Is Language a Bridge or Barrier? Impact of linguistic distance on woman and child health in India

Advaith Jayakumar[a] & Anisha Sharma[b]

September 15, 2023

Abstract

Using National Family Health Survey – 4 (2015-16) and 5 (2019-21) data, and the language tree on Ethnologue, this paper aims to understand the impact of speaking in an unfamiliar language as a barrier to health outcomes among mothers and their children in India. We employ the concept of linguistic distance between the mother tongue and the dominant language of the region to capture the distinctness between languages. We find evidence that the growing linguistic distance (i.e. the more distinct languages are, the bigger the language barrier becomes) between the mother tongue and the dominant language of the woman results in poorer health outcomes including morbidities and reduced health-seeking behaviour among women/mothers and this has a spillover effect on their children in terms of their under-5 immunisation status. We find evidence of reduced health-seeking behaviour and exposure to health information, and decreased autonomy among women as mechanisms that lead to worsened health outcomes. We also explore heterogeneity across wealth and length of migration.

1 Introduction

Language is the cornerstone of all social and interpersonal relations. The evolution of languages over time has resulted in considerable variation in their structure and composition, even within the same geographic regions (Bromham et al., 2015). This variation can generate friction in multiple economic markets as learning a new language is costly and dependent on innate ability. On the other hand, investments in language skills can reap large benefits in terms of improved social integration (Chiswick, 2008), (Ginsburgh and Weber, 2020).

For an individual, the cost of acquiring fluency in a new language depends on the distinctness of this language from the native language spoken by the individual. Rising costs of acquiring languages affect an individual's social and economic participation in the local economy, as well as their ability to gain access to public services, that are often conditional on communication in the dominant language of the region. This can impose significant economic costs on migrants in particular, whose native language typically varies from the dominant language of the place they reside.

[a]Independent Researcher
[b]Ashoka University

In this paper, we examine the consequences of this friction on access to healthcare, and consequently on health outcomes. Our study is set in India, a large developing country with considerable linguistic diversity across 22 officially sanctioned languages, and thousands of dialects in those languages. The official languages are the languages used by the government and in providing public services and access to services usually requires some fluency in the local official language. Proficiency in the local official language enables people to get access to the labour market, education and healthcare (Laitin and Ramachandran, 2016). Given the scale of internal migration in India (Tumbe, 2018) with over 51 million rural-urban migrants (Census 2011), and the diversity in languages across the country, it is likely that a migrant's mother tongue is different from the dominant language of the region, imposing costs on them in terms of human capital outcomes.

Using data from two rounds of a nationally representative survey, the National Family Health Survey, and data on the distinctness between languages from Ethnologue (Eberhard et al., 2023), we estimate the effect of increasing costs of acquiring a local dominant language on observed health outcomes and health-seeking behaviour. To quantify the costs of acquiring a language, we use a measure of 'Linguistic Distance' developed by (Fearon, 2003). The greater the linguistic distance between any two languages, the greater the cost of learning the language. We find evidence that increasing linguistic distance between a woman's native language and the dominant language of the district she resides in results in poorer health outcomes for her. We focus on health outcomes which are likely to be responsive to receiving information on prevention and treatment from health services (for example, anaemia and blood sugar). Our specification controls for a range of household and mother-level characteristics and district fixed effects, and is robust to the use of matching techniques so as to compare similar households that vary only in their linguistic distance from the dominant language of the region. This is particularly important to identification since our analysis effectively compares migrants to non-migrants, which is also affected by differential selection into migration.

We find that a one-unit increase in linguistic distance leads to a 1-2.5 percentage point increase in the probability of having either morbidity. We also find evidence of reduced access to healthcare services for children: again, a one-unit increase in linguistic distance is associated with a reduced probability of receiving a vaccination by 1.5-2.5 percentage points. Our results are robust to using three different measures of linguistic distance. In terms of mechanisms, we find evidence of reduced health-seeking behaviour by women, reduced exposure to public health information, as well as reduced autonomy of women to go outside the house and visit health facilities on their own – all of which increase with linguistic distance. Additionally, we explore heterogeneous treatment effects by wealth and length of residence in a district.

This paper contributes to the literature on the adverse effects of linguistic barriers on human capital outcomes. A large literature has identified the negative effects of distance between the native language of a child and the language of instruction in school on educational attainment (see Ginsburgh and Weber (2020) for a review). (Kumar et al., 2020). For health outcomes, the evidence is somewhat mixed: some studies find a negative association between dominant language fluency and health outcomes (Ponce et al., 2006; Schachter et al., 2012; Pottie et al., 2008; Nguyen and Reardon, 2013), particularly for women (Guven and Islam, 2015), while others find no impact at all (Aoki and Santiago, 2018). Much of this literature uses variation in linguistic distance across immigrants to developed countries from the dominant languages spoken there. In the developing country context, Laitin and Ramachandran (2016) use microdata from India

to find that linguistic distance from the language of government reduces awareness about health-improving behaviours such as the use of bed nets to protect against malaria, and awareness of AIDS. Gomes (2020) finds that in sub-Saharan Africa, increased linguistic distance from one's neighbours is associated with higher mortality and malnutrition for children. Our study also uses microdata from a large country and studies a wider range of health outcomes and health-seeking behaviours.

We are also able to provide clear evidence on the mechanisms that likely explain our results. Prior research has shown that linguistic distance between healthcare providers and patients can lead to reduced trust and the quality of healthcare received (Street and Haidet, 2011). Language barriers also make it less likely that individuals, especially mothers, receive important information about healthcare (Laitin and Ramachandran, 2016; Gomes, 2020). We too find support for the hypothesis that linguistic distance is correlated with less exposure to media and information for mothers, and we also provide evidence that this leads to lower health-seeking behaviour by these mothers. We also show that language barriers are correlated with reduced autonomy of women and willingness to seek out health care outside the home, which compounds the difficulties in accessing information.

Section 2 describes different methods of computing linguistic distance. Section 3 describes the data and provides descriptive statistics. Section 4 discusses the estimating strategy and Section 5 presents the results. Section 6 discusses mechanisms that explain our results, Section 7 provides further results and a number of robustness checks. Section 8 concludes.

2 Measuring linguistic distance

To proxy the costs incurred by a native speaker of language A in learning another language B, we need a measure of linguistic distance or distinctness between the two languages. Such a measure would allow us to distinguish the distance between languages like Tamil and Kannada (languages that are similar in structure, reducing the cost for the speaker of one language to learn the other), on the one hand, and Tamil and Nepali (languages that are very distinct with high costs of learning).

Measures of linguistic distance rely on 'language trees', which classify and group languages based on ancestry, origin, and structure, among other parameters. One such language tree is the Ethnologue (Lewis et al., 2014), which shows the relationship between different related languages (see Figure 1 for an example). Several methods have been developed to compute the linguistic distance between any two languages in the language tree. We use the following measure, based on Fearon (2003):

$$ld_{ij} = 1 - \left[\frac{\text{no. of common nodes between } i \text{ and } j}{\frac{\text{no. of nodes for } i + \text{no. of nodes for } j}{2}} \right]^{\lambda} \tag{1}$$

where ld_{ij} is the linguistic distance between between two languages, i and j. Here, if the languages belong to completely different language families, then the number of common nodes is 0 and the distance between the two languages is 1. The value of λ determines the relative distance between two languages that belong to the same family compared to two languages that belong to distinct language families. In the absence of any strong theoretical justification for λ, we follow Fearon (2003) and assume $\lambda = 0.5$.

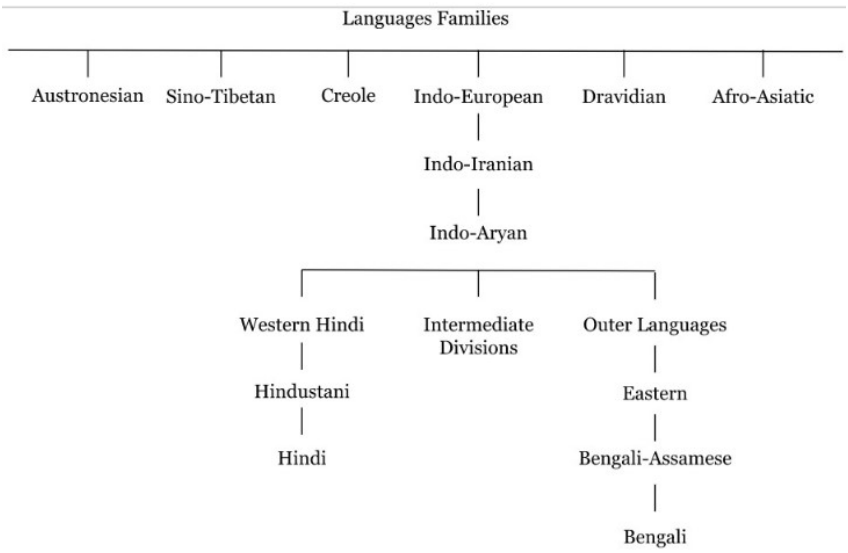

Figure 1: Language Families Tree from the Ethnologue

We additionally show that our results are robust to the use of alternative measures of linguistic distance, such as those developed by Lewis et al. (2014) and used in studies such as Jain (2017), and the ASJP method developed by Wichmann et al. (2011). Further details on these measures are discussed in the appendix.

3 Data and Descriptive Statistics

To test the hypothesis that increasing linguistic distance is correlated with worsened health outcomes, we used pooled survey data from two rounds of the National Family Health Survey conducted in 2015-16 (round 4) and 2019-21 (round 5). The pooled dataset includes health outcomes on 1,415,675 unique women between the ages of 15-49 years. The survey This dataset includes over 2.5 million observations. We look at health investments such as under-5 immunisations of 130,000–370,000 children as well as anthropometric outcomes of approximately 400,000 children. The data also includes a rich set of mother and household characteristics including the mother's age, whether the mother has completed primary education, total children ever born to the mother, religion, caste, whether the household is located in an urban area, household wealth index and total number of members in the household. We present the summary statistics of the various measures of linguistic distance in the sample in Table 1.

To compute linguistic distance, we use the Fearon (2003) method based on data from the Ethnologue (Lewis et al., 2014). We compute the linguistic distance (LD) between the respondent's mother tongue and the language of the questionnaire, which reflects the dominant language of the region. 21 official languages are recorded for both the respondent's mother tongue and the language of the questionnaire. Observations that list the mother tongue as "other" are dropped since linguistic distance cannot be calculated. Thus, we get a 21x21 matrix of pairwise linguistic distances for any combination of languages to get all possible cases of language mismatches. Table 10.1 in the appendix throws light on the variation in the computed linguistic distance data. A majority of the sampled women have LD = 0 i.e. no mismatch between the mother tongue and the dominant language of the district. This means that a large portion of our sample (85.36%)

are women who are native language speakers. The remaining 14.64% of women in our sample have some mismatch between their mother tongue and the dominant language of the region and the size of this mismatch i.e. distinctness varies from 0.13 (lowest LD between, for example, Hindi and Urdu) to 1 (highest LD between, for example, Malayalam and Marathi).[a]

For women, we aim to look at a comprehensive review of their health characteristics, morbidities, and health-seeking behaviour. We look at the prevalence of anaemia, high blood pressure, and high blood sugar. These outcomes are chosen because they can be easily influenced by dietary choices and other environmental factors, and are therefore amenable to change based on receiving health information from public and medical sources.

The data includes a variable that captures a woman's haemoglobin levels: we create an indicator variable that takes the value 1 if the haemoglobin level is less than 12.0 gm/dL, and 0 otherwise.[b]. Similarly, a woman is categorized as having 'high' blood pressure if her systolic reading is at least 140 mmHg and has 'high' blood sugar if her blood sugar level is at least 141 mg/dL. [c]

For children, we examine the immunisation status of the child. There are two reasons for this. First, vaccine-preventable diseases can cause child stunting and long-term poor mental and physical health among adults (Nandi et al., 2020). Thus, routine childhood vaccinations can significantly reduce the disease burden among children and improve their health outcomes. In India, free immunizations are provided to children, largely under the age of 5 years, to protect against 12 diseases. However, the process of getting a child vaccinated requires communication with healthcare providers about follow-up dates, and health cards that document vaccine status are usually in the dominant language of the region. We thus consider the impact of linguistic distance on routine vaccination for a number of vaccines for which information is available in the survey (DPT, Polio, Measles, Pentavalent, Rotavirus, Hepatitis B, Vitamin A1 and Vitamin A2 supplementation).

Table 2 shows differences between mismatched mothers (having linguistic distance >= 1) and matched mothers (having linguistic distance = 0) for a number of variables. In general, we see that mismatched individuals show lower individual health outcomes and lower uptake of immunizations for their children. There are also some statistically significant differences across socioeconomic variables that could affect health outcomes. For this reason, we control for these variables in the analysis that follows and also implement a matching estimator.

To control for time-invariant and time-variant differences, district-fixed effects and year-fixed effects are included respectively which are discussed in the empirical strategy.

4 Empirical Strategy

We estimate the impact of the growing linguistic distance between the dominant language and mother tongue on the incidence of poorer health outcomes among women and their

[a]The NFHS only captures languages and not dialects. For example, the mother tongue of the North Indian belt is recorded as 'Hindi' even though there exist differences in the dialects in Hindi, which are distinct from Hindi. However, we cannot measure these distances in the data.

[b]This standard is based on that of the American Hematology Society and WHO

[c]These standards are based on NFHS – 5 reports.

children using the following specification:

$$Y_{idt} = \beta_0 + \beta_1 \text{LinguisticDistance}_{id} + \beta_2 X_{id} + \delta_d + \eta_t + \varepsilon_{idt} \quad (2)$$

Where Y_{idt} is the outcome variable for woman/child i currently residing in district d, surveyed in the year t. The independent variable is $LinguisticDistance_{id}$, which is the computed linguistic distance between the dominant language of the region and the mother tongue of woman/child i. The identifying assumption here is that the linguistic distance between the mother tongue and the dominant language of the region is uncorrelated, conditional on controls, with other characteristics that could explain health. This means that there is no correlation between the explanatory variable and any unobservable factors that affect our dependent variable. One could argue that there could be several socioeconomic and other demographic characteristics that are correlated to linguistic distance and health outcomes. To tackle this, we employ X_{id}, which are covariates that capture socioeconomic and demographic characteristics, including Age of the woman/mother, Religion, Caste, Educational level, Wealth Index, number of family members, number of children alive, sex of the household head, and number of years in the current residence. We also include a set of fixed effects such as district, year of interview (to account for the fact that NFHS-5 was surveyed during the COVID-19 Pandemic and that we are using 2 rounds of NFHS spanning over 5 years), and birth-order fixed effects (for regressions being run for children). Where Y_{idt} is the outcome variable for woman/child i currently residing in district d, surveyed in the year t. The independent variable is $LinguisticDistance_{id}$, which is the computed linguistic distance between the dominant language of the region and the mother tongue of woman/child i. The identifying assumption here is that the linguistic distance between the mother tongue and the dominant language of the region is uncorrelated, conditional on controls, with other characteristics that could explain health. In X_{id}, we control for socioeconomic and demographic characteristics, including age of the woman, religion, caste, educational level, wealth quintile, number of family members, number of children alive, sex of the household head, and number of years spent in the current residence. We also include a set of fixed effects such as district, year of interview (to account for the fact that NFHS-5 was surveyed during the COVID-19 Pandemic and that we are using 2 rounds of NFHS spanning over 5 years), and birth-order fixed effects (for regressions being run for children). For regressions of child outcomes, we also include the current age of the child and the sex of the child as covariates.

5 Results

5.1 Impact on Morbidities of Women

The primary results are presented in Table 3. In Column 1, the coefficient for linguistic distance is positive and significant at the 5% level for anaemia, indicating that increasing linguistic distance is associated with an increase in the probability that a woman is anaemic. Moving from a linguistic distance of 0 to 1 between the mother tongue and the dominant language of the region leads to a 1.1 pp increase in the probability of a woman being anaemic (statistically significant at the 5% level). Similarly, the incidence of high blood sugar also increases with linguistic distance: as linguistic distance increases from 0 to 1, the probability of a woman having high blood pressure increases by 0.9 pp (significant at the 1% level). These results are consistent with the hypothesis that

a growing linguistic barrier is leading to an increase in the incidence of poorer health outcomes among women. However, there is no impact on the incidence of hypertension.

To establish robustness of our results, we re-estimate our main specification on a matched sample obtained using propensity score weights. We use the obtained propensity score weights to weight observations based on their likelihood of belonging to the treated group. Since we match on a binary variable – treatment – we redefine linguistic distance to equal 1 when it is greater than 0 (i.e. there is a mismatch in the language spoken at home and outside the home) and 0 if the linguistic distance is 0 (i.e. there is no mismatch). These results are presented in Table 4. We predict language mismatch using the following household characteristics: source of drinking water, household ownership of fridge, toilet, type of cooking fuel, type of floor, wall and roof material. Our results are similar to those obtained in the unmatched sample, suggesting that differential selection into migration is not driving our results.

5.2 Impact on Child Immunizations

Table 5 provides the results for the take-up of vaccinations and supplements. The sign on the estimated coefficient on linguistic distance is negative across all columns, implying that as linguistic distance increases, the probability that the child has received all doses of that particular immunization decreases. These coefficients are significantly different from 0 for Vitamin A1, Vitamin A2, Measles and Pentavalent by 3.8 pp (significant at 1%), 1.1 pp (significant at 5% level), 2.7 pp (significant at 1%) and 4.5 pp (significant at 1%) respectively. DPT takeup does increase, however, though this result is statistically significant at the 10% level.

We repeat the same process of matching by propensity score to establish the robustness of these results in Table 6. No, polio, hepatitis and DPT are negative and significantly different from 0 as well, though the effects on Vitamin A2 and measles drop out.

5.3 Impact on Child Anthropometric Outcomes

Another way to ascertain the health status of the child is through anthropometric parameters such as the Height-for-Age (HAZ), Weight-for-Age (WAZ), and Weight-for-Height (WHZ) ratios. Anthropometric measures are accurate indicators of nutritional intake which could also be affected due to the language barrier (Jayachandran and Pandey, 2015). We create variables for the incidence of stunting, wasting and whether a child is underweight from the child's recorded anthropometric values and regress these outcomes on our measure of linguistic distance. [d] The results are in Table 7. In Columns 1 and 2, we see negative and statistically significant coefficients on linguistic distance, suggesting that children are less likely to be stunted and underweight as linguistic distance increases, which is a surprising result. When we examine these results with a matched sample in Table 8, however, the coefficients become smaller and insignificantly different from 0, emphasising the importance of migrant selection into specific household characteristics.

Nonetheless, why do we not see an increased incidence of malnutrition among children as linguistic distance increases? A potentially mitigating effect could be the number of

[d] These are computed from the child's z-scores. Based on WHO Child Growth Standards. The child is considered stunted/underweight/wasted if their height-for-age/weight-for-age/ weight-for-height z-score is 2 SD below the mean. These are binary variables (0 if not stunted/underweight/wasted and 1 if stunted/underweight/wasted)

children born to a woman. We examine the fertility preferences of women by linguistic distance in Table 9. We find that increasing linguistic distance is correlated with the total number of children born to a woman and the number of children of a woman who are still alive. This is also reflected in a lower ideal number of children. While we do control for number of children born in all our specifications, there could still be unobserved differences across smaller and larger families that lead to improved health outcomes: such as improved maternal health due to fewer pregnancies, and hence improved health status of children born, and increased parental resources and attention expended on children.

6 Mechanisms: Health-seeking behaviour and female autonomy

We next turn to an explanation of the specific channels through linguistic distance that can disadvantage people who are seeking out health care services. One possibility is that increased linguistic distance is correlated with reduced or inefficient engagement and communication with healthcare workers and services due to the language barrier. In the context of India, (Laitin and Ramachandran, 2016) use data from the NFHS-3 to find that women are less likely to be aware of AIDS or use bed nets to reduce the incidence of malaria. We find the same results continue to hold even in our more recent dataset of pooled NFHS-4 and NFHS-5 observations (Table 12. We extend this analysis by looking at a number of different outcomes as well.

6.1 Engagement with the healthcare process

We first examine the impact of increasing linguistic distance on a woman's engagement with healthcare workers and her experience with medical procedures. For this, we look at whether the woman has met with an ANM/LHV (Lady Heath Visitors) worker in the last 3 months, the quality of care received during and after a medical operation (this information is provided for the female sterilization procedure) and whether the woman is covered by health insurance. We additionally consider whether pregnant women have accessed healthcare. These variables of interest are proxies for willingness to engage with healthcare workers as well as experiences during medical treatment and post-treatment procedures.

The results in Table 9 and Table 11 suggest that as linguistic distance increases, the probability of a woman having met a healthcare worker reduces by 1.3 pp (significant at the 1% level), and women are 0.8 pp less likely to report the quality of their healthcare as "good" (significant at 5%). There is no effect on being covered by health insurance and whether the woman is informed about the side effects of a medical procedure. For pregnant women, they are less likely to have a health card, less likely to have visited an ANC worker during pregnancy, to have been told about pregnancy complications, or have received any pregnancy supplements.

These results, all significant at the 5% or 1% level, taken together suggest reduced access to healthcare and engagement with healthcare service providers as linguistic distance increases.

6.2 Individual autonomy

Another channel to understand worsened health status and healthcare engagement is the reduced autonomy of women. Linguistic distance could be correlated with reduced female autonomy if a woman is restricted from leaving the house in an unfamiliar environment where she is less able to communicate with ease. We test this hypothesis by regressing different measures of women's autonomy on linguistic distance. The results are presented in Table 10. The results indicate that as linguistic distance increases from 0 to 1, the probability of the woman facing difficulties in accessing medical help for herself decreases by 2.7 pp (Column 1, significant at 1% level). Similarly, the probability of a woman being "allowed" to get medical help for herself alone decreases by 3.9 pp (Column 3, significant at the 1% level) and for her family by 3.6 percentage points (Column 3, significant at the 1% level). Finally, we examine whether linguistic distance is correlated with other measures of autonomy as well: we find that women are less likely to be allowed to go to a market alone if linguistic distance increases from 0 to 1.

6.3 Exposure to health-pertinent information through media

We next consider exposure to health-pertinent information through the media. Previous literature has found that increased familiarity with dominant languages improves the knowledge of health through exposure to the media (Ruiz et al., 1992). The variables we consider all relate to family planning. We estimate the impact of the growing linguistic distance between the mother tongue and the dominant language of the region on whether a woman has heard of family planning methods through radio, television and newspapers. The rationale behind this is that the language of communication for these forms of media is primarily in the dominant language and women who do not speak this language well may be less likely to access this information. Table 12 presents these results; the estimated coefficients on linguistic distance are negative across a range of media exposure outcomes. As linguistic distance increases from 0 to 1, the probability of the woman having heard of family planning through radio reduces by 5.9 pp, having heard from television reduces by 1.5 pp and having heard from newspapers reduces by 5.7 pp, all significant at the 1% level. This provides some suggestive evidence that women facing language barriers may be less likely to access health-related information from the media.

Taken together, linguistic distance presents a barrier to accessing healthcare on multiple fronts. Women are less likely to leave the home by themselves, they are less likely to engage with healthcare services, and they are less likely to receive health-relevant information through the media.

7 Robustness & further results

7.1 Other measures of distinctness

As explained before, there are diverse ways to measure distinctness between languages and each measure captures different differences across languages. We test to see if other ways of looking at the language barrier affect the effects and results. For this, the (Lewis et al., 2014) method used by (Jain, 2017) and the ASJP method used by Wichmann et al. (2011) are used in place of the (Fearon, 2003) measure. This gives us opportunities to test for robustness in different ways of measuring distinctness between languages. For this

sample, given all the combinations of language mismatch, the linguistic distance (Lewis et al., 2014) varies between 3-16 (0 being no mismatch) and for the ASJP method, the linguistic distance varies between 34.84-104.16 (0 being no mismatch). Given the variation in the measure of linguistic distance, the size of the estimated effects may be different but the sign of the coefficient for the outcome variables should be consistent with the results we get from using the (Fearon, 2003) linguistic distance. The results in Table 18 and Table 19 show that even with changing the measure of linguistic distance, for morbidities on women, across all columns, the sign of the coefficients remains positive which is the same as the results from (Fearon, 2003). Similarly for children, Table 20 and Table 21 present consistent results to the main findings.

7.2 Heterogeneity by wealth

We next disaggregate the results by wealth. Poorer families tend to rely more on public health officials including government doctors and nurses, while people in higher quintiles are more likely to access private healthcare. A priori, it is not clear whether public or private healthcare will be better equipped to deal with linguistic mismatches. If doctors speaking the same language as a patient or a patient's parent are more likely to be found in the private sector, then the adverse effects of linguistic distance should decline with wealth. On the other hand, if these private options are simply not available, public health care can provide a basic level of health care to all those who access it. For this analysis, we use the five wealth quintiles that are derived from the wealth index in NFHS [e]. We create a variable that takes the value 1 if the woman belongs to the top 2 quintiles, and 0 if not. In Table 14, for morbidities of women, we see the magnitude of the disadvantage faced by the language barrier is more pronounced for the richer wealth group. The coefficient on the interaction between rich and linguistic distance is positive and significant. For vaccine take-up, the positive coefficient on the interaction between rich and linguistic distance is positive, suggesting that linguistic distance is more costly for the rich (Table 15). This suggests that public healthcare services could play an important role in mitigating the effect of a language barrier, and the more people access high-quality public healthcare services, the smaller the cost of the language barrier will be.

7.3 Heterogeneity by length of residence

An important point to note is while languages are costly to learn, they are usually acquired over time. Thus, the language barrier should decline with time. The longer one stays in a particular region, the more likely their language skills are to improve and thus the size of the barrier posed by language can be reduced. We hypothesize that the magnitude of the negative health effect will likely become smaller the longer a person has stayed in a particular region. To test this, we divide the sample based on the year of current residence for the respondent. Women who have spent above 18 years (the median value in the sample) in their current place of residence take the value 1 and 0 if they have spent below 18 years in their current place of residence. In this sample, this varies from 0 to

[e]This is a composite measure of the household's cumulative living standard. Using principal component analysis, householders are attributed scores according to the number and kinds of consumer goods they own, and the characteristics of their housing (such as the source of drinking water, toilet facilities, and flooring materials). Then, national wealth quintiles are formed by assigning the score of the household to each member, ranking all people in the household population by their score, and dividing the distribution into 5 equal bands (of 20% of the total population).

49 years of residence. We interact this dummy with linguistic distance and test the size of the coefficient on the interaction term.

Table 16 presents the results for women and Table Table 17 for children, the coefficient across the anemia, and blood sugar columns is positive. This means that the marginal effect of increasing linguistic distance on anaemia and blood sugar increases with each additional year in residence, but with blood pressure, the same effect is negative. For child vaccination outcomes, the impact of linguistic distance on vaccine take-up weakens, rather than increases. It could be that for women's morbidities, the impacts of early decline in health outcomes persist and are not reversed over time, but women are more open to accessing health services for their children as time spent in a new place increases.

8 Conclusion

This paper contributes to the literature on the adverse effects of linguistic distance on human capital accumulation, specifically through access to the healthcare system. Using data from the National Family Health Survey, we find that increased linguistic distance between a woman's mother tongue and the dominant language in her place of residence is associated with poorer health outcomes for mothers and their children, including higher rates of morbidities and lower immunization status. The study identifies reduced health-seeking behaviour, reduced exposure to health information obtained through the media, and diminished autonomy among women as mechanisms driving these effects. Our results are robust to using three different measures of linguistic distance.

Language acquisition is a critical channel for social, economic and cultural mobility, and effective inclusion of linguistically diverse groups into society through policies would play a pivotal role in reducing disparities not only in health but other socio-economic indicators of development as well. It is incumbent on policy-makers to make additional efforts to reach out to linguistically diverse groups to improve their health outcomes. Alternatively, ensuring access to learning dominant languages could also reduce the costs of acquiring new languages and mitigate the adverse effects of linguistic distance. This is especially important in linguistically diverse countries like India with large flows of internal migrants.

References

Aoki, Y. and L. Santiago (2018). Speak better, do better? education and health of migrants in the uk. *Labour Economics 52*, 1–17.

Bromham, L., X. Hua, T. G. Fitzpatrick, and S. J. Greenhill (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences 112*(7), 2097–2102.

Chiswick, B. R. (2008). The economics of language: An introduction and overview.

Eberhard, D. M., G. F. Simons, and C. D. Fennig (2023). *Ethnologue: Languages of the World* (Twenty-sixth ed.). Dallas, Texas: SIL International.

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

Ginsburgh, V. and S. Weber (2020). The economics of language. *Journal of Economic Literature 58*(2), 348–404.

Gomes, J. F. (2020). The health costs of ethnic distance: evidence from sub-saharan africa. *Journal of Economic Growth 25*(2), 195–226.

Guven, C. and A. Islam (2015). Age at migration, language proficiency, and socioeconomic outcomes: evidence from australia. *Demography 52*(2), 513–542.

Jain, T. (2017). Common tongue: The impact of language on educational outcomes. *The Journal of Economic History 77*(2), 473–510.

Jayachandran, S. and R. Pandey (2015). Son preference drives india's high child malnutrition rates. *VoxEU*.

Kumar, H., R. Somanathan, M. Vasishth, et al. (2020). Language and learning in ethically mixed communities: A study of school children in an indian village. Technical report.

Laitin, D. D. and R. Ramachandran (2016). Language policy and human development. *American Political Science Review 110*(3), 457–480.

Lewis, M. P., G. Simon, and P. Fennig (2014). *Ethnologue: Languages of the world*. Dallas, Texas: SIL International.

Nandi, A., S. Kumar, A. Shet, D. E. Bloom, and R. Laxminarayan (2020). Childhood vaccinations and adult schooling attainment: Long-term evidence from india's universal immunization programme. *Social Science Medicine 250*, 112885.

Nguyen, D. and L. J. Reardon (2013). The role of race and english proficiency on the health of older immigrants. *Social Work in Health Care 52*(6), 599–617.

Petroni, F. and M. Serva (2010). Measures of lexical distance between languages. *Physica A: Statistical Mechanics and its Applications 389*(11), 2280–2283.

Ponce, N. A., R. D. Hays, and W. E. Cunningham (2006). Linguistic disparities in health care access and health status among older adults. *Journal of General Internal Medicine 21*(7), 786–791.

Pottie, K., E. Ng, D. Spitzer, A. Mohammed, and R. Glazier (2008). Language proficiency, gender and self-reported health: an analysis of the first two waves of the longitudinal survey of immigrants to canada. *Canadian Journal of Public Health 99*, 505–510.

Ruiz, M. S., G. Marks, and J. L. Richardson (1992). Language acculturation and screening practices of elderly hispanic women: The role of exposure to health-related information from the media. *Journal of Aging and Health 4*(2), 268–281.

Schachter, A., R. T. Kimbro, and B. K. Gorman (2012). Language proficiency and health status: are bilingual immigrants healthier? *Journal of health and social behavior 53*(1), 124–145.

Street, R. L. and P. Haidet (2011). How well do doctors know their patients? factors affecting physician understanding of patients' health beliefs. *Journal of General Internal Medicine 26*(1), 21–27.

Tumbe, C. (2018). *India moving: A history of migration*. Penguin Random House India Private Limited.

Wichmann, S., T. Rama, and E. W. Holman (2011). Phonological diversity, word length, and population sizes across languages: The asjp evidence. pp. 177–197.

9 Tables

Table 1: Summary Statistics

	Mean	SD	Min	Max	N
Linguistic Distance (Fearon 2003)	0.05	0.18	0	1	1620768
Linguistic Distance (Jain 2017)	0.74	2.73	0	16	1620821
Linguistic Distance (ASJP)	6.18	22.65	0	104	1619150

Table 2: Balance Tables

	Treatment (Language Mismatch)	Control (No Language Mismatch)	Difference	SE	N
Anemia	0.549	0.465	0.085***	(0.001)	1367367
High Blood Pressure	0.034	0.055	-0.021***	(0.000)	1602434
High Blood Sugar	0.062	0.054	0.009***	(0.001)	1579511
Polio	0.546	0.486	0.060***	(0.002)	378027
Hepatitis B	0.427	0.381	0.046***	(0.002)	373520
Vitamin A1	0.664	0.613	0.051***	(0.002)	374868
Vitamin A2	0.199	0.204	-0.005***	(0.002)	374394
DPT	0.716	0.679	0.038***	(0.002)	376808
Measles	0.638	0.589	0.049***	(0.002)	375648
Pentavalent	0.703	0.660	0.043***	(0.004)	130275
Rotavirus	0.338	0.211	0.126***	(0.004)	129757
Stunting (-2SD)	0.376	0.337	0.039***	(0.002)	428951
Underweight (-2SD)	0.340	0.258	0.082***	(0.002)	433404
Wasting (-2SD)	0.200	0.171	0.029***	(0.002)	424672
Stunting (-3SD)	0.161	0.141	0.020***	(0.002)	428951
Underweight (-3SD)	0.110	0.082	0.027***	(0.001)	433404
Wasting (-3SD)	0.076	0.072	0.003***	(0.001)	424672
No. of Household Members	5.625	5.259	0.366***	(0.006)	1740471
Total No. of Children Ever Born	1.769	1.737	0.032***	(0.004)	1740471
Caste	2.642	2.516	0.126***	(0.003)	1352177
Respondent's Current Age	30.454	30.392	0.062***	(0.023)	1740471
Wealth Index	2.925	2.859	0.066***	(0.003)	1728629
Religion	1.328	2.037	-0.709***	(0.002)	1740471
Highest Education Level	1.550	1.526	0.025***	(0.002)	1740471
Male Household Head	0.860	0.865	-0.005***	(0.001)	1740477
Urban/Rural	1.727	1.760	-0.033***	(0.001)	1516361
Years Lived in Place of Residence	45.165	40.890	4.275***	(0.123)	1041393
Round of Survey	4.461	4.526	-0.065***	(0.001)	1740471

Table 3: Morbidities of Women

	Anemia	High Blood Pressure	High Blood Sugar
Linguistic Distance	0.011**	0.002	0.009***
	(0.005)	(0.002)	(0.003)
Observations	593,399	567,501	591,736
Mean	0.566	0.041	0.064

All estimations include a binary indicator for whether the woman is Anemic, has high blood pressure and high blood sugar. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: PSM Matching for Women

	Anemia	High Blood Pressure	High Blood Sugar
Treatment (Language Mismatch)	0.013***	0.000	0.009***
	(0.003)	(0.001)	(0.002)
Observations	205,005	205,005	205,005
Mean	0.519	0.044	0.060

All estimations include a binary indicator for whether the woman is Anemic, has high blood pressure and high blood sugar. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Impact on Child Immunisations

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
Linguistic Distance	-0.007	-0.006	-0.038***	-0.011**	0.011*	-0.027***	-0.045***	-0.004
	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)	(0.010)	(0.011)	(0.010)
Observations	332,531	328,711	329,769	329,358	331,634	113,718	113,749	113,258
Mean	0.549	0.430	0.665	0.201	0.718	0.307	0.705	0.333

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Propensity Score Matching for Child Immunisations

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
Treatment (Language Mismatch)	-0.029***	-0.017**	-0.028***	0.002	-0.015**	0.000	-0.014**	0.006
	(0.007)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
Observations	32,843	32,843	32,843	32,843	32,843	32,843	32,843	32,843
Mean	0.602	0.521	0.684	0.238	0.734	0.287	0.682	0.261

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Child Anthropometry

	Stunted	Underweight	Wasted
Linguistic Distance	-0.027***	-0.010*	0.008
	(0.006)	(0.006)	(0.005)
Observations	376,622	380,410	372,975
Mean	0.374	0.336	0.198

All estimations include a binary indicator for whether the child is stunted, underweight or wasted based on their z-scores. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: PSM - Severe Child Anthropometry

	Stunted	Underweight	Wasted
Treatment (Language Mismatch)	-0.006	-0.001	0.005
	(0.004)	(0.004)	(0.003)
Observations	115,775	115,775	115,775
Mean	0.351	0.274	0.181

All estimations include a binary indicator for whether the child is stunted, underweight or wasted based on their z-scores. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Engagement with the Healthcare Process

	Met with an ANM/LHV Worker	Health Insurance	Informed about Side Effects	Quality of Care
Linguistic Distance	-0.013***	-0.005	0.001	-0.008**
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	623,821	623,829	623,829	172,420
Mean	0.200	0.325	0.262	0.971

All estimations include a binary indicator for whether the woman has met with an ANM/LHV worker, covered by health insurance, informed about side effects of a medical procedure and whether the women self-reports that she received 'good' quality care after a medical procedure. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Channels: Autonomy of Women

	Allowed to go Alone and get Medical Help for Self	Allowed to get Medical Help for Self	Allowed to go to a Medical Facility Alone	Allowed to go Outside this Village Alone	Allowed to go Market Alone
Linguistic Distance	-0.027***	-0.039***	-0.036***	0.010	-0.028**
	(0.004)	(0.004)	(0.013)	(0.013)	(0.013)
Observations	623,829	623,829	93,474	93,474	93,474
Mean	0.176	0.140	0.510	0.565	0.498

All regressions include Covariates, District and Year fixed effects. In Column 1, the question asked to the respondent is whether going alone to get Medical Help for self is a problem or not. In Column 2, the question asked to the respondent is whether they are allowed to get medical help for self. In Column 3, the respondent is asked whether they are allowed to go to a medical facility alone. In Column 4, the respondent is asked whether they are allowed to go to the market alone. In Column 5, the respondent is asked whether they are allowed to go outside their village alone. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Other Health Channels

	Has Health Card	Received ANC for Pregnancy	Told about Pregnancy Complications	Given Supplements during Pregnancy
Linguistic Distance	-0.031***	-0.025***	-0.018***	-0.016**
	(0.006)	(0.005)	(0.007)	(0.008)
Observations	114,482	153,338	275,754	310,880
Mean	0.947	0.945	0.707	0.849

All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Exposure to Family Planning through Media & Other Health

	Source of FP: Radio	Source of FP: Newspaper	Source of FP: TV	Respondent Slept Under Bed Net	Heard of Aids
Linguistic Distance	-0.059***	-0.017***	-0.059***	-0.013***	-0.026***
	(0.004)	(0.005)	(0.005)	(0.003)	(0.008)
Observations	623,821	623,821	623,821	623,829	93,474
Mean	0.151	0.585	0.349	0.222	0.877

All estimations include a binary indicator for whether the woman has heard of family planning methods through various channels of media such as Radio, Television or Newspaper, whether the respondent slept under a bed net and whether the respondent has heard of Aids. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Fertility Preferences

	Total No. of Children Ever Born	No. of Children Alive	Ideal No. of Children	Ideal No. of Boys	Ideal No. of Girls	Ideal Son Preference
Linguistic Distance	-0.033***	-0.029***	-0.513***	-0.900***	-0.902***	-0.274***
	(0.006)	(0.006)	(0.103)	(0.120)	(0.120)	(0.047)
Observations	672,353	672,353	623,829	623,829	623,829	594,692
Mean	2.571	2.703	3.200	2.404	2.268	0.701

All regressions include Covariates, District and Year fixed effects. In Columns 1 and 2, the specification is run on the Birth Recode and the sample is restricted to children born up to 5 years preceding the year of the survey. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Women: Heterogeneity by Wealth

	Anemia	High Blood Pressure	High Blood Sugar
Rich X LD	0.015**	0.000	-0.001
	(0.008)	(0.003)	(0.004)
Observations	593,399	567,501	591,736
Mean	0.566	0.041	0.064

The dummy variable Rich takes the value 1 when the woman belongs the the 'richer' and 'richest' quintiles and is equal to 0 when the woman belongs to the 'poorest', 'poorer' and 'middle' quintiles. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Children: Heterogeneity by Wealth

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
Rich X LD	0.007	0.001	0.008	0.019***	0.016**	0.026*	0.029*	-0.005
	(0.009)	(0.009)	(0.009)	(0.007)	(0.008)	(0.015)	(0.016)	(0.015)
Observations	332,531	328,711	329,769	329,358	331,634	113,718	113,749	113,258
Mean	0.549	0.430	0.665	0.201	0.718	0.307	0.705	0.333

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. The dummy variable Rich takes the value 1 when the woman belongs the the 'richer' and 'richest' quintiles and is equal to 0 when the woman belongs to the 'poorest', 'poorer' and 'middle' quintiles. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Women: Years in Residence Interaction

	Anemia	High Blood Pressure	High Blood Sugar
Years X LD	0.015**	-0.004*	0.008***
	(0.006)	(0.002)	(0.003)
Observations	1,185,342	1,143,688	1,182,166
Mean	0.546	0.037	0.058

All estimations include a binary indicator for whether the woman is Anemic, has high blood pressure or high blood sugar. All regressions include Covariates, District and Year fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Immunisations : Years in Residence Interaction

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
Years X LD	0.027**	0.025**	0.014	0.003	0.007	-0.021	-0.043**	-0.014
	(0.011)	(0.011)	(0.011)	(0.009)	(0.010)	(0.020)	(0.021)	(0.019)
Observations	329,977	326,198	327,248	326,840	329,090	113,001	113,030	112,543
Mean	0.550	0.431	0.665	0.201	0.718	0.307	0.705	0.333

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Morbidities of Only Women - Lewis (2014)

	Anemia	High Blood Pressure	High Blood Sugar
LD (Lewis 2014)	0.800**	0.262	0.675***
	(0.385)	(0.164)	(0.192)
Observations	593,414	567,516	591,751
Mean	0.566	0.041	0.064

All regressions include Covariates, District and Year fixed effects. The Independent variable Linguistic Distance as computed by this method is divided by 1000. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Morbidities of Women - ASJP

	Anemia	High Blood Pressure	High Blood Sugar
LD (ASJP)	0.117**	0.026	0.083***
	(0.046)	(0.020)	(0.023)
Observations	593,002	567,113	591,341
Mean	0.566	0.041	0.064

All regressions include Covariates, District and Year fixed effects. The Independent variable Linguistic Distance as computed by this method is divided by 1000. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: All Child Immunizations - Lewis (2014)

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
LD (Lewis 2014)	-0.899*	-0.634	-2.528***	-0.899**	0.752*	1.195***	-2.938***	-0.228
	(0.477)	(0.470)	(0.455)	(0.388)	(0.430)	(0.434)	(0.785)	(0.722)
Observations	332,544	328,724	329,782	329,371	331,647	330,742	113,751	113,260
Mean	0.549	0.430	0.665	0.201	0.718	0.640	0.705	0.333

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. The Independent variable Linguistic Distance as computed by this method is divided by 1000. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: All Child Immunizations - ASJP

	Polio	Hepatitis B	Vitamin A1	Vitamin A2	DPT	Measles	Pentavalent	Rotavirus
LD (ASJP)	-0.072	-0.044	-0.332***	-0.097**	0.121**	0.177***	-0.341***	-0.055
	(0.057)	(0.057)	(0.055)	(0.047)	(0.052)	(0.052)	(0.093)	(0.086)
Observations	332,039	328,225	329,283	328,872	331,143	330,237	113,667	113,176
Mean	0.549	0.430	0.665	0.201	0.718	0.640	0.705	0.333

All estimations include a binary indicator for whether the child has received all doses of a particular immunisation. The Independent variable Linguistic Distance as computed by this method is divided by 1000. All regressions include Covariates, District,Year and Birth-Order fixed effects. Standard errors in parenthesis.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

10 Appendix

10.1 Alternative measures of linguistic distance

In addition to the method proposed by Fearon (2003), we show our results are robust to using alternative measures proposed by Jain (2017) and Wichmann et al. (2011). These are discussed below.

The (Lewis et al., 2014) method used by (Jain, 2017) method computes the number of nodes between each language pair on the language tree. We trace the nodes from the end note of one language, then to the common nodes that the language pair shares till the end node of the other language. In this approach, the more nodes between the two languages (i.e. the farther the languages are on the tree), the more distinct they are in terms of structure, grammar and other linguistic parameters. Thus, to calculate the linguistic distance between two languages, we count the total number of nodes between them, including the end nodes. As an example, in Figure 1, the distance between Hindi and Bengali can be traced from Hindi (1 node) → Hindustani (2) → Western Hindi (3) → Indo-Aryan (4) → Outer Languages (5) → Eastern (6) → Bengali-Assamese (7) → Bengali (8). Thus, the resulting linguistic distance is 8.

Another method of measuring distinctness between languages uses the concept of 'lexical or lexicostatistical distance' between languages. This is based on identifying similarities between common roots of words and shared vocabularies between languages. The measure of distance is based on the percentage of shared cognates between two languages (cognate words are words in any two languages that share similar meaning, spelling and pronunciation). An automated method proposed by Petroni and Serva (2010) uses a normalisation of a Levenshtein Distance (LV) – the minimum number of insertions, deletions, or substitutions of a single character needed to transform one word into the other. Wichmann et al. (2011) develop an ASJP (Automated Similarity Judgment Program) software to calculate this distance, based on the LV to calculate the lexical distance between any two languages.

Table 22: Summary Statistics

	Mean	SD	Min	Max	N
Anemia	0.54	0.50	0	1	1367367
High Blood Pressure	0.04	0.19	0	1	1602434
High Blood Sugar	0.06	0.24	0	1	1579511
Polio	0.54	0.50	0	1	378027
Hepatitis B	0.42	0.49	0	1	373520
Vitamin A1	0.66	0.47	0	1	374868
Vitamin A2	0.20	0.40	0	1	374394
DPT	0.71	0.45	0	1	376808
Measles	0.63	0.48	0	1	375648
Pentavalent	0.70	0.46	0	1	130275
Rotavirus	0.32	0.47	0	1	129757
Stunting (-2SD)	0.37	0.48	0	1	428951
Underweight (-2SD)	0.33	0.47	0	1	433404
Wasting (-2SD)	0.20	0.40	0	1	424672
Stunting (-3SD)	0.16	0.36	0	1	428951
Underweight (-3SD)	0.11	0.31	0	1	433404
Wasting (-3SD)	0.08	0.26	0	1	424672
No. of Household Members	5.58	2.53	1	41	1740471
Total No. of Children Ever Born	1.77	1.76	0	45	1740471
Caste	2.62	1.02	1	4	1352177
Respondent's Current Age	30.45	10.11	15	54	1740471
Wealth Index	2.92	1.36	1	8	1728629
Religion	1.42	0.82	1	4	1740471
Highest Education Level	1.55	0.99	0	3	1740471
Male Household Head	0.86	0.35	0	1	1740477
Urban/Rural	1.73	0.44	1	2	1516361
Years Lived in Place of Residence	13.90	10.16	0	54	661219
Round of Survey	4.47	0.50	4	5	1740471

Table 23: Variation in Linguistic Distance (Fearon 2003)

Linguistic Distance (Fearon 2003)	Observations	Percent
0	1,191,155	91.37
0.13	6,047	0.46
0.21	2,656	0.20
0.29	383	0.03
0.39	25,491	1.96
0.5	4	0.00
0.65	46,139	3.54
1	31,746	2.44
Total	1,303,620	100.00