleland, J. G. & J. K. van Ginneken 1988. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Science & Medicine,* 27**,** 1357-1368.

Commission on Social Determinants of Health 2008. CSDH final report: closing the gap in a generation: health equity through action on the social determinants of health. Geneva: World Health Organization.

Connor, C., Denise Averbug, and Maria Miralles 2010. Angola Health System Assessment 2010. Bethesda: USAID.

Cutler, D., A. Deaton & A. Lleras-Muney 2006. The Determinants of Mortality. *The Journal of Economic Perspectives,* 20**,** 97-120.

de Oliveira, R. S. 2007. Business success, Angola-style: postcolonial politics and the rise and rise of Sonangol. *The Journal of Modern African Studies,* 45**,** 595-619.

Dunteman, G. H. 1989. *Principal components analysis,* Thousand Oaks, Calif., Sage.

Feng, X. L., S. Guo, D. Hipgrave, J. Zhu, L. Zhang, L. Song, Q. Yang, Y. Guo & C. Ronsmans 2011. China's facility-based birth strategy and neonatal mortality: a population-based epidemiological study. *The Lancet,* 378**,** 1493-1500.

Frøystad, M., O. Mæstad & N. Villamil Forthcoming 2011. Health Services in Angola: Availability and utilization. Bergen: CMI.

Global Witness 2002. All the President's men : The devastating story of oil and banking in Angola's privatised war.

Global Witness 2004. Time for Transparency: Coming clean oil, mining and gas revenues.

Hill, K. 1991. Approaches to the Measurement of Childhood Mortality: A Comparative Review. *Population Index,* 57**,** 368-382.

Hill, K. 2003. Frameworks for studying the determinants of child survival. *Bulletin of the World Health Organization,* 81**,** 138-139.

Hobcraft, J. 1993. Women's education, child welfare and child survival: a review of the evidence. *Health Transition Review,* 3**,** 159-173.

Hobcraft, J. N., J. W. McDonald & S. O. Rutstein 1984. Socio-Economic Factors in Infant and Child Mortality: A Cross-National Comparison. *Population Studies,* 38**,** 193-223.

Hodges, T. 2004. *Angola: anatomy of an oil state,* Lysaker, Fridtjof Nansen Institute.

Houweling, T. A. J. & A. E. Kunst 2010. Socio-economic inequalities in childhood mortality in low- and middle-income countries: a review of the international evidence. *British Medical Bulletin,* 93**,** 7-26.

Huynen, M., P. Martens & H. Hilderink 2005a. The health impacts of globaliation: a conceptual framework. Bilthoven: Netherlands Environmental Assesment Agency (MNP).

Huynen, M., P. Martens & H. Hilderink 2005b. The health impacts of globalisation: a conceptual framework. *Globalization and Health,* 1**,** 14.

Instituto Nacional de Estatística 2010. Principais Resultados Definitivos Versão Resumida 2008-09. *Inquérito Integrado Sobre o Bem Estar da Populacão.* Luanda.

International Classification of Diseases 9 2000. Millennium Edition, International Classification of Diseases, Ninth Revision. Los Angeles: Practice Management Information Corporation [PMIC].

Kapuscinski, R. 1976. *Atter ein dag: Angola 1975,* Aschehough.

Ki-moon, B. 2010. Global Strategy for Women's and Children's Health. The Partnership for Maternal, Newborn and Child Health.

Kiros, G.-E. & D. P. Hogan 2001. War, famine and excess child mortality in Africa: the role of parental education. *International Journal of Epidemiology,* 30**,** 447-455.

Labonté, R. & T. Schrecker 2007. Globalization and social detreminants of health: Introduction and metodological background *Globalization and Health,* 3.

Lavy, V., J. Strauss, D. Thomas & P. de Vreyer 1996. Quality of health care, survival and health outcomes in Ghana. *Journal of Health Economics,* 15**,** 333-357.

Lewis, J. J. C., C. A. Donnelly, P. Mare, Z. Mupambireyi, G. P. Garnett & S. Gregson 2007. Evaluating the proximate determinants framework for HIV infection in rural Zimbabwe. *Sexually Transmitted Infections,* 83**,** i61-i69.

Lindkvist, I. 2011. *Health Worker Motivation and Effort in a Low-Income Context.* Philosophiae doctor, University of Bergen.

Lozano, R., H. Wang, K. J. Foreman, J. K. Rajaratnam, M. Naghavi, J. R. Marcus, L. Dwyer-Lindgren, K. T. Lofgren, D. Phillips, C. Atkinson, A. D. Lopez & C. J. L. Murray 2011. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. *The Lancet,* 378**,** 1139-1165.

Macassa, G., J. Hallqvist & J. W. Lynch 2011. Inequalities in child mortaliy in sub-Sahran Africa: A social epidemologic framework. *African Journal of Health Sciences* 18**,** 14-26.

Manda, S. O. M. 1999. Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. *Social Science & Medicine,* 48**,** 301-312.

Mogstad, M. & M. Wiswall 2010. Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects. *NYU Working Paper*.

Mosley, W. H. & L. C. Chen 1984. An Analytical Framework for the Study of Child Survival in Developing Countries. *Population and Development Review,* 10**,** 25-45.

Norwegian Embassy in Angola. 2011. *Stedsrapport* [Online]. The Norwegian Embassy in Angola. Available: http://www.noruega.ao/Norsk/Bosatt-i-Angola/stedsrapport/#anchor4 [Accessed 24th July 2011].

Palloni, A. & M. Tienda 1986. The effects of breastfeeding and pace of childbearing on mortality at early ages. *Demography,* 23**,** 31-52.

Pandey, A. 1998. Infant and child mortality in India. *National Family Health Survey Subject Reports; no 11.* Mumbai, India.

Rajaratnam, J. K., J. R. Marcus, A. D. Flaxman, H. Wang, A. Levin-Rector, L. Dwyer, M. Costa, A. D. Lopez & C. J. L. Murray 2010. Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. *The Lancet,* 375**,** 1988-2008.

Sastry, N. 1997. What explains rural-urban differentials in child mortality in Brazil? *Social Science & Medicine,* 44**,** 989-1002.

Simon, C. P. & L. Blume 1994. *Mathematics for economists,* New York, Norton.

Skolnik, R. 2008. *Essentials of global health,* Boston, Jones and Bartlett.

Smith, L. I. 2002. *A tutorial on Principal Components Analysis* [Online]. University of Otago, New Zealand. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf [Accessed 26th September 2011].

Spector, J. M. & S. Daga 2008. Preventing those so-called stillbirths. *Bulletin of the World Health Organization,* 86**,** 315-316.

The Global Campaign for the Health Millenium Development Goals 2011. Innovating for Every Woman, Every Child. *Thematic Report.* Oslo.

The World Bank 2010. *World development indicators. 2010,* Washington, D.C., World Bank Publications.

Todaro, M. P. & S. C. Smith 2006. *Economic development,* Harlow, Pearson/Addison Wesley.

Transperency International 2010. Corruption Perceptions Index 2010.

UN. 2011. *Millenium Development Goals* [Online]. Available: http://www.un.org/millenniumgoals/ [Accessed 22th March 2011].

UNDP 2010. *The real wealth of nations: pathways to human development,* New York, Oxford University Press.

United Nations High Commissioner for Human Rights 2008. The Right to Health, Fact Sheet No. 31. *In:* UNITED NATIONS (ed.). Geneva.
USAID 2009. USAID Country Health Statistical report: Angola.
USAID. 2011. *Demographic and Health Surveys* [Online]. Available: http://www.measuredhs.com/countries/country_main.cfm?ctry_id=76&c=Angola [Accessed 26th March 2011].
Verbeek, M. 2008. *A guide to modern econometrics,* Chichester, Wiley.
Vyas, S. & L. Kumaranayake 2006. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning,* 21**,** 459-468.
Wagstaff, A. 2002. Poverty and health sector inequalities. *Bulletin of the World Health Organization,* 80**,** 97-105.
Walker, J. T. & S. Maddan 2005. *Statistics in criminology and criminal justice: analysis and interpretation,* Sudbury, Mass., Jones and Bartlett Publishers.
Wang, L. 2003. Determinants of child mortality in LDCs: Empirical findings from demographic and health surveys. *Health Policy,* 65**,** 277-299.
Wooldridge, J. M. 2006. *Introductory econometrics: a modern approach,* Mason, Ohio, Thomson South-Western.
World Bank 2009. *World Development Report 2010: Development and Climate Change,* Washington, DC.
World Bank 2010. *World development indicators. 2010,* Washington, D.C., World Bank Publications.
World Bank 2011. *World Development Indicators 2011,* Washington, D.C., The World Bank.
World Health Organization 1998. Postpartum care of the mother and newborn: a practical guide.
World Health Organization 2009a. Infant and young child feeding: Model Chapter for textbooks for medical students and allied health professional.
World Health Organization 2009b. Women and health: today's evidence tomorrow's agenda. World Health Organization.
World Health Organization 2010a. Trends in maternal mortality: 1990 to 2008. Geneva: World Health Organization, World Bank, UNICEF, United Nations Population Fund.
World Health Organization 2010b. World Health Statistics 2010: Indicator compendium, Interim version.
World Health Organization. 2011a. *10 facts on breastfeeding* [Online]. Available: http://www.who.int/features/factfiles/breastfeeding/en/index.html [Accessed 8th September 2011].
World Health Organization. 2011b. *Antenatal care coverage* [Online]. [Accessed 28th November 2011].
World Health Organization. 2011c. *Children: reducing mortality* [Online]. Available: http://www.who.int/mediacentre/factsheets/fs178/en/index.html [Accessed 20th November 2011].
World Health Organization. 2011d. *Infectious diseases* [Online]. Available: http://www.who.int/topics/infectious_diseases/en/ [Accessed 20th November 2011].
World Health Organization. 2011e. *Social Determinants of Health* [Online]. Available: http://www.who.int/social_determinants/en/ [Accessed 18. May 2011].
World Health Organization 2011f. World Health Statistics 2011: Indicator code book. *World Health Statistics Indicators.*
World Health Organization & UNICEF 2010. Countdown to 2015 decade report (2000–2010): taking stock of maternal, newborn and child survival. *WHO Library Cataloguing-in-Publication Data*.

Appendix A: Descriptive statistics

Table A.1: Name, description, number of observations, mean, standard deviation, min and max values for all variables in the analysis

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Dependent variables						
under5_dead	Number of children under-five dead in the household	931	.1976369	.5424306	0	5
infant_dead	Number of infants dead in the household	946	.115222	.3657865	0	3
Household characteristics						
land_pc1	Wealth index derived from PCA	943	-.0456762	3.394099	-5.403555	7.427008
children_born	Number of children born alive in the household	946	2.108879	1.0905	1	9
children_616	Number of children aged 6-16 years in the household	946	3.418605	1.59968	1	6
youths_1721	Number of youths aged 17-21 years in the household	946	1.506342	.5002242	1	2
noeducation	Education dummy for households where no adults have attained any education	946	.0306554	.1724734	0	1
primary	Education dummy for households where no adults have attained a higher level than primary education	946	.0634249	.2438546	0	1
secondary	Education dummy for households where no adults have attained more than secondary level of education	946	.1490486	.3563248	0	1
professional	Education dummy for households where no adults have attained more than professional training	946	.3171247	.4656026	0	1
medium	Education dummy where no adults have attained more than medium level education	946	.320296	.4668371	0	1
university	Education reference dummy	946	.1194503	.3244892	0	1
age_mother	Age of the woman in the household who was the last to give birth	942	28.31741	7.865507	1	62
age_mother_sq	(age_mother)^2	942	863.6762	490.5211	1	3844
alwaysfood	Dummy for households that always have enough food for the household members	945	.7904762	.4071843	0	1
a_control4	Dummy for households where the last woman to give birth attended 4 or	941	.7577046	.4287	0	1

	more antenatal controls					
del_home_post_other	Dummy for households where the last woman to give birth delivered home, in a health post or "other"	939	.4100106	.4920974	0	1
del_centre_other_fac	Dummy for households where the last woman to give birth delivered in a hospital	939	.1906283	.3930059	0	1
del_hospital	Dummy for households where the last woman to give birth delivered in a health centre or other type of health facility	939	.399361	.4900281	0	1
att_doctor	Dummy for households where the last woman to give birth was attended by a medical doctor, nurse or midwife	936	.667735	.4712772	0	1
checkup	Dummy for households where the last woman to give birth was checked by a professional health worker during the first week after delivery	938	.6012793	.4898963	0	1
breastf	Dummy for households where the last woman to give birth breastfed her child	933	.9549839	.2074506	0	1
qual_very_low_low	Dummy for households that perceived quality of nearest public health facility as very low or low	946	.3255814	.4688395	0	1
qual_medium	Dummy for households that perceived quality of nearest public health facility as medium	946	.538055	.4988134	0	1
qual_high_very_high	Dummy for households that perceived quality of nearest public health facility as high or very high	946	.1331924	.3399623	0	1
urban_uige	Dummy for households situated in the Luanda region	946	.2008457	.4008448	0	1
luanda	Dummy for households classified as urban in the Uíge region	946	.4936575	.5002242	0	1
rural	Dummy for households classified as rural	946	.3054968	.4608612	0	1
Health facility characteristics						
power	Dummy for health facilities where electricity was continously available at the times the facility was open for services the last week before the survey	946	.6765328	.4680467	0	1
im_facility_outreach	Dummy for facilities that offers immunization services either at the	921	.8990228	.3014622	0	1

	facility or as outreach, or both					
im_no (reference dummy)	Dummy for facilities that do not offer any immunization services	921	.1009772	.3014622	0	1
antenatal_serv	Dummy for health facilities offering antenatal services	946	.6987315	.4590518	0	1
del_serv	Dummy for health facilities offering delivery services	946	.4556025	.4982884	0	1
malaria_serv	Dummy for health facilities offering malaria services	946	.8255814	.3796698	0	1
stethoscope	Dummy for health facilities where a stethoscope was available and functioning the day the survey was conducted	946	.8498943	.3573639	0	1
thermometer	Dummy for health facilities where a thermometer was available and functioning the day the survey was conducted	946	.9270613	.2601734	0	1
hf_fridge	Dummy for healt facilities where a refrigarator was available and functional the day of the survey was conducted	946	.6733615	.4692319	0	1
midwife	Dummy for health facilities where at least one midwife was employed at the time of the survey	946	.372093	.4836187	0	1
spes_doctor	Dummy for health facilities where at least one specialized medical doctor was employed at the time of the survey	946	.2019027	.4016324	0	1
Health worker characteristics						
nquest_tot	Number of questions on list that health worker asked during the consultations. Highest attainable score: 53	946	29.34567	11.88436	5	46
nexam_tot	Numberof examinations on the list the health worker did during consultations. Highest attainable score: 36	946	15.20085	9.527117	1	35
diag_tot	Number of the diagnoses the doctor made correctly. Highest attainable score: 5	946	.4306554	.2711352	0	5

Table A.2: Per household values of number of children born, under-five dead, infants dead, under-five mortality rate and infant mortality rate

Facility	Region	Born	Under-5 dead	Infant dead	Rate(<5)	Rate(<1)

C.S. Paz	Luanda	2,20	0,00	0,00	**0,00**	**0,00**
C.S. Vila da Mata	Luanda	2,10	0,50	0,20	**245,37**	100,00
Centro da Pedeira	Uíge	2,04	0,00	0,00	**0,00**	**0,00**
Centro de saude de Quitexe	Uíge	2,12	0,16	0,12	76,67	63,33
Centro de saude de Sousa	Uíge	2,05	0,14	0,05	39,68	15,87
Centro de saude de Kifutila	Uíge	2,40	0,24	0,20	82,67	74,67
Centro materno infantile de Puri	Uíge	2,57	0,39	0,22	113,77	79,71
CS 11 de Novembro	Luanda	2,22	0,13	0,13	52,17	52,17
CS Boa vista	Luanda	1,84	0,24	0,16	126,98	93,33
CS Cariango	Luanda	1,55	0,29	0,14	178,57	68,18
CS Palanca	Luanda	1,92	0,08	0,08	21,33	21,33
CS Siga	Luanda	1,92	0,29	0,25	159,72	**138,89**
CS progresso	Luanda	1,71	0,04	0,04	13,89	13,89
Camama	Luanda	1,96	0,09	0,04	32,61	10,87
Divina providencia	Luanda	1,72	0,04	0,00	20,00	0,00
H. G. A. Ngangula	Luanda	1,37	0,11	0,05	78,95	26,32
H. G. Cajueiro	Luanda	2,35	0,17	0,13	40,58	26,09
H. G. Maternidad Kilamba Kiaxins pasted	Luanda	1,64	0,04	0,04	13,33	13,33
Hospital muncipla do Puri	Uíge	2,44	0,44	0,12	137,33	40,00
Hospital Provincal do Uige	Uíge	2,00	0,12	0,04	33,33	10,00
Ilha do cabo	Luanda	2,28	0,12	0,04	41,33	13,33
Kandonbe velho	Uíge	1,91	0,13	0,09	50,72	28,99
Malangino	Luanda	1,86	0,00	0,00	**0,00**	**0,00**
Nossa senhora da Paz	Luanda	2,74	0,13	0,09	43,48	28,99
Posto do saude de Cassexe	Uíge	2,88	0,60	0,24	184,76	120,00
Posto de saude de Kibaba	Uíge	2,25	0,21	0,13	65,97	41,67
Posto de saude de Quitoque	Uíge	2,25	0,29	0,21	104,17	76,39
Posto de saude de Povo Mateus02	Uíge	1,80	0,12	0,12	46,67	46,67
Ps Tala Hady	Luanda	2,00	0,04	0,00	14,49	0,00
Santa Catarina	Luanda	2,22	0,30	0,17	78,26	52,17
Santa Terezinha	Luanda	1,46	0,04	0,04	20,83	20,83
Vista Alegere	Uíge	2,05	0,32	0,05	54,92	15,15
Wegi Maca	Luanda	2,08	0,13	0,04	41,67	13,33
Centro de saude materno infantil do uige	Uíge	2,16	0,20	0,12	80,00	46,67
Posto de saude de camancoco	Uíge	2,38	0,42	0,29	106,15	73,41
Posto de saude de cambamba	Uíge	2,00	0,15	0,15	87,50	87,50
Posto de saude de cambila	Uíge	2,57	0,30	0,17	112,32	79,71
Posto de saude do mbanza pombo	Uíge	2,32	0,28	0,24	84,67	71,33
Posto de saude kizambi	Uíge	2,40	0,36	0,20	110,00	66,67
Posto de saude ngunga cruz	Uíge	2,44	0,32	0,20	84,00	44,00

Table A.3: Facility dummies, names and region

Dummy	Facility	Region
REFERENCE	C.S. Paz	Luanda

_Ifacility_2	C.S. Vila da Mata	Luanda
_Ifacility_3	Centro da Pedreira	Uige
_Ifacility_4	Centro de saude de Quitexe	Uige
_Ifacility_5	Centro de saude de Sousa	Uige
_Ifacility_6	Centro de saude de Kifutila	Uige
_Ifacility_7	Centro materno infantile de Puri	Uige
_Ifacility_8	CS 11 de Novembro	Luanda
_Ifacility_9	CS Boa vista	Luanda
_Ifacility_10	CS Cariango	Luanda
_Ifacility_11	CS Palanca	Luanda
_Ifacility_12	CS Siga	Luanda
_Ifacility_13	CS progresso	Luanda
_Ifacility_14	Camama	Luanda
_Ifacility_15	Divina providencia	Luanda
_Ifacility_16	H. G. A. Ngangula	Luanda
_Ifacility_17	H. G. Cajueiro	Luanda
_Ifacility_18	H. G. Maternidad Kilamba Kiaxins pasted	Luanda
_Ifacility_19	Hospital municipal do Puri	Uige
_Ifacility_20	Hospital Provincial do Uige	Uige
_Ifacility_21	Ilha do cabo	Luanda
_Ifacility_22	Kandonbe velho	Uige
_Ifacility_23	Malangino	Luanda
_Ifacility_24	Nossa senhora da Paz	Luanda
_Ifacility_25	Posto de saude de cassexe	Uige
_Ifacility_26	Posto de saude de Kibaba	Uige
_Ifacility_27	Posto de saude de Quitoque	Uige
_Ifacility_28	Posto de saude de Povo Mateus02	Uige
_Ifacility_29	Ps Tala Hady	Luanda
_Ifacility_30	Santa Catarina	Luanda
_Ifacility_31	Santa Terezinha	Luanda
_Ifacility_32	Vista Alegre	Uige
_Ifacility_33	Wegi Maca	Luanda
_Ifacility_34	centro de saude materno infantil do uige	Uige
_Ifacility_35	posto de saude de camancoco	Uige
_Ifacility_36	posto de saude de cambamba	Uige
_Ifacility_37	posto de saude de cambila	Uige
_Ifacility_38	posto de saude do mbanza pombo	Uige
_Ifacility_39	posto de saude kizambi	Uige
_Ifacility_40	posto de saude ngunga cruz	Uige

Appendix B: Poisson tests

Table B.1: Regression results from negative binomial regression

	Under-five		Infant	
	Coefficient (Std. Err)	Marginal Effects (Std. Err)	Coefficient (Std. Err)	Marginal Effects (Std. Err)
children_born	0.682***	0.077***	0.594***	0.042***
	(0.042)	(0.008)	(0.081)	(0.005)
land_pc1	0.015	0.002	-0.001	-0.000
	(0.048)	(0.005)	(0.067)	(0.005)
noeduc_primary_secondary (d)	0.803***	0.114**	0.780*	0.069
	(0.286)	(0.050)	(0.404)	(0.046)
professional (d)	0.615**	0.079*	0.614	0.049
	(0.303)	(0.045)	(0.408)	(0.038)
medium (d)	0.477**	0.059*	0.280	0.021
	(0.238)	(0.033)	(0.350)	(0.028)
a_control4 (d)	0.005	0.001	-0.359	-0.028
	(0.216)	(0.024)	(0.302)	(0.025)
del_home_post_other (d)	0.501**	0.060*	0.567**	0.043*
	(0.255)	(0.033)	(0.287)	(0.024)
del_centre_other_fac (d)	0.020	0.002	-0.039	-0.003
	(0.323)	(0.037)	(0.345)	(0.024)
att_doctor (d)	0.045	0.005	0.310	0.021
	(0.199)	(0.022)	(0.275)	(0.018)
breastf (d)	-0.526	-0.076	-0.938**	-0.105
	(0.407)	(0.072)	(0.415)	(0.066)
urban_uige (d)	-0.149	-0.016	-0.403	-0.026
	(0.404)	(0.042)	(0.475)	(0.027)
luanda (d)	-0.505	-0.057	-0.480	-0.034
	(0.484)	(0.056)	(0.647)	(0.047)
power (d)	0.379	0.040	0.591*	0.038*
	(0.343)	(0.033)	(0.355)	(0.020)
im_facility_outreach (d)	0.555**	0.051**	0.206	0.013
	(0.276)	(0.020)	(0.377)	(0.022)
antenatal_serv (d)	0.083	0.009	0.138	0.010
	(0.157)	(0.017)	(0.281)	(0.019)
del_serv (d)	-0.570**	-0.064**	-0.803***	-0.057***
	(0.274)	(0.030)	(0.306)	(0.021)
malaria_serv (d)	0.182	0.019	0.236	0.015
	(0.184)	(0.019)	(0.206)	(0.012)
thermometer (d)	0.000	0.000	0.105	0.007
	(0.207)	(0.023)	(0.204)	(0.013)
midwife (d)	0.386	0.046	0.683**	0.054*
	(0.323)	(0.043)	(0.300)	(0.028)
spes_doctor (d)	-0.098	-0.011	0.158	0.012
	(0.285)	(0.031)	(0.282)	(0.021)

nquest_tot	-0.032**	-0.004**	-0.022	-0.002
	(0.014)	(0.002)	(0.022)	(0.002)
nexam_tot	0.065**	0.007**	0.045	0.003
	(0.027)	(0.003)	(0.037)	(0.003)
diag_tot	-0.220**	-0.025**	-0.211*	-0.015**
	(0.102)	(0.010)	(0.110)	(0.007)
α	***3.88e-07***		***0.065241***	
	(8.50e-07)		*(0.3748314)*	
Log pseudolikelihood	*-365.43*		*-274.68*	
Wald chi2(23)	*1765.80*		*338.16*	
Prob > chi2	*0.0000*		*0.0000*	
N	*879*		*893*	

*: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table B.2: Goodness of fit test for Poisson models

	Under-five	Infants
Goodness-of-fit chi2	455.042	366.7412
Prob > chi2 (observations)	1.0000 (855)	1.0000 (869)

Appendix C: Facility dummies

Table C.1: OLS, Poisson and binary logit regression with under-five deaths as dependent variable and facility dummies

	OLS (Std. Err)	Poisson		Binary Logit	
		Coefficients (Std. Err)	Marginal effects (Std. Err)	Coefficients (Std. Err)	Marginal effects (Std. Err)
children_born	0.239***	0.631***	0.021***	1.123***	0.062***
	(0.039)	(0.084)	(0.002)	(0.127)	(0.003)
land_pc1	0.010	0.032	0.001	0.089	0.005
	(0.012)	(0.062)	(0.002)	(0.076)	(0.004)
noeduc_primary_secondary (d)	0.130*	0.554*	0.021	0.752*	0.049
	(0.064)	(0.331)	(0.015)	(0.457)	(0.036)
professional (d)	0.084	0.696**	0.026*	0.639	0.039
	(0.055)	(0.333)	(0.015)	(0.500)	(0.034)
medium (d)	0.062	0.325	0.011	0.443	0.026
	(0.037)	(0.289)	(0.011)	(0.424)	(0.027)
a_control4 (d)	-0.012	0.057	0.002	-0.085	-0.005
	(0.048)	(0.228)	(0.007)	(0.298)	(0.017)
del_home_post_other (d)	0.125**	0.641***	0.023**	1.002***	0.061**
	(0.060)	(0.236)	(0.009)	(0.389)	(0.026)
del_centre_other_fac (d)	0.006	0.072	0.002	0.063	0.004
	(0.050)	(0.331)	(0.011)	(0.498)	(0.028)
att_doctor (d)	0.123	0.293	0.009	0.655**	0.033**
	(0.075)	(0.206)	(0.007)	(0.313)	(0.015)
breastf (d)	-0.091	-0.400	-0.016	-0.867	-0.068
	(0.081)	(0.397)	(0.019)	(0.646)	(0.068)
urban_uige (d)	-0.330	0.040	0.001	-0.671	-0.031
	(0.292)	(0.427)	(0.014)	(0.681)	(0.027)
luanda (d)	-0.508*	-16.903***	-126.297*	-19.359***	-0.999***
	(0.298)	(1.113)	(69.005)	(0.732)	(0.001)
_Ifac_numbe_2 (d)	0.488***	18.260***	1955446.451	20.874***	0.961***
	(0.024)	(1.025)	(1912137.269)	(0.126)	(0.004)
_Ifac_numbe_3 (d)	-0.119***	-16.183***	-0.052***		
	(0.037)	(1.037)	(0.005)		
_Ifac_numbe_4 (d)	0.022	0.149	0.005	0.368	0.024
	(0.043)	(0.208)	(0.008)	(0.263)	(0.019)
_Ifac_numbe_5 (d)	-0.130	-0.134	-0.004	-0.254	-0.013
	(0.124)	(0.239)	(0.007)	(0.372)	(0.017)
_Ifac_numbe_6 (d)	-0.287	0.093	0.003	-0.551	-0.024
	(0.304)	(0.415)	(0.015)	(0.708)	(0.024)
_Ifac_numbe_7 (d)	-0.142	0.399	0.016	-0.033	-0.002
	(0.254)	(0.361)	(0.017)	(0.585)	(0.031)
_Ifac_numbe_8 (d)	0.109***	16.739***	407789.343	18.655***	0.963***

	(0.025)	(1.027)	(400652.466)	(0.240)	(0.003)
_Ifac_numbe_9 (d)	0.263***	17.403***	795115.424	19.439***	0.963***
	(0.034)	(1.032)	(782825.837)	(0.231)	(0.003)
_Ifac_numbe_10 (d)	0.442***	18.153***	1912526.349	20.757***	0.957***
	(0.055)	(1.063)	(1936544.300)	(0.171)	(0.004)
_Ifac_numbe_11 (d)	0.205***	16.526***	318996.149	18.704***	0.964***
	(0.035)	(1.031)	(313100.924)	(0.173)	(0.003)
_Ifac_numbe_12 (d)	0.409***	17.987***	1378356.138	20.438***	0.964***
	(0.033)	(1.029)	(1343215.185)	(0.144)	(0.003)
_Ifac_numbe_13 (d)	0.142***	16.001***	187856.192	17.954	0.964***
	(0.040)	(1.041)	(184630.373)	.	(0.003)
_Ifac_numbe_14 (d)	0.151***	16.470***	307777.194	18.471***	0.963***
	(0.026)	(1.030)	(301814.353)	(0.179)	(0.003)
_Ifac_numbe_15 (d)	0.165***	16.062***	199363.715	18.371***	0.965***
	(0.035)	(1.030)	(193805.752)	(0.183)	(0.003)
_Ifac_numbe_16 (d)	0.269***	16.896***	503200.036	19.163***	0.961***
	(0.073)	(1.100)	(523936.660)	(0.372)	(0.004)
_Ifac_numbe_17 (d)	0.130***	16.853***	438490.219	17.849***	0.964***
	(0.022)	(1.018)	(425404.658)	(0.268)	(0.003)
_Ifac_numbe_18 (d)	0.163***	16.090***	204729.492	18.111***	0.965***
	(0.052)	(1.053)	(203438.513)	(0.176)	(0.003)
_Ifac_numbe_19 (d)	0.256***	0.821***	0.041**	1.846***	0.215***
	(0.053)	(0.238)	(0.020)	(0.281)	(0.054)
_Ifac_numbe_20	0.000				
	(0.000)				
_Ifac_numbe_21 (d)	0.102*	16.762***	401132.699	18.520***	0.964***
	(0.051)	(1.069)	(407162.928)	(0.284)	(0.003)
_Ifac_numbe_22 (d)	0.037	0.180**	0.006*	0.747***	0.056***
	(0.026)	(0.084)	(0.003)	(0.137)	(0.014)
_Ifac_numbe_23 (d)	0.058*	0.479	0.020		
	(0.033)	(1.443)	(0.075)		
_Ifac_numbe_24 (d)	0.007	16.430***	290563.343	17.250***	0.963***
	(0.062)	(1.094)	(303062.513)	(0.458)	(0.003)
_Ifac_numbe_25 (d)	-0.068	-0.107	-0.003	0.138	0.008
	(0.312)	(0.521)	(0.015)	(0.700)	(0.043)
_Ifac_numbe_26 (d)	-0.284	0.302	0.011	0.416	0.027
	(0.305)	(0.401)	(0.017)	(0.689)	(0.053)
_Ifac_numbe_27 (d)	-0.220	0.312	0.012	0.714	0.053
	(0.301)	(0.417)	(0.018)	(0.688)	(0.066)
_Ifac_numbe_28 (d)	-0.252	0.002	0.000	0.105	0.006
	(0.273)	(0.389)	(0.013)	(0.603)	(0.036)
_Ifac_numbe_29 (d)	0.077**	15.983***	191440.736	17.848***	0.963***
	(0.038)	(1.045)	(189786.979)	(0.189)	(0.003)
_Ifac_numbe_30 (d)	0.279***	17.468***	814403.777	18.779***	0.964***
	(0.033)	(1.022)	(793288.828)	(0.176)	(0.003)
_Ifac_numbe_31 (d)	0.220***	16.451***	302108.357	18.668***	0.964***
	(0.058)	(1.069)	(304434.232)	(0.197)	(0.003)
_Ifac_numbe_32 (d)	0.169***	-0.274	-0.008	0.210	0.013
	(0.052)	(0.336)	(0.008)	(0.353)	(0.023)
_Ifac_numbe_33 (d)	0.147***	16.779***	408140.726	18.326***	0.964***

	(0.025)	(1.023)	(397958.015)	(0.178)	(0.003)
_Ifac_numbe_34 (d)	0.048	0.349**	0.014**	1.077***	0.093***
	(0.033)	(0.138)	(0.007)	(0.175)	(0.022)
_Ifac_numbe_35 (d)	-0.115	0.339	0.013	0.527	0.036
	(0.300)	(0.442)	(0.020)	(0.724)	(0.060)
_Ifac_numbe_37 (d)	-0.229	0.313	0.012	0.052	0.003
	(0.268)	(0.350)	(0.015)	(0.578)	(0.033)
_Ifac_numbe_38 (d)	-0.241	0.394	0.016	0.308	0.019
	(0.304)	(0.414)	(0.019)	(0.689)	(0.049)
_Ifac_numbe_39 (d)	-0.175	0.704*	0.033	0.576	0.040
	(0.306)	(0.414)	(0.026)	(0.696)	(0.060)
_Ifac_numbe_40 (d)	-0.313	0.014	0.000	-0.023	-0.001
	(0.296)	(0.497)	(0.016)	(0.755)	(0.041)
Log pseudolikelihood		*-353.90*		*-268.03*	
R^2	*0.2906*				
pseudo R^2				*0.257*	
N	*878*	*878*		*833*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table C.2: Regression results for OLS, Poisson and binary logit with infant deaths as dependent variable and facility dummies

	OLS (Std. Err)	Poisson		Binary Logit	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
children_born	0.099***	0.528***	0.007***	0.755***	0.034***
	(0.022)	(0.108)	(0.001)	(0.171)	(0.005)
land_pc1	0.006	0.011	0.000	0.026	0.001
	(0.009)	(0.082)	(0.001)	(0.105)	(0.005)
noeduc_primary_secondary (d)	0.089*	0.698	0.011	0.648	0.034
	(0.048)	(0.451)	(0.008)	(0.578)	(0.035)
professional (d)	0.061	0.748*	0.011	0.671	0.034
	(0.041)	(0.432)	(0.007)	(0.555)	(0.031)
medium (d)	0.029	0.255	0.003	0.235	0.011
	(0.028)	(0.399)	(0.005)	(0.504)	(0.025)
a_control4 (d)	-0.042	-0.258	-0.003	-0.233	-0.011
	(0.040)	(0.313)	(0.004)	(0.349)	(0.017)
del_home_post_other (d)	0.105**	0.774***	0.011**	1.078***	0.053**
	(0.041)	(0.290)	(0.005)	(0.359)	(0.021)
del_centre_other_fac (d)	0.017	0.126	0.002	-0.003	-0.000
	(0.033)	(0.384)	(0.005)	(0.459)	(0.021)
att_doctor (d)	0.103**	0.491*	0.006*	0.887**	0.036***
	(0.046)	(0.256)	(0.003)	(0.353)	(0.014)
breastf (d)	-0.112	-0.857**	-0.016	-1.230**	-0.093
	(0.067)	(0.431)	(0.012)	(0.604)	(0.067)
urban_uige (d)	0.030	0.427	0.006	0.144	0.007
	(0.103)	(0.524)	(0.008)	(0.755)	(0.037)

luanda (d)	-0.041	-15.273***	-19.908*	-16.854***	-0.991***
	(0.105)	(1.121)	(10.844)	(0.813)	(0.004)
_Ifacility_2 (d)	0.170***	15.980***	77802.356	18.602***	0.969***
	(0.019)	(1.033)	(76681.504)	(0.216)	(0.004)
_Ifacility_3 (d)	-0.047	-15.370***	-0.019***		
	(0.028)	(1.136)	(0.002)		
_Ifacility_4 (d)	0.062*	-0.271	-0.003	1.026***	0.072**
	(0.031)	(0.519)	(0.005)	(0.338)	(0.034)
_Ifacility_5 (d)	-0.015	-1.350***	-0.009***	-0.198	-0.008
	(0.047)	(0.425)	(0.002)	(0.394)	(0.015)
_Ifacility_6 (d)	0.143	0.189	0.003	1.134	0.084
	(0.103)	(0.160)	(0.002)	(0.789)	(0.088)
_Ifacility_7 (d)	0.130	0.068	0.001	0.645	0.038
	(0.085)	(0.247)	(0.003)	(0.618)	(0.048)
_Ifacility_8 (d)	0.137***	15.668***	55394.505	18.231***	0.970***
	(0.016)	(1.029)	(54550.756)	(0.289)	(0.003)
_Ifacility_9 (d)	0.145***	15.583***	48452.563	18.264***	0.972***
	(0.024)	(1.044)	(48098.432)	(0.260)	(0.003)
_Ifacility_10 (d)	0.201***	16.095***	85564.584	18.738***	0.970***
	(0.033)	(1.062)	(86037.860)	(0.324)	(0.004)
_Ifacility_11 (d)	0.143***	15.284***	36817.664	17.897***	0.971***
	(0.025)	(1.037)	(36622.644)	(0.118)	(0.003)
_Ifacility_12 (d)	0.325***	16.645***	143877.384	19.282***	0.971***
	(0.022)	(1.037)	(141209.256)	(0.266)	(0.004)
_Ifacility_13 (d)	0.088***	14.784***	22235.115	17.247***	0.971***
	(0.025)	(1.050)	(22096.019)	(0.205)	(0.003)
_Ifacility_14 (d)	0.081***	14.575***	18723.923	16.985	0.970***
	(0.017)	(1.033)	(18505.708)	.	(0.004)
_Ifacility_15 (d)	0.058**	0.240	0.003		
	(0.024)	(1.451)	(0.023)		
_Ifacility_16 (d)	0.155***	15.626***	55932.130	18.130***	0.968***
	(0.051)	(1.156)	(61319.859)	(0.558)	(0.004)
_Ifacility_17 (d)	0.103***	15.270***	36319.821	16.472***	0.970***
	(0.016)	(1.027)	(35825.998)	(0.224)	(0.003)
_Ifacility_18 (d)	0.100***	14.882***	24467.565	17.273***	0.971***
	(0.034)	(1.063)	(24556.799)	(0.271)	(0.004)
_Ifacility_19 (d)	0.059*	-0.403	-0.004	0.935**	0.063*
	(0.033)	(0.542)	(0.005)	(0.373)	(0.036)
_Ifacility_20 (d)	0.000	-1.161*	-0.009***		
	(0.000)	(0.594)	(0.003)		
_Ifacility_21 (d)	0.036	14.543***	17869.513	16.806***	0.970***
	(0.037)	(1.104)	(18757.118)	(0.459)	(0.004)
_Ifacility_22 (d)	0.071***	-0.222	-0.002	1.108***	0.081***
	(0.017)	(0.543)	(0.006)	(0.179)	(0.018)
_Ifacility_23 (d)	0.020	0.325	0.005		
	(0.023)	(1.451)	(0.025)		
_Ifacility_24 (d)	0.042	15.060***	29573.384	17.217***	0.970***
	(0.044)	(1.136)	(32053.420)	(0.530)	(0.004)
_Ifacility_25 (d)	0.104	-0.615	-0.006**	1.092	0.079

	(0.107)	(0.442)	(0.003)	(0.781)	(0.086)
_Ifacility_26 (d)	0.100	0.103	0.001	1.239*	0.096
	(0.104)	(0.254)	(0.003)	(0.752)	(0.090)
_Ifacility_27 (d)	0.155	0.176	0.002	1.525**	0.133
	(0.102)	(0.181)	(0.003)	(0.744)	(0.108)
_Ifacility_28 (d)	0.117	0.017	0.000	1.285**	0.101
	(0.095)	(0.172)	(0.002)	(0.636)	(0.079)
_Ifacility_29 (d)	0.020	0.281	0.004		
	(0.026)	(1.466)	(0.024)		
_Ifacility_30 (d)	0.129***	15.435***	43384.623	17.467***	0.970***
	(0.021)	(1.037)	(43123.015)	(0.220)	(0.004)
_Ifacility_31 (d)	0.129***	15.230***	35509.564	17.750***	0.970***
	(0.039)	(1.090)	(36535.771)	(0.351)	(0.004)
_Ifacility_32 (d)	-0.015	-1.765***	-0.011***	-0.808	-0.026***
	(0.035)	(0.423)	(0.001)	(0.505)	(0.010)
_Ifacility_33 (d)	0.059***	14.435***	15833.675	16.835***	0.971***
	(0.019)	(1.023)	(15460.913)	(0.157)	(0.003)
_Ifacility_34 (d)	0.064***	-0.220	-0.002	1.129***	0.083***
	(0.021)	(0.523)	(0.005)	(0.212)	(0.024)
_Ifacility_35 (d)	0.226**	0.198	0.003	1.556**	0.138
	(0.102)	(0.242)	(0.004)	(0.776)	(0.115)
_Ifacility_36	0.094			1.202	0.092
	(0.102)			(0.806)	(0.095)
_Ifacility_37 (d)	0.096	0.027	0.000	1.195*	0.091
	(0.091)	(0.213)	(0.003)	(0.642)	(0.076)
_Ifacility_38 (d)	0.187*	0.494**	0.008**	1.752**	0.168
	(0.106)	(0.198)	(0.004)	(0.755)	(0.126)
_Ifacility_39 (d)	0.146	0.370*	0.006	1.281*	0.101
	(0.104)	(0.219)	(0.004)	(0.753)	(0.093)
_Ifacility_40 (d)	0.139	0.224	0.003	0.727	0.045
	(0.102)	(0.261)	(0.004)	(0.742)	(0.061)
Log pseudolikelihood		*-276.10*		*-239.94*	
R^2	*0.1406*				
pseudo R^2				*0.175*	
N	*917*	*917*		*826*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Appendix D: Alternative wealth measures

Table D.1: OLS, Poisson and binary logit regressions with under-five deaths as dependent variable and wealth quintiles

	OLS (Std. Err)	Poisson		Binary Logit	
		Coefficient (Std. Err)	Marginal effects (Std. Err)	Coefficient (Std. Err)	Marginal effects (Std. Err)
children_born	0.226***	0.692***	0.078***	1.017***	0.088***
	(0.037)	(0.046)	(0.008)	(0.117)	(0.011)
land_q2 (d)	0.076	0.023	0.003	-0.124	-0.010
	(0.064)	(0.283)	(0.032)	(0.324)	(0.026)
land_q3 (d)	0.065	-0.072	-0.008	0.049	0.004
	(0.088)	(0.382)	(0.041)	(0.471)	(0.042)
land_q4 (d)	0.150	0.343	0.043	0.875*	0.094
	(0.100)	(0.427)	(0.061)	(0.517)	(0.070)
land_q5 (d)	0.091	-0.001	-0.000	0.391	0.037
	(0.100)	(0.495)	(0.056)	(0.553)	(0.059)
noeduc_primary_secondary (d)	0.115*	0.714**	0.098**	0.680	0.068
	(0.063)	(0.284)	(0.048)	(0.456)	(0.051)
professional (d)	0.056	0.512*	0.064	0.586	0.056
	(0.049)	(0.293)	(0.041)	(0.448)	(0.046)
medium (d)	0.033	0.371	0.045	0.475	0.044
	(0.033)	(0.253)	(0.033)	(0.378)	(0.037)
a_control4 (d)	0.009	-0.003	-0.000	0.025	0.002
	(0.044)	(0.235)	(0.027)	(0.287)	(0.025)
del_home_post_other (d)	0.128**	0.501**	0.060*	0.792**	0.074*
	(0.054)	(0.254)	(0.033)	(0.363)	(0.038)
del_centre_other_fac (d)	-0.009	0.030	0.003	-0.048	-0.004
	(0.046)	(0.318)	(0.036)	(0.442)	(0.037)
att_doctor (d)	0.101	0.051	0.006	0.307	0.025
	(0.074)	(0.204)	(0.023)	(0.312)	(0.026)
breastf (d)	-0.084	-0.549	-0.080	-0.894	-0.107
	(0.076)	(0.400)	(0.072)	(0.565)	(0.085)
urban_uige (d)	-0.119	-0.123	-0.013	-0.457	-0.036
	(0.082)	(0.397)	(0.042)	(0.512)	(0.036)
luanda (d)	-0.136	-0.490	-0.055	-1.160**	-0.102*
	(0.090)	(0.483)	(0.056)	(0.568)	(0.053)
power (d)	0.035	0.349	0.037	0.390	0.032
	(0.042)	(0.339)	(0.033)	(0.461)	(0.035)
im_facility_outreach (d)	0.147**	0.536*	0.050**	0.693	0.048**
	(0.058)	(0.276)	(0.020)	(0.428)	(0.024)
antenatal_serv (d)	0.014	0.079	0.009	0.154	0.013
	(0.040)	(0.171)	(0.019)	(0.279)	(0.023)
del_serv (d)	-0.000	-0.540**	-0.060**	-0.528	-0.045
	(0.054)	(0.271)	(0.029)	(0.459)	(0.038)
malaria_serv (d)	0.015	0.167	0.018	0.331	0.026

	(0.036)	(0.191)	(0.019)	(0.302)	(0.022)
thermometer (d)	-0.067	-0.042	-0.005	-0.173	-0.016
	(0.041)	(0.204)	(0.024)	(0.304)	(0.030)
midwife (d)	0.005	0.390	0.047	0.401	0.036
	(0.058)	(0.324)	(0.043)	(0.461)	(0.045)
spes_doctor (d)	-0.017	-0.094	-0.010	-0.177	-0.015
	(0.043)	(0.295)	(0.032)	(0.432)	(0.035)
nquest_tot	-0.005	-0.030**	-0.003**	-0.023	-0.002
	(0.003)	(0.014)	(0.002)	(0.022)	(0.002)
nexam_tot	0.006	0.061**	0.007**	0.039	0.003
	(0.005)	(0.027)	(0.003)	(0.043)	(0.004)
diag_tot	-0.013	-0.206**	-0.023**	-0.125	-0.011
	(0.018)	(0.101)	(0.010)	(0.138)	(0.011)
F(26, 38)	*6.92*				
Prob > F	*0.0000*				
Log pseudolikelihood		*-364.48459*		*-291.73487*	
Wald chi2(26)		*2009.15*		*171.12*	
Prob > chi2		*0.0000*		*0.0000*	
R^2	*0.241*				
Pseudo R2				*0.2003*	
N	*879*	*879*		*879*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table D.2: OLS, Poisson and binary logit regression results with number of infant deaths as dependent variable and quintile dummies

	OLS (Std. Err)	Poisson		Binary Logit	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
children_born	0.109***	0.610***	0.042***	0.839***	0.049***
	(0.020)	(0.063)	(0.005)	(0.116)	(0.006)
land_q2 (d)	0.016	-0.065	-0.004	-0.139	-0.008
	(0.046)	(0.319)	(0.021)	(0.372)	(0.020)
land_q3 (d)	0.045	-0.136	-0.009	-0.056	-0.003
	(0.059)	(0.410)	(0.026)	(0.611)	(0.035)
land_q4 (d)	0.123	0.617	0.052	0.988	0.076
	(0.081)	(0.568)	(0.059)	(0.778)	(0.077)
land_q5 (d)	0.048	-0.238	-0.015	0.064	0.004
	(0.072)	(0.655)	(0.038)	(0.807)	(0.049)
noeduc_primary_secondary (d)	0.075	0.603	0.049	0.556	0.037
	(0.045)	(0.434)	(0.043)	(0.546)	(0.042)
professional (d)	0.034	0.427	0.032	0.464	0.029
	(0.035)	(0.423)	(0.035)	(0.537)	(0.037)
medium (d)	0.001	0.057	0.004	0.162	0.010
	(0.026)	(0.415)	(0.029)	(0.514)	(0.032)
a_control4 (d)	-0.049	-0.377	-0.029	-0.222	-0.014
	(0.040)	(0.318)	(0.026)	(0.351)	(0.023)
del_home_post_other (d)	0.100**	0.575**	0.042*	0.850**	0.054**
	(0.040)	(0.280)	(0.023)	(0.376)	(0.027)

del_centre_other_fac (d)	-0.003	-0.000	-0.000	-0.262	-0.014
	(0.029)	(0.346)	(0.024)	(0.407)	(0.021)
att_doctor (d)	0.103**	0.307	0.020	0.683*	0.036*
	(0.044)	(0.271)	(0.018)	(0.354)	(0.019)
breastf (d)	-0.110*	-0.988**	-0.111*	-1.259**	-0.122
	(0.064)	(0.403)	(0.066)	(0.537)	(0.074)
urban_uige (d)	-0.070	-0.364	-0.023	-0.646	-0.032
	(0.044)	(0.472)	(0.027)	(0.566)	(0.024)
luanda (d)	-0.084	-0.590	-0.041	-1.177	-0.071
	(0.062)	(0.682)	(0.050)	(0.850)	(0.054)
power (d)	0.042	0.539	0.034*	0.382	0.021
	(0.027)	(0.347)	(0.020)	(0.426)	(0.022)
im_facility_outreach (d)	0.043	0.133	0.009	-0.068	-0.004
	(0.034)	(0.372)	(0.023)	(0.443)	(0.027)
antenatal_serv (d)	-0.011	0.139	0.009	0.207	0.012
	(0.025)	(0.290)	(0.019)	(0.357)	(0.019)
del_serv (d)	-0.034	-0.739***	-0.051***	-0.713*	-0.041*
	(0.028)	(0.284)	(0.018)	(0.409)	(0.023)
malaria_serv (d)	0.016	0.183	0.012	0.383	0.020
	(0.024)	(0.194)	(0.012)	(0.305)	(0.014)
thermometer (d)	-0.043	0.069	0.005	-0.151	-0.009
	(0.028)	(0.194)	(0.013)	(0.254)	(0.017)
midwife (d)	0.041	0.684**	0.052**	0.745**	0.048*
	(0.030)	(0.291)	(0.025)	(0.367)	(0.026)
spes_doctor (d)	0.012	0.156	0.011	-0.025	-0.001
	(0.023)	(0.288)	(0.021)	(0.389)	(0.022)
nquest_tot	-0.001	-0.016	-0.001	-0.005	-0.000
	(0.002)	(0.022)	(0.002)	(0.026)	(0.002)
nexam_tot	0.001	0.035	0.002	0.015	0.001
	(0.003)	(0.037)	(0.003)	(0.047)	(0.003)
diag_tot	-0.011	-0.189*	-0.013*	-0.126	-0.007
	(0.012)	(0.107)	(0.007)	(0.133)	(0.008)
F(26, 38)	*5.97*				
Prob > F	*0.0000*				
Log pseudolikelihood		*-271.86*		*-237.71*	
Wald chi2(26)		*2477.31*		*357.56*	
Prob > chi2		*0.0000*		*0.0000*	
Pseudo R^2				*0.1730*	
R^2	*0.137*				
N	*893*	*893*		*893*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table D.3: OLS, Poisson and binary logit results for under-five deaths in Luanda with Luanda wealth index

	OLS Luanda (Std. Err)	Poisson Luanda		Binary Logit Luanda	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
children_born	0.110***	0.682***	0.046***	0.725***	0.042***
	(0.032)	(0.132)	(0.009)	(0.149)	(0.008)
luanda_pc1	0.006	0.042	0.003	0.123	0.007
	(0.013)	(0.075)	(0.005)	(0.097)	(0.006)
noeduc_primary_secondary (d)	0.075	0.808*	0.079	0.736	0.056
	(0.067)	(0.473)	(0.051)	(0.661)	(0.057)
professional (d)	0.100	0.865***	0.077**	0.923*	0.067
	(0.058)	(0.324)	(0.036)	(0.478)	(0.041)
medium (d)	0.039	0.499	0.035	0.686*	0.041*
	(0.031)	(0.316)	(0.022)	(0.390)	(0.023)
a_control4 (d)	-0.088	-0.435	-0.035	-0.736	-0.055
	(0.065)	(0.334)	(0.033)	(0.493)	(0.048)
del_home_post_other (d)	0.106*	1.093*	0.108*	1.582**	0.143**
	(0.052)	(0.566)	(0.062)	(0.629)	(0.067)
del_centre_other_fac (d)	0.061	0.593	0.046	0.811	0.055
	(0.066)	(0.719)	(0.056)	(0.842)	(0.059)
att_doctor (d)	-0.035	0.043	0.003	0.178	0.010
	(0.053)	(0.272)	(0.018)	(0.438)	(0.023)
breastf (d)	0.026	0.301	0.018	0.297	0.015
	(0.063)	(0.627)	(0.033)	(0.713)	(0.033)
urban_uige	0.000				
	(0.000)				
luanda	0.000				
	(0.000)				
power (d)	0.077	0.599	0.040	0.766	0.044
	(0.065)	(0.632)	(0.038)	(0.838)	(0.042)
im_facility_outreach	0.000				
	(0.000)				
antenatal_serv (d)	-0.088**	-0.500	-0.042	-0.941**	-0.076**
	(0.040)	(0.444)	(0.040)	(0.435)	(0.038)
del_serv (d)	-0.080	-0.564	-0.039	-0.581	-0.034
	(0.110)	(0.751)	(0.059)	(0.897)	(0.058)
malaria_serv (d)	-0.091	-0.428	-0.035	-1.161	-0.106
	(0.084)	(0.773)	(0.075)	(1.094)	(0.132)
thermometer	0.000				
	(0.000)				
midwife (d)	0.118	0.918	0.060	1.155	0.064
	(0.095)	(0.732)	(0.060)	(0.878)	(0.061)
spes_doctor (d)	-0.030	-0.445	-0.028	-0.767	-0.040
	(0.085)	(0.663)	(0.037)	(0.809)	(0.034)
nquest_tot	-0.005	-0.052	-0.004	-0.056	-0.003
	(0.004)	(0.039)	(0.003)	(0.053)	(0.003)
nexam_tot	0.011*	0.110*	0.007	0.104	0.006

	(0.006)	(0.056)	(0.005)	(0.077)	(0.005)
diag_tot	-0.035***	-0.275***	-0.019**	-0.242*	-0.014*
	(0.012)	(0.103)	(0.008)	(0.127)	(0.007)
Log pseudolikelihood		*-141.43182*		*-116.85649*	
R^2	*0.134*				
Pseudo R^2				*0.1886*	
N	*429*			*429*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table D.4: OLS, Poisson and Binary logit results for Luanda with infant deaths as dependent variable and Luanda wealth index

	OLS Luanda (Std. Err)	Poisson Luanda		Binary Logit Luanda	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
children_born	0.065***	0.654***	0.024***	0.714***	0.026***
	(0.017)	(0.119)	(0.005)	(0.136)	(0.005)
luanda_pc1	0.003	0.039	0.001	0.061	0.002
	(0.009)	(0.105)	(0.004)	(0.124)	(0.005)
noeduc_primary_secondary (d)	0.036	0.672	0.034	0.708	0.034
	(0.049)	(0.933)	(0.062)	(1.029)	(0.063)
professional (d)	0.066	0.742	0.034	0.527	0.022
	(0.050)	(0.512)	(0.027)	(0.692)	(0.031)
medium (d)	0.042*	0.655*	0.025*	0.727	0.027
	(0.024)	(0.395)	(0.014)	(0.487)	(0.017)
a_control4 (d)	-0.148**	-1.198***	-0.076*	-1.242**	-0.072
	(0.069)	(0.449)	(0.040)	(0.608)	(0.050)
del_home_post_other (d)	0.098**	1.399***	0.085**	1.620***	0.095**
	(0.044)	(0.443)	(0.039)	(0.538)	(0.047)
del_centre_other_fac (d)	0.033	0.590	0.025	0.515	0.021
	(0.034)	(0.503)	(0.023)	(0.536)	(0.024)
att_doctor (d)	0.029	0.483	0.015	0.564	0.017
	(0.056)	(0.365)	(0.009)	(0.463)	(0.012)
breastf (d)	-0.025	-0.258	-0.011	-0.371	-0.015
	(0.056)	(0.728)	(0.034)	(0.889)	(0.042)
urban_uige	0.000				
	(0.000)				
luanda	0.000				
	(0.000)				
power (d)	0.083**	1.018*	0.038**	0.942	0.033*
	(0.033)	(0.529)	(0.016)	(0.617)	(0.018)
im_facility_outreach	0.000				
	(0.000)				
antenatal_serv (d)	-0.084*	-0.892	-0.049	-0.981	-0.051
	(0.042)	(0.891)	(0.066)	(0.796)	(0.054)
del_serv (d)	-0.079	-0.951	-0.037	-0.870	-0.032
	(0.055)	(0.619)	(0.028)	(0.739)	(0.029)
malaria_serv (d)	-0.103**	-0.950	-0.056	-1.627*	-0.119

	(0.046)	(0.728)	(0.061)	(0.935)	(0.102)
thermometer	0.000				
	(0.000)				
midwife (d)	0.107**	1.378***	0.049*	1.418**	0.048*
	(0.043)	(0.533)	(0.026)	(0.651)	(0.027)
spes_doctor (d)	-0.001	-0.189	-0.007	-0.621	-0.020
	(0.050)	(0.625)	(0.021)	(0.825)	(0.024)
nquest_tot	-0.003	-0.053	-0.002	-0.060	-0.002
	(0.002)	(0.042)	(0.001)	(0.055)	(0.002)
nexam_tot	0.007**	0.115*	0.004*	0.110	0.004
	(0.003)	(0.062)	(0.002)	(0.081)	(0.003)
diag_tot	-0.027**	-0.347**	-0.013***	-0.297*	-0.011*
	(0.011)	(0.135)	(0.005)	(0.164)	(0.005)
Log pseudolikelihood		*-100.63*		*-91.24*	
R^2	*0.114*				
Pseudo R^2				*0.1878*	
N	*443*	*443*	*443*	*443*	*443*

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table D.5: OLS, Poisson and binary logit for Uíge sample with number of under-five deaths as dependent variable and Uíge wealth index

	OLS Uíge (Std. Err)	Poisson Uíge		Binary Logit Uíge	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
uige_pc1	0.018	0.041	0.005	0.074	0.007
	(0.012)	(0.057)	(0.007)	(0.066)	(0.006)
children_born	0.304***	0.659***	0.087***	1.231***	0.117***
	(0.054)	(0.038)	(0.009)	(0.181)	(0.013)
noeduc_primary_secondary (d)	0.076	0.381	0.053	0.190	0.018
	(0.137)	(0.558)	(0.081)	(0.970)	(0.095)
professional (d)	0.009	-0.018	-0.002	-0.131	-0.012
	(0.133)	(0.575)	(0.076)	(1.032)	(0.096)
medium (d)	-0.010	0.126	0.017	0.045	0.004
	(0.124)	(0.633)	(0.091)	(1.010)	(0.098)
a_control4 (d)	0.011	0.179	0.023	0.184	0.017
	(0.058)	(0.270)	(0.034)	(0.352)	(0.032)
del_home_post_other (d)	0.157	0.197	0.026	0.262	0.024
	(0.101)	(0.300)	(0.038)	(0.583)	(0.053)
del_centre_other_fac (d)	-0.101	-1.366**	-0.114***	-1.633**	-0.097***
	(0.060)	(0.644)	(0.030)	(0.708)	(0.024)
att_doctor (d)	0.220*	0.242	0.032	0.588	0.057
	(0.115)	(0.272)	(0.037)	(0.462)	(0.047)
breastf (d)	-0.223	-0.912*	-0.191	-1.695*	-0.278
	(0.169)	(0.522)	(0.158)	(1.018)	(0.229)
urban_uige (d)	-0.100	-0.051	-0.007	-0.382	-0.035
	(0.096)	(0.273)	(0.036)	(0.430)	(0.038)

luanda	0.000				
	(0.000)				
power (d)	0.179	1.850***	0.164***	2.320***	0.142***
	(0.116)	(0.579)	(0.035)	(0.782)	(0.037)
im_facility_outreach (d)	0.270**	1.471***	0.138***	2.292***	0.141***
	(0.100)	(0.501)	(0.035)	(0.725)	(0.033)
antenatal_serv (d)	0.007	-0.250	-0.034	-0.324	-0.031
	(0.046)	(0.228)	(0.031)	(0.406)	(0.040)
del_serv (d)	-0.040	-0.948**	-0.123**	-1.048**	-0.096**
	(0.061)	(0.407)	(0.049)	(0.448)	(0.041)
malaria_serv (d)	0.090**	0.576**	0.067**	0.943***	0.075***
	(0.036)	(0.273)	(0.027)	(0.324)	(0.024)
thermometer (d)	-0.047	0.155	0.019	-0.113	-0.011
	(0.045)	(0.248)	(0.029)	(0.295)	(0.029)
midwife (d)	-0.227***	-1.075**	-0.104***	-2.377***	-0.134***
	(0.062)	(0.544)	(0.039)	(0.777)	(0.030)
spes_doctor (d)	-0.087	0.333	0.050	0.160	0.016
	(0.065)	(0.361)	(0.063)	(0.541)	(0.056)
nquest_tot	-0.011*	-0.060**	-0.008**	-0.084**	-0.008**
	(0.005)	(0.028)	(0.004)	(0.036)	(0.003)
nexam_tot	0.025**	0.181***	0.024***	0.247***	0.023***
	(0.011)	(0.070)	(0.009)	(0.086)	(0.008)
diag_tot	-0.029	-0.561***	-0.074***	-0.463	-0.044
	(0.041)	(0.206)	(0.025)	(0.305)	(0.029)
Log pseudolikelihood		*-208.18*		*-154.41*	
R^2	*0.349*				
Pseudo R2				*0.2822*	
N	*450*	*450*		*450*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

Table D.6: OLS, Poisson and binary logit for Uíge with infant deaths as dependent variable and Uíge wealth index

	OLS Uíge (Std. Err)	Poisson Uíge		Binary Logit Uíge	
		Coefficients (Std. Err)	Marginal Effects (Std. Err)	Coefficients (Std. Err)	Marginal Effects (Std. Err)
uige_pc1	0.011	0.054	0.005	0.074	0.007
	(0.011)	(0.072)	(0.006)	(0.066)	(0.006)
children_born	0.134***	0.549***	0.046***	1.231***	0.117***
	(0.032)	(0.067)	(0.007)	(0.181)	(0.013)
noeduc_primary_secondary (d)	0.064	0.504	0.045	0.190	0.018
	(0.090)	(0.700)	(0.068)	(0.970)	(0.095)
professional (d)	0.006	0.114	0.010	-0.131	-0.012
	(0.087)	(0.724)	(0.062)	(1.032)	(0.096)
medium (d)	-0.070	-0.556	-0.039	0.045	0.004
	(0.083)	(0.832)	(0.049)	(1.010)	(0.098)
a_control4 (d)	-0.011	0.084	0.007	0.184	0.017
	(0.047)	(0.351)	(0.029)	(0.352)	(0.032)

del_home_post_other (d)	0.088	0.096	0.008	0.262	0.024
	(0.071)	(0.311)	(0.026)	(0.583)	(0.053)
del_centre_other_fac (d)	-0.085*	-1.687	-0.082***	-1.633**	-0.097***
	(0.044)	(1.055)	(0.025)	(0.708)	(0.024)
att_doctor (d)	0.139*	0.195	0.016	0.588	0.057
	(0.071)	(0.323)	(0.028)	(0.462)	(0.047)
breastf (d)	-0.237	-1.249**	-0.200	-1.695*	-0.278
	(0.144)	(0.527)	(0.139)	(1.018)	(0.229)
urban_uige (d)	-0.037	0.004	0.000	-0.382	-0.035
	(0.043)	(0.385)	(0.032)	(0.430)	(0.038)
luanda	0.000				
	(0.000)				
power (d)	0.140*	2.109***	0.114***	2.320***	0.142***
	(0.070)	(0.803)	(0.028)	(0.782)	(0.037)
im_facility_outreach (d)	0.047	0.585	0.042	2.292***	0.141***
	(0.060)	(0.716)	(0.043)	(0.725)	(0.033)
antenatal_serv (d)	0.010	0.072	0.006	-0.324	-0.031
	(0.030)	(0.348)	(0.029)	(0.406)	(0.040)
del_serv (d)	-0.085**	-1.287***	-0.106***	-1.048**	-0.096**
	(0.032)	(0.410)	(0.031)	(0.448)	(0.041)
malaria_serv (d)	0.074***	0.640**	0.047***	0.943***	0.075***
	(0.021)	(0.271)	(0.016)	(0.324)	(0.024)
thermometer (d)	-0.022	0.187	0.015	-0.113	-0.011
	(0.030)	(0.262)	(0.019)	(0.295)	(0.029)
midwife (d)	-0.047	-0.356	-0.027	-2.377***	-0.134***
	(0.041)	(0.670)	(0.044)	(0.777)	(0.030)
spes_doctor (d)	0.023	0.665	0.073	0.160	0.016
	(0.052)	(0.528)	(0.075)	(0.541)	(0.056)
nquest_tot	-0.002	-0.012	-0.001	-0.084**	-0.008**
	(0.003)	(0.040)	(0.003)	(0.036)	(0.003)
nexam_tot	0.005	0.091	0.008	0.247***	0.023***
	(0.007)	(0.088)	(0.007)	(0.086)	(0.008)
diag_tot	-0.021	-0.611**	-0.051***	-0.463	-0.044
	(0.021)	(0.238)	(0.018)	(0.305)	(0.029)
Log pseudolikelihood		*-160.32*		*-135.80*	
R^2	*0.180*				
Pseudo R2				*0.2059*	
N	*450*	*450*		*450*	

(d): dummy variables. *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level