International Journal of Population Studies

RESEARCH ARTICLE

Multilevel analysis of infant mortality and its risk factors in South Africa

Samuel Abera Zewdie[1*] **and Vissého Adjiwanou**[2]

[1] Ethiopian Development Research Institute, Addis Ababa, Ethiopia

[2] Centre for Actuarial Research, University of Cape Town

Abstract: The study analyzed infant mortality and its risk factors in South Africa. It aimed to examine infant mortality in the country by taking into account the hierarchical nature of the problem and investigate the with-in country variation in modeling. In addition to the usual individual level risk factors of infant mortality, living standard, mother's education, and income inequality were defined at municipal level, while HIV prevalence was fixed at province level. A multilevel logistic regression model was then fitted with Bayesian MCMC parameter estimation procedure using the 2011 South African census data. Most of the demographic and socioeconomic variables identified at individual level were found significant. More remarkably, the result indicated that communities with better living standard and women's education were associated with lower infant mortality rates, while higher income inequality and HIV prevalence in the communities were associated higher levels of infant mortality. The changes in infants' odds of death were estimated to be 26%, -21%, 13% and 8% respectively for HIV, women's education, income inequality and level of the living standard. In addition, unobservable municipal and province level random effects significantly affected the level of infant mortality rates.

Keywords: Infant mortality; multilevel; poverty; inequality

ARTICLE INFO
Received: June 21, 2017
Accepted: August 25, 2017
Published Online: September 3, 2017

***CORRESPONDING AUTHOR**
Samuel Abera Zewdie, Ethiopian Development Research Institute, Addis Ababa, Ethiopia; samuel_abera@yahoo.com

CITATION
Zewdie SA and Adjiwanou V (2017). Multilevel analysis of infant mortality and its risk factors in South Africa. *International Journal of Population Studies*, 3(2): 43-56.
doi: 10.18063/ijps.v3.i2.330

1. Introduction

Infant mortality rate is an important indicator of health and development. Biologically, infants have much weaker immune systems than adults and are therefore far more vulnerable to environmental or social complications (Caldwell, 1996). In addition, they are unable to care for themselves and are hence completely dependent on others. As a result, children are generally the group first and most strongly affected by poor living standards. Likewise, advances in health or social conditions are often first observed in improvements in infant or child mortality (Omran, 1971). Studies on infant mortality have accumulated a huge list of determinants or associates, including individual- and community-level factors such as maternal age, race, income, sanitation, water source, electricity, urban/rural residence, region of residence, household composition, occupation, female education, access to health care, and so forth (Caldwell, 1979; Hobcraft, McDonald and Rutstein, 1985; Kembo and Ginneken, 2009; Omariba, Beaujot and Rajulton, 2007; Victora, Wagstaff, Schellenberg *et al*., 2003; Wang, 2003).

Within-country variation in child mortality has been well documented, with rates often varying substantially across different regions and social groups (Moser, Leon and Gwatkin, 2005; Mosley and Chen, 1984). It is said that although global or national level results are vital for assisting policy makers to better prepare for the emerging health needs of populations, they constitute a less effective guide for refocusing health priorities because efforts to reduce health disparities would be more successful if they are based on evidences from lower administrative units (Heuveline, Guillot and Gwatkin, 2002). This research, therefore, studies infant mortality at national, provincial and municipality levels to highlight concentration at lower levels of geography that in turn underscores the ineffectiveness of national level indicators for monitoring progress in health achievement.

The objective of this research is to analyze infant mortality using hierarchical model and to identify important risk factors for infant mortality in South Africa. The study uses the 2011 South African census data and all the risk factors are defined at three levels: Individual, municipal and province level. A three-level logistic regression model is fitted using data on children born twelve months before the census date where child, municipality and province are the first, second, and third levels.

Factors affecting infant mortality are investigated by fitting multilevel logistic regression models in order to quantify the impacts of socioeconomic factors, including poverty and inequality, on infant mortality. The hypothesis is that there are significant spatial variations of child mortality, which are associated with socioeconomic differentials in the country, and hence multilevel modelling helps to measure the impacts of the risk factor at different administrative levels.

Mortality of children has been explained by different theories such as the social and economic explanation, the public health explanation, and the Mosley and Chen analytical framework. Mosley *et al*., (1984) identified a set of proximate determinants that directly have an impact on the morbidity and mortality of children. The factors are then grouped into 5 sets: Maternal factors (age, parity and birth interval), environmental contamination (air, food/water/fingers, skin/soil/inanimate objects, insect vectors), nutrient deficiency (calories, protein, micronutrient, vitamins and minerals), injury (accidental, intentional); the last set is personal illness control (personal preventive measures, medical treatment). It is within this framework that many studies on child mortality and its correlates have been carried out. This study also follows this framework but based on the following classification of the determinants: New-born demographics (Boco, 2010; Hill and Upchurch, 1995; Kembo and Ginneken, 2009; Mustafa and Odimegwu, 2008), Maternal factors (Boco, 2010; Hobcraft, McDonald *et al*., 1985; Kabir, Islam, Ahmed *et al*., 2001; Kembo and Ginneken, 2009; Omariba, Beaujot *et al*., 2007), socioeconomic factors (Bawah and Zuberi, 2005; Cleland, 1990; Hobcraft, 1993; Mustafa and Odimegwu, 2008; Sastry, 1996; Wagstaff, 2000), environmental factors (Bartlett, 2005; Kabir, Islam, Ahmed *et al*., 2001; Kazembe, Clarke and Kandala, 2012; Kembo and Ginneken, 2009) and HIV/AIDS (Dorrington, Johnson, Bradshaw *et al*., 2006; Ng'weshemi, Urassa, Usingo *et al*., 2003; Wang, Liddell, Coates *et al*., 2014; Zaba, Marston and Floyd, 2003). These factors are defined at individual level, which could be child, mother or household, municipal level or province level depending on the nature of the variable as well as the availability of data.

2. Data and Methods

2.1 Data

The study used data from the 10%-unit record of the 2011 de facto population and housing census of South Africa (StatsSA 2014). Children born within 12 months before the census date are considered for this research. The sample data, among other things, consisted of data on children's demographics,general health functioning, income, educational attainment, employment status, fertility, household characteristics and mortality variables. After removing missing, unknown and inconsistent cases, there are 86 877 (un-weighted) children with valid survival status for analysis. These children can be viewed as they are nested in a structure under the 234 municipalities and 9 provinces of South Africa. The mortality status of these children was the outcome measure of the study, and it was assumed that the rate of under-reporting of births in the past 12 months is the same as that of under-reporting of deaths in the past 12 months so that the effect on mortality estimates is negligible.

HIV prevalence rates were taken from the 2012 South African National HIV Prevalence, Incidence and Behaviour Survey conducted by the Human Science Research Council (HSRC, 2014). More details of the survey can be found elsewhere (HSRC, 2014)

2.2 Risk Factors Considered in the Study

The individual level risk factors considered are: sex, age and birth order of the child; age, racial group, marital status and employment status of the mother as well as the living standard of the household where the infant resides. Living standard was computed by constructing an index from different variables which are supposed to be related with the living

Table 1. Summary of variables used for LS index construction

Variable	Category (code)	Mean	SD	Factor loading	Coeff-icient
Dwelling Type	House (1), Other (0)	0.66	0.48	0.384	0.066
Room per person	Greater or equal to 1 (1), less than 1 (0)	0.69	0.46	0.257	0.030
Roof made of	Tiles (3), Concrete/Block (2) Other (1)	1.98	0.66	0.431	0.058
Wall made of	Brick (3), Concrete/Block (2) Other (1)	1.93	0.60	0.388	0.067
Energy used for lighting	Electricity (1), Other (0)	0.85	0.36	0.631	0.128
Energy used for cooking	Electricity/Gas (1), Other (0)	0.77	0.42	0.674	0.123
Piped water on premises	Available (1), Not available (0)	0.73	0.44	0.667	0.106
Flush Toilet	Available (1), Not available (0)	0.60	0.49	0.717	0.179
Television	Available (1), Not available (0)	0.76	0.43	0.595	0.096
Satellite Dish	Available (1), Not available (0)	0.26	0.44	0.554	0.092
Refrigerator	Available (1), Not available (0)	0.70	0.46	0.641	0.118
Washing Machine	Available (1), Not available (0)	0.32	0.47	0.645	0.120
Vacuum Cleaner	Available (1), Not available (0)	0.17	0.38	0.536	0.097
Computer	Available (1), Not available (0)	0.22	0.41	0.555	0.105
Internet access	Available (1), Not available (0)	0.36	0.48	0.436	0.057
Rubbish collected by local authority	Yes (1), No (0)	0.62	0.49	0.625	0.104

Source: Stats SA census 2011

standard (LS) of people. The LS index was constructed based on different indicators of wellbeing from the 2011 census data, specifically on those variables which measure how good the environment is for the infant to live in. Factor analysis (FA) was chosen for constructing the index (Hair, Black, Babin *et al*., 2010). FA mainly involves extracting the factor(s) by partitioning the total variance in each of the variables into variances which are shared and have unique variance. The detail theory and application of FA can be found in any standard multivariate text like Hair, Black, Babin *et al*., (2010). The descriptions of the variables used for constructing the index including some summary statistics of the variables are shown in **Table 1**. As a measure of internal consistency of the scale, Cronbach Alpha–a known measure of reliability – is computed giving a scale reliability coefficient of 0.8597.

The first factor was found to be enough to explain about 80% of the variance in the dataset and hence it was used to construct the index. The factor loadings and the coefficients of each variable used to generate the index are also given on **Table 1**. For ease of understanding, the constructed index was divided into 5 quintiles which can be used as ranking the level of living standard to households. A household lying in the first quintile was categorised as to have the poorest living standard, whereas a household lying in the fifth quintile was categorised to have the best living standard.

Community level variables used in this study were: Poverty and inequality levels of the municipality; education level of the municipality and HIV-prevalence rate of the provinces. The LS index was used to determine whether the municipality was poor or not, whereas Gini-index (GI) was used to determine the level of income inequality. GI is expected to be positively correlated with infant mortality as greater inequality in income within communities reflects unequal access to healthcare, nutrition and other services which is likely to reduce the health of the poor (Waldmann, 1992; Rodgers, 2002). GI is a number between 0 and 1, where 0 corresponds with perfect equality and 1 corresponds with perfect inequality. It is computed from a Lorenz curve (LC) which is literally a plot of the cumulative percentage of population versus the cumulative percentage of wealth/income.

All the independent variables were defined as categorical and hence, the odds ratio of death given in the last column of the tables measures the odds of the category compared to the reference group. Note that the independent variables are listed according to their level such that the variables on proportion of poor mean mothers' years of education and income inequality as measured by Gini index, are identified at municipal level, while HIV prevalence rate is at province level. All these four variables were classified into two categories: lower and higher magnitude of the respective measures. The lower and higher values dictate that the respective quantity in the area is less than and greater than the national estimate. For instance, about 49% of the children live in municipalities where the level of income poverty is higher than the national poverty head count ratio of 41%. Note also that among the child level variables, age of the child is an indicator variable showing whether the child has age of less than one month (neonatal) or not.

2.3 Multilevel Models

Multilevel analysis is a suitable approach to take into account community level contexts at different levels, like at municipal and province levels, as well as individual subjects. A three-level random intercept logistic regression model was considered where the first level is children born 12 months before the census, whereas the municipalities and provinces in which the children live are the second and third levels respectively. Let π_{ijk} be the probability that child *i* living in municipality *j* and province *k* died before reaching age one. Then, the three-level random intercept logistic regression model in question with the predicator variables described above can, therefore, be expressed as

$$\ln[\pi_{ijk}/(1-\pi_{ijk})] = \beta_{0jk} + \sum_{l=1}^{16} \beta_l X_{lijk} \qquad \text{[Level 1]}$$

$$\beta_{0jk} = \beta_{00k} + \sum_{l=17}^{22} \beta_l X_{ljk} + u_{0jk} \qquad \text{[Level 2]}$$

$$\beta_{00k} = \beta_{000} + \sum_{l=23}^{24} \beta_l X_{lk} + v_{00k} \qquad \text{[Level 3]}$$

where $v_{00k} \sim N(0, \sigma^2_{v0})$, $u_{0jk} \sim N(0, \sigma^2_{u0})$, and the notations of the independent variables are as given in **Table 2**. The coefficients $\beta_1, \beta_2, \ldots, \beta_{24}$, called fixed effects, measure the impact of the corresponding predictor variable on the log of odds of death, whereas β_{0jk}, the random intercept, measures the combination of municipal and provincial level effects as defined in the second and third level of the model. Unlike ordinary logistic regression, there are two types of residual terms, u_{0jk} and v_{00k}, defined at level 2 and level 3 respectively and assumed to be normally distributed with mean zero and constant variance. Bayesian approach with Markov Chain Monte Carlo (MCMC) was implemented to the parameters of the above model. Further information regarding methods of parameter estimation is given in the appendix.

3. Results

3.1 Descriptive Statistics of Variables

The descriptive statistics of all individual, municipal and province level variables chosen for the analysis including the bivariate odds of infant death are shown in **Table 2**. It shows that some of the variables, such as race and education of the mother, living standard, birth order and HIV prevalence contribute to greater odds of death of the infant than others.

3.2 Multilevel Model Outputs

The final results of the regression are shown in **Table 3**. All parameter estimates were measured on the log-odds (logit) scale. In order to make more specific and meaningful inference about the effect of the risk factors on the infant mortality, the odds ratios (ORs) were given corresponding to each coefficient estimate in the same table. Note that among the independent variables, proportion of poor people, income inequality and mean years of mother's education were measured at municipality level, whereas HIV prevalence rate was computed at province level. All these four variables were dichotomised as higher and lower values of the respective quantities.

All coefficients of the living standard dummy variables are negative and their 95% confidence intervals exclude zero. Compared to infants who were in the first quintile of living standard, those who were in the second to fifth quintiles had 6%, 7%, 14 % and 24% lower odds to die, respectively. Likewise, the income poverty has a positive and significant coefficient, entailing that children living in a household whose members earned a per capita income of less than the South African poverty line were more likely to die than those who were above the poverty line.

Most of the municipal level indicator variables are significant, which implies that the level of poverty, women education and inequality of the municipality affected the survival status of infants. An infant was more likely to die in a highly poor and more unequal municipality compared to municipalities where the levels of poverty and inequality were lower after controlling for other risk factors. Considering the magnitude of the impact, it seems that the income inequality mattered more for infant mortality risk than the size of poverty in that more unequal municipalities were associated with 13% higher odds ratio of infant death than less unequal municipalities, whereas municipalities where poverty was high were

Table 2. Summary statistics of the variables in the regression model

Label	Variable		Mean	Std. Dev.	Odds Ratios of Dying
Y	Child died		0.0187	0.1353	0.0190
	Individual level				
X1	Sex of child		0.5066	0.5000	1.2108
X2	Age less than 1 month		0.1139	0.3177	0.8152
	Mother's age at birth	<20 yrs	0.1852	0.3885	1.0551
X3		20–34 years	0.1524	0.3594	0.8552
X4		>34 years	0.4000	0.4899	1.2212
	Birth order	1	0.2984	0.4575	0.8636
X5		2	0.1623	0.3687	0.8342
X6		3	0.1393	0.3463	1.0893
X7		4+	0.1312	0.3377	1.5342
	Mother's educ	No/primary educ	0.7722	0.4194	1.5116
X8		Secondary educ	0.0966	0.2954	0.9552
X9		Higher educ	0.5250	0.4994	0.5006
X10		Mother never married	0.2028	0.4021	1.1803
X11		Mother works	0.8592	0.3478	0.9151
X12		Mother is Black African	0.2317	0.4219	2.0703
	Living Standard Quintiles	Q1	0.2040	0.4030	1.4821
X13		Q2	0.1900	0.3923	1.1295
X14		Q3	0.1320	0.3385	1.0532
X15		Q4	0.3371	0.4727	0.7591
X16		Q5	0.3865	0.4869	0.4331
	Municipal level variables				
	Municipal poverty level	Low	0.3371	0.4727	1.2865
X17		Medium	0.3865	0.4869	1.0610
X18		High	0.2764	0.4472	0.6718
	Municipal Inequality	Low	0.3335	0.4715	0.8963
X19		Medium	0.3421	0.4744	1.0924
X20		High	0.3244	0.4682	1.0173
	Municipal women educ	Low	0.3362	0.4724	1.4668
X21		Medium	0.3507	0.4772	0.7895
X22		High	0.3131	0.4638	0.8325
	Province level variables				
	HIV prevalence rate	Low	0.3710	0.4831	0.7494
X23		Medium	0.3047	0.4603	0.9941
X24		High	0.3243	0.4681	1.3334

Note: Odds ratios were bivariate.

Table 3. Outputs from the multilevel logistic regression model

Variables		Coefficient	p	Odds Ratios
Fixed Effect parameters				
Cons		-2.2274	0.000	0.11
Sex of child (Male)		0.0790	0.000	1.08
Age of child (months)		-0.0774	0.010	0.93
Birth order	1 (Reference)			
	2	-0.0019	0.4760	1.00
	3	0.0736	0.0150	1.08
	4+	0.1440	0.0000	1.15
Mother's age at birth	20-34 years (Reference)			
	< 20 years	0.0297	0.125	1.03
	> 34 years	0.0225	0.261	1.02
Mother's years of education		-0.0708	0.0080	0.93
		-0.1922	0.0000	0.83
Mother never married		0.0518	0.0130	1.05
Mother works		0.0891	0.0010	1.09
Mother is Black African		0.1252	0.0020	1.13
Living Standard Quintile	Q1(Reference group)			
	Q2	-0.0572	0.0270	0.94
	Q3	-0.0716	0.0130	0.93
	Q4	-0.1471	0.0001	0.86
	Q5	-0.2808	0.0001	0.76
Municipal-level variables				
Level of poverty	Low Poverty (Reference group)			
	Medium	0.0466	0.1160	1.05
	High	0.0264	0.0450	1.03
Inequality	Low Inequality (Reference group)			
	Medium	0.0737	0.0210	1.08
	High	0.0530	0.0690	1.05
Proportion of educated women	Low Proportion (Reference group)			
	Medium	-0.1023	0.0010	0.90
	High	-0.1167	0.0200	0.89
Province level variable				
HIV prevalence rate	Low Prevalence (Reference group)			
	Medium	0.1352	0.0420	1.14
	High	0.1152	0.0420	1.12
Random effect parameters				
Province effect (level 3)		0.0048		
Municipality effect (level 2)		0.0085		

only associated with 3.5% higher odds of infant death than those municipalities where poverty was low. Similarly, infants living in municipalities where the average years of education of women was higher had a better likelihood of survival irrespective of the education level of their own mother. They were about 16% less likely to die than infants living in areas where there was less education of mothers.

Considering the effect of HIV on mortality of infants, it is apparent that its coefficient is positive and significant. Infants in provinces with high HIV prevalence were 12% more likely to die than infants in other provinces after controlling other factors in the model.

The sex effect is also positive and significant at p<0.001 level. More specifically, boys were 8% more likely to die during their first year of life than girls, all else being equal. Similarly, age of the child significantly affected the mortality of the child in that children who survived the first month after birth were less likely to die than those who were less than one month old by about 7%.

In terms of mother characteristics, the results in **Table 3** show that mothers' age at the birth of the child was not statistically significant. All the other independent variables associated with mothers were significant, which include years of education, employment status, birth order, marital status and population group.

In accordance with the literature, the more education a mother gets, the less likely the infant dies. Infant whose mothers completed secondary and higher education were 7% and 17% respectively less likely to die than those children whose mothers had no education or only had primary education. In addition, infants of higher birth orders, infants born to mother who were never married or single mothers, and infant born to mothers from African population group had a greater chance of dying than their counterparts of lower birth order, born to mothers who were not single and born to mothers from non-African population group. Specifically, an infant of the fourth or higher birth orders was associated with 15% higher odds of dying than first born infants, whereas an infant of the third birth order was associated with 8% of higher odds of dying. No difference was found in infants mortality between the second birth order and the first birth order. Similarly, children born to Black women or born to not-married women were associated with 13% and 5% higher odds of dying before age 1 in comparison with children born to non-Black or born to married women.

It is also worthy to see the distribution of the predicted log-odds of infant deaths in the country to see how the multilevel model performs in estimating infant mortality. The predicted log-odds of death of the municipalities and provinces are shown in **Figure 1** and **2** below. It can be observed that municipalities such as City of Cape Town, Liansburg, Mossel Bay and Nama Khoi had the lowest level of infant odds of death, while Disobotla, Mamusa, Msukaligwa and Ingwe scored

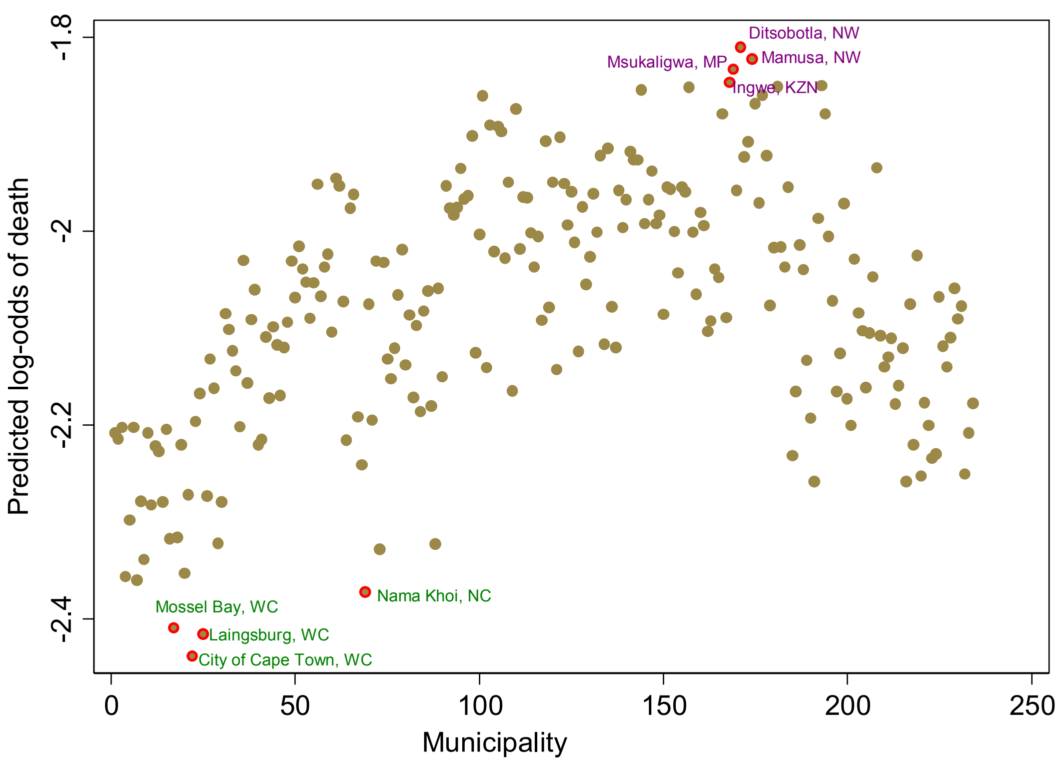


Figure 1. Predicated log-odds of infant death across the municipalities of South Africa

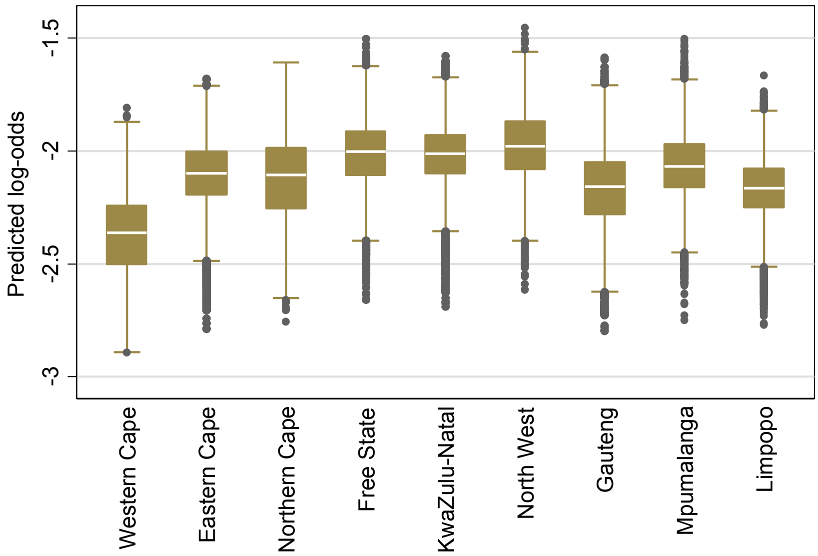


Figure 2. Predicated log-odds of infant death across the provinces South Africa

the highest mortality. Among the province, Western Cape showed the lowest odds of death followed by Gauteng and Limpopo.

The random effects terms included in the model were also significant. Hence, there was a unique effect for each province (level 3) and for each municipality (level 2) in addition to the fixed effects discussed above. The reported mean values in **Table 3** are the average random effect estimates of all provinces and all municipalities. The addition of the municipality specific effects as well as province specific effects made the model more accurate than the fixed only model. Specifically, controlling for the municipal and province level variations had on average increased the odds of death of infants by 1% and 0.5% respectively.

4. Discussion

In order to study the determinants or associates of infant mortality in South Africa, a multilevel logit model was fitted using data on the survival status of children born twelve months before the census and its risk factors which include several demographic, socioeconomic and environmental variables. The hierarchical nature of the data was taken into account in the process by considering provinces, municipalities and children as third, second and first levels respectively– hence, fitting a three-level logistic regression model employing the Bayesian MCMC procedure for estimating the parameters of the model. The results obtained from the regression model suggest that child mortality in South Africa is jointly determined by the observed individual demographic and socioeconomic characteristics of the child and the mother, and by municipal and province level covariates, as well as unobserved municipal and province level effects.

Among the demographic factors incorporated in the model, only age of mother at birth is not significant. Specifically, the odds of death for infants of four or more birth order is 15% higher than first born infants, whereas third born infants are 8% more likely to die. It shows that mortality risk is increasing with the addition of successive births, which implies that last-born (fourth and higher order) infants are at significantly greater risk of dying compared with earlier-born children. These findings are highly consistent with the Resource Depletion Hypothesis (Ashiabi and O'Neal, 2007), according to which, with the addition of successive children in the family, parental resources – both material and emotional – become diminished and consequently last-borns suffer the most. This is particularly true in the post-neonatal period since resource depletion is so heavily influenced by socioeconomic and environmental conditions that come into play in the later phases of infancy (Kembo and Ginneken, 2009). Similarly, male infants are 8% more likely to die than female, where as each month of life of infants contributes to 7% reduction on the odds of death of the infant. Genetic components result in newborn females being biologically advantaged when it comes to surviving their first birthday. Biologically, males, have lower chances of surviving infancy in comparison to female babies (Boko, 2010; Hill and Upchurch, 1995).

All the socioeconomic variables considered at level-one of the model are significant at $p<0.05$ levels and are in agreement with the results from other researches (Mosley and Chen, 1984; Hobcraft, 1993; Sastry, 1996; Kabir, Islam *et al*., 2001) as well as our child mortality estimation results presented above. For instance, children of black African mothers have a higher risk of death as compared to other population groups, while those who are from better educated mother have much lower risk of death. As expected, the living standard (LS) index, which is used as a measure of poverty, is highly significant too in that the higher the LS index of the household where the child lives the less likely the risk of dying. Note that because of multicollinearity issue the household income poverty indicator variable is not included in the model–the LS index and income poverty have a strong correlation coefficient of 0.85. Hence, in relation to poverty the result could be interpreted as, for example, children in the least poor and the second least poor household have more than 24% and 14% chance of dying respectively as compared to those living in the poorest households. Children living in municipalities where there is higher level of income poverty and inequality have greater likelihoods of death. Similarly, average years of schooling of women at municipal-level significantly affects child survival positively, whereas higher women HIV prevalence rate of provinces is highly related with higher risk of death of children as one expects. Poverty, inequality and women's education of the municipalities affected the odds of infant death by 8% 13% and -21% respectively, while the impact of province level HIV/AIDS is estimated to be 26. The results of the regression model also indicate small but statistically significant residuals which can convey province-level and municipal-level random effects on the risk of dying, even after controlling for a range of child-level, municipal-level, and province-level variables.

The study examined a comprehensive array of multilevel risk factors for infant mortality in South Africa by giving more attention to provincial and municipal variation. The modelling and estimation strategies utilized are appropriate and supported by large number of observation from the latest available census data. The use of multilevel analysis helped to understand the impact of some community level infant mortality risk factors of infant. Developmental indicators are likely to still vary significantly across the municipalities and also their effects on infant mortality. Hence, for further improvements, there is a need to focus on municipal level planning as well instead of only at the national or province level. Taking into account the findings under the present study, for a data involving hierarchical structure, there is a need to emphasize the use of the possible highest levels in hierarchical models. To further emphasize, such optimal considerations may provide additional important clues to policy planners leading to optimal use of available resources regarding public health programs.

Although the study achieved its objectives, there were some unavoidable limitations. First, our analyses were primarily based on the 2011 South African census data. Therefore, the significance and reliability of the results depends on the quality of the census data which includes the quality of enumeration and data processing. Any defect in the census data might seriously impact the results in the research. Second, the cross-sectional nature of this study design cannot us determine the causal relationship between independent variables and the outcome variable. The use of longitudinal data would have been much better. Third, the assumption that under reports of birth and deaths in the census data are the same. Fourth, the unavailability of municipal level HIV prevalence rates, and hence the assumption that these prevalence rates are the same as the rates at the respective provinces. Ignoring the HIV prevalence variation within provinces might especially impact the results of the regression model to some extent.

5. Conclusions

The main objective of the research was to investigate infant mortality risk factors with special emphasis on the impact of poverty and inequality. The results from the multilevel logistic regression model suggest that most of the demographic and socioeconomic factors as well as the province and municipal level random effects are significant. The significant predictors at individual-level include birth order and sex of the child, education, employment status, race, and marital status of the mother, and living standard of the family. These factors can bring from -24% to 15% change on the odds of death of infants. The changes in odds of infant death due to municipal and province level effects are estimated to be 12%, -11%, 7%, and 3% respectively for the HIV prevalence women's education, inequality and LS poverty. This implies that communities with better living standard and women education are associated with lower infant mortality rates, while higher income inequality and HIV prevalence in the communities depict higher levels of infant mortality. In addition, unobservable municipal and province level random effects significantly affect the level of infant mortality rates. The study helps to analyze infant mortality in the country by taking the hierarchical nature of the theme into account and to investigate the with-in country variation of infant mortality.

Authors' Contribution

SA Zewdie designed the study, prepared the data, performed the analysis, drafted and revised the manuscript, and interpreted the results. V Adjiwanou supervised the study design and analysis, revised the manuscript and interpreted the results.

Conflict of Interest

This paper is part of an MPhil degree thesis at the university of Cape Town in 2014. The authors declare that there is no conflict of interest.

Ethics

Consent was obtained from all participating persons in the study.

Appendix

Methods of Parameter Estimation

There are two commonly used estimation methods for multilevel logistic regression models: quasi-likelihood (QL) approach and Bayesian approach with Markov Chain Monte Carlo (MCMC) methods (Goldstein, 2011). In QL approach, the non-linear logistic regression equation is estimated first using a Taylor series expansion which approximates a nonlinear function by an infinite series of terms (Breslow *et al*., 1993). If the Taylor series is expanded about the fixed and the random parameters, then the estimation is known as penalised quasi-likelihood (PQL) (Breslow *et al*.). Once the quasi-likelihood has been formed, unbiased estimates of the random parameters can be found by applying either iterative generalised least squares (IGLS) or restricted generalised least square (RGLS) which are estimation procedures in the case of continuous response variables (Goldstein, 2011). On the other hand, the Bayesian approach using MCMC estimation methods can be used by first specifying starting values prior distributions for each of the model parameters and then sequentially sampling subsets of parameters from their conditional posterior distributions using Markov chain. A discussion and technical details of MCMC estimation methods for multilevel models can be found from in Browne (2003) and Goldstein (2011)The MCMC procedure followed by MLwiN–software dedicated for multilevel modelling and used by this research–by default assigns flat prior distributions to the parameters of the model. That is, for fixed terms $p(\beta) \propto 1$ and for random terms, $p(1/\sigma^2) \sim Gamma\ (\varepsilon\varepsilon, \varepsilon\varepsilon)$ where $\varepsilon\varepsilon$ is a very small number. After assigning initial values, usually estimates from QL methods, the MCMC procedure in MLwiN then performs the simulations in two phases. In the initial burn-in period it runs until the chain converges to its stationary distribution; and in the next stage (monitoring period) it runs so that the means and standard errors of the parameters are estimated. The 2.5th and 97.5th quantiles of the chains provide Bayesian 95% credible intervals in order to make inferences concerning the estimated parameters, serving the same purpose as 95% confidence intervals. For fitting the aforementioned model, the number of iterations run is 1000 in the burn-in period and 10 000 for the monitoring period.

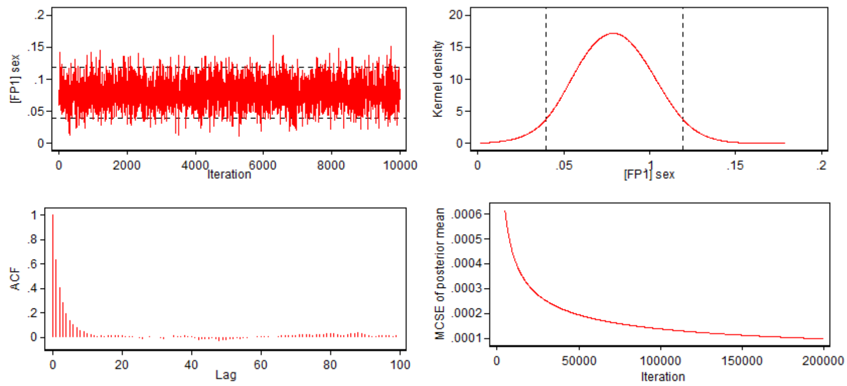
After running the model, residuals at municipal and province level (estimates of u_{0jk} and v_{00k}) are calculated so that the underlying assumptions, such as normality and constant variance of residuals, be investigated with the help diagnostic plots. Furthermore, as part of model diagnostics, the trace of the chains, autocorrelations (AC) and partial autocorrelations (PAC) functions at iteration t and t-k having accounted for iterations $t-1, \ldots, t-(k-1)$, and Monte Carlo standard errors (MCSE) are investigated for each of the posterior distributions of the parameter in the model. For the model to be good it is expected that the traces be not skewed, the AC and PAC functions be less correlated and the MCSE be close to zero. Increasing the number of iterations produces better results in all these dimensions. A comprehensive detail of parameter estimation and model diagnostics using MCMC simulations methods can be found from MLwiN manual (Rasbash, Charlton, Browne et al., 2012).

Model Diagnostics

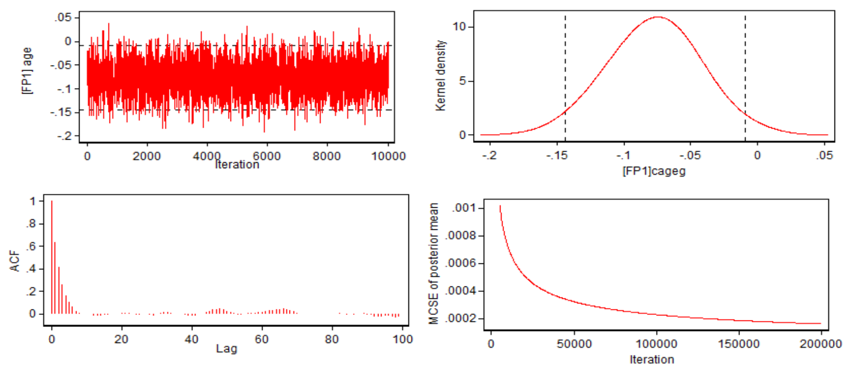
A three-level logistic regression model is fitted on the survival status of children born twelve months before the census. The parameters of the model are estimated using the Bayesian MCMC procedure by running the simulation for 1000 burn-in and 10000 monitoring period. After fitting the model, the reasonableness of the parameter estimates are assessed by looking at some diagnostics plots including the autocorrelation plots of successive iterations of the chains and Monte Carlo standard error plots for checking convergence of the posterior distributions. These are done for each of the fixed and random terms in the model. Some of these plots are given in the annex from **Figure (1)** to **Figure (f)**. The assumptions of normality of the residual terms at municipality and province level are approximately maintained.

MCMC Diagnostic Plots

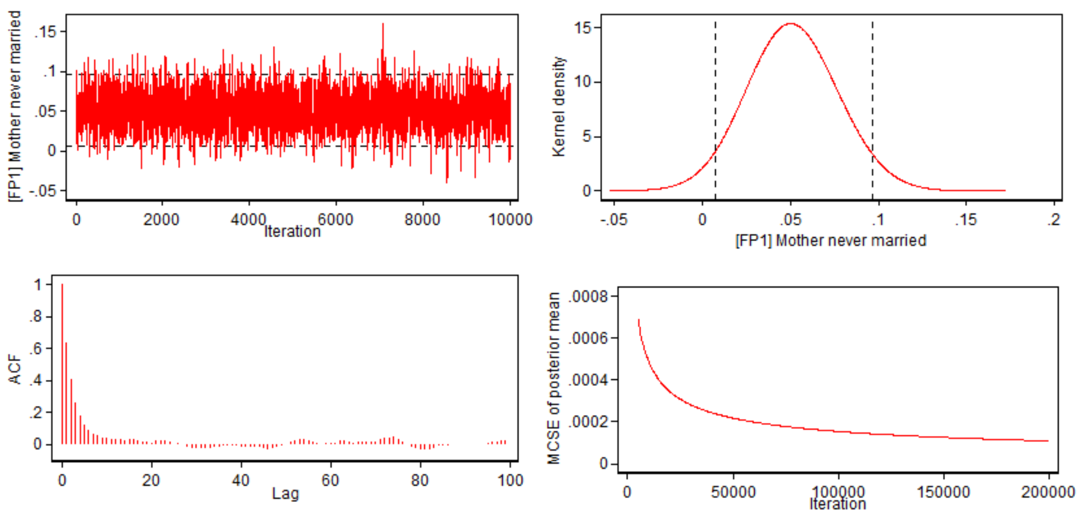
a)Fixed effect - Child sex



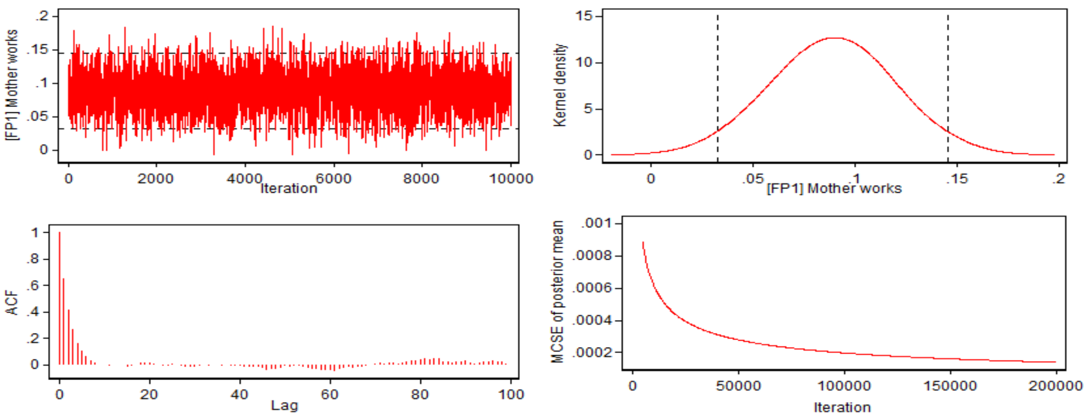
b)Fixed effect - Child age



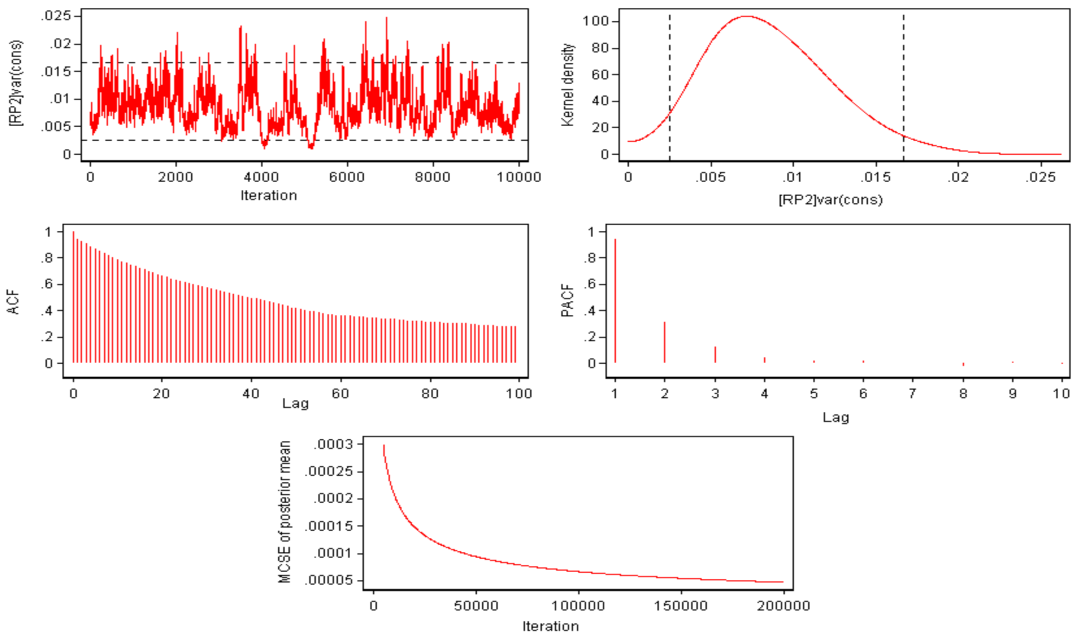
c) Fixed effect - Mother never married



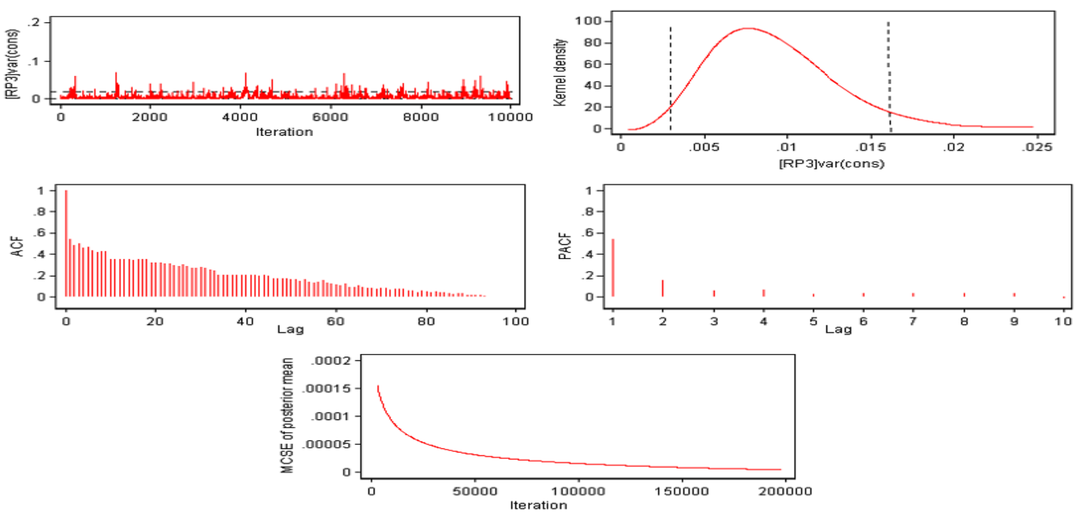
d) Fixed effect - Mother works



e) Level 2 random effect



f) Level 3 random effect



References

Ashiabi GS, O'Neal KK (2007). Children's health status: Examining the associations among income poverty, material hardship, and parental factors. *PLoS ONE*, e940; 2: 1–9. http://doi.org/10.1371/journal.pone.0000940

Bartlett S (2005). Water, sanitation and urban children: The need to go beyond "improved" provision. *Children, Youth and Environments*, 15(1): 115–137. http://doi.org/10.7721/chilyoutenvi.15.1.0115.

Bawah AA and Zuberi T (2005). Socioeconomic status and child survival in southern Africa. *Genus*, 61(2): 55–83. http://doi.org/29788852.

Boco AG (2010). Individual and community level effects on child mortality: An analysis of 28 demographic and health surveys in sub-saharan Africa. Calverton, Maryland, USA, ICF Macro.

Breslow NE and Clayton DG (1993). Approximate inference in generalised linear mixed models. *Journal of the American Statistical Association*, 88(421): 9–25.http://doi.org/ 10.1080/01621459.1993.10594284.

Browne WJ (2003). MCMC estimation in MLwiN. Avaliable from: http://www.bristol.ac.uk/cmm/media/migrated/2-31/mcmc-print.pdf.

Caldwell JC (1979). Education as a factor in mortality decline an examination of Nigerian data. *Population Studies*, 33(3): 395–413. http://doi.org/stable/2173888.

Caldwell P (1996). Child survival: Physical vulnerability and resilience in adversity in the European past and the contemporary Third World. *Social Science and Medicine*, 43(5): 609. https://doi.org/10.1016/0277-9536(96)00109-8.

Cleland JC (1990). Maternal education and child survival: Further evidence and explanations. What we know about health transition: The cultural, social and behavioural determinants of health. Caldwell JC *et al*. (editorss), Health Transition Center, The Australian National University.

Dorrington R, Johnson L, Bradshaw D, *et al*. (2006). The demographic impact of HIV/AIDS in south Africa. National and Provincial Indicators for 2006. Cape Town, Centre for Actuarial Research, South African Medical Research Council, Actuarial Society of South Africa.

Goldstein H (2011). Multilevel Statistical Models, John Wiley & Sons Ltd.

Hair J, Black C, Babin J, *et al*. (2010). Multivariate data analysis, Pearson Prentice Hall.

Heuveline P, Guillot M, Gwatkin DR (2002). The uneven tides of the health transition. *Social Science and Medicine*, 55(2): 313–322. https://doi.org/10.1016/S0277-9536(01)00172-1.

Hill K, Upchurch DM (1995). Gender differences in child health: Evidence from the demographic and health surveys. *Population and Development Review*, 21(1): 127–151. https://doi.org/10.2307/2137416.

Hobcraft J (1993). Women's education, child welfare and child survival: A review of the evidence. *Health Transition Review*, 3(2): 159–175.

Hobcraft JN, McDonald JW and Rutstein SO (1985). Demographic determinants of infant and child mortality: A comaparative analysis. *Population Studies*, 39(3): 363–385. Avaliable from: http://www.jstor.org/stable/2174100.

HSRC (2014). South African national HIV prevalence, incidence and behaviour survey. Human Science Research Coucil.

Kabir A, Islam MS, Ahmed MS, *et al*. (2001). Factors influencing infant and child mortality in Bangladesh: Research paper. *The Sciences*, 1(5): 292–295. Avaliable from: http://docsdrive.com/pdfs/ansinet/jms/2001/292-295.pdf.

Kazembe L, Clarke A and Kandala NB(2012). Childhood mortality in sub-Saharan Africa: cross-sectional insight into small-scale geographical inequalities from Census data. *BMJ Open*, 2(5): e001421. https://doi.org/10.1136/bmjopen-2012-001421.

Kembo J and Ginneken JKV (2009). Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis. *Demographic Research*, 21(3): 1255–1267. https://doi.org/10.4054/DemRes.2009.21.13.

Moser KA, Leon DA and Gwatkin DR (2005). How does progress towards the child mortality millennium development goal affect inequalities between the poorest and least poor? Analysis of Demographic and Health Survey data. *BMJ*, 331(7526): 1180–1182. https://doi.org/10.1136/bmj.38659.588125.79.

Mosley WH and Chen L (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, 10: 25–45.

Mustafa EH and Odimegwu C (2008). Socioeconomic determinants of infant mortality in Kenya: Analysis of Kenya DHS 2003.

Journal of Humanities and Social Sciences, 2(2): 1–16. Avaliable from: https://www.scientificjournals.org/journals2008/articles/1409.pdf.

Ng'weshemi J, Urassa M, Usingo R, et al. (2003). HIV impact on mother and child mortality in rural Tanzania. *Journal of Acquired Immune Deficiency Syndromes*, 33: 393–404. https://doi.org/10.1097/00126334-200307010-00015.

Omariba DWR, Beaujot R and Rajulton F (2007). Determinants of infant and child mortality in Kenya: An analysis controlling for frailty effects. *Population Research and Policy Review*, 26(3): 299–321. https://doi.org/10.1097/00126334-200307010-00015.

Omran AR (1971). The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 49(4): 509–538. http:// doi.org/10.2307/3349375.

Rasbash J, Charlton C, Browne WJ, *et al*. (2012). MLwiN Version 2.26, Centre for Multilevel Modelling, University of Bristol.

Rodgers G B (2002). Income and inequality as determinants of mortality: An international cross-section analysis. *International Journal of Epidemiology*, 31: 533–538.

Sastry N (1996). Community characteristics, individual and household attributes, and child survival in Brazil. *Demography*, 33(2): 211–229. http://doi.org/10.2307/2061873.

Victora CG, Wagstaff A, Schellenberg JA, *et al*. (2003). Applying an equity lens to child health and mortality: More of the same is not enough. *The Lancet*, 362: 233–241. http://doi.org/10.1016/S0140-6736(03)13917-7.

Wagstaff A (2000). Socio-economic inequalities in child mortality: Comparisons among nine developing countries. Bulletin of World Health Organization, WHO. 78.

Waldmann RJ (1992). Income distribution and infant mortality. *The Quarterly Journal of Economics*, 107(4): 1283–1302. http://doi.org/10.2307/2118389.

Wang H, Liddell CA, Coates MM, *et al*. (2014). Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 384(9947): 957–979. http://doi.org/10.1016/S0140-6736(14)60497-9.

Wang L (2003). Determinants of child mortality in LDCs. Empirical findings from demographic and health surveys. *Health Policy*, 65(3): 227–299. http://doi.org/10.1016/S0168-8510(03)00039-3.

Zaba B, Marston M and Floyd S (2003). The effect of HIV on child mortality trends in sub-Saharan Africa. Training Workshop on HIV/AIDS and Adult Mortality in Developing Countries, New York, Population Division Department of Economic and Social Affairs United Nations Secretariat.