The Health Costs of Ethnic Distance: Evidence from Sub-Saharan Africa.

Joseph Flavian Gomes*
University of Essex

October 2014

Abstract

We show that ethnic distances can explain the ethnic inequalities in child mortality rates in Africa. Using individual level micro data from the DHS surveys for fourteen Sub-Saharan African countries combined with a novel high resolution dataset on the spatial distribution of ethnic groups we show that children whose mothers have a higher linguistic distance from their neighbours have a higher probability of dying. In contrast, fractionalization reduces the probability of child death. We argue that fractionalization reflects a higher stock of knowledge and information leading on average to better health outcomes, but that access to that information is impaired for mothers that are linguistically distant from their neighbours. Consistent with this interpretation, linguistically distant mothers also have a lower probability of knowing about the oral rehydration product (ORS) for treating children with diarrhoea.

Keywords: child mortality, ethnic inequality, ethnic/linguistic distance, ethno-linguistic diversity, African development, health inequalities.

---

*Institute for Social and Economic Research, University of Essex, Colchester, UK. E-mail: joseph.dse@gmail.com, jgomes@essex.ac.uk. A previous version of this paper was submitted as part of the author's PhD thesis at the Department of Economics, Universidad Carlos III de Madrid. The author is grateful to Klaus Desmet, Ignacio Ortuno-Ortin, Ulrich Wagner and Jesus Carro for their guidance and support; to Sonia Bhalotra, Irma Clots-Figueras, Juan Jose Dolado, Joan Maria Esteban, Jim Fearon, James Fenske, Jed Friedman, Paola Giuliano, Oded Galor, Saumitra Jha, Eliana La Ferrara, Edward Leamer, Matilde Machado, Stelios Michalopoulos, Ricardo Mora, Dan Posner, Diego Puga, Richard Scheffler, Nico Voigtlaender, Romain Wacziarg, and all seminar/conference participants at the UC3M, ISER (U. Essex), III Development Week at NCID (U. Navarra), Warwick Summer Workshop in Economic Growth (U. Warwick), and the Nottingham School of Economics, University of Essex (Government department), University of Kent (Economics) for their comments and suggestions; to Jim Fearon for generously sharing the data on ethnicity and languages matching; to Romain Wacziarg and UCLA Anderson for their hospitality.

1 Introduction

Nineteen thousand children die worldwide every day before reaching the age of five. The highest rates of child mortality are still concentrated in Sub-Saharan Africa, where 1 in 9 children die before reaching the age of five, which is not only more than 16 times the average for developed regions (1 in 152) but also a lot higher than in South Asia (1 in 16) which has the second highest rates of child mortality (UNICEF, 2012). The striking feature of child mortality rates in Africa is that while all of Africa is poor, there is a huge disparity in child mortality rates across ethnic groups. We argue that ethnic distances can explain the existence of such disparities in child mortality rates across ethnic groups in Africa. The ethnic distance between any two ethnic groups is measured by how different the languages are that the two groups speak. We show that children of mothers who are ethnically distant to their neighbours face a higher probability of dying before reaching the age of five. One possible explanation for our finding is that information does not flow smoothly across ethnic lines and individuals who are ethnically distant to their neighbours lose out.

Child mortality or under-five mortality is an important measure of development and reducing it is a Millennium Developmental Goal (MDG no. 4). Our main focus is to explain why child mortality rates vary across ethnic groups in Africa. The literature has attributed such differences in child mortality rates across ethnic groups in Africa to different practices by different ethnic groups which affect the demographic behaviour and cultural status of women; the geographical location of different groups, i.e. whether the groups live close to cities, or in environmentally better regions with better climate that supports better crops, and less diseases like malaria or if they live in economically important regions etc. Gyimah (2002) for example, particularly stresses the differences in socio-cultural practices such as dietary taboos and food avoidances on mothers and infants, as well as perceptions of disease aetiology and treatment patterns across ethnic groups.[2]

In this paper we make a distinction between two different concepts of ethnic diversity. The first is an aggregated measure of diversity like fractionalization. Fractionalization is defined as the probability that two randomly selected individuals from a given region speak two different languages. All individuals, regardless of their ethnicity, face the same level of ethnic

[2] "Among the Mole-Dagbani groups in northern Ghana, for instance, women are denied eggs and other protein food during pregnancy which is likely to affect their nutritional status hence the birth weight of the child. Similarly, pregnant Akan women are encouraged to avoid rich food such as mangoes and ripe plantain for fear of miscarriages, particularly in the early months of pregnancy" (Gyimah, 2002).

fractionalization in a region. However, our primary focus is on the concept of individual level ethnic distance, which measures how ethnically different an individual is from others living in the same region. Using high quality individual level micro data from the Demographic and Health Surveys (DHS) combined with a novel dataset on the spatial distribution of ethnic groups at the level of approximately 1 square km for fourteen Sub-Saharan African countries we highlight the importance of ethnic distance in explaining the disparities in child health outcomes between ethnic groups while controlling for the commonly provided explanations including geography, location, cultural differences between ethnic groups etc. The ethnic distance variable is calculated using the average linguistic distance of the mother from individuals living around her in circles of different radii.

There are two primary findings of our study. First, children of mothers who are ethnically distant from their neighbours have a higher probability of dying as children. This result is robust to the inclusion of the commonly used measures of ethnic diversity, as well as several individual specific controls, apart from ethnicity, region, and country-time fixed effects. This finding holds as long as ethnic distances are calculated using circles of radii ranging from 25 km to 125 km around the mother, but not in bigger circles. Our second key finding is that children of mothers living in more ethnically fractionalized places have a lower probability of child death.[3]

In terms of average marginal effects, if we consider a circle of 75km around the mother, a one standard deviation (0.267) increase in the linguistic distance of the mother from her neighbours increases the probability of her children dying by around 0.73%, which is about 3.2% of all the child deaths in our sample. By itself this marginal effect looks small. But to put things in perspective, a one year increase in the mother's education, which is a very important variable, leads to only a 0.5% decrease in the probability of child death.

While there are several possible alternative explanations, we use the recent insights of Ashraf and Galor (2013) to interpret our empirical findings. Ashraf and Galor (2013) point out that diversity could have both positive and negative impacts on economic outcomes. In the same vein, we argue that ethnic diversity leads to a higher stock of knowledge in society about how to rear one's children and thus improves the health outcomes of children, whereas individual ethnic distances act as barriers to accessing such knowledge and thus lead to worse health outcomes.[4] Using the same DHS data we find some evidence in favour of this hypothesis. We see that,

[3]Controlling for polarization instead does not change the results. Polarization measures how far the distribution of the linguistic groups is from the bipolar distribution (1/2, 0, 0, ... , 0, 1/2).

[4]We discuss some of the other possible explanations in the next section.

linguistically distant individuals are less likely to have heard of the oral rehydration product (ORS) for treating children with diarrhoea, whereas individuals living in more fractionalized places are more likely to have heard of ORS.

Since we control for ethnicity and religion specific fixed effects our results are not driven by heterogeneity in unobservable characteristics across ethnic groups. This lets us abstract from previously provided explanations of differences in child mortality rates between ethnic groups including differences in cultural practices between different ethnic groups etc. Also, the use of region and country-time fixed effects allow us to discard the location and geography explanations. However, there still remain concerns about endogeneity of our results due to unobserved differences between individuals who live in places where they are ethnically more or less distant to others. In order to address such endogeneity we control for a host of variables that would reduce the possibilities of omitted variable bias. We also explicitly control for migration in some of the specifications. Then, using recently developed methods by Altonji et al. (2005) we show that our results are unlikely to be driven by selection on unobservables.

Apart from our two main results we also find some evidence that wealthier mothers can mitigate some of the negative effects of linguistic distance. But we find no heterogeneity in the effects of linguistic distance by education, child gender, place of residence (urban or rural), or distance from the capital. Neither do we find any evidence of nonlinearities in the effects of linguistic distance. However, in line with Ashraf and Galor (2013) we do find some weak evidence that ethnic diversity has a nonlinear effect on child mortality. We also undertake several robustness checks and show that our results remain qualitatively unchanged to different specifications, controls etc.

Our paper contributes to several different strands of the literature. One of the primary strands is the literature that finds ethnic diversity to have a negative effect on the provision of public goods and hence health outcomes.[5] However, most of the literature is at the cross country level. Interestingly, Platas (2010) finds ethnic diversity to be either insignificant or have a positive and significant effect on some health outcomes including child mortality in Africa. The literature has yet to provide an explanation for this seemingly puzzling finding. In contrast we take this literature to an individual level analysis with a rich set of controls and point out that while ethnic diversity might indeed have a positive effect on child mortality rates, ethnic

[5] Miguel and Gugerty (2005), Alesina et al. (1999), Vigdor (2004), Habyarimana et al. (2007), Alesina et al. (2003), Desmet et al. (2012), La Porta et al. (1999), Ahlerup (2009), Ghobarah et al. (2004), Lieberman (2007), Platas (2010)

distance has a negative effect.

The second strand that we contribute to is the literature that views child mortality as an indicator of individual welfare or development and tries to understand its determinants.[6] Some of this literature particularly focuses on ethnic favouritism (Kudamatsu, 2009; Franck and Rainer, 2012). However, these papers usually attempt to identify the effect of the ethnicity of the countries' leader on mortality rates in different groups. So far none of these papers have looked at the effects of individual ethnic distance on child health outcomes.

The third strand to which we contribute is the small but growing literature that emphasizes the role of ethnic distances in explaining different socio-economic outcomes.[7] We take this literature a step forward. To the best of our knowledge, we are the first to relate individual level health outcomes to the ethnic distance of the individual from her neighbours.[8]

Finally, we contribute to the literature that tries to explain the emergence of ethnic inequality. Ethnic inequality defined as the inequality in well-being across ethnic groups that coexist, is bad for economic growth (Alesina et al., 2012), provision of public goods (Baldwin and Huber, 2010), and can lead to civil conflicts (Mitra and Ray, 2010; Gomes, 2012) etc. We show how ethnic distances might lead to ethnic inequality in health outcomes in Africa. While some previous papers have attempted to explain ethnic inequality in child mortality rates (Gyimah, 2002; Brockerhoff and Hewett, 2000), we are the first to underscore the importance of ethnic distance.[9]

The rest of the paper is organized as follows. In Section 2, we explain why ethnic distance and diversity might matter for child mortality. In Section 3, we discuss the data sources and how the different variables are constructed. In Section 4 we present our empirical analysis and results. In Section 5 we provide some evidence of how ethnic distance might affect access to information. In Section 7 we conclude.

[6] Besley and Kudamatsu (2006), Kudamatsu (2009), and Kudamatsu et al. (2012), Franck and Rainer (2012).

[7] Economic development via human capital accumulation (Laitin and Ramachandran, 2014; Shastry, 2012), trade flows (Isphording and Otten, 2013), literacy and labour market outcomes of immigrants (Isphording, 2013). See also Desmet et al. (2012), Desmet et al. (2009), Gomes (2013), Spolaore and Wacziarg (2009), Esteban et al. (2012a), Esteban et al. (2012b) for aggregate cross-country measures incorporating ethnic distances.

[8] Kumar et al. (2012) find geographic distance to be an important barrier to good maternal and child health outcomes in India. We on the other hand investigate the effects of ethnic distance on child health outcomes while controlling for geographic distance.

[9] Ethnic inequality is not just a feature of Africa. See Banerjee and Somanathan (2007), Sethi and Somanathan (2010) for evidence of ethnic inequality in India.

2 Why Does Ethnic Distance Matter?

There are several cross country papers showing how ethnic diversity might have negative effects on different health outcomes including infant mortality rates (Alesina et al., 2003); public health spending, life expectancy (Ghobarah et al., 2004); AIDS expenditures, anti-retroviral treatment coverage, etc. (Lieberman, 2007). This literature however, is in sharp contrast to the literature that studies the effects of diversity at a more localized level and find diversity to have a positive effect on different socio-economic outcomes including, creativity in groups (McLeod et al., 1996), economic growth in cities (Glaeser et al., 1995; Ottaviano and Peri, 2006), public goods (Desmet et al., 2014). Moreover, Platas (2010) focusing on a wide range of health outcomes including infant and child mortality shows how local ethnic diversity might in fact have a positive effect on health outcomes.

Ashraf and Galor (2013), on the other hand underscore that diversity could have both beneficial and detrimental effects on the economy. On the one hand diversity enhances knowledge creation and accumulation and fosters technological progress in the economy. On the other hand diversity leads to more inefficiency by increasing the possibilities of disarray and mistrust and thus leading to reduced cooperation and disrupting socioeconomic order. We use these insights of Ashraf and Galor (2013) and their framework with minor modifications and argue that diversity reflects a higher stock of knowledge and information about how to rear one's children. Overall child mortality is thus lower in ethnically diverse localities. However, such knowledge does not flow smoothly across ethnic groups and especially to groups which are ethnically very distant and thus such groups lose out.

Formally, we consider an economy where the level of ethnic diversity affects the level of productivity. The level of technology is given by $A = A(z, \omega)$, where z denotes the institutional, geographical and human capital factors and $\omega = [0, 1]$, which is the degree of ethnic diversity, has a positive but diminishing effect on the level of technology.[10] However, d_i, which is the ethnic distance of an individual to her neighbours, reduces the individual production. Let x be the individual labour input. Thus, the individual health production function is given by,

$$y_i = (1 - d_i)A(z, \omega)f(x_i) \tag{1}$$

The individual ethnic distance, d_i, represents a barrier to knowledge and impedes indi-

[10] $A(z; \omega) > 0$, $A_\omega(z; \omega) > 0$, and $A_{\omega\omega}(z; \omega) < 0$ for all $\omega \in [0; 1]$.

vidual access to information about good health.[11] On the other hand overall diversity actually improves health outcomes via the technology term technology $A = A(z, \omega)$.

The role of information for healthcare cannot be over emphasized. Malhotra (2012) for example points out how lack of information on feeding practices or nutritional knowledge amongst families plays a key role in the persistence of chronic child malnutrition in India. More than a third of under-five deaths worldwide are attributable to under nutrition (UNICEF, 2012). Again, the Audience Scape National Surveys highlight the role of information for health care for several African countries. For example, they find that in Zambia people who have received information about HIV/AIDS, malaria, or family planning within a month prior to the survey were more likely to be in better health (Zhou, 2010). "Word of mouth" is found to be important for health information, with friends or family members acting as key channels of such information. The top three most trustworthy sources for health issues were found to be medical doctors, radio, and friends or family members.[12]

While ours is the first paper to underscore the negative health costs of linguistic distant via the information channel, there is some evidence in the literature on ethnicity being a barrier for information. Fisman et al. (2012) for example show how cultural proximity (or a lower ethnic distance) between lenders and borrowers in India mitigates problems of asymmetric information in lending. Similarly, Pongou (2009) points out that information circulates more easily within ethnic groups than across and highlights the implications for HIV/AIDS in Africa. Again, in Kenya there is evidence on targeting of spread of health information via maximum language use, Kiswahili being the language of the majority (Bowen, 2010). However, this makes it harder for fringe groups to get access to such information.[13]

While we use the Ashraf and Galor (2013) framework to explain the role of linguistic distance and ethnic diversity on healthcare, our framework can be easily modified and related to the literature on ethnic diversity and public goods. If like the Alesina et al. (1999) model we

[11] If we take $d_i = 0$, then we get a simplified version of the model in Ashraf and Galor (2013). In Ashraf and Galor (2013) additionally "a fraction, $\alpha\omega$, of the economy's potential productivity, $A(z;\omega)$, is lost due to lack of cooperation and resultant inefficiencies in the production process." Hence, they show that $y = (1 - \alpha\omega)A(z,\omega)f(x) \equiv y(x, z, \omega)$ is a strictly concave hump-shaped function of ω.

[12] Similar surveys with similar findings exist for other countries as well. See Montez (2011) for Tanzania, for example.

[13] Again, Singleton and Krause (2009) points out how Spanish speaking patients face barriers to accessing health care even in the US. In Mexico indigenous people don't go to the hospital in fear that their language and customs will not be understood and due to lack of trust between groups. http://www.nytimes.com/video/2013/08/13/world/americas/100000002373842/a-chiapas-medicine-man.html

assume that the median voter decides which public goods get provided, then as an individual's distance from the median voter (the average person around her) increases, the worse off she is. Again, individuals who are very different from others can be easily identified and thus can be easily discriminated against reducing their access to public goods (Blattman and Miguel, 2010; Caselli and Coleman, 2013; Fearon and Laitin, 1996; Miguel and Gugerty, 2005). However, using these models we cannot explain why ethnic diversity has a positive effect on health outcomes while ethnic distances have a negative effect. We thus stick to our simple framework based on Ashraf and Galor (2013).

3 Data

3.1 Spatial Distribution of Ethnic groups

In order to construct the ethnic distance of the mother from people living around her, we need the distribution of ethnic groups across space. Until very recently there was no comprehensive database on the spatial distribution of ethnic groups available. Desmet et al. (2014) (work in progress) fill this gap by constructing the most comprehensive database on the spatial distribution of ethnic groups for the whole world at a resolution of approximately 1 sq. km. We exploit this new database.[14]

Desmet et al. (2014) construct their database using two different sources of data. For the spatial distribution of population they use the Landscan data. Landscan is the finest resolution global population distribution data available for the entire world. The resolution is 30 arc seconds by 30 arc seconds, which is approximately 1 sq. km at the equator.[15] For the information on ethnic groups they use the 15^{th} edition of Ethnologue which maps over 7600 linguistic groups for the whole world. The linguistic groups are represented in the form of polygons across space where each polygon represents the homeland of a particular linguistic group. Areas where multiple languages are spoken are represented via overlapping polygons. The total population pertaining to a particular linguistic polygon within a particular political boundary is also provided.

For example, let us consider the case of Ethiopia whose population distribution from Landscan can be seen in the Figure 1.[16] Figure 2 gives the languages in Ethiopia once it is over-

[14] Alesina and Zhuravskaya (2011) construct a database of ethnic diversity at the sub-national level but their database only goes down to the district level and only for 92 countries.

[15] For details see http://web.ornl.gov/sci/landscan/

[16] In Figure A.3, we zoom in into the capital Addis Ababa and can clearly notice how the population is concentrated around the capital.

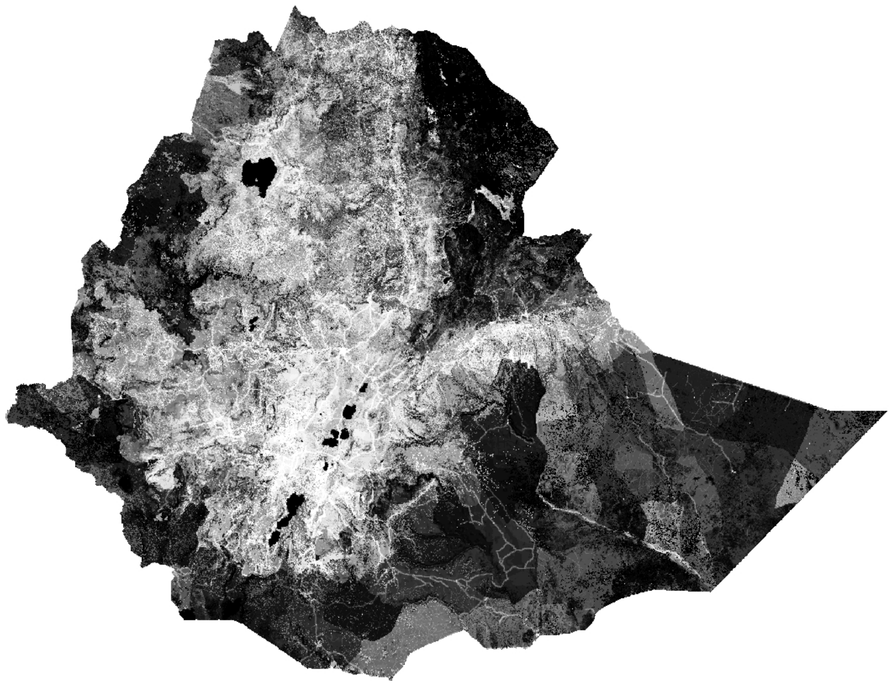

Figure 1: Ethiopia Population distribution based on Land Scan data. Each pixel roughly represents 1x1 sq. km of land area and darker the pixel the less populated the corresponding sq. km is.

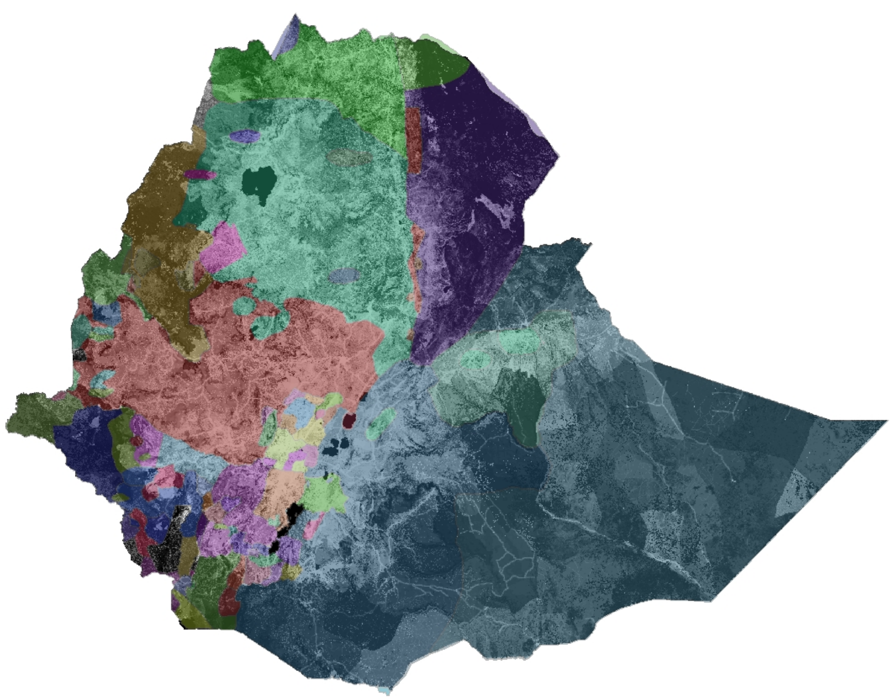

Figure 2: Ethiopia language distribution based on Ethnologue data overlying the population distribution. Each colour polygon represents a different language group.

laid on the population distribution. Polygons of different colours represent different languages. While in Ethiopia, there are no overlapping languages according to the Ethnologue data, in some countries, we have areas where multiple languages are spoken. Such areas where multiple languages are spoken are represented by overlapping polygons. For instance, Figure 3, maps the distribution of languages for Mali. While it is not obvious from the map in Figure 3, some areas in Mali have overlapping languages. The small area highlighted in blue in the South East corner of Mali is one such area. In Figure 4 we zoom into that area. Now we can clearly see an area demarcated by blue borders.

The blue border demarcating the outermost polygon in Figure 4, is the area where the Senufo Mamara language is spoken. This entire polygon can be dived into 3 zones. The leftmost zone in dark pink is where Maasina Fulfulde and Bamanankan languages are spoken in addition to the Senufo Mamara language. The middle zone, which is greenish in colour, is where Maasina Fulfulde and Senufo Mamara are spoken, and finally, the rightmost zone, in light pink is the zone where only the Senufo Mamara language is spoken.[17]

Thus, we see that within the same country there could be zones where one, two or multiple languages are spoken. The language allocation procedure for arriving at the distribution of languages, takes this into account in the allocation algorithm. Depending on the number of language polygons overlying a particular population cell, the total language population, and the total population of the country, each cell gets a particular distribution of languages via programming in ArcGIS, Python, Matlab and Stata. The final data on the spatial distribution of ethnic groups gives us for each square km of the world, the languages spoken and how many people speak each of those languages in that square km.[18][19]

3.2 Linguistic Distance

We measure ethnic distances using the linguistic distance between the languages that the different ethnic groups speak. We follow Fearon (2003), Desmet et al. (2009), Desmet et al. (2012) and several other recent papers which use linguistic tree diagrams to measure distances between

[17]Appendix Figure A.1 gives the distribution of all African languages along with their corresponding populations. The blank areas have no information on languages but are almost always sparsely populated (or unpopulated) desert areas. If there is some population living in these areas then they are assigned to the language of the nearest polygon. Figure A.2 gives the population distribution at the 1 sq. km level coming from LandScan which gives us the number of people living in each square km of the world.

[18]The reader is directed to the data section of Desmet et al. (2014) for further details.

[19]Alesina et al. (2013) use a similar methodology using the same data sources and calculate district and country level averages of historical plough use by ancestors of different ethnic groups.

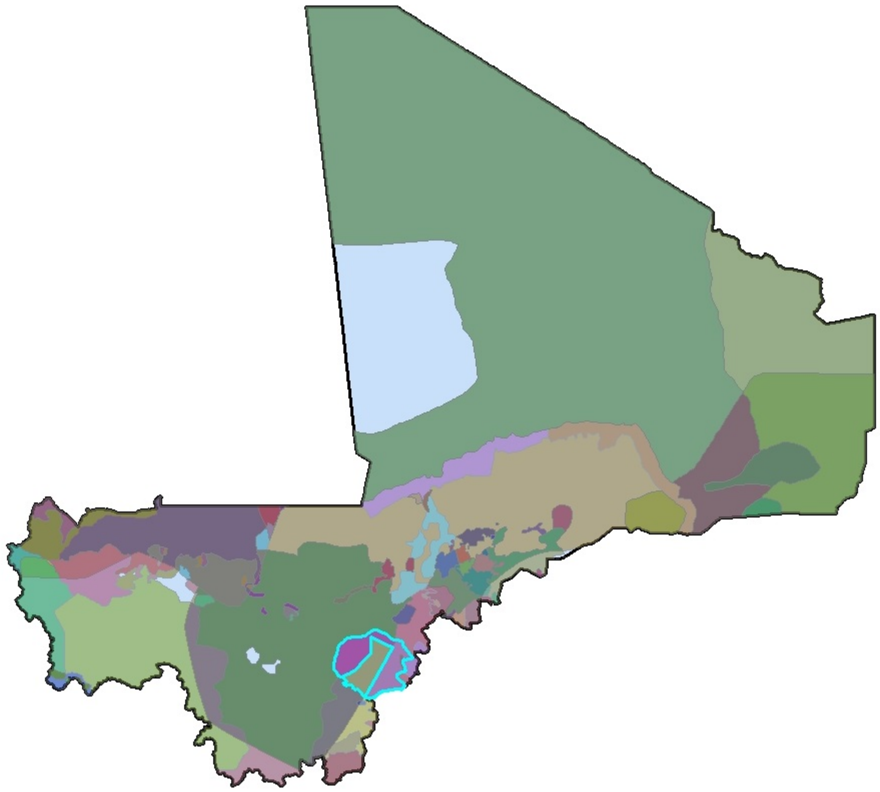

Figure 3: Mali Language distribution based on Ethnologue data. Each colour polygon represents a different language group.

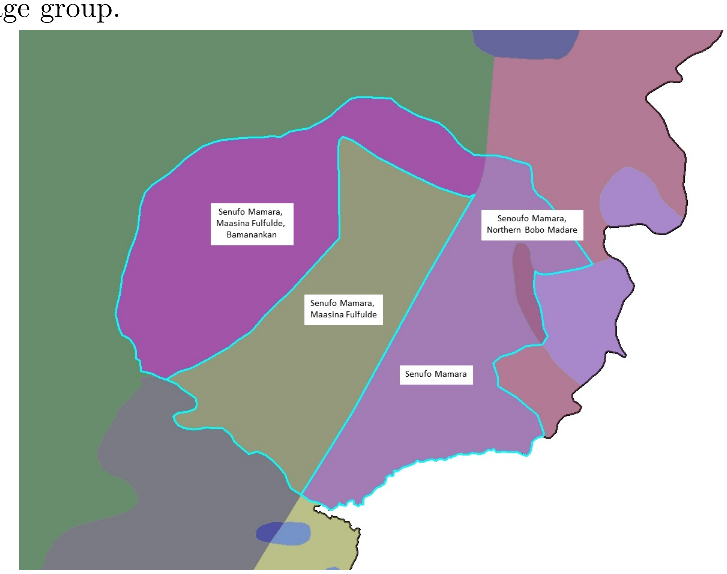

Figure 4: Mali overlapping language polygons example. The polygons demarcated by blue denote an area where multiple languages are spoken. The leftmost zone in dark pink is where Maasina Fulfulde and Bamanankan languages are spoken in addition to the Senufo Mamara language. The middle zone, which is greenish in colour, is where Maasina Fulfulde and Senufo Mamara are spoken, and finally, the rightmost zone, in light pink is the zone where only the Senufo Mamara language is spoken.

languages. The distance between two languages j and k using this approach is defined as:

$$\tau_{jk} = 1 - \left(\frac{l}{m}\right)^{\delta} \tag{2}$$

where l is the number of shared branches between j and k, m is the maximum number of branches between any two languages, and δ is the decay factor, which is a parameter that determines how fast the distance declines as the number of shared branches increases. Data on language trees come from the Ethnologue data.[20]

For our final analysis we need to calculate the average linguistic distance of each mother in our sample to all other individuals living around her in circles of different radii. The linguistic distance for each mother j (who speaks language j) to all other individuals in the circle is given by,

$$\sum_{k=1}^{n} \tau_{jk} \tag{3}$$

where there are n individuals living in the circle and k represents the language of each of those n individuals. The function τ_{jk} is defined by the formula 2.

The decay factor δ measures, "how much more distant should we consider two languages from different families to be relative to languages that belong to the same family" (Desmet et al., 2009). There is no consensus in the literature on what the value of δ should be. While, in their empirical exploration, Desmet et al. (2009) find that values of δ between 0.04 and 0.10 perform well and choose a δ of 0.05, Fearon (2003) uses a δ of 0.5. Since, there is no theoretical basis for choosing one value of delta or the other, we let the data tell us which values of δ perform better than others and find that lower values of δ perform better than higher values and fix δ at 0.0025. Choosing a δ of 0.05 like Desmet et al. (2009) leads to qualitatively similar results.

In order to understand what the different values of δ imply, let us consider the two Indo-European languages Greek and Italian. Following the language tree from Ethnologue, these two languages share one common branch. Taking a δ of 0.5 like Fearon (2003), the distance between them is 0.74. Again, if we consider Chinese and Italian which belong to completely different families and thus share no branches in common, the distance between them is one. On the other

[20] There are other ways of measuring linguistic distances. For example, Dyen et al. (1992) use the proportion of cognates in any two languages. Again, Isphording (2013) uses not only cognates but also the number of sounds that need to be changed between two words that have the same meaning (say, Tu and You) in two different languages. However, distance calculated using language tree diagrams are more useful since the data is a lot more comprehensive and exists for all countries. See Desmet et al. (2009) for a discussion.

hand if we take a δ of 0.05 following Desmet et al. (2009), the distance between Greek and Italian becomes 0.13, whereas that between Chinese and Italian continues to be one. Finally, if we choose a δ of 0.0025, the distance between Greek and Italian is 0.007 while that between Chinese and Italian is still one.[21]

3.3 Linguistic Diversity

Our primary measure of ethnic diversity is the commonly used measure of ethnic fractionalization. Ethnic fractionalization has been found to be bad for a host of socio economic outcomes (Alesina et al., 2003) and has often been blamed for Africa's poor economic performance (Easterly and Levine, 1997). However, recent literature has emphasized that for certain outcomes like civil conflicts etc. ethnic polarization rather than fractionalization is more relevant (Montalvo and Reynal-Querol, 2005). In some of our specifications we also control for polarization. If the level of analysis is the country level, then there is a distinction between these two measures of diversity. However, at very fine levels of disaggregation, as is our case, both fractionalization and polarization are highly correlated with a correlation of above 0.82, and yield similar results.

The fractionalization measure $frac(j)$ gives the probability that two randomly selected individuals from a given country speak two different languages. The polarization measure $pol(j)$ on the other hand measures how far the distribution of the linguistic groups is from the bipolar distribution (i.e. the (1/2, 0, 0, ... , 0, 1/2) distribution) which represents the highest level of polarization (Montalvo and Reynal-Querol, 2005). The fractionalization index is maximized when each individual in the region belongs to a different linguistic group, while the polarization index is maximized when there are only two groups in the region and they of equal size.[22] Formally, the two measures are defined as follows:

$$\text{Fractionalization: } frac(j) = 1 - \Sigma[S_{i(j)}]^2. \tag{4}$$

$$\text{Polarization: } pol(j) = 4\Sigma[S_{i(j)}]^2[1 - S_{i(j)}]. \tag{5}$$

where $S_{i(j)}$ is the proportion of the population speaking language i at geographic region j. The disaggregated nature of our data allows us to calculate diversity at different levels of aggregation.

[21] Example from Desmet et al. (2009).

[22] The reader is directed to Montalvo and Reynal-Querol (2005) for a detailed discussion and comparison of the two measures.

We thus calculate our measures of diversity at both the circle and district levels. Geographic region in the different specifications could refer to either circles or districts. However, as we will see later this does not affect our results.

3.4 Child Mortality

Child mortality is the death of a child before reaching the age of five. If the child dies before reaching the age of one then it is termed as infant mortality while if the child does not survive for a month after its birth, then it is termed as neo-natal mortality. The Demographic and Health surveys (DHS) make available data on child mortality at the individual level for many developing countries from across the world. Funded by the the U.S. Agency for International Development (USAID), the DHS has been conducting surveys in several developing countries since the 1980s. By interviewing a nationally representative sample of women of child bearing age (15 to 49), the DHS collects data on all the children they have ever given birth to in the past including the children who did not survive till the time of the interview. The standardized components of the DHS questionnaire can be used to compile cross country micro data sets.

Each DHS survey provides information on life and health outcomes of individuals which allow us to construct measures of child mortality and several other individual level variables including the child's gender, birth-order, their mother's weight, stature, years of education, and occupation, wealth level etc. Moreover, in many of the surveys the mother's ethnicity is also provided. Also, using a GPS receiver, the geographic coordinate of each geographic cluster (village or town) is also collected. In Figures 5 and 6 for example, we see the spatial location of all DHS clusters for Mali and Ethiopia respectively. We draw circles of 25 km radius around the DHS clusters for the DHS surveys conducted in 1996 in the case of Mali and 2000 in the case of Ethiopia, in order to calculate average ethnic distances.

We use fourteen countries in our analysis viz. Kenya, Uganda, Ethiopia, Burkina Faso, Malawi, Senegal, Zambia, Sierra Leone, Mali, Guinea, Ghana, Benin, Namibia and Niger. To start with we consider all Sub-Saharan African countries for which DHS data is available. Our analysis requires along with the regular DHS data on health outcomes, the ethnicity and GPS coordinates of the mother's location. Thus, countries which have either of these two information missing have to be discarded. Then in order to calculate the linguistic distance we need to be able to map the ethnicities to the languages. For the matching between ethnicity and languages we closely follow Fearon (2003). But our matching is a lot more comprehensive. Fearon (2003) provides the matching for the major ethnic groups in the country whereas we construct the matching for each and every ethnic group in the country regardless of its size. We have also had to discard countries for which this mapping is not good enough.[23]

3.5 Geographic Distance

The geographic isolation and distance from the capital are calculated by using the formula for the great circle distance. The geographic distance between any two points in space ℓ and k, denoted by $|\ell, k|$, is computed as the great circle distance:

$$|\ell, k| = r_E \arccos(\sin(lat_\ell)\sin(lat_k) + \cos(lat_\ell)\cos(lat_k)\cos(long_\ell - long_k) \tag{6}$$

Using the above formula we calculate the geographic isolation of any individual as the average geographic distance betweeen that individual and all the other people in the country. The highest distance between any two individuals in the country is normalized to 1. The distance to the capital is simply the distance of the individual to the capital calculated using formula 6, where we take ℓ to be the individual's location and k to be the location of the capital.

3.6 Summary Statistics

Figure A.4 in the appendix gives the maps of all the countries included in the analysis. We have fourteen countries and a total of thirty surveys with information on the births and deaths of over 825,000 children. However, we can consider only the children who would have already reached the age of five by the day of the sampling, since we do not know if the others are going to survive till the age of five or not. Thus, as can be seen from Table B.1, in the child mortality sample we have information on the births of 658,755 children out of who about 23% do not survive till their fifth birthday. About 12% do not survive till their first birthday. There is huge variability in the data over time and space. From Table B.2 we can see that in Niger 35% of the children die before reaching the age of five, whereas in Kenya the corresponding number is 13%. The linguistic distance and fractionalization variables all lie between 0 and 1. 22% of the sample is urban and 49% of the sample is female.

4 Empirical Analysis

4.1 Econometric Specification

Our primary relationship of interest is that between child mortality and the ethno-linguistic distance of the mother from her neighbours. We would also like to understand how ethnic

[23]We have cross-checked our data using Fearon (2003).

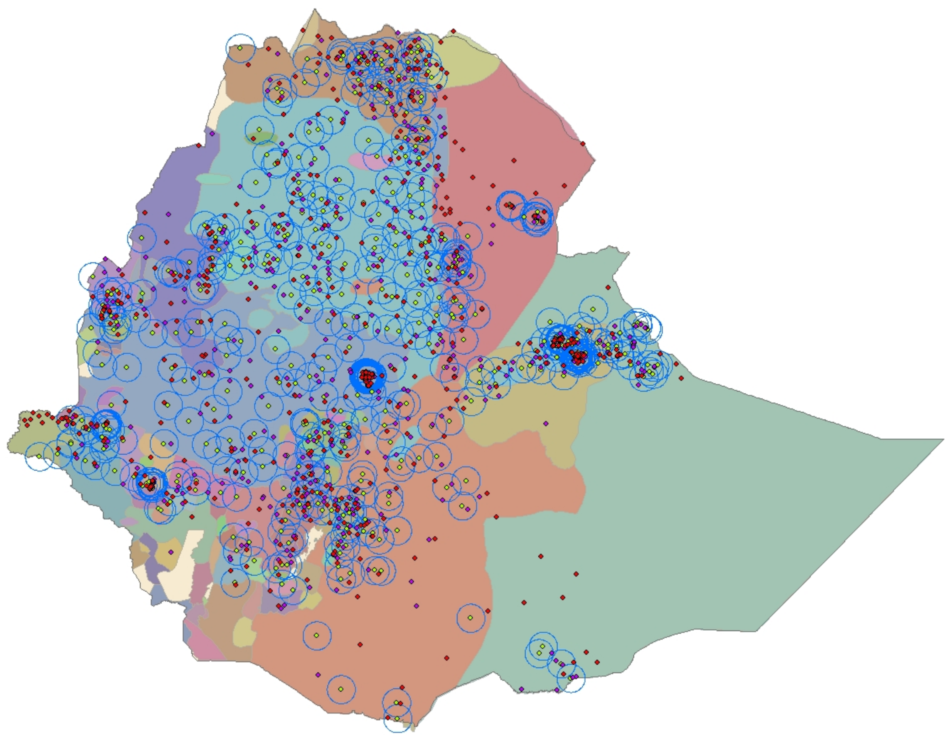

Figure 5: Ethiopia distance example. Circles of 25 km radius have been drawn around DHS clusters from the DHS survey for Ethiopia in 2000.

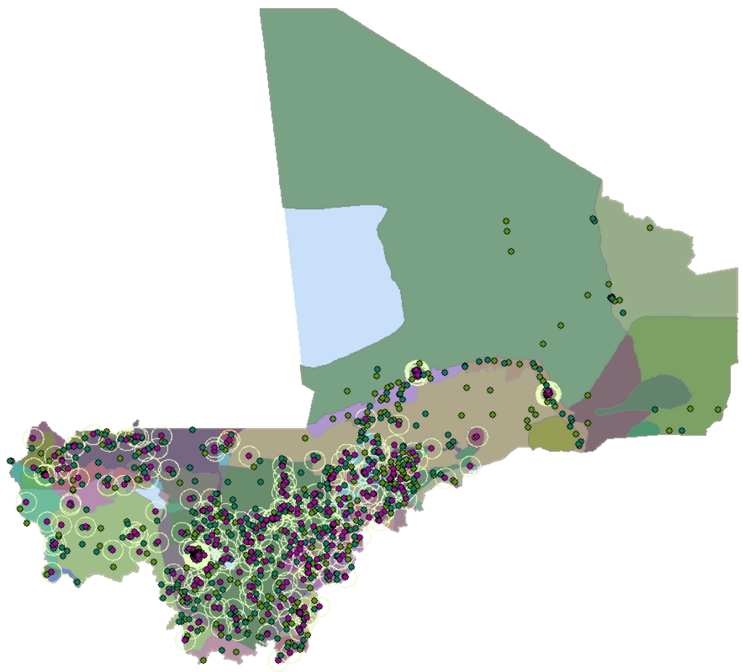

Figure 6: Mali distance example. Circles of 25 km radius have been drawn around DHS clusters from the DHS survey for Mali in 1996.

diversity of the neighbourhood affects child mortality. The baseline specification is given by equation 7. In our baseline specification we have child mortality on the left hand side and our primary variables of interest on the right hand side along with a host of control variables that have been found to be important for child mortality.

$$\begin{aligned} Prob(death_{iet}) = \alpha_R \ + \ \alpha_{ethnicity} \ + \ \alpha_{religion} \ + \ \alpha_{c*t} \ + \ \beta_1 \ ethnic\_distance_{ie} \\ + \ \beta_2 \ ELF_i \ + \ \beta_3 \ X_{it} \ + \ \beta_4 \ X_i \ + \ \epsilon_{iet} \end{aligned} \tag{7}$$

where $Prob(death_{iet})$ is the probability of death of child 'i' born to mother belonging to ethnicity 'e' in year 't'. The $ethnic\_distance_{ie}$ variable is our primary variable of interest and it gives the linguistic distance of the mother of child 'i' belonging to ethnicity 'e' from people living within circles of different radii around her. The ELF_i variable gives the ethno-linguistic fractionalization in the circles of different radii around the mother. For calculating both the linguistic distance and the ELF we have used circles of different radii, viz. 25, 50, 75, 100, 125, 150, 175, 200, 250 km around the mother.[24]

The variables X_{it} and X_i come from the literature on child mortality and have been found to be important for child mortality.[25] X_{it} includes birth specific variables viz. female child dummy, age at birth, age at birth squared, multiple birth, birth order, birth order squared, short birth space prior, short birth space post. X_i includes mother specific variables viz. urban dummy, education years, and dummies for the wealth index. We also control for the geographic distance, which gives the geographic isolation of the mother from everybody else in the country and the distance of the mother's location from the capital, which are both included in X_i.

We include region fixed effects α_R apart from country specific time effects α_{c*t} which allows for the non-parametric evolution of time effects differently for each country. By including ethnicity fixed effects, $\alpha_{ethnicity}$, we control for unobserved heterogeneity across ethnic groups, which allows us to identify the effect of ethnic distance on child mortality that is not driven by ethnicity specific characteristics like ethnic dominance of certain groups, cultural differences leading to differences in health practices between different groups, etc. Likewise $\alpha_{religion}$, the religion dummy controls for differences in religious beliefs and practices across different individuals.

[24]In some specifications we also calculate ELF at the district level. As we will discuss later, controlling for polarization instead does not affect our results.

[25]See Kudamatsu (2009), Baird et al. (2011), Franck and Rainer (2012) for example.

Since child mortality is a 0-1 binary variable we estimate it via pooled probit regressions.[26] β_1 is our coefficient of interest since it gives the effect of linguistic distance of the mother on the probability of death of the child. Given the possibilities of endogeneity giving a causal interpretation to β_1 is not straightforward. Moreover, we cannot use mother specific fixed effects since the ethnic distance variable does not vary across time for the same mother. However, we are able to control for a host of maternal and birth characteristics which alleviate endogeneity concerns to a great extent. Moreover, we later do some analysis to gauge how much selection on unobservables is taking place and this increases our faith in the causal interpretation of β_1. The standard errors are clustered at the Survey-Country level in the baseline specifications, but we do robustness checks by clustering standard errors at different levels.

4.2 Results- Ethnic Distance and Child mortality

Most studies look at how ethnic diversity affects different economic outcomes including infant/child mortality at the aggregate country/ district/ or town level. In this paper we study the effects of individual level ethnic distance on child mortality at the individual level while controlling for the more aggregated measures of ethnic diversity. We are able to construct measures of ethnic diversity taking into account the exact location of the individual.

Table 1 gives our baseline specification. In Table 1 we try to explain how the linguistic distance of the mother from people living around her, affects the probability of her child dying before reaching the age of five. In this baseline specification we consider the average distance of the mother from all individuals living in a radius of 75 km around her.

In the different columns of Table 1 we keep adding different control variables. In column 1 we do not control for any other variables apart from the linguistic distance variable. In column 2, we add a control for linguistic fractionalization. We notice that without any other controls neither linguistic distance nor fractionalization is significant. In column 3 we add individual level controls including wealth and education of the mother, sex of the child, location of the mother, and other individual specific variables that affect child mortality like birth order etc.[27] In column 4 we add region fixed effects and country-time fixed effects. In column 5 we add religion and ethnicity fixed effects and in column 6 we add controls for geographic isolation and

[26] As a robustness check we also use linear probability models, but as we will see later our results remain unchanged.

[27] The wealth index variable is an ordinal index that takes five different values with a higher value indicating a higher level of wealth.

Table 1: Child mortality: Baseline

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 75	0.0502	0.0511	0.0862*	0.0280	0.0995***	0.100***
	(0.0559)	(0.0580)	(0.0470)	(0.0272)	(0.0339)	(0.0340)
fractionalization 75		-0.0113	-0.0646	-0.0313	-0.0461*	-0.0473*
		(0.0741)	(0.0642)	(0.0248)	(0.0235)	(0.0245)
urban			-0.0903***	-0.0674***	-0.0692***	-0.0686***
			(0.0317)	(0.0160)	(0.0153)	(0.0153)
female			-0.0654***	-0.0675***	-0.0679***	-0.0679***
			(0.00435)	(0.00429)	(0.00429)	(0.00430)
education_years			-0.0363***	-0.0219***	-0.0210***	-0.0210***
			(0.00532)	(0.00253)	(0.00248)	(0.00248)
wealth index 2			0.0313	0.0102	0.00725	0.00693
			(0.0190)	(0.0143)	(0.0141)	(0.0139)
wealth index 3			0.0129	-0.00825	-0.0112	-0.0110
			(0.0210)	(0.0217)	(0.0209)	(0.0209)
wealth index 4			-0.0235	-0.0569**	-0.0588**	-0.0586**
			(0.0251)	(0.0256)	(0.0246)	(0.0246)
wealth index 5			-0.0996***	-0.156***	-0.156***	-0.155***
			(0.0365)	(0.0278)	(0.0257)	(0.0255)
log distance to capital						0.00891
						(0.0101)
geographic isolation						-0.0816
						(0.0525)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	658755	658755	658505	649182	648468	648468
pseudo R^2	0.000	0.000	0.062	0.084	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated. The individual controls include age at birth, age at birth squared, multiple births (twins, triplets etc.), birth order, birth order squared, short birth space prior to birth, and short birth space post birth.

distance from the capital.

Column 6 is our most complete and preferred specification. The inclusion of ethnicity fixed effects allows us to control for heterogeneity in unobservable characteristics across ethnic groups including heterogeneity in health outcomes across ethnic groups, and differences in cultural practices between different groups etc. We notice that once religion and ethnicity fixed effects are controlled for (in both columns 5 and 6), our linguistic distance variable becomes highly significant. The linguistic distance of the mother from people living around her significantly increases the probability of her child dying before reaching the age of five. Contrary to previous findings in the literature, the fractionalization variable on the other hand significantly reduces the probability of a child dying before reaching the age of five.

The rest of the variables all have the expected signs and effects. Female children, urban children and children whose mothers are more educated have a lower probability of dying. Again, children whose mothers belong to either of the two highest wealth quintiles have a lower probability of dying. Geographic distance to the capital increases the probability of dying, but geographic isolation does not significantly affect the probability of death.

In Table 2 we verify if our results are robust for circles of alternate radii around the mother. In the different panels of Table 2 we have exactly the identical specifications as in Table 1 except that in each panel we have a different radius of the circle around the mother. To start with in panel 1 we investigate the effects of linguistic distance of the mother from people living in a radius of 25 km around her. Subsequently in panel 2, 3, 4, 5, 6, 7, 8, 9 the radii are 50, 75, 100, 125, 150, 175, 200, 250 km respectively.

As in Table 1, in Table 2 we see that the linguistic distance of the mother from people living around her significantly increases the probability of her children dying. This holds true for linguistic distance calculated in circles ranging from radius 25 km to 125 km, beyond which the linguistic distance variable is not significant. The results are robust to the inclusion of region, ethnicity, country-time fixed effects, individual specific controls, geographic isolation and the popularly used measure of ethno-linguistic fractionalization. Fractionalization whenever it is significant continues to reduce the probability of child death.

In the tables above, so far we have considered the coefficients from the regressions. It is clear that the coefficient for the linguistic distance variable is statistically significant, but is it economically significant as well? In order to gauge the magnitude of the effect of linguistic distance on the probability of child death we now present the average marginal effects. The average marginal effect of an increase in linguistic distance of the mother from her neighbours within a circle of 75 km is of 2.7%. A one standard deviation (0.27) increase in the linguistic distance of the mother from her neighbours increases the probability of her children dying by around 0.73%. Thus linguistic distance explains about 3.2% of the mean child deaths in our sample. Looking at the average marginal effects of fractionalization in the radius of 75 km we see that if we move from the completely homogenous (fractionalization = 0) to a completely heterogeneous (fractionalization = 1) location, the probability of child death falls by 1.3%. And a one standard deviation (0.27) increase in fractionalization leads to a fall of 0.35% in the probability of child death.

By themselves these marginal effects look very small. But to put them in perspective let us compare these average marginal effects with those of some of the other important variables. For instance, a one year increase in the mother's education leads to only a 0.56% decrease in

Table 2: Child mortality: Alternative radii

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 25	0.0439	0.0437	0.0833*	0.0242	0.0605***	0.0610***
	(0.0536)	(0.0573)	(0.0472)	(0.0259)	(0.0219)	(0.0221)
fractionalization 25		0.00163	-0.0490	-0.0270	-0.0386**	-0.0369**
		(0.0514)	(0.0443)	(0.0182)	(0.0165)	(0.0160)
linguistic distance 50	0.0476	0.0473	0.0856*	0.0264	0.0823***	0.0828***
	(0.0555)	(0.0584)	(0.0476)	(0.0268)	(0.0270)	(0.0271)
fractionalization 50		0.00275	-0.0544	-0.0219	-0.0343*	-0.0339*
		(0.0630)	(0.0547)	(0.0204)	(0.0195)	(0.0199)
linguistic distance 75	0.0502	0.0511	0.0862*	0.0280	0.0995***	0.100***
	(0.0559)	(0.0580)	(0.0470)	(0.0272)	(0.0339)	(0.0340)
fractionalization 75		-0.0113	-0.0646	-0.0313	-0.0461*	-0.0473*
		(0.0741)	(0.0642)	(0.0248)	(0.0235)	(0.0245)
linguistic distance 100	0.0519	0.0533	0.0846*	0.0262	0.106***	0.106***
	(0.0562)	(0.0576)	(0.0465)	(0.0275)	(0.0388)	(0.0385)
fractionalization 100		-0.0225	-0.0748	-0.0406*	-0.0582***	-0.0611***
		(0.0865)	(0.0733)	(0.0239)	(0.0214)	(0.0230)
linguistic distance 125	0.0526	0.0545	0.0826*	0.0253	0.112**	0.111**
	(0.0557)	(0.0567)	(0.0456)	(0.0273)	(0.0486)	(0.0479)
fractionalization 125		-0.0419	-0.0873	-0.0441	-0.0638***	-0.0692**
		(0.100)	(0.0843)	(0.0291)	(0.0246)	(0.0281)
linguistic distance 150	0.0510	0.0530	0.0798*	0.0224	0.102	0.0997
	(0.0560)	(0.0566)	(0.0454)	(0.0274)	(0.0631)	(0.0617)
fractionalization 150		-0.0537	-0.0971	-0.0262	-0.0504	-0.0582
		(0.117)	(0.0975)	(0.0378)	(0.0315)	(0.0367)
linguistic distance 175	0.0496	0.0520	0.0779*	0.0207	0.104	0.100
	(0.0563)	(0.0565)	(0.0450)	(0.0274)	(0.0776)	(0.0760)
fractionalization 175		-0.0695	-0.109	-0.0152	-0.0448	-0.0525
		(0.137)	(0.113)	(0.0465)	(0.0373)	(0.0429)
linguistic distance 200	0.0489	0.0515	0.0768*	0.0190	0.104	0.0979
	(0.0569)	(0.0566)	(0.0451)	(0.0272)	(0.0889)	(0.0871)
fractionalization 200		-0.0835	-0.124	0.00314	-0.0341	-0.0410
		(0.159)	(0.130)	(0.0562)	(0.0437)	(0.0484)
linguistic distance 250	0.0476	0.0487	0.0720	0.0226	0.112	0.102
	(0.0591)	(0.0575)	(0.0459)	(0.0274)	(0.104)	(0.101)
fractionalization 250		-0.123	-0.183	-0.0939	-0.149**	-0.154**
		(0.203)	(0.162)	(0.0842)	(0.0684)	(0.0659)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	658755	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

the probability of child death. Again, moving from a rural to an urban location reduces the probability of child death by only 1.8%. Thus, relatively speaking the average marginal effect of linguistic distance is economically important when compared to some of the other variables. The marginal effect of fractionalization on the other hand is much smaller.

From the above analysis it is clear that the linguistic distance of the mother from people living around her is a significant and robust predictor of child mortality. Fractionalization on the other hand actually reduces the probability of child death.

4.3 Heterogeneous Effects

Up to this point we have assumed that the linguistic distance variable has a homogenous effect on the children of all mothers. However, there are several reasons why this might not be the case. For example, wealthier and more educated mothers might be better able to insulate their children from the negative effects of linguistic distance. Again linguistic distance might have different implications for male and female children. In this section we thus try to identify the heterogeneity in impacts of linguistic distance by focusing on different variables like education, wealth, gender, place of residence (urban or rural), and distance from the capital.

Table 3: Heterogeneous effects of linguistic distance by wealth level

	ling dist 25	ling dist 50	ling dist 75	ling dist 100	ling dist 125
linguistic distance	0.118***	0.141***	0.155***	0.157***	0.159***
	(0.0365)	(0.0405)	(0.0465)	(0.0506)	(0.0598)
wealth index					
2	0.0105	0.0107	0.00975	0.00932	0.00893
	(0.0150)	(0.0151)	(0.0150)	(0.0150)	(0.0150)
3	4.31e-05	0.000634	4.46e-05	-0.000430	-0.000961
	(0.0223)	(0.0223)	(0.0221)	(0.0220)	(0.0219)
4	-0.0519**	-0.0508**	-0.0503**	-0.0502**	-0.0507**
	(0.0251)	(0.0253)	(0.0253)	(0.0253)	(0.0253)
5	-0.144***	-0.144***	-0.144***	-0.145***	-0.145***
	(0.0261)	(0.0261)	(0.0260)	(0.0260)	(0.0260)
Interaction					
2	-0.0274	-0.0284	-0.0208	-0.0173	-0.0136
	(0.0267)	(0.0259)	(0.0250)	(0.0244)	(0.0242)
3	-0.0921**	-0.0931**	-0.0864**	-0.0812*	-0.0754*
	(0.0461)	(0.0443)	(0.0427)	(0.0415)	(0.0404)
4	-0.0560	-0.0618	-0.0630	-0.0614	-0.0564
	(0.0470)	(0.0478)	(0.0485)	(0.0481)	(0.0469)
5	-0.0912*	-0.0873*	-0.0801	-0.0733	-0.0694
	(0.0475)	(0.0494)	(0.0496)	(0.0495)	(0.0498)
N	648,468	648,468	648,468	648,468	648,468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

In Table 3, we verify if linguistic distance variable has a heterogeneous impact on the probability of child death by the wealth level of the mother. The wealth index variable takes five different values with a higher value indicating a higher level of wealth. In this table, we use our baseline specification of column 6 from Table 1 but add to it the interaction term of linguistic distance and wealth level. We notice that linguistic distance continues to be highly significant and positive in sign. Also, as we have already seen before, a higher level of wealth of the mother reduces child mortality. From the interaction term we see that the negative effects of linguistic distance are partly reduced by higher levels of wealth. This is particularly true for the wealth levels 3 and 5 compared to the omitted wealth category 1. In other words, wealthier mothers are partially able to mitigate the disadvantage of being linguistically distant.

We also have various specifications in which we try to identify the heterogeneity in the effects of linguistic distance by fractionalization, education, child gender, place of residence (urban or rural), and distance from the capital. We do not find any heterogeneity in terms of these variables.[28]

4.4 Migration

A possible concern in estimating the effects of ethnic distance on child mortality is the possibility of individuals to migrate. Given the choice to migrate if individuals know that they incur a cost by living with people who are ethnically distant to them then they might choose to move to neighbourhoods which have people who are ethnically more similar to them. However, given various barriers to movement (eg. transportation costs), perfect sorting is not observed in reality.

Michalopoulos and Papaioannou (2014) point out that in spite of population movements ethnic populations tend to reside in their respective historical homelands. In fact, even in the face of large scale population displacements caused by civil wars, individuals tend to try and return to their historical ethnic homeland.[29] However, if individuals actually are able to move to places where they are less distant to others then if anything we are underestimating the effects of ethnic distance on child mortality and the effects of distance would be stronger.

However, in order to ensure that migration is not driving our results, in Table 4 we provide specifications restricting our sample to mothers who have always lived in their current place of residence. Thus we are looking at the variation in linguistic distance within the sample of those individuals who have never migrated.[30] However, this information is not available for all countries and we lose a lot of our observations when we use the variable that indicates for how

[28]Results not provided and are available upon request.

[29]Glennerster et al. (2013) document this for Sierra Leone. Also, Nunn and Wantchekon (2009) document that around 55% of the Afrobarometer Survey respondents currently live in their ethnic group's ancestral homeland.

[30]Of course this sample itself might not be random since if there is the possibility of migration, this sample represents individuals who have not been able to make use of that possibility.

Table 4: Child mortality: No migration sample

	ling dist 25	ling dist 50	ling dist 75	ling dist 100	ling dist 125
linguistic distance	0.0871***	0.115***	0.151***	0.154***	0.183**
	(0.0338)	(0.0397)	(0.0506)	(0.0565)	(0.0751)
fractionalization	-0.0272	-0.0255	-0.0331	-0.0361	-0.0524
	(0.0246)	(0.0266)	(0.0307)	(0.0322)	(0.0386)
N	235311	235311	235311	235311	235311
pseudo R^2	0.085	0.085	0.085	0.085	0.085

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The column headings indicate the radius of the circles in which the linguistic distance and fractionalization variables have been calculated. In these regressions only the sample of mothers who have resided in their place of residence always have been included.

long the individuals have been residing in their current place of residence. Our results remain qualitatively similar. Linguistic distance of the mother continues to significantly increase the probability of child mortality. Fractionalization on the other hand becomes insignificant.

4.5 Selection on Unobservables

One of the main shortcomings of the literature on ethnic diversity has been to establish a causal relationship of ethnic diversity with the different socio-economic variables of interest. Like in our case, ethnic diversity is almost always a time invariant measure and is only observed at a particular point of time. Moreover, there are issues of measurement and omitted variable bias leading to non-causal estimates. Thus, most of the literature has to rest content with finding a correlation.[31]

Unlike most of the previous literature our measure of ethnic distance is individual specific. We are thus able to rely on controlling for an exhaustive set of controls, including individual specific controls, for identification. However, we still cannot completely rule out the possibility of endogeneity arising out of omitted variable bias due to unobserved maternal and family characteristics. We cannot incorporate mother fixed effects since our main variable of interest "linguistic distance" remains unchanged for the mother over time. We thus follow the methodology developed by Altonji et al. (2005) and Bellows and Miguel (2009)who present new estimation strategies that can be used when strong prior information regarding the exogeneity of the variable of interest is unavailable. Exploiting the measure and intuition provided by them, we use the strength of selection on observables to assess the potential bias arising from selection

[31]See Ahlerup (2009) for a discussion on the endogeneity of ethnic diversity.

on unobservables.

Following Altonji et al. (2005), we calculate a ratio that tells us how much stronger the selection on unobservables must be, relative to selection on observables, to explain away the full estimated effect of linguistic distance on child mortality. To calculate such a ratio we need two sets of regressions. We need one regression with a restricted set of controls (or no controls) and another one with a full set of controls. Let β^R be the coefficient of linguistic distance from the restricted regression and β^F be the coefficient from the regression with the full set of controls. Then the ratio $\beta^F/(\beta^R - \beta^F)$ is our ratio of interest. The smaller the denominator, the less the selection on observables and thus the selection on unobservables need to be stronger in order to explain away the entire effect of linguistic distance. Again, the selection on unobservables has to explain away more the higher the numerator is in magnitude (Nunn and Wantchekon, 2009).

Table 5 gives us the corresponding ratios for circles of radii 25 km to 125 km for two distinct sets of restricted regressions. In one of the restricted regressions we consider no controls while regressing child mortality on linguistic distance. However, it makes little sense to not control for at least the individual controls in the restricted regression. Thus, in the next restricted regression we control for the individual controls. In the unrestricted regression on the other hand we consider the full set of controls from our baseline specification.

Table 5: Selection on Unobservables/Observables

	No Controls	Individual Controls
linguistic distance 25	3.57	2.74
linguistic distance 50	2.35	29.57
linguistic distance 75	2.01	7.25
linguistic distance 100	1.96	4.95
linguistic distance 125	1.90	3.91

In this table we provide Altonji ratios ($\beta^F/(\beta^R - \beta^F)$) for two different set of restricted regressions. In the restricted regression for column 1, we consider no controls. For column 2, we control for the individual controls. In the unrestricted regression we consider the full set of controls from our baseline regression.

All the ratios in Table 5 are quite big. Even when we do not control for the individual controls in the restricted regression, the selection on unobservables has to be at least two times larger than that on observables, while in some cases it would need to be at least more than three and a half times larger. Again, if we consider the ratios calculated using the restricted regression with individual controls, we see that the selection on unobservables has to be at least more than around three times larger than that on observables while in some cases it would need to be at least more than around thirty times. Thus, it is highly unlikely that selection on unobservables will explain away the entire effect that we are currently attributing to our linguistic distance variable. We thus argue that linguistic distance indeed increases the probability of child death and this relation is not driven by omitted variable bias.

4.6 Other Robustness Checks

In this section we subject our results to several additional robustness checks. Our primary variable of interest is the probability of child death which is a 0-1 binary variable. Thus, so far we have used probit regressions in order to estimate the effects of linguistic distance of the mother on child mortality. First, we verify if our results hold true if we use Linear Probability Model (LPM) instead. Hence in Table B.9, we rerun our baseline regressions from Table 1 using OLS instead of probit regressions. We notice that our results remain qualitatively unchanged. Linguistic distance continues to significantly increase the probability of child death whereas fractionalization continues to reduce it.

Ethnic diversity usually measured by Ethno-linguistic fractionalization or ELF has often been found to have a negative effect on different socio-economic outcomes.[32] However, we show that fractionalization if anything has a positive effect on child mortality, i.e. it reduces child mortality. Following Ashraf and Galor (2013), we argue that that diversity could potentially reflect a higher stock of knowledge and information, and thus lead to better health outcomes. In Appendix Table B.5 we use the identical specifications from the previous sections but do not control for linguistic distance and show that even when we do not control for linguistic distance, fractionalization continues to significantly reduce child mortality. In all but the last panel of Table B.5 we control for fractionalization calculated for circles of different radii. As a robustness test in the last panel of Table B.5 we also control for fractionalization at the district level.

Some recent papers like Montalvo and Reynal-Querol (2005) stress the importance of polarization rather than fractionalization particularly in the context of intergroup conflict. However, in our context the polarization and fractionalization measures are highly correlated and in Appendix Table B.6 we show that controlling for polarization instead of fractionalization does not change our results.[33]

[32]See Alesina et al. (2003), Alesina et al. (1999), Easterly and Levine (1997) for example. See Alesina and La Ferrara (2000) for a review of the literature.

[33]In Appendix Tables B.8 and B.7 we control for fractionalization and polarization together in the same specification once with the linguistic distance variable and once without respectively. When both the measures of diversity enter the specifications together, polarization becomes insignificant, while fractionalization continues to have a negative and significant effect on child mortality.

Next, we try to identify if either the linguistic distance variable or the fractionalization variable have a non-linear effect on child mortality. In Table B.10 we present the identical specification used in the preferred specification from the baseline table, but add to it nonlinear terms for both fractionalization and linguistic distance in different combinations. We notice that on including nonlinear terms the results become less significant in some of the specifications, but overall the results remain similar to what we found in the previous sections. Linguistic distance continues to increase the probability of child death and there is no robust evidence of any nonlinearity of this effect. As far as fractionalization is concerned there is some weak evidence of nonlinearity. Initially, fractionalization reduces child mortality, but eventually after a certain level it increases child mortality. This finding is in line with the findings of Ashraf and Galor (2013).

So far our focus has been on child mortality which is the event that the child dies before reaching the age of five. However, a lot of the literature has focused on Infant mortality which is defined as the child dying before reaching the age of one. In some specifications we also try out our baseline specifications of child mortality on infant mortality. Qualitatively the impact of the linguistic distance of the mother on the probability of infant mortality is similar to that on the probability of child mortality. However, the results are not statistically significant. There might be two interpretations for this. First, "Infant mortality is a rare event and estimating it requires a large number of observations" (Kudamatsu, 2009). Second, infant mortality has a lot to do with conditions during birth and need not have anything to do with ethnic distance. However, as the child grows the access to certain information becomes crucial, which linguistic distance might impede, thus linguistic distance affects child mortality more than it affects infant mortality.[34]

In the analysis so far we have included all the births in the maternal history of the mother. One possible concern in using retrospective data is recall bias which stems from the fact that women might be less likely to accurately remember more distant births and deaths. To minimize recall bias in appendix Table B.3 we replicate our baseline Table 1 using births and deaths occurring in the preceding ten years from the date of the survey. The results become slightly less significant but overall they remain qualitatively unchanged.[35]

Finally, we show that our results are robust to alternative clustering of standard errors.

[34]Results not provided and are available upon request.

[35]Both Baird et al. (2011) and Kudamatsu et al. (2012) resort to this strategy as well. See Table B.4 for alternative radii.

In all our regressions so far we have clustered standard errors at the survey- country level. In Table B.11 we cluster the standard errors at the ethnicity level instead. Our results remain unchanged.

5 Some Evidence on Channels: Information or Access?

Do linguistically distant mothers indeed have lower access to information or do they have lower access to public goods in general? Using the same sample of mothers from the DHS surveys, whose child mortality outcomes we have investigated in the previous sections, in this section we provide some evidence in favour of the former. In order to understand whether linguistic distance acts as a barrier to information we exploit the DHS question about whether the respondent has heard of the oral rehydration product (ORS) for treating children with diarrhoea. We create a 0-1 variable called ORS which takes the value 1 if the individual has either heard or used ORS, and 0 if the individual has never heard of ORS. Then we exploit the information on whether the household has access to electricity, how long it takes the household to get to a source of water and whether the individual is literate or not. These latter variables allow us to measure access to public goods.[36]

We present the results from the regressions in Table 6. The dependent variables in Panel 1 is the knowledge of ORS, in Panel 2 it is the access to electricity, in Panel 3 it is the access to water (time taken to get to a source of water) and in Panel 4, it is the literacy of the individual . Looking at the results from Panel 1, it does look like that those individuals who are linguistically more distant from their neighbours are less likely to have heard of ORS. On the other hand, linguistic distance has no effects on either access to electricity, water or literacy. Individuals living in more fractionalized neighbourhoods are more likely to have heard of ORS. While fractionalization has no robust effects on the other access variables. In Appendix Table B.12, we run similar regressions for different variables like immunization, antenatal visits etc. which might indicate lack of access to health care facilities. However, we find no evidence that linguistic distance reduces individual access to health services.

Thus, the results from this section do indicate that individuals who are linguistically distant to others living around them do have lower access to information which leads to higher rates of mortality for their children. On the other hand, lingusitically distant individuals do not necessarlily face lower access to public goods in general.

[36]For e.g. lower literacy among linguistically distant mothers might indicate lower access to schools.

Table 6: Information vs. Access

	(1)	(2)	(3)	(4)	(5)
Radius	25	50	75	100	125
ORS					
linguistic distance	-0.0323**	-0.0401**	-0.0407*	-0.0438*	-0.0486*
	(0.0149)	(0.0192)	(0.0225)	(0.0249)	(0.0282)
fractionalization	0.0321*	0.0374*	0.0381*	0.0489**	0.0631***
	(0.0184)	(0.0197)	(0.0202)	(0.0206)	(0.0223)
N	204684	204684	204684	204684	204684
R^2	0.158	0.158	0.158	0.159	0.159
Electricity					
linguistic distance	-0.00935	-0.00256	-0.00283	-0.00777	-0.00845
	(0.0341)	(0.0409)	(0.0487)	(0.0512)	(0.0520)
fractionalization	0.0295**	0.0190	0.0192	0.0255	0.0303
	(0.0138)	(0.0166)	(0.0194)	(0.0221)	(0.0240)
N	203812	203812	203812	203812	203812
R^2	0.533	0.533	0.533	0.533	0.533
Water					
linguistic distance	-18.39	-25.00	-34.08	-39.93	-40.74
	(16.91)	(19.62)	(22.84)	(26.12)	(28.74)
fractionalization	8.812	-8.097	-6.515	-0.587	8.373
	(10.68)	(11.40)	(14.10)	(19.62)	(26.25)
N	181402	181402	181402	181402	181402
R^2	0.208	0.208	0.208	0.208	0.208
Literacy					
linguistic distance	-0.0157	-0.0137	-0.0111	-0.0150	-0.0210
	(0.00948)	(0.0105)	(0.0109)	(0.0125)	(0.0135)
fractionalization	0.0115*	0.0141*	0.0160*	0.0231**	0.0367***
	(0.00598)	(0.00779)	(0.00904)	(0.0108)	(0.0127)
N	169547	169547	169547	169547	169547
R^2	0.681	0.681	0.681	0.681	0.681

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable in Panel 1 is a 0-1 variable indicating whether the mother has heard of the oral rehydration product (ORS) for treating children with diarrhoea, in Panel 2 it indicates whether the individual has access to electricity, in Panel 3 it is the time taken by the individual to get to a source of water and in Panel 4 it the whether the individual is literate or not.. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated. All the regressions control for location dummy, distance from capital, geographic isolation, years of education, wealth index dummy, religion dummy, ethnicity dummy, and year of birth dummy, and country dummy.

6 Conclusion

There exist huge disparities in child mortality rates across ethnic groups in Africa. The literature so far has not convincingly addressed the reasons behind such ethnic inequality in child mortality rates in Africa. Using high quality DHS data on individual level health outcomes and combining it with a novel dataset on the spatial distribution of ethnic groups at the 1 sq. km level, we estimate the effects individual level ethnic distance and diversity on individual level child health outcomes. The exact ethnic distance variable is calculated using the linguistic distance between individuals. We find that children of mothers who are ethnically distant from their neighbours have a higher probability of dying before reaching the age of five, whereas children of mothers living in more ethnically fractionalized places on the other hand have a lower probability of dying before reaching the age of five. Since we control for ethnicity specific fixed effects our results are not driven by heterogeneity in unobservable characteristics across ethnic groups including heterogeneity in health outcomes across ethnic groups or cultural differences between ethnic groups. To alleviate endogeneity concerns we control for a host of variables that reduce the possibilities of omitted variable bias. Then using recently developed methods by Altonji et al. (2005) we use the selection on the observables to gauge how strong the selection is on unobservables. We do not find evidence of a strong selection on unobservables.

While there are several possible explanations for our results, we follow Ashraf and Galor (2013) and argue that on the one hand, diversity implies a higher stock of knowledge and information, and thus leads to better health outcomes. On the other hand, such knowledge does not flow smoothly to groups which are linguistically distant and thus such groups lose out. One of the more crucial issues in child rearing and child deaths is access to information about how to care for the child. If the probability of survival of a child depended on the number of mistakes a mother makes in taking care of the child, and if the mother has less access to the information on how to take care of her child due to her linguistic distance, then she would be prone to committing more mistakes, resulting in a higher probability of her child not surviving. Exploiting the DHS question about whether the mother has heard of ORS, we show that linguistically distant mothers are more likely to have not heard of ORS, whereas they are not more likely to have lower access to other public goods in general.

Given the findings of our paper, a clear policy implication is to have more information campaigns about child health and health care in general particularly targeting linguistically distant minorities.

References

Ahlerup, P. (2009). The causal effects of ethnic diversity:an instrumental variables approach. *Working Papers in Economics No 386, September.*

Alesina, A., R. Bakir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics, MIT Press 114(4) November*, 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8, no. 2, June*, 155–194.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics, MIT Press 115(3) August*, 847–904.

Alesina, A. F., S. Michalopoulos, and E. Papaioannou (2012). Ethnic inequality. *NBER Working Paper No. 18512, November.*

Alesina, A. F. and E. Zhuravskaya (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review, American Economic Association vol. 101(5), August.*

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy Vol. 113, No. 1, February*, 151–184.

Ashraf, Q. and O. Galor (2013). The 'out of africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review Vol. 103 No. 1, February*, 1–46.

Baird, S., J. Friedman, and N. Schady (2011). Aggregate income shocks and infant mortality in the developing world. *Review of Economics and Statistics 93*(3), 847–856.

Baldwin, K. and J. D. Huber (2010). Ecnomic versus cultural differences: Forms of ethnic diversity and public good provision. *American Political Science Review 104, No. 4, November.*

Banerjee, A. and R. Somanathan (2007). The political economy of public goods: Some evidence from india. *Journal of Development Economics, Elsevier 82(2)*, 287–314.

Bellows, J. and E. Miguel (2009). War and local collective action in sierra leone. *Journal of Public Economics 93*, 1144–1157.

Besley, T. and M. Kudamatsu (2006). Health and democracy. *The American Economic Eeview 96*(2), 313–318.

Blattman, C. and E. Miguel (2010). Civil war. *The Journal of Economic Literature 48(1)*, 3–57(55).

Bowen, H. (2010). Information at the grassroots: Analyzing the media use and communication habits of kenyans to support effective development. *Africa Development Research Series, available at: http://www.audiencescapes.org/africa-research-survey-quantitative-analysis-ghana-kenya.*

Brockerhoff, M. and P. Hewett (2000). Inequality of child mortality among ethnic groups in sub-saharan africa. *Bulletin of the World Health Organization 78*(1), 30–41.

Caselli, F. and W. J. Coleman (2013). On the theory of ethnic conflict. *Journal of the European Economic Association January*.

Desmet, K., J. F. Gomes, and I. Ortuno-Ortin (2014). The geography of linguistic diversity and the provision of public goods. *Work in Progress*.

Desmet, K., I. Ortuno-Ortin, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of Development Economics 97*, 322–338.

Desmet, K., I. Ortuno-Ortin, and I. Weber (2009). Linguistic diversity and redistribution. *Journal of European Economic Association 7(6)*, 1291–1318.

Dyen, I., J. B. Kruskal, and P. Black (1992). An indo-european classification, a lexicostatistical experiment. 1. *Transactions of the American Philosophical Society 82*, 1–132.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics 112, no.4 November*, 1203–1250.

Esteban, J., L. Mayoral, and D. Ray (2012a). Ethnicity and conflict: An empirical study. *American Economic Review 102, No.4*, 1310–1342.

Esteban, J., L. Mayoral, and D. Ray (2012b). Ethnicity and conflict: Theory and facts. *Science 336, 858.*

Fearon, J. and D. Laitin (1996). Explaining interethnic cooperation. *American Political Science Review 90 (4)*, 715–35.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222.

Fisman, R., D. Paravisini, and V. Vig (2012). Cultural proximity and loan outcomes. Technical report, National Bureau of Economic Research.

Franck, R. and I. Rainer (2012). Does the leaders ethnicity matter? ethnic favoritism, education and health in sub-saharan africa. *American Political Science Review 106*(2, May).

Ghobarah, H. A., P. Huth, and B. Russett (2004). Comparative public health: The political economy of human misery and well-being. *International Studies Quarterly 48*(1), 73–94.

Glaeser, E. L., J. Scheinkman, and A. Shleifer (1995). Economic growth in a cross-section of cities. *Journal of monetary economics 36*(1), 117–143.

Glennerster, R., E. Miguel, and A. D. Rothenberg (2013). Collective action in diverse sierra leone communities. *The Economic Journal 123*(568), 285–316.

Gomes, J. F. (2012). The political economy of the maoist conflict in india: An empirical analysis. *UC3M Working Paper 12-18, Economic Series, June.*

Gomes, J. F. (2013). Religious diversity, intolerance and civil conflicts. *UC3M Working Paper 13-11, Economic Series, May.*

Gyimah, S. O. (2002). Ethnicity and infant mortality in sub-saharan africa: The case of ghana. *PSC Discussion Papers Series 16*(10), 1.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(04), 709–725.

Isphording, I. E. (2013). Disadvantages of linguistic origin: Evidence from immigrant literacy scores. Technical report, IZA Discussion Paper.

Isphording, I. E. and S. Otten (2013). The costs of babylonlinguistic distance in applied economics. *Review of International Economics 21*(2), 354–369.

Kudamatsu, M. (2009). *Ethnic Favoritism: Micro Evidence from Guinea.* unpublished.

Kudamatsu, M., T. Persson, and D. Strmberg (2012). Weather and infant mortality in africa. *Prepared for the conference on Climate and the Economy in Stockholm, September 5-8*.

Kumar, S., E. Dansereau, and C. Murray (2012). Does distance matter for institutional delivery in rural india: An instrumental variable approach. *Available at SSRN 2243709*.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny (1999). The quality of government. *Journal of Law, Economics, and Organization 15 (1)*, 222–79.

Laitin, D. and R. Ramachandran (2014). Language policy and economic development. *Available at http://www.econ.brown.edu/econ/events/laitin.pdf*.

Lieberman, E. S. (2007). Ethnic politics, risk, and policy-making a cross-national statistical analysis of government responses to hiv/aids. *Comparative Political Studies 40*(12), 1407–1432.

Malhotra, N. (2012). Inadequate feeding of infant and young children in india: lack of nutritional information or food affordability? *Public Health Nutrition 1*(1), 1–9.

McLeod, P. L., S. A. Lobel, and T. H. Cox (1996). Ethnic diversity and creativity in small groups. *Small group research 27*(2), 248–264.

Michalopoulos, S. and E. Papaioannou (2014). National institutions and subnational development in africa. *The Quarterly Journal of Economics 129*(1), 151–213.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of Public Economics 89*(11), 2325–2368.

Mitra, A. and D. Ray (2010). Implications of an economic thory of conflict: Hindu-muslim violence in india. *unpublished*.

Montalvo, J. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict and civil war. *The American Economic Review 95(3) June*, 796–816.

Montez, D. (2011). Identifying health information needs in tanzania evidence from the audiencescapes national survey. *AudienceScapes Development Research Report*.

Nunn, N. and L. Wantchekon (2009). The slave trade and the origins of mistrust in africa. *American Economic Review*.

Ottaviano, G. I. and G. Peri (2006). The economic value of cultural diversity: evidence from us cities. *Journal of Economic Geography 6*(1), 9–44.

Platas, M. R. (2010). Africa's health tragedy? ethnic diversity and health outcomes. *Prepared for delivery at the Winter 2010 Working Group on African Political Economy No. 7360, April*, December 17–18.

Pongou, R. (2009). Anonymity and infidelity: Ethnic identity, strategic cross-ethnic sexual network formation, and hiv/aids in africa. *Unpublished paper, Department of Economics, Brown University*.

Sethi, R. and R. Somanathan (2010). Caste hierarchies and social mobility in india. *Columbia University, document non publié*.

Shastry, G. K. (2012). Human capital response to globalization education and information technology in india. *Journal of Human Resources 47*(2), 287–330.

Singleton, K. and E. Krause (2009). Understanding cultural and linguistic barriers to health literacy. *OJIN: The online journal of issues in nursing 14*(3).

Spolaore, E. and R. Wacziarg (2009). The diffusion of development. *The Quarterly Journal of Economics 124*(2), 469–529.

UNICEF (2012). Levels trends in child mortality: Report 2012. *UN Inter-agency Group for Child Mortality Estimation, United Nations Childrens Fund, UNICEF*.

Vigdor, J. L. (2004). Community composition and collective action: Analyzing initial mail response to the 2000 census. *Review of Economics and Statistics 86*(1), 303–312.

Zhou, Y. (2010). Health information gaps in zambia evidence from the audiencescapes national survey. *Africa Development Research Series*.

A Figures

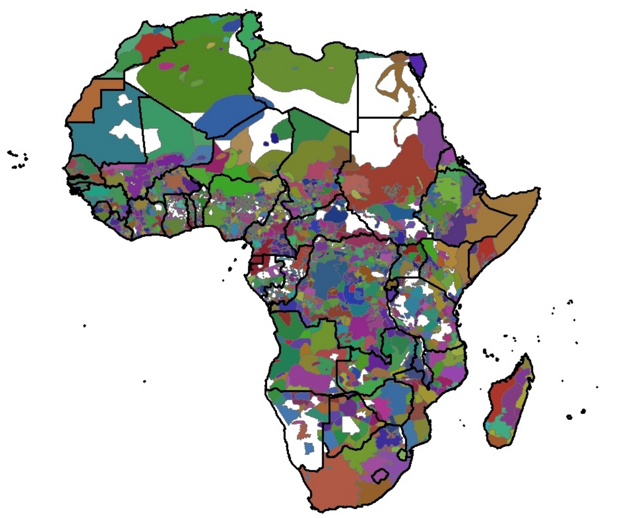

Figure A.1: African Languages distribution based on Ethnologue data. Each colour polygon represents a different language group. Mixed language areas cannot be seen in this map.

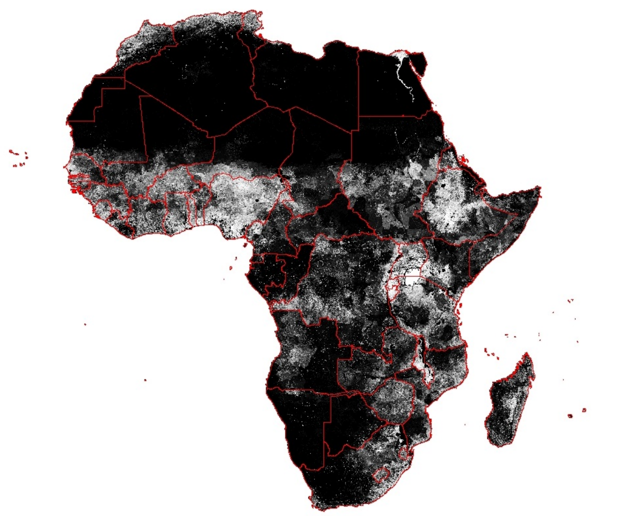

Figure A.2: Africa Population distribution based on Land Scan data. Each pixel roughly represents 1x1 sq. km of and darker the pixel the less populated the corresponding sq. km is.

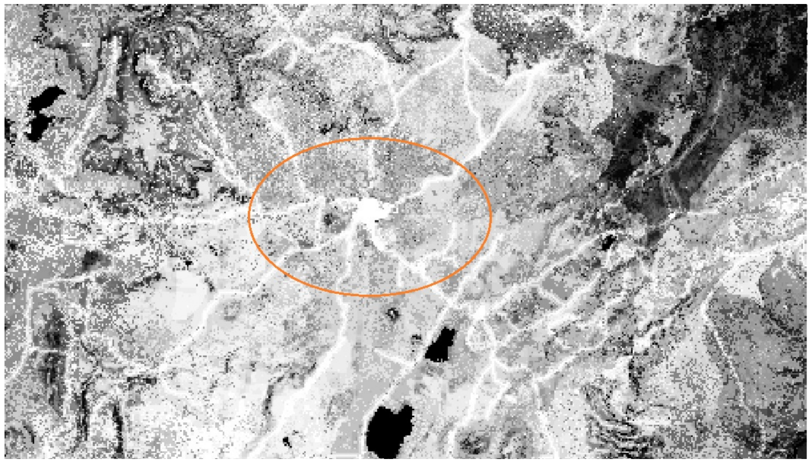

Figure A.3: Addis Abbaba Population from LandScan. The bright white area in the middle of the map encircled in red reprensents the area in and around Addis Abbaba.

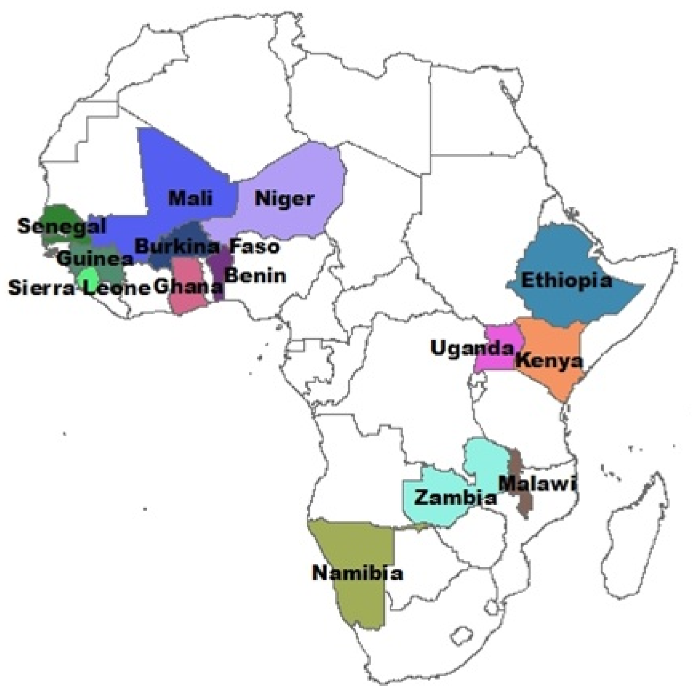

Figure A.4: The 14 Countries used in the study.

B Tables

Table B.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
child death	0.229	0.42	0	1	658755
infant death	0.12	0.325	0	1	821918
linguistic distance 25	0.121	0.266	0	1	868629
linguistic distance 50	0.125	0.265	0	1	868629
linguistic distance 75	0.128	0.266	0	1	868629
linguistic distance 100	0.131	0.267	0	1	868629
linguistic distance 125	0.133	0.27	0	1	868629
linguistic distance 150	0.135	0.271	0	1	868629
linguistic distance 175	0.138	0.272	0	1	868629
linguistic distance 200	0.14	0.274	0	1	868629
linguistic distance 250	0.145	0.276	0	1	868629
fractionalization district	0.399	0.289	0	0.997	868629
fractionalization 25	0.388	0.298	0	0.905	868761
fractionalization 50	0.467	0.284	0	0.921	868761
fractionalization 75	0.52	0.268	0	0.935	868761
fractionalization 100	0.556	0.252	0	0.929	868761
fractionalization 125	0.584	0.238	0	0.933	868761
fractionalization 150	0.608	0.225	0	0.935	868761
fractionalization 175	0.629	0.21	0	0.933	868761
fractionalization 200	0.648	0.195	0	0.933	868761
fractionalization 250	0.676	0.172	0.093	0.930	868761
polarization district	0.464	0.299	0	1	868629
urban	0.225	0.417	0	1	868629
female	0.49	0.5	0	1	868629
age at birth	24.991	6.422	8	50	868629
age at birth sqaured	665.786	348.351	64	2500	868629
multiple birth	0.032	0.177	0	1	868629
birth order	3.442	2.316	1	18	868629
birth order squared	17.216	22.533	1	324	868629
short birth space prior	0.209	0.407	0	1	868629
short birth space post	0.209	0.406	0	1	868629
education years	2.013	3.415	0	26	868306
wealth index	2.871	1.401	1	5	868629
birth year	1992.261	9.364	1955	2011	868629
log of distance to capital	5.123	1.217	-2.614	7.188	868629
log of geographic distance	5.656	0.39	4.662	7.028	868629

Table B.2: Summary statistics Child mortality

Variable	Mean	Std. Dev.	Obs
Benin	0.246	0.43	28,184
Burkina Faso	0.241	0.428	93,972
Ethiopia	0.215	0.411	95,686
Ghana	0.162	0.369	35,612
Guinea	0.265	0.442	37,924
Kenya	0.130	0.337	31,867
Malawi	0.228	0.420	103,901
Mali	0.310	0.462	96,365
Namibia	0.089	0.285	10,646
Niger	0.358	0.479	21,879
Senegal	0.168	0.374	54,651
Sierra Leone	0.226	0.419	13,780
Uganda	0.165	0.371	18,253
Zambia	0.185	0.389	15,038
All	0.229	0.420	657758

Table B.3: Child mortality: Baseline-last 10 years

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 75	0.0599	0.0598	0.0992*	0.0571*	0.0747*	0.0760*
	(0.0621)	(0.0653)	(0.0571)	(0.0313)	(0.0400)	(0.0421)
fractionalization 75		0.000408	-0.0608	-0.0101	-0.0243	-0.0233
		(0.0871)	(0.0744)	(0.0207)	(0.0209)	(0.0244)
urban			-0.0642*	-0.0451*	-0.0481**	-0.0474*
			(0.0366)	(0.0242)	(0.0243)	(0.0246)
female			-0.0675***	-0.0697***	-0.0701***	-0.0701***
			(0.00709)	(0.00655)	(0.00660)	(0.00660)
education_years			-0.0292***	-0.0179***	-0.0180***	-0.0180***
			(0.00569)	(0.00282)	(0.00279)	(0.00280)
wealth index 2			0.0587**	0.0321	0.0255	0.0253
			(0.0229)	(0.0204)	(0.0195)	(0.0195)
wealth index 3			0.0505**	0.0218	0.0128	0.0131
			(0.0246)	(0.0301)	(0.0284)	(0.0283)
wealth index 4			0.0229	-0.0142	-0.0219	-0.0218
			(0.0291)	(0.0329)	(0.0314)	(0.0313)
wealth index 5			-0.0459	-0.0941***	-0.106***	-0.105***
			(0.0423)	(0.0354)	(0.0326)	(0.0317)
lndist2cap						0.00699
						(0.0164)
ln_geog_dist						-0.0895*
						(0.0507)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	273822	273822	273706	264046	263730	263730
pseudo R^2	0.000	0.000	0.059	0.088	0.090	0.090

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.4: Child mortality: Alternative radii, last 10 years

	(1)	(2)	(3)	(4)	(5)	(6)
llinguistic distance 25	0.0505	0.0473	0.0921	0.0478	0.0393	0.0415
	(0.0610)	(0.0656)	(0.0584)	(0.0296)	(0.0333)	(0.0349)
fractionalization 25		0.0198	-0.0386	-0.0181	-0.0355	-0.0314
		(0.0608)	(0.0503)	(0.0217)	(0.0232)	(0.0235)
linguistic distance 50	0.0566	0.0545	0.0973*	0.0522*	0.0555	0.0575
	(0.0627)	(0.0666)	(0.0587)	(0.0298)	(0.0363)	(0.0377)
fractionalization 50		0.0184	-0.0459	-0.00561	-0.0204	-0.0165
		(0.0739)	(0.0627)	(0.0219)	(0.0249)	(0.0261)
linguistic distance 75	0.0599	0.0598	0.0992*	0.0571*	0.0747*	0.0760*
	(0.0621)	(0.0653)	(0.0571)	(0.0313)	(0.0400)	(0.0421)
fractionalization 75		0.000408	-0.0608	-0.0254	-0.0403	-0.0364
		(0.0871)	(0.0744)	(0.0220)	(0.0257)	(0.0283)
linguistic distance 100	0.0607	0.0615	0.0970*	0.0556*	0.0757*	0.0756*
	(0.0615)	(0.0641)	(0.0556)	(0.0316)	(0.0426)	(0.0440)
fractionalization 100		-0.0113	-0.0733	-0.0512**	-0.0681***	-0.0642**
		(0.102)	(0.0859)	(0.0233)	(0.0228)	(0.0269)
linguistic distance 125	0.0606	0.0618	0.0941*	0.0533*	0.0698	0.0675
	(0.0606)	(0.0628)	(0.0541)	(0.0312)	(0.0538)	(0.0541)
fractionalization 125		-0.0252	-0.0821	-0.0531*	-0.0713**	-0.0692**
		(0.118)	(0.0997)	(0.0308)	(0.0284)	(0.0351)
linguistic distance 150	0.0588	0.0599	0.0909*	0.0501	0.0627	0.0584
	(0.0600)	(0.0620)	(0.0530)	(0.0313)	(0.0735)	(0.0726)
fractionalization 150		-0.0289	-0.0857	-0.0211	-0.0458	-0.0455
		(0.139)	(0.116)	(0.0403)	(0.0353)	(0.0437)
linguistic distance 175	0.0575	0.0589	0.0890*	0.0491	0.0755	0.0681
	(0.0596)	(0.0613)	(0.0520)	(0.0313)	(0.0943)	(0.0930)
fractionalization 175		-0.0378	-0.0932	-0.00672	-0.0384	-0.0376
		(0.163)	(0.135)	(0.0516)	(0.0427)	(0.0506)
linguistic distance 200	0.0570	0.0584	0.0877*	0.0469	0.0762	0.0655
	(0.0596)	(0.0611)	(0.0515)	(0.0310)	(0.112)	(0.111)
fractionalization 200		-0.0456	-0.105	0.0233	-0.0133	-0.0105
		(0.190)	(0.157)	(0.0635)	(0.0499)	(0.0570)
linguistic distance 250	0.0567	0.0573	0.0846*	0.0521*	0.116	0.0982
	(0.0609)	(0.0606)	(0.0507)	(0.0310)	(0.137)	(0.136)
fractionalization 250		-0.0807	-0.163	-0.0696	-0.113	-0.109
		(0.244)	(0.198)	(0.106)	(0.0905)	(0.0874)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	273822	273822	273706	264046	263730	263730

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.5: Child mortality: Fractionalization in different radii

	(1)	(2)	(3)	(4)	(5)
fractionalization 25	0.00726	-0.0380	-0.0228	-0.0346**	-0.0328**
	(0.0484)	(0.0418)	(0.0189)	(0.0164)	(0.0158)
fractionalization 50	0.00762	-0.0453	-0.0174	-0.0288	-0.0284
	(0.0610)	(0.0534)	(0.0215)	(0.0189)	(0.0192)
fractionalization 75	-0.00697	-0.0571	-0.0271	-0.0385*	-0.0397*
	(0.0729)	(0.0636)	(0.0258)	(0.0226)	(0.0235)
fractionalization 100	-0.0187	-0.0687	-0.0373	-0.0498**	-0.0527**
	(0.0858)	(0.0732)	(0.0239)	(0.0204)	(0.0219)
fractionalization 125	-0.0386	-0.0821	-0.0407	-0.0546**	-0.0601**
	(0.0999)	(0.0845)	(0.0282)	(0.0235)	(0.0269)
fractionalization 150	-0.0506	-0.0923	-0.0222	-0.0417	-0.0497
	(0.117)	(0.0979)	(0.0364)	(0.0305)	(0.0355)
fractionalization 175	-0.0664	-0.104	-0.0101	-0.0357	-0.0438
	(0.137)	(0.114)	(0.0446)	(0.0374)	(0.0426)
fractionalization 200	-0.0804	-0.119	0.00924	-0.0244	-0.0320
	(0.159)	(0.131)	(0.0539)	(0.0445)	(0.0486)
fractionalization 250	-0.121	-0.181	-0.0852	-0.138*	-0.144**
	(0.203)	(0.163)	(0.0813)	(0.0706)	(0.0676)
fractionalization district	-0.00129	-0.0464	-0.0404	-0.0560**	-0.0534**
	(0.0445)	(0.0408)	(0.0258)	(0.0244)	(0.0238)
polarization district	0.0662	0.00681	-0.0296	-0.0372*	-0.0355*
	(0.0494)	(0.0414)	(0.0235)	(0.0210)	(0.0205)
N	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.6: Child mortality: Polarization

	(1)	(2)	(3)	(4)	(5)
linguistic distance 25	0.0359	0.0755	0.0219	0.0570***	0.0577***
	(0.0559)	(0.0466)	(0.0248)	(0.0213)	(0.0214)
polarization district	0.0625	-0.00109	-0.0314	-0.0393*	-0.0376*
	(0.0511)	(0.0429)	(0.0233)	(0.0210)	(0.0204)
linguistic distance 50	0.0399	0.0793*	0.0262	0.0805***	0.0809***
	(0.0576)	(0.0475)	(0.0264)	(0.0256)	(0.0257)
polarization district	0.0622	-0.00128	-0.0316	-0.0396*	-0.0379*
	(0.0511)	(0.0428)	(0.0233)	(0.0210)	(0.0204)
linguistic distance 75	0.0428	0.0808*	0.0278	0.0939***	0.0940***
	(0.0580)	(0.0474)	(0.0271)	(0.0321)	(0.0321)
polarization district	0.0621	-0.00114	-0.0314	-0.0393*	-0.0375*
	(0.0510)	(0.0427)	(0.0233)	(0.0211)	(0.0205)
linguistic distance 100	0.0449	0.0800*	0.0262	0.0979***	0.0972***
	(0.0582)	(0.0473)	(0.0274)	(0.0371)	(0.0366)
polarization district	0.0621	-0.000780	-0.0312	-0.0388*	-0.0371*
	(0.0509)	(0.0426)	(0.0234)	(0.0211)	(0.0206)
linguistic distance 125	0.0458	0.0784*	0.0254	0.103**	0.101**
	(0.0577)	(0.0468)	(0.0271)	(0.0476)	(0.0466)
polarization district	0.0621	-0.000492	-0.0311	-0.0387*	-0.0369*
	(0.0508)	(0.0426)	(0.0234)	(0.0211)	(0.0206)
linguistic distance 150	0.0444	0.0759	0.0232	0.0971	0.0934
	(0.0579)	(0.0469)	(0.0271)	(0.0625)	(0.0609)
polarization district	0.0623	-0.000157	-0.0309	-0.0385*	-0.0368*
	(0.0508)	(0.0425)	(0.0234)	(0.0212)	(0.0206)
linguistic distance 175	0.0431	0.0739	0.0220	0.101	0.0956
	(0.0581)	(0.0469)	(0.0270)	(0.0781)	(0.0762)
polarization district	0.0625	0.000135	-0.0308	-0.0383*	-0.0366*
	(0.0507)	(0.0425)	(0.0233)	(0.0211)	(0.0206)
linguistic distance 200	0.0427	0.0728	0.0210	0.102	0.0953
	(0.0586)	(0.0474)	(0.0268)	(0.0902)	(0.0881)
polarization district	0.0626	0.000367	-0.0307	-0.0380*	-0.0363*
	(0.0507)	(0.0424)	(0.0233)	(0.0210)	(0.0205)
linguistic distance 250	0.0419	0.0700	0.0209	0.0930	0.0826
	(0.0606)	(0.0494)	(0.0268)	(0.109)	(0.107)
polarization district	0.0629	0.000998	-0.0306	-0.0374*	-0.0358*
	(0.0506)	(0.0423)	(0.0233)	(0.0209)	(0.0204)
N	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after the linguistic distance variable indicate the radius of the circle around the mother in which these variables have been calculated. Polarization is calculated at the district level.

Table B.7: Child mortality: Fractionalization & Polarization

	(1)	(2)	(3)	(4)	(5)
fractionalization district	-0.173*	-0.156**	-0.0401	-0.0620**	-0.0589**
	(0.0950)	(0.0783)	(0.0299)	(0.0286)	(0.0283)
polarization district	0.202**	0.129	-0.000286	0.00716	0.00655
	(0.102)	(0.0795)	(0.0258)	(0.0219)	(0.0217)
urban		-0.0905***	-0.0672***	-0.0691***	-0.0685***
		(0.0325)	(0.0160)	(0.0153)	(0.0152)
female		-0.0655***	-0.0675***	-0.0679***	-0.0679***
		(0.00433)	(0.00429)	(0.00430)	(0.00430)
education_years		-0.0351***	-0.0219***	-0.0210***	-0.0209***
		(0.00514)	(0.00254)	(0.00248)	(0.00247)
_Iwealth_in_2		0.0296	0.0102	0.00768	0.00739
		(0.0188)	(0.0143)	(0.0140)	(0.0139)
_Iwealth_in_3		0.0125	-0.00799	-0.0107	-0.0106
		(0.0211)	(0.0218)	(0.0208)	(0.0209)
_Iwealth_in_4		-0.0238	-0.0569**	-0.0584**	-0.0583**
		(0.0252)	(0.0257)	(0.0245)	(0.0245)
_Iwealth_in_5		-0.101***	-0.156***	-0.156***	-0.155***
		(0.0377)	(0.0278)	(0.0256)	(0.0255)
lndist2cap					0.00800
					(0.00987)
ln_geog_dist					-0.0738
					(0.0512)
N	658755	658505	649182	648468	648468
pseudo R^2	0.001	0.062	0.084	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level.

Table B.8: Child mortality: Fractionalization & Polarization with distance

	(1)	(2)	(3)	(4)	(5)
linguistic distance 25	0.0502	0.0893*	0.0258	0.0615***	0.0619***
	(0.0576)	(0.0476)	(0.0257)	(0.0215)	(0.0217)
fractionalization district	-0.183*	-0.174**	-0.0454	-0.0651**	-0.0621**
	(0.0973)	(0.0795)	(0.0284)	(0.0288)	(0.0285)
polarization district	0.204**	0.134*	0.00146	0.00716	0.00654
	(0.102)	(0.0788)	(0.0254)	(0.0214)	(0.0212)
linguistic distance 50	0.0538	0.0927*	0.0301	0.0859***	0.0859***
	(0.0592)	(0.0485)	(0.0275)	(0.0265)	(0.0266)
fractionalization district	-0.183*	-0.174**	-0.0459	-0.0658**	-0.0627**
	(0.0970)	(0.0793)	(0.0285)	(0.0287)	(0.0284)
polarization district	0.205**	0.134*	0.00169	0.00737	0.00676
	(0.102)	(0.0787)	(0.0255)	(0.0213)	(0.0211)
linguistic distance 75	0.0561	0.0936*	0.0314	0.0997***	0.0995***
	(0.0593)	(0.0483)	(0.0282)	(0.0332)	(0.0332)
fractionalization district	-0.183*	-0.173**	-0.0456	-0.0659**	-0.0628**
	(0.0966)	(0.0790)	(0.0286)	(0.0286)	(0.0283)
polarization district	0.205**	0.134*	0.00166	0.00780	0.00718
	(0.102)	(0.0787)	(0.0255)	(0.0213)	(0.0212)
linguistic distance 100	0.0577	0.0922*	0.0294	0.104***	0.103***
	(0.0592)	(0.0479)	(0.0284)	(0.0386)	(0.0381)
fractionalization district	-0.183*	-0.173**	-0.0449	-0.0655**	-0.0624**
	(0.0964)	(0.0788)	(0.0286)	(0.0288)	(0.0285)
polarization district	0.205**	0.134*	0.00141	0.00802	0.00740
	(0.102)	(0.0787)	(0.0255)	(0.0215)	(0.0213)
linguistic distance 125	0.0580	0.0900*	0.0284	0.109**	0.107**
	(0.0584)	(0.0471)	(0.0280)	(0.0487)	(0.0477)
fractionalization district	-0.183*	-0.172**	-0.0443	-0.0650**	-0.0620**
	(0.0961)	(0.0786)	(0.0285)	(0.0288)	(0.0285)
polarization district	0.204**	0.133*	0.00115	0.00785	0.00724
	(0.102)	(0.0788)	(0.0255)	(0.0215)	(0.0214)
linguistic distance 150	0.0559	0.0868*	0.0259	0.103	0.0988
	(0.0584)	(0.0470)	(0.0279)	(0.0631)	(0.0616)
fractionalization district	-0.182*	-0.171**	-0.0437	-0.0643**	-0.0612**
	(0.0959)	(0.0783)	(0.0286)	(0.0289)	(0.0286)
polarization district	0.204**	0.133*	0.000861	0.00749	0.00688
	(0.102)	(0.0788)	(0.0256)	(0.0217)	(0.0215)
linguistic distance 175	0.0541	0.0843*	0.0244	0.105	0.0999
	(0.0585)	(0.0469)	(0.0278)	(0.0779)	(0.0761)
fractionalization district	-0.181*	-0.170**	-0.0432	-0.0636**	-0.0605**
	(0.0957)	(0.0782)	(0.0285)	(0.0287)	(0.0284)
polarization district	0.204**	0.132*	0.000669	0.00721	0.00661
	(0.102)	(0.0788)	(0.0256)	(0.0217)	(0.0216)
linguistic distance 200	0.0532	0.0827*	0.0233	0.105	0.0981
	(0.0588)	(0.0472)	(0.0275)	(0.0894)	(0.0875)
fractionalization district	-0.181*	-0.169**	-0.0429	-0.0628**	-0.0598**
	(0.0955)	(0.0780)	(0.0285)	(0.0286)	(0.0283)
polarization district	0.204**	0.132*	0.000551	0.00697	0.00639
	(0.102)	(0.0787)	(0.0256)	(0.0217)	(0.0216)
linguistic distance 250	0.0518	0.0793	0.0229	0.0919	0.0820
	(0.0605)	(0.0490)	(0.0275)	(0.107)	(0.104)
fractionalization district	-0.181*	-0.168**	-0.0427	-0.0617**	-0.0588**
	(0.0954)	(0.0778)	(0.0284)	(0.0283)	(0.0280)
polarization district	0.204**	0.132*	0.000526	0.00678	0.00623
	(0.102)	(0.0787)	(0.0256)	(0.0218)	(0.0216)
N	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.9: Child mortality: Linear Probability Model

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
linguistic distance 25	0.0134	0.0133	0.0233*	0.00576	0.0137**	0.0140**
	(0.0165)	(0.0177)	(0.0135)	(0.00677)	(0.00539)	(0.00544)
fractionalization 25		0.000458	-0.0139	-0.00780	-0.0107**	-0.0102**
		(0.0155)	(0.0125)	(0.00520)	(0.00477)	(0.00460)
linguistic distance 50	0.0145	0.0144	0.0238*	0.00647	0.0188***	0.0191***
	(0.0171)	(0.0180)	(0.0137)	(0.00702)	(0.00667)	(0.00672)
fractionalization 50		0.000790	-0.0155	-0.00637	-0.00959*	-0.00935*
		(0.0190)	(0.0156)	(0.00560)	(0.00541)	(0.00550)
linguistic distance 75	0.0153	0.0156	0.0239*	0.00691	0.0221**	0.0224**
	(0.0173)	(0.0179)	(0.0136)	(0.00712)	(0.00856)	(0.00862)
fractionalization 75		-0.00345	-0.0186	-0.00942	-0.0132*	-0.0134*
		(0.0224)	(0.0184)	(0.00694)	(0.00665)	(0.00690)
linguistic distance 100	0.0159	0.0163	0.0234*	0.00652	0.0232**	0.0233**
	(0.0174)	(0.0178)	(0.0135)	(0.00721)	(0.0101)	(0.0100)
fractionalization 100		-0.00685	-0.0217	-0.0124*	-0.0170**	-0.0176**
		(0.0263)	(0.0212)	(0.00689)	(0.00623)	(0.00661)
linguistic distance 125	0.0161	0.0167	0.0228*	0.00633	0.0241*	0.0239*
	(0.0173)	(0.0176)	(0.0132)	(0.00716)	(0.0123)	(0.0120)
fractionalization 125		-0.0128	-0.0255	-0.0135	-0.0186**	-0.0198**
		(0.0306)	(0.0246)	(0.00827)	(0.00708)	(0.00794)
linguistic distance 150	0.0156	0.0162	0.0220	0.00563	0.0211	0.0207
	(0.0173)	(0.0175)	(0.0132)	(0.00722)	(0.0162)	(0.0158)
fractionalization 150		-0.0164	-0.0285	-0.00870	-0.0149	-0.0169
		(0.0358)	(0.0286)	(0.0107)	(0.00907)	(0.0104)
linguistic distance 175	0.0152	0.0159	0.0215	0.00523	0.0213	0.0203
	(0.0174)	(0.0175)	(0.0131)	(0.00725)	(0.0200)	(0.0195)
fractionalization 175		-0.0213	-0.0319	-0.00564	-0.0134	-0.0154
		(0.0419)	(0.0333)	(0.0131)	(0.0107)	(0.0122)
linguistic distance 200	0.0150	0.0158	0.0212	0.00477	0.0212	0.0196
	(0.0176)	(0.0176)	(0.0132)	(0.00721)	(0.0228)	(0.0222)
fractionalization 200		-0.0256	-0.0363	-0.000227	-0.0100	-0.0119
		(0.0489)	(0.0386)	(0.0159)	(0.0127)	(0.0139)
linguistic distance 250	0.0145	0.0150	0.0199	0.00582	0.0238	0.0208
	(0.0183)	(0.0178)	(0.0134)	(0.00722)	(0.0272)	(0.0264)
fractionalization 250		-0.0377	-0.0530	-0.0268	-0.0419**	-0.0432**
		(0.0626)	(0.0483)	(0.0243)	(0.0202)	(0.0193)
N	658755	658755	658505	658505	657795	657795

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. OLS regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.10: Child mortality: Non-linearities

	(1)	(2)	(3)	(4)	(5)
linguistic distance 25	0.0452	0.164		0.0613***	0.178
	(0.102)	(0.116)		(0.0219)	(0.113)
linguistic distance 25_2	0.00779	-0.110			-0.125
	(0.108)	(0.120)			(0.117)
fractionalization 25		-0.0458***	0.0320	0.0286	0.0244
		(0.0169)	(0.0537)	(0.0528)	(0.0533)
fractionalization 25_2			-0.0849	-0.0858	-0.0936
			(0.0697)	(0.0682)	(0.0652)
linguistic distance 50	0.174	0.273**		0.0849***	0.298**
	(0.110)	(0.130)		(0.0275)	(0.132)
linguistic distance 50_2	-0.109	-0.206			-0.230*
	(0.113)	(0.130)			(0.132)
fractionalization 50		-0.0483**	0.145**	0.142**	0.133**
		(0.0217)	(0.0622)	(0.0606)	(0.0587)
fractionalization 50_2			-0.215***	-0.218***	-0.227***
			(0.0755)	(0.0736)	(0.0737)
linguistic distance 75	0.132	0.225*		0.105***	0.240*
	(0.111)	(0.128)		(0.0347)	(0.128)
linguistic distance 75_2	-0.0474	-0.135			-0.145
	(0.119)	(0.131)			(0.131)
fractionalization 75		-0.0554**	0.218***	0.216***	0.208***
		(0.0258)	(0.0657)	(0.0632)	(0.0620)
fractionalization 75_2			-0.305***	-0.311***	-0.313***
			(0.0833)	(0.0807)	(0.0807)
linguistic distance 100	0.115	0.217*		0.109***	0.227*
	(0.107)	(0.123)		(0.0385)	(0.124)
linguistic distance 100_2	-0.0248	-0.119			-0.126
	(0.115)	(0.126)			(0.127)
fractionalization 100		-0.0679***	0.177**	0.173**	0.167**
		(0.0244)	(0.0742)	(0.0722)	(0.0710)
fractionalization 100_2			-0.258***	-0.263***	-0.264***
			(0.0900)	(0.0884)	(0.0879)
linguistic distance 125	-0.0122	0.0919		0.111**	0.0968
	(0.0962)	(0.119)		(0.0471)	(0.120)
linguistic distance 125_2	0.116	0.0201			0.0146
	(0.102)	(0.118)			(0.120)
fractionalization 125		-0.0679**	0.0998	0.0901	0.0908
		(0.0299)	(0.0844)	(0.0827)	(0.0823)
fractionalization 125_2			-0.171*	-0.171*	-0.170*
			(0.0932)	(0.0915)	(0.0918)
N	648468	648468	648468	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level.

Table B.11: Child mortality: Alternative CLustering of SE

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
linguistic distance 25	0.0439	0.0437	0.0833**	0.0242	0.0605**	0.0610**
	(0.0478)	(0.0488)	(0.0413)	(0.0270)	(0.0251)	(0.0247)
fractionalization 25		0.00163	-0.0490	-0.0270	-0.0386**	-0.0369*
		(0.0425)	(0.0353)	(0.0170)	(0.0190)	(0.0193)
linguistic distance 50	0.0476	0.0473	0.0856**	0.0264	0.0823**	0.0828**
	(0.0486)	(0.0492)	(0.0416)	(0.0279)	(0.0354)	(0.0346)
fractionalization 50		0.00275	-0.0544	-0.0219	-0.0343	-0.0339
		(0.0479)	(0.0406)	(0.0207)	(0.0225)	(0.0230)
linguistic distance 75	0.0502	0.0511	0.0862**	0.0280	0.0995**	0.100**
	(0.0490)	(0.0495)	(0.0419)	(0.0291)	(0.0459)	(0.0447)
fractionalization 75		-0.0113	-0.0646	-0.0313	-0.0461*	-0.0473*
		(0.0545)	(0.0461)	(0.0257)	(0.0265)	(0.0273)
linguistic distance 100	0.0519	0.0533	0.0846**	0.0262	0.106**	0.106**
	(0.0492)	(0.0497)	(0.0421)	(0.0295)	(0.0520)	(0.0506)
fractionalization 100		-0.0225	-0.0748	-0.0406	-0.0582**	-0.0611**
		(0.0624)	(0.0522)	(0.0274)	(0.0290)	(0.0307)
linguistic distance 125	0.0526	0.0545	0.0826*	0.0253	0.112*	0.111*
	(0.0494)	(0.0502)	(0.0423)	(0.0299)	(0.0594)	(0.0575)
fractionalization 125		-0.0419	-0.0873	-0.0441	-0.0638*	-0.0692*
		(0.0705)	(0.0591)	(0.0332)	(0.0344)	(0.0379)
linguistic distance 150	0.0510	0.0530	0.0798*	0.0224	0.102	0.0997
	(0.0499)	(0.0508)	(0.0427)	(0.0305)	(0.0742)	(0.0714)
fractionalization 150		-0.0537	-0.0971	-0.0262	-0.0504	-0.0582
		(0.0806)	(0.0676)	(0.0408)	(0.0420)	(0.0465)
linguistic distance 175	0.0496	0.0520	0.0779*	0.0207	0.104	0.100
	(0.0504)	(0.0512)	(0.0428)	(0.0308)	(0.0878)	(0.0843)
fractionalization 175		-0.0695	-0.109	-0.0152	-0.0448	-0.0525
		(0.0927)	(0.0780)	(0.0461)	(0.0480)	(0.0527)
linguistic distance 200	0.0489	0.0515	0.0768*	0.0190	0.104	0.0979
	(0.0508)	(0.0516)	(0.0427)	(0.0309)	(0.0999)	(0.0955)
fractionalization 200		-0.0835	-0.124	0.00314	-0.0341	-0.0410
		(0.107)	(0.0900)	(0.0509)	(0.0524)	(0.0556)
linguistic distance 250	0.0476	0.0487	0.0720*	0.0226	0.112	0.102
	(0.0516)	(0.0522)	(0.0423)	(0.0309)	(0.121)	(0.115)
fractionalization 250		-0.123	-0.183*	-0.0939	-0.149**	-0.154**
		(0.136)	(0.111)	(0.0657)	(0.0679)	(0.0657)
N	658755	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the ethnicity level. The numbers after linguistic distance and fractionalizationtionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table B.12: Some Other Variables- Access

Linguistic Distance	25	50	75	100	125
Full Immunization					
Linguistic Distance	-0.00131	0.00177	0.00321	0.00203	0.000277
	(0.00262)	(0.00273)	(0.00325)	(0.00313)	(0.00354)
(0.00411)					
Fractionalization	0.000467	0.00146	0.000573	-0.000590	-0.00126
	(0.00377)	(0.00421)	(0.00422)	(0.00424)	(0.00545)
N	867376	867376	867376	867376	867376
BCG					
Linguistic Distance	-0.000168	0.00120	0.00308	0.00332	0.00325
	(0.00459)	(0.00443)	(0.00541)	(0.00592)	(0.00692)
Fractionalization	0.00288	0.00494	0.00379	0.00354	
	(0.00494)	(0.00599)	(0.00628)	(0.00617)	(0.00678)
N	849344	849344	849344	849344	849344
Polio All					
Linguistic Distance	-0.000268	0.00361	0.00577	0.00476	0.00296
	(0.00352)	(0.00365)	(0.00425)	(0.00416)	(0.00488)
Fractionalization	0.000105	0.000347	-0.00119	-0.00269	-0.00345
	(0.00426)	(0.00491)	(0.00513)	(0.00518)	(0.00635)
N	867376	867376	867376	867376	867376
Measles					
Linguistic Distance	0.0000601	0.00295	0.00307	0.00215	0.00142
	(0.00398)	(0.00388)	(0.00446)	(0.00458)	(0.00502)
Fractionalization	0.00334	0.00582	0.00501	0.00498	0.00331
	(0.00385)	(0.00452)	(0.00455)	(0.00456)	(0.00529)
N	849344	849344	849344	849344	849344
Skilled Delivery					
Linguistic Distance	0.00225	0.00658	0.00453	0.00313	0.000838
	(0.00493)	(0.00602)	(0.00620)	(0.00649)	(0.00731)
Fractionalization	0.00640	0.00832	0.00867	0.00685	0.00544
	(0.00526)	(0.00644)	(0.00563)	(0.00478)	(0.00497)
N	867376	867376	867376	867376	867376
Tetanus					
Linguistic Distance	0.00620	0.00720	0.00695	0.00601	0.00410
	(0.00508)	(0.00504)	(0.00506)	(0.00584)	(0.00730)
Fractionalization	0.00589	0.0126**	0.0178**	0.0175**	0.0144**
	(0.00390)	(0.00475)	(0.00665)	(0.00714)	(0.00617)
(0.00713)					
N	867376	867376	867376	867376	867376
Antenatal Vists					
Linguistic Distance	0.00495	0.00587	0.00439	0.00440	0.00343
	(0.00527)	(0.00479)	(0.00474)	(0.00478)	(0.00627)
Fractionalization	0.00637	0.0109**	0.0149**	0.0130*	0.00893
	(0.00397)	(0.00503)	(0.00670)	(0.00716)	(0.00740)
N	867376	867376	867376	867376	867376
adj. R^2	0.430	0.430	0.430	0.430	0.430
Iron Tablets					
Linguistic Distance	0.00536*	0.00123	-0.00233	-0.00160	0.00114
	(0.00312)	(0.00297)	(0.00303)	(0.00326)	(0.00363)
Fractionalization	0.00347	0.00258	0.00403	0.00666	0.0115**
	(0.00474)	(0.00507)	(0.00533)	(0.00428)	(0.00452)
N	702515	702515	702515	702515	702515

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. All the regressions control for urban location dummy, years of education, wealth index dummy, religion dummy, ethnicity dummy, and year of birth dummy, etc.