Local social inequality, economic inequality, and disparities in child height in India

Diane Coffey* Ashwini Deshpande† Jeffrey Hammer‡ Dean Spears§ ¶

January 9, 2019

Abstract

This paper investigates disparities in child height — an important marker of population-level health — among population groups in rural India. India is an informative context in which to study processes of health disparities because there is wide heterogeneity in the degree of local segregation or integration among caste groups. Building on a literature that identifies discrimination by quantifying whether differences in socioeconomic status (SES) can account for differences in health, we decompose height differences between rural children from higher castes and rural children from three disadvantaged groups. We find that socioeconomic differences can explain the height gap for Scheduled Tribe children, who tend to live in geographically isolated places. However, SES does not fully explain height gaps for children from the Scheduled Castes (SC) and Other Backward Classes (OBC). Among SC and OBC children, local processes of discrimination also matter: the fraction of households in a child's locality that outrank her household in the caste hierarchy predicts her height. SC and OBC children who are surrounded by other lower caste households are no shorter than higher caste children of the same SES. Our results contrast with studies from other populations where segregation or apartheid are negatively associated with health.

*Indian Statistical Institute – Delhi Centre and University of Texas, Austin. Email: coffey@utexas.edu

†Delhi School of Economics. Email: ashwini@econdse.org

‡Princeton University. Email: jhammer@princeton.edu

§Indian Statistical Institute – Delhi Centre and University of Texas, Austin. Email: dspears@utexas.edu

¶We appreciate helpful suggestions from Sarah Burgard, Kenzie Lantham, Aashish Gupta, Mark Hayward, Narayan Sastry, and participants at the 2017 Population Association of America session on Economic Inequality and Health. This research was supported by grant P2CHD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. All errors are our own.

1 Introduction

Inequality in child health among population groups can be caused by inequality in socioeconomic status, can cause socioeconomic inequality, or can be independent of socioeconomic status. Examples of each of these patterns have been found in both developed and developing countries (Elo, 2009). In developing countries, because malnutrition and disease among children are common and often severe, unequal child health outcomes among population groups may contribute even more to perpetuating inequality than they do in developed countries (Palloni, 2006).

In order to better understand relationships between inequality and health, a large literature in demography has focused on quantifying the extent to which differences in socioeconomic status (SES) between population groups can explain differences in health outcomes (Williams and Collins, 1995). In this literature, evidence of disparate outcomes at the same level of SES is a common indicator of discrimination (Rogers, 1992; Cramer, 1995; Elo and Preston, 1996; Williams, 1999). For example, Burgard (2002) finds that the distribution of socioeconomic resources across households and geographic areas can explain differences in rates of stunting between children of different racial groups in Brazil, but that, in South Africa, differences in stunting between white and non-white children are not fully explained by these factors.

Literature on discrimination has increasingly progressed beyond documenting the existence of disparities among people with the same SES, towards identifying processes and asking how discrimination has its effects. In the health literature, Sastry and Hussey (2003) study differences in birth weight between non-Hispanic whites and Mexican-origin Hispanics; they find that not only do inputs such as SES and neighborhood factors differ between these groups, but that the effects of these inputs also differ between groups. More generally, the relationships between socioeconomic variables and child outcomes within a society may change over time; they may depend on neighborhood-level factors as well as household-level factors; or they may interact with social group (Hummer, 1993; Sastry, 1996, 2004; Reichman, Teitler and Hamilton, 2009). Demographers have particularly focused on the effect of segregation: they have often asked whether racial or other types of segregation contribute to health disparities (Massey, 1990; Williams and Collins, 2001; LaVeist et al., 2011; Leung and Takeuchi, 2011).

In this paper, we advance both literatures, making two contributions to understanding disparities in child height in rural India. First, we show that some disparities can be only partially explained by

SES, with an important remaining role for discrimination. Second, we study the process of discrimination's effect by documenting the role of the local social context. Rural India's combination of mixed and segregated localities offers an informative contrast with research from other societies, such as the United States and South Africa, where scholars have posited a relationship between segregation and health. We find that in rural India lower caste children only have worse height outcomes than higher caste children of similar SES if they are surrounded by higher caste households in the same locality. This result has implications for theories of discrimination: in this context, discrimination is, in part, channeled through one's rank within the local area.[1] Our results contrast with studies of other populations, where segregation or apartheid harm health, in part by denying access to public services or other resources. In rural India, being surrounded by other households from lower castes appears to advantage lower caste children.

Beyond the ways in which this case contributes to a broader understanding of processes of discrimination in health, both caste and child height are important demographic topics in their own right. Caste is strongly associated with poverty, education, land ownership, consumption, use of health services, and even subjective well-being (Deshpande, 2000; Borooah, 2005; Desai and Dubey, 2012; Roy, Kulkarni and Vaidehi, 2004; Spears, 2016). Child height is an important marker of population health and a predictor of adult human capital and mortality (Deaton, 2007). Within populations, children whose early life health allows them to grow taller go on to reach higher levels of cognitive achievement, on average (Case and Paxson, 2010). The association between childhood height and cognitive achievement is considerably steeper in India than in the United States (Spears, 2012).

Because height is such an important marker of health and well-being, and because the Demographic and Health Survey that we study shows that the average child under five in rural India is about two standard deviations shorter than the WHO reference population for healthy growth, better understanding disparities of height among Indian children is important for human development. To our knowledge, this paper is the first to investigate explanations for height disparities and processes of discrimination among different population groups in India.[2]

[1] In this study, as in the 'relational approach' described by Cummins et al. (2007), place has a different relationship with health for higher caste children than it does for lower caste children.

[2] The literature on health disparities by caste and tribe has largely focused on disparities in the use of health services (Borooah, 2012; Acharya, 2012; Baru et al., 2010). There is a larger literature on gender disparities in health outcomes (Das Gupta, 1987; Pande, 2003; Murthi, Guio and Dreze, 1995; Arnold, Choe and Roy, 1998; Barcellos, Carvalho and Lleras-Muney, 2014). Some examples of research on health disparities by religion include Guillot and Allendorf (2010), Bhalotra, Valente and van Soest (2010), Desai and Temsah (2014), Brainerd and Menon (2015), and Geruso and Spears (2018).

We present results in two parts. First, we conduct a reweighting decomposition of average child height, constructing counterfactual average child height among three disadvantaged population groups – Scheduled Tribes (ST), Scheduled Castes (SC) and Other Backward Classes (OBC).[3,4] We ask: what would the mean height for age of children in each group be if they had the same distribution of socioeconomic characteristics as higher caste Indian children? In contrast with our results for lower caste SC and OBC children, SES variables can explain height disparities for ST children, a geographically isolated and socioeconomically disadvantaged population group that is not comparably ranked by the caste system.

Although the average gap in height between higher caste and lower caste (SC and OBC) children in rural India cannot fully be explained by household-level SES variables, these variables can fully explain the caste height gap in those localities where SC and OBC children do not live with higher caste neighbors. This suggests a process of local discrimination that is the focus of the second set of results. We use regression analysis to document that SC and OBC children are shorter in contexts where a larger fraction of the households in their primary sampling unit is from a higher caste. Together, SES and local caste rank can account for the entire height gap between lower caste and higher caste children.

This paper proceeds as follows: section 2 provides background about why rural India is an informative context in which to study health inequalities. Section 3 presents a conceptual framework linking caste, tribal status, and local processes of discrimination to child height in rural India. Section 4 presents our data and empirical strategies. Section 5 presents results of the reweighting decomposition and then uses regression to understand the role of local caste rank. Section 6 discusses possible mechanisms for our results and section 7 concludes by considering the implications for policies to reduce inequality.

[3] The Government of India categorizes caste and tribal groups in this way for the purpose of affirmative action programs. We discuss these programs in section 2.

[4] We classify households as they report themselves to the DHS surveyor. These categories are reported for both Hindus and non-Hindus. Table A1 in the Supplementary Appendix tabulates SC, ST, OBC and general caste status separately for Hindus, Muslims, and other religions.

2 Background: Caste, tribe, and inequality in India

Caste and tribal status are important dimensions of social stratification in India.[5] The contemporary manifestation of caste comprises 6000 endogamous social groups called *jatis*. These groups were traditionally occupation-specific and hereditary.[6] Caste is also relevant to a person's social networks and voting choices (Deshpande, 2017).[7] People from India's tribal communities, referred to as adivasis, or indigenous people, are considered to have a social identity outside the caste system. Tribal communities speak many different languages and have many different cultural practices, including tribe-specific religions. According to the 2011 census, adivasis make up about 9% of India's population. Ninety percent of adivasis live in rural areas, often in isolated villages that are not well served by public resources. People from adivasi backgrounds are among the most economically and educationally deprived in India (Mitra, 2008; Maharatna, 2000).

Recent research investigates whether disparities in wages, employment, and consumption between lower castes and adivasis and higher castes have been narrowing in recent decades (Hnatkovska, Lahiri and Paul, 2012; Desai and Kulkarni, 2008). There is debate about the extent to which these gaps have narrowed since independence, and also clear evidence that many people from disadvantaged social groups continue to face discrimination (Deshpande and Ramachandran, 2016; Deshpande, 2017; Thorat and Newman, 2012).

India's affirmative action program has been designed to address disparities and discrimination along these axes of disadvantage.[8] The government "reserves" seats in schools, government jobs, and public office for people from Scheduled Tribes (STs), who belong to the adivasi groups described above, and from Scheduled Castes (SCs), whose members were once, and in some circumstances still are, treated as "untouchable." Reservations or "quotas" for these two groups are constitutionally mandated and the names of *jatis* and tribes are listed in a government schedule. A third group, the Other Backward Classes (OBCs), is also eligible for some affirmative action programs. OBCs are

[5]The caste system is traditionally associated with India, but aspects of caste also operate in other South Asian countries, including Nepal, Pakistan and Bangladesh (Jodhka and Shah, 2010; International Dalit Solidarity Network., 2009).

[6]A large number of occupations in the modern economy do not have a caste counterpart, i.e. these are not hereditary caste-specific occupations. Thus, in an obvious sense, the caste-occupation overlap has weakened. However, the overlap between caste and occupational status continues in that higher ranked castes are disproportionately represented in more prestigious and better paying occupations.

[7]Although the caste system has its origins in Hinduism, many non-Hindus, including Muslims and Christians whose ancestors converted to these religions, also have a *jati* identity.

[8]One important affirmative action program that is not related to caste or tribe is the reservation of some seats in public office for women.

castes and communities that have been identified by the government as disadvantaged, even though they were not considered untouchable, and so are not stigmatized in the same way as people from Scheduled Castes. When we refer to "lower castes," we are referring to SCs and OBCs collectively. People who do not fit into any one of the three broad and internally diverse categories are not eligible for affirmative action programs. In the rest of the paper, we will use the term "general caste," instead of "higher caste," when referring to people who, by virtue of their caste identifty, are ineligible for affirmative action programs. This is the term that the National Family Health Survey, our data source, uses in its survey documentation.

3 Conceptual framework: Child height and the social context

Processes of discrimination affect people in the ST, OBC, and SC categories differently. STs are the poorest of these four groups, and, as we will show, tend to live in localities with other STs. In government circles and in the popular imagination, the regions where STs live are considered "remote" and difficult to serve. Although the government has many on-paper laws and policies to promote development in predominately ST regions, in practice infrastructure and public services in these regions are sorely lacking (Jones, 1978; Mosse, 2005). Literature on health and human development outcomes among STs focuses on how STs suffer from exclusion from government services as well as from displacement from productive land (Nathan and Xaxa, 2012).

People from lower castes, on the other hand, face discrimination of a different nature. They are not stereotyped as "remote" – many SCs and OBCs live in heterogenous localities, alongside people from higher castes. Many lower caste households have historically been expected to work for higher caste households. Sociologists of India have described the ways in which these cross-caste economic and social relationships have changed over time, but nevertheless remain an important part of village life (Srinivas, 1955; Jeffrey, 2001; Kumar, 2016). Therefore, although lower caste people certainly face institutional discrimination, they are also exposed to day-to-day discrimination in their interactions with people from higher castes.

This is particularly true of people from the Scheduled Castes, who were once considered "untouchable." Untouchability status is often justified with reference to hereditary occupation: people from castes that performed menial and ritually "impure" jobs, such as slaughtering animals for meat and leather, and manual scavenging (which involves physically carrying human excreta for disposal) are considered to themselves be unclean. The fact that people from Scheduled Castes do or have done dirty work has been used to justify excluding them from water sources, temples, social events, and at one time, even schools (see Valmiki, 2003). It was also used to justify a number of discriminatory practices of daily interaction. For example, in some villages people from untouchable castes were not allowed to sit in front of people from higher castes, or were not allowed to wear shoes in village lanes. In most places, untouchables and non-untouchables were prohibited from eating together (Ambedkar, 2014; Srinivas, 1976). The specific ways in which untouchability is enforced vary from place to place; what these practices have in common is that they are intended to exploit, exclude, and humiliate. Although there is evidence that the most severe practices of untouchability are less common in rural India than they were a few decades ago (Shah et al., 2006), this discrimination nevertheless persists: Thorat and Joshi (2015) analyze the nationally-representative 2012 India Human Development Survey and find that about a third of households admit that at least one member practices untouchability.

How do these processes of discrimination shape child height? We hypothesize that the neglect and social exclusion of STs will be reflected in their socioeconomic status, and that this will explain the height disadvantage of ST children relative to general caste children. There is, after all, considerable evidence that variables capturing economic status and education predict child height, likely in part because they are correlated with the quality and quantity of food and other health inputs that young children need to grow (Desai and Alva, 1998; Case, Lubotsky and Paxson, 2002). For lower castes, we hypothesize that exposure to higher caste people in their locality will also matter. This is because higher caste neighbors might enforce the social rank of lower caste households, especially SCs, in ways that could create stress, and limit access to common resources, like clean water, which would matter for child health, but would not show up in household economic status. We would not expect neighbors of similar caste-rank, however, to have the same detrimental effects on stress or access to common resources. Although our data will not permit us to definitively pin down the mechanisms linking local exposure to higher caste neighbors and child height, we discuss possible mechanisms and areas for future research in Section 6. Maternal stress, which has been shown to matter for child health in other contexts (Torche, 2011; Lauderdale, 2006), and differential disease exposure, which has been shown to be particularly relevant for height in the South Asian context (Hathi et al., 2017), may be promising areas for future research.

4 Data & empirical strategy

4.1 Data

Our sample includes children in rural households measured by the 2005 National Family Health Survey-3 (NFHS-3), India's Demographic and Health Survey (DHS). The NFHS-3 is a nationally representative two-stage random sample survey. Throughout the paper, we use sampling weights provided by the NFHS.

Health outcome. The NFHS-3 measures the heights of children under 5 years old. Throughout the paper, we scale child height according to the 2006 WHO international reference population (Onis, 2006).[9] This transforms measured child height into height-for-age z-scores, or differences between Indian children and children of the same age and sex in a healthy reference population. Extensive field verification has documented that the WHO norms are appropriate for Indian children. For example, children raised in affluent south Delhi grow, on average, to the international norms (Bhandari et al., 2002).

Population group data. We follow the division of Indian children into the four population groups as described above and recorded in the NFHS-3: SC, ST, OBC, and general. Data on population group are missing for 3.5 percent of the rural children whose heights were measured by the DHS.[10] Summary statistics in Table 1 are presented separately for each of the four population groups. In the weighted sample of children we study, 21.7 percent are SC, 11.9 percent are ST, 42.7 percent of children are OBC, and the remaining 23.7 percent are from a general caste background.

SES variables used in the demographic decomposition. Below, we ask what the average height of ST, SC, and OBC children would be if the distribution of two core socioeconomic characteristics among each of these three groups of children matched the distribution of these characteristics among general caste children. The two characteristics we use in the reweighting decomposition are the type of floor that the child's household has (4 categories),[11] and the education level of her mother (6 categories). Summary statistics for each floor type and mother's education level are given in Table 1.

[9] As the WHO recommends, we exclude children more than 6 standard deviations from the mean of the reference population.

[10] Table A1 in the Supplementary Appendix shows the fraction of children with measured heights for whom social group data are missing. Muslims are more likely than Hindus to have missing caste data: 17% of rural Muslim children do not have a caste designation in the NFHS-3.

[11] Table A2 in the Supplementary Appendix presents the results of OLS regressions showing that floor type is associated with height among children in rural India.

As we discuss below, the number of socioeconomic variables used in a reweighting decomposition is necessarily small because reweighting is done non-parametrically, using the 24 intersecting bins of these two variables.

SES variables used in regression results. We conduct a further analysis of the process of discrimination using a regression analysis that permits us to control for a wider range of SES variables in addition to floor type and mother's education, which are used in the demographic decomposition. In the regression analysis, we additionally control for indicators for every combination of mother's education; household electricity; phone, radio, TV, refrigerator, bicycle, motorcycle, car ownership; land ownership; and whether the household uses a toilet or latrine. These variables are summarized in Table 1. These summary statistics show that SC, OBC, and ST children grow up in poorer households, on average, than general caste children.

Measuring the local social context. We construct a local social context variable that depends on both a child's caste group and the caste composition of her locality. The NFHS data permit measurement of local context variables at the Primary Sampling Unit (PSU) level.[12] The sampling frame for rural PSUs in the NFHS-3 was villages in the 2001 Census.[13] In large villages (greater than 500 households), PSUs were selected from clusters of 100-200 households within the village.[14]

PSUs are a useful measure of place for studying local caste interactions because they capture the caste status of those households that live in closest proximity to the observed child. Yet, considering how important the measurement of place can be to our understanding of the extent, causes, and consequences of segregation (Lee et al., 2008; Reardon et al., 2008), future research using higher geographic levels of aggregation, such as the village or the block, may produce additional informative results.

The NFHS-3 randomly selected approximately 20 households in each PSU to interview.[15] We

[12] Another recent paper which examines the effect of PSU-level variables on child health is Hathi et al. (2017); they use the fraction of households that do not use a toilet or latrine in a PSU as a measure of disease externalities.

[13] According to the Office of the Registrar General and the Census Commissioner of India, for the purposes of the 2001 Census, a village was defined as "the smallest area of habitation, viz., the village generally follows the limits of a revenue village that is recognised by the normal district administration" (Government of India, 2001). This geographic unit is different from a *gram panchayat*, which is a village or cluster of villages represented by an elected local leader.

[14] To further elaborate on the NFHS-3 study design, the survey manual explains that PSUs and households are selected in the following way: "A uniform sample design was adopted in all states. In each state, the rural sample was selected in two stages, with the selection of Primary Sampling Units (PSUs), which are villages, with probability proportional to population size (PPS) at the first stage, followed by the random selection of households within each PSU in the second stage." (IIPS & Macro International, 2005, pp.12-23).

[15] Because not all households in a PSU are surveyed, any constructed PSU mean (such as these fractions) is an unbiased estimate with sampling error of the true PSU mean. This sampling error will tend to attenuate estimated effects of PSU means, with the consequence that the importance of local caste composition for child health outcomes could be even greater

use the caste composition of the sampled households to construct a measure of her household's local caste rank. This variable is only defined for SC, OBC, and general caste children; ST children are omitted from this analysis because their households are not comparably ranked by the caste system. We operationalize local caste rank by constructing a *fraction higher ranking* variable, which is defined as follows:

- For all general caste children, this variable is zero.
- For OBC children, this variable is defined as the fraction of non-ST households in their PSU that are general caste.
- For SC children, this variable is defined as the fraction of non-ST households in their PSU that are general caste or OBC.

This variable reflects an interaction between household and village properties: two children living in the same village will have different values if they belong to different caste groups. Similarly, two SC children in different villages may have different values ranging from close to 0 (if their village is almost all SC) to close to 1 (if their village is almost all OBC or general caste). We do not summarize the *fraction higher ranking* variable in Table 1 but instead show the cumulative density of this variable for OBCs and SCs in Figure 1.

Variables used in placebo test & robustness checks. In order to verify that our main result – that SES together with local caste rank explains height gaps between general and lower caste children – is not driven by other variables that influence height, we present a placebo test and several robustness checks. For the placebo test, instead of using the *fraction higher ranking* variable described above, we use the fraction of households in child's PSU that have a higher NFHS asset index (also known as the "wealth index") score than that child's household.[16]

For the robustness checks, we add variables about local infrastructure and the caste composition of the child's village. Local infrastructure variables are the fraction of households in the child's PSU that have electricity and the fraction of last births to a given mother in a child's PSU that received prenatal care. Caste composition variables are the fraction of households in the child's PSU belonging to Scheduled Castes, the fraction belonging to Other Backward Classes, and the fraction belonging

than we document.

[16]For more on the construction of the NFHS "wealth index," see IIPS and Macro International, 2005.

to general castes. We also add controls for demographic variables including: number of household members, birth order of the child, sex of the child, and whether the child lives with her paternal grandparents in a joint family.[17] Means of these variables by caste group can be found in Table 1.

4.2 Empirical strategy

We first ask: what would the average height of rural ST, SC, and OBC children be if the distribution of socioeconomic characteristics among these groups of children matched the distribution of socioeconomic characteristics among rural general caste children? We find that SES variables largely explain the height gap for STs, but not for SCs and OBCs. For these groups, the remaining height gap is explained by local processes related to social inequality.

We use two complementary empirical strategies to arrive at these results. First, we use a reweighting decomposition to quantify the fraction of height disparities that can be explained by SES. Then, we use regression to show that the remaining gap can be explained by the fraction of a child's locality that outranks her family in the caste hierarchy.

Reweighting decomposition method. We apply a reweighting decomposition similar to that proposed by DiNardo, Fortin and Lemieux (1996) and implemented by Geruso (2012) and Coffey (2015) in prior research on health disparities. The reweighting function, Ψ, that we use to produce the means and confidence intervals presented in figure 3 is defined as

$$\Psi(\mathbf{x}) = \frac{f(\mathbf{x}|g=1)}{f(\mathbf{x}|g=0)}, \tag{1}$$

where $\mathbf{x}$ is a single set of indicators for the intersections of 4 categories of floor type in the child's households, and 6 categories for the educational attainment of her mother. In this case, reweighting is done over 24 floor type/education category bins. The function $f(x|g)$ is the empirical probability mass function for bin x among the general caste population ($g = 1$) or a disadvantaged population ($g = 0$). In other words, $f(x|g)$ is the fraction of the population group g sample in SES bin x, computed using survey sampling weights. The reweighting function $\Psi(\mathbf{x})$ is multiplied be the sampling weight of each observation in the disadvantaged population, so that sample means can be computed for a counterfactual disadvantaged population that matches the distribution of SES of the general

[17] See Allendorf (2013) for an investigation of the association between living in a nuclear or joint family on Indian women's health.

caste population. Thus, the counterfacutal reweighted mean height is computed as

$$\bar{h}^{RW} = \frac{\Sigma_i \Psi(\mathbf{x_i}) w_i h_i}{\Sigma_i \Psi(\mathbf{x_i}) w_i}, \tag{2}$$

where h_i is the height of child i; x_i is the SES bin of child i; and w_i is the survey sampling weight of child i.

Unlike a regression approach which only matches the mean, this reweighting strategy has the advantage of matching the full distribution of these SES variables among general caste children. Further, it flexibly allows any non-parametric interaction between the education and floor type variables, and does not require any *ad hoc* combination of variables into a single SES index.[18] Instead, for example, the few lower caste children who have highly educated mothers and good housing (as measured by floor material) would receive a large weight in the reweighted calculation of mean SC height, because these are the children with a large Ψ, meaning that their SES matches a larger fraction of higher caste children than lower caste children. Therefore, if these children are much taller than the average SC child, then the reweighed average SC height will increase. The extent to which it increases is the measure of the amount of height difference that is due to SES.

However, the non-parametric nature of the approach inherently limits the number of SES variables that can be used: if the sample is partitioned into many bins, computing reweighted mean heights for disadvantaged groups would not be possible without dropping some general caste children from the sample. This is because the denominator in equation 1 would be zero if there are general caste children who have no counterparts among the sample for the relevant disadvantaged group. When we reweight over only floor type and mother's education, no general caste children need to be dropped in order to compute reweighted mean height. This dimensionality limit is one motivation for our further regression strategy.

Modified reweighting decomposition using local caste rank. The reweighting method reports what the mean heights of rural ST, SC, and OBC children would be if they were exposed to the same distribution of SES variables as rural general caste children. The remaining gap can be interpreted as the average consequence for height of discrimination. But it may be that not all disadvantaged children are exposed to the *average* level of discrimination. If effects of discrimination on height in part reflect the role of *local* processes, then we might expect post-SES differences to differ across

[18]See Filmer and Pritchett (2001) for more on creating an index with asset variables in DHS data.

localities.

We use the *fraction higher ranking* variable which was described above to measure a child's exposure to people of a higher caste rank in her localilty. However, we cannot use the *fraction higher ranking* variable in the reweighting decomposition because *all* general caste children have a *fraction higher ranking* of zero, while only a very few SC and OBC children live in sufficiently segregated localities to have such a very low *fraction higher ranking*.

Therefore, in order to investigate the role of local discriminatory processes, we compute many replications of the entire reweighting decomposition, each time progressively restricting the SC and OBC samples to those with smaller and smaller values of *fraction higher ranking*. That is, we restrict the SC and OBC samples to children with *fraction higher ranking* less than 0.9, and compute reweighted heights. Then, we restrict the SC and OBC samples to children with *fraction higher ranking* less than 0.88, and so on,[19] until eventually only those SC and OBC children with the smallest *fraction higher ranking* values – that is, children who live in PSUs in which they are surrounded by almost entirely SC, or SC and OBC neighbors, respectively – are included in the reweighting decomposition. This produces a *sequence* of counterfactual SC and OBC average heights, each reweighted to match the general caste distribution of SES, but concentrating on SC and OBC children exposed to different levels of local social rank.

Regression method using local caste rank. To verify our results using another method, we run ordinary least squares regressions in which we control for bins of intersecting SES indicators. In contrast with the reweighting decomposition, the regression method allows us to control for a larger set of SES indicators, as well as to control for other variables that might influence child height.

We run regressions of the following form, where each observation is a child under 5 years old:

$$height_{ip} = \beta_0 + \underbrace{\beta_1 SC_{ip} + \beta_2 OBC_{ip} + \beta_3 ST_{ip}}_{\text{group indicators}} + \alpha_{ip} + X_{ip}\theta + \varepsilon_{ip}. \quad (3)$$

height is a child's height-for-age *z*-score; *i* indexes children; and *p* indexes survey PSUs. The variables *SC*, *OBC*, and *ST* are each indicator variables for group membership, with general caste children as the omitted category. α_{ip} are dummy-variable indicators of socioeconomic status that vary across specifications to explore the roles of different types of explanatory factors. The coefficient estimates

[19] We proceed in steps of 0.02.

$\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ can be interpreted as the remaining unexplained discrimination after controlling for these explanatory factors. X_{ip} is a vector of village caste composition and demographic controls, described above, that will be added as a robustness check. We cluster standard errors by survey PSU.

We estimate several regressions of this form. First, we compute average differences in child height by running a regression with no explanatory factors. Then, to estimate differences in child height at the same level of SES, we control for a set of dummy variables about the household's SES. Unlike the decomposition approach, which matches on the entire distribution of included variables, the regression approach matches only on means, so it allows us to control for larger number of SES variables. We use indicators that control for every combination of mother's education category, floor type, use of a toilet or latrine, household electricity, and ownership of a phone, radio, TV, refrigerator, bicycle, motorcycle, car, and land. Any height gap that remains after controlling for these indicators reflects a difference that persists even after very detailed SES information has been accounted for.

The further regression analysis investigates the role of local discriminatory processes for SCs and OBCs, using bins constructed from the *fraction higher ranking* variable. From this point forward, ST children are omitted from the regression because their height gap has already been fully statistically explained by the SES variables. We continue our regression analysis by replacing the SES indicators with a new set of indicators that combine SES and local caste rank: each is an indicator for a bin at the intersection of the prior SES bins and the deciles of the *fraction higher ranking* variable. Thus, with these new control variables, we account for the consequences of both SES and local caste rank.

We also perform a placebo test and robustness checks using other variables that may influence child height. For the placebo test, we interact a child's SES bins with deciles of the *local wealth rank* variable described above. Controlling for *local wealth rank* does not account for gaps in child height in the same way that controlling for *fraction higher ranking* does. For the robustness checks, we add the neighborhood composition and demographic controls described above.

5 Results

5.1 Descriptive statistics

The apparent similarity of the SC and ST height deficits. It is common in studies of social disadvantage in India to refer simultaneously to SCs and STs as a collective set of the most disadvantaged people in India.[20] Although one conclusion of this paper is that the disadvantaged health outcomes of SCs and STs reflect different *processes*, Table 2 documents that their quantitative *levels* of height disadvantage are strikingly similar. This can also be seen is Supplementary Appendix Figure A1, which plots the empirical cumulative distribution of height-for-age among children in each population group.

Column 1 of Table 2 shows that rural SC and ST children are shorter than rural general caste children by about four-tenths of a height-for-age standard deviation. Columns 2 through 4 add an "SC or ST" indicator variable, as is commonly found in the literature. So similar are the magnitudes of the ST and SC height gaps that neither a separate SC indicator nor a separate ST indicator appears to add explanatory power to the models in columns 3 and 4.[21] A researcher who does not investigate the processes of discrimination may conclude from statistical tests like those in Table 2 that there is no reason to separate these categories. However, we will see that the unconditional correlations between membership in these categories and child height reflect different processes.

The distribution of local caste contexts. Figure 1 shows the cumulative distribution of the fraction of households in a child's PSU that are the *same* caste as hers by population groups. It is evident from Figure 1 that ST households are more likely than members of other population groups to live in separate localities, rather than in localities with households belonging to other groups. This is because a child whose value of the plotted variable is 1 lives in a locality where all households are of the same population group as she is. Many ST children are found massed at 1 – that is, they live in ST-only localities. Further, the ST distribution stochastically dominates the other distributions. SC children, in contrast, are generally found in localities where the majority of households are not SC.

Figure 2 plots the distribution of *fraction higher ranking*, our key explanatory variable, for OBC and SC children. It is clear from both figures that there is wide variation across rural India, and that different population groups experience different patterns of local caste composition.

[20]For example, several states, including Delhi and Bihar, have a department of "SC and ST welfare." Employing what is a standard practice in the literature, both Burgess, Pande and Wong (2005) and Azam and Bhatt (2015) (selected as examples of high-quality research) employ a single, combined categorical variable for "SC or ST." Of course, whether or not such a combined indicator is inappropriate for studying height would not necessarily imply it is inappropriate for studying other outcomes.

[21]This is visible in the fact that neither the coefficient on SC in column 3, nor the coefficient on ST in column 4 are statistically significant. Note also that R^2 is unchanged.

5.2 Reweighting decomposition results

The fraction explained by SES. Figure 3 presents the first of our reweighting decomposition results. It plots the fractions of height gaps between general and ST, SC, and OBC children that can be explained by SES variables.[22] The reweighting function recomputes the height gap after producing a counterfactual sample of disadvantaged children, reweighted to match the SES distribution of general caste children. Here, the apparently similar SC and ST gaps diverge. The ST-general gap can be completely accounted for by SES, in the sense that the reweighted gap is small, and the confidence interval of the reweighted gap includes zero. This is likely because ST households are particularly poor, and are likely to live in isolated rural villages that are underserved by public resources, such as schools. Indeed, in a further result unreported in the figure, ST children who live in villages that are the most segregated, that is where the fraction of ST households in the village is larger, are shorter, on average. This difference, too, is accounted for by differences in SES.

In contrast, for SC and OBC children, the unexplained gap remains large and statistically significantly different from zero when the same SES variables are used in the reweighting decomposition. The point estimate of the SC gap unexplained by SES is approximately twice the point estimate of the remaining ST gap unexplained by SES. This suggests that SC children are not short merely because they are poor, but perhaps also due to a form of discrimination that is not measured by their SES. The remainder of this paper further investigates the remaining SC and OBC height gaps.

The role of local caste rank. We now turn to an investigation of the role of local caste rank. This will principally be pursued through the parametric regression analysis described in section 5.3. However, we first present results from a modified reweighting decomposition that tells a story that is similar to what we find in the regression analysis. ST children are excluded from these analyses because their height gap has been fully accounted for by SES.

Figure 4 presents the results of a series of reweighting decompositions. The horizontal axis of Figure 4 represents progressive restriction of the samples of SC and OBC children. At the far left of the figure, all SC and OBC children are included in the reweighting decomposition. Therefore, the point estimates match those in Figure 3. As the lines move along the horizontal axis to the right, the SC and OBC samples are restricted to include only those children exposed to that *fraction higher*

[22] If we use father's education instead of mother's education in the reweighting decomposition, the results are qualitatively similar. Figure A2 in the Supplementary Appendix shows the same comparisons as the ones shown in Figure 3 using father's education instead of mother's education.

ranking level or lower. At the far right end of the figure, the only SC and OBC children included in the SES reweighting are those with few or no households of higher caste rank in their localities.

Two aspects of the figure point to an important role for local discrimination. First, for both SC and OBC children, the curves slope downwards. This indicates that the post-SES gap between lower caste and higher caste children is increasing in the fraction of the neighborhood that locally outranks a lower caste child. Second, the point estimates at the far right end of the graph are at or close to zero.[23] This suggests that lower caste children who are not outranked are no shorter, on average, than similarly low-SES general caste children living in other localities.

5.3 Explaining the process of discrimination with regression

Kernel-weighted local regression. Figure 4 suggests that SC and OBC children do not experience height disadvantages that cannot be explained by their SES in neighborhoods where they are not locally outranked. Figure 5 moves towards testing this hypothesis directly. The vertical axis of figure 4 plots the residuals of height-for-age, after they are regressed on a detailed set of SES controls, which are described in section 4.1. The horizontal axis is the explanatory variable *fraction higher ranking*. Because SES has already been residualized out of the vertical axis, any apparent association is an association with local social rank net of SES. This semi-parametric technique makes no assumption about the functional form of the relationship between the two variables. For both SC and OBC children, the lines slope down: children whose households are outranked by a larger fraction of households in their locality are shorter, on average, net of SES.[24]

Results of OLS regression analysis. Testing the robustness and statistical significance of these conclusions requires parametric regression analysis, introduced in equation 3. Estimates are presented in Table 3.[25] The first two columns test the robustness of the conclusions of the decomposition analysis by applying ordinary least squares regression with controls. Column 1 presents the apparently similar SC and ST gaps. Column 2 shows that controlling for SES as described in Section 4 eliminates the ST gap but not the SC or the OBC gap, and leaves the point estimate for the post-SES SC gap

[23] The SC line ends before the OBC line because very few SC children live in localities with no higher ranking households — beyond this point, there are too few SC children to reweight on a set of SES bins.

[24] Note that this figure, drawn only for the disadvantaged SC and OBC groups, includes mainly negative averages of height for age residuals because the residualizing regression includes general caste children, who all have a *fraction higher ranking* of zero.

[25] Motivated by the concern that the caste rank and composition of OBC categorization varies throughout India, but that SCs are always of the lowest-ranking castes, Appendix Table A3 verifies that our regression results are robust – quantitatively and in their qualitative pattern – to excluding OBC children and focusing only on the SC-general height gap.

twice as large as the point estimate for the post-SES ST gap.

The remaining columns investigate the role of local discriminatory processes. Column 3 verifies that the coefficients in column 2 on SC and OBC are unchanged after dropping ST children from the regression. Column 4 adds controls for *fraction higher ranking* by replacing the SES indicators with a new set of indicators for the intersection of each of the SES bins and decile categories of *fraction higher ranking*. Column 4 shows that adding local caste rank completely accounts for the remaining height gap.

The remaining columns present robustness and placebo tests. Column 5 is a placebo test which repeats the regression from column 4, except it interacts deciles for a child's household's *wealth* rank in the neighborhood, rather than deciles of *fraction higher ranking*, with her SES bin. This placebo test verifies that it is local *caste* rank that explains the height gap, rather than *economic* rank or local rank more generally. The coefficients in column 5 are very similar to those in column 3, which verifies that our finding is specific to local caste rank.

Columns 6 and 7 are a robustness check that replicates columns 3 and 4, but that include two further sets of regression controls: a set of neighborhood composition controls and demographic controls, as described in section 4.1.[26] The purpose of the neighborhood composition controls, which control for the fraction of the neighborhood's households belonging to each population group, is to demonstrate that there is not a spurious correlation between *fraction higher ranking* of neighborhood caste composition.[27] The conclusions of columns 3 and 4 are qualitatively confirmed in columns 6 and 7: large SC and OBC gaps persist even after the SES indicators and additional controls (column 6) are added, but this gap is fully accounted for by adding controls for *fraction higher ranking* (column 7).

The results of three further robustness checks – one related to child sex and one related to the local disease environment – are presented in the Supplementary Appendix. Because child health in India often differs by sex, Table A4 presents our regression results separately for boys and girls. We find similar patterns to what we found in the main results presented in Table 3. Table A5 in the Supplementary Appendix shows the results of Table 3 with an additional control for the fraction of

[26]Each of these, including quantitative counts such as household size and birth order, is implemented as a semi-parametric set of indicators.

[27]Note that such an omitted variable bias concern would be anyway misguided because *fraction higher ranking* is an *interaction* between neighborhood composition and the population group membership of the child: increasing the fraction of a village that is OBC at the expense of SCs and generals, for example, would increase *fraction higher ranking* for SC children and decrease it for OBC children.

households in a child's primary sampling unit that defecate in the open, rather than use a toilet or latrine. This is an important control because prior literature has shown that this measure of the local sanitation and the disease environment is strongly predictive of child height in South Asia (Spears, 2013; Hathi et al., 2017). Although the fraction of households that defecate in the open statistically significantly predicts child height in this sample as well (children living in PSUs where all households defecate in the open are a quarter of a standard deviation shorter, on average, than children living in PSUs where no household defecates in the open), the additional of this control does not change the main results from Table 3. Finally, Table A6 shows the results of Table A5 with additional controls for the number of siblings a child has, her household's religion, and additional asset controls.

The results of different statistical methods tell a consistent story: although SES can account for the ST-general height gap, it cannot account for the height deficits that SC and OBC children suffer. SC and OBC children appear to face a process of *local* discrimination based on the caste rank of the households in their locality.

6 Discussion

In India, hundreds of millions of people have been born into disadvantaged ST, OBC, and SC households. This disadvantage is evidenced not only in their economic and social lives, but also in the heights of their children. We find that the SC–general and ST–general height gaps are quantitatively similar, but in fact reflect different processes of SES deficits and local experiences of discrimination. These results suggest that, at least for some outcomes, empirical studies should not use "SC or ST" as a single indicator of disadvantage.[28]

The finding that the ST–general caste height gap can be explained by differences in socioeconomic status coheres with prior literature that finds that household socioeconomic status is strongly correlated with child height Currie (2009). It also highlights a further consequence of the neglect of ST regions discussed in section 3. Although encouraging government investment in ST regions, and ensuring that such investments reach the poor may not be straightforward (Mosse, 2005), this finding suggests that investment that improves the economic and educational status of parents could be effective at reducing the height gap.

[28]We do not mean to suggest that there are not important differences *within* SC and ST populations. Indeed, documenting health disparities by sub-caste within SCs, or by tribe within STs, would be useful in further elucidating both the extent of and processes behind health inequality in India.

The finding that, for lower caste children, it is an *advantage* to live in SC-dominated localities presents an informative contrast to research from other countries that finds that segregation is associated with worse health outcomes among disadvantaged groups (Massey, 1990; Williams and Collins, 1995, 2001; Kawachi and Berkman, 2003; LaVeist et al., 2011; Leung and Takeuchi, 2011). Although it is beyond the scope of this paper to present a comparative analysis of the effects of segregation, we note that the literature on racial segregation in the United States emphasizes how segregation concentrates poverty and creates what Massey (2004) calls a "uniquely disadvantaged social environment characterized by high rates of joblessness, welfare, dependency, substance abuse, and single parenthood" (p. 17).[29]

In section 3, we proposed that, in the Indian context, the kind of day-to-day discrimination that arises from interaction between lower and higher castes may be useful in understanding our results. Here, we will discuss what additional data would be useful to further explore this hypothesis. We also discuss the possible mechanisms that we propose can be ruled out.

Ethnographic accounts provide evidence that lower caste people who live near households from higher castes experience high levels of stress and violence (Valmiki, 2003; Srinivas, 1976). This is in part because, where SCs and higher castes live close together, SC households have historically worked for higher caste households as part of an exploitative economic system supported by discriminatory rules limiting their use of common resources, including water (Ambedkar, 2014; Shah et al., 2006). Future research might document stress levels among lower caste people living in different kinds of localities. In order to better understand the effects of local context on child height in particular, this research should focus on measuring stress among pregnant women.

Although one of the robustness checks presented in Section 5 found that controlling for village-level mean open defecation in a child's primary sampling unit did not change our main results, future research may identify a mechanism for these results related to differential exposure to enteric infection. This is because the negative effects of a PSU's open defecation may not be equally distributed: survey data on rural sanitation from five states in north India finds that SC respondents are more likely than non-SC respondents to report seeing people defecating near their houses.[30,31] Table A7 in

[29] We note that literature on ethnic and immigrant enclaves, a different form of residential segregation, finds mixed associations between living in segregated neighborhoods and health outcomes (Xie and Gough, 2011; Osypuk, Bates and Acevedo-Garcia, 2010).

[30] Coffey et al. (2014) present details on about the goals, methods, and main findings of this survey.

[31] We note that Om Prakash Valmiki, author *Joothan: A dalit's life*, opens his autobiography by describing how, when he was growing up, his neighbors would defecate on the shores of a pond next to his and other SC people's houses (Valmiki,

the Supplementary Appendix summarizes this result. If open defecation is more likely to occur near SC homes in mixed localities than in homogenous ones, then the same level of open defecation may harm SC children by more than it harms other children. But this evidence can only take us so far: ideally, we would have data on whether SC and OBC children in mixed localities experience more enteric infection than SC and OBC children at the same level of SES in homogeneous localities.

Although possible, we find it unlikely that selection of unhealthy lower caste children into mixed localities could explain our results. This is because migration in this context is very low. Although migration for marriage is common for women (Rosenzweig and Stark, 1989), and men in many parts of India engage in temporary labor migration (Deshingkar and Farrington, 2009), rates of permanent internal migration are considerably lower than in other developing countries (Deshingkar and Anderson, 2004). Anderson (2011) reviews the literature on the caste composition of villages in Uttar Pradesh and Bihar, two of India's largest states and concludes that in most cases "the origins of the distribution of caste groups at the village level go back hundreds of years" (242).[32]

Another possible, but, we think, unlikely explanation for our results relates to the measurement of the local caste context. For most, but not all, of the children we study, the child's PSU and her village are one and the same. As we discuss in Section 4, for children in villages of over 500 households, PSUs are geographically proximate groups of 100-200 households. It is therefore possible that fraction of households in a child's PSU that outranks hers is correlated with village size, which prior research has shown to be correlated with village development indicators (Singh, Chakraborty and Roy, 2008). Unfortunately, the NFHS does not report village size. However, other findings, which we present in the Supplementary Appendix, suggest that this way of measuring the local context is unlikely to be responsible for our results. First, Figure A3 uses India Human Development Survey (IHDS) data (Desai et al., 2005) to show that SC children in villages with more than 500 households are not taller than SC children in villages with less than 500 households. Table A8 verifies this result is robust to state fixed effects and other controls. Second, Figure A4 plots both the cumulative distri-

2003, p. 1).

[32] Anderson (2011) finds that agricultural yields of lower caste households are higher in villages with only SCs, or SCs and OBCs, than in villages where general castes live as well. This difference can be explained by the fact that lower caste households in homogenous villages are better able to negotiate irrigation for their crops than those living in villages with general caste households. Although this study suggests that the economic variables measured by the DHS may not be able to paint a full picture of differences in households' economic situations across village types – the DHS does not measure, for example, agricultural yields – research documenting tenuous links between agriculture and child anthropometry (Gillespie, Harris and Kadiyala, 2012) suggest to us that if we had data on agricultural production, Anderson (2011)'s finding about caste-heterogenous villages would be unlikely to be able to explain our results.

bution of the fraction of *sampled* households in an NFHS PSU that are SC and then — using Census data — the cumulate distribution of the village-level fraction of *all* households in a village that are SC. The two CDFs are very similar, suggesting that results based on measuring caste composition at the village level, rather than at the PSU level, would be similar. Nevertheless, we hope that these findings will generate further study – using different measures and aggregations of local context – to understand of exactly how social inequality contributes to child health outcomes in rural India.

7 Conclusion

Our results are informative for enduring debates in India about how social policy can best respond to discrimination against lower castes and adivasis.[33] Much of the debate about addressing social inequality has been focused on national or state-level policies of affirmative action, such as admission at government universities, or allocation of government jobs. Although research highlights the success of these policies, including in improving the representation of marginalized communities without reducing the efficiency of public services (Deshpande and Weisskopf, 2014), and of in improving class-composition of higher education as well as the diversity of social backgrounds (Bertrand, Hanna and Mullainathan, 2010), we note that these affirmative action policies may be insufficient to respond to local discrimination that appears to have effects early in life. We recommend that future research seek to understand exactly how social inequality gets under the skin in order to inform policies that target local processes and weaken the link between social group and health.

[33]It is important to note that local processes of caste discrimination may limit the effectiveness of existing programs intended to promote health and nutrition, such as India's public distribution of food, or its rural sanitation programs (Thorat and Lee, 2005; Lamba and Spears, 2013).

References

Acharya, Sanghmitra. 2012. "Caste and Patterns of Discrimination in Rural Public Health Care Services." In *Blocked by caste: Economic Discrimination in Modern India.* , ed. Sukhdeo Thorat and Katherine Newman. Oxford University Press.

Allendorf, Keera. 2013. "Going nuclear? Family structure and young women's health in India, 1992–2006." *Demography*, 50(3): 853–880.

Ambedkar, Bhimrao Ramji. 2014. *Annihilation of Caste.* Verso Books.

Anderson, Siwan. 2011. "Caste as an Impediment to Trade." *American Economic Journal: Applied Economics*, 3(1): 239–263.

Arnold, Fred, Minja Kim Choe, and Tarun K Roy. 1998. "Son preference, the family-building process and child mortality in India." *Population Studies*, 52(3): 301–315.

Azam, Mehtabul, and Vipul Bhatt. 2015. "Like Father, Like Son? Intergenerational Educational Mobility in India." *Demography*, 52(6): 1929–1959.

Barcellos, Silvia Helena, Leandro S Carvalho, and Adriana Lleras-Muney. 2014. "Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics*, 6(1): 157–189.

Baru, Rama, Arnab Acharya, Sanghmitra Acharya, AK Shiva Kumar, and K Nagaraj. 2010. "Inequities in access to health services in India: Caste, class and region." *Economic and Political Weekly*, 49–58.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. 2010. "Affirmative action in education: Evidence from engineering college admissions in India." *Journal of Public Economics*, 94(1-2): 16–29.

Bhalotra, Sonia, Christine Valente, and Arthur van Soest. 2010. "The puzzle of Muslim advantage in child survival in India." *Journal of Health Economics*, 29(2): 191–204.

Bhandari, Nita, Rajiv Bahl, Sunita Taneja, Mercedes de Onis, and Maharaj K. Bhan. 2002. "Growth performance of affluent Indian children is similar to that in developed countries." *Bulletin of the World Health Organization*, 80(3): 189–195.

Borooah, Vani. 2012. "Inequality in Health Outcomes in India: The Role of Caste and Religion." In *Blocked by caste: Economic discrimination in modern India.* , ed. Sukhdeo Thorat and Katherine Newman. Oxford University Press.

Borooah, Vani K. 2005. "Caste, inequality, and poverty in India." *Review of Development Economics*, 9(3): 399–414.

Brainerd, Elizabeth, and Nidhiya Menon. 2015. "Religion and health in early childhood: Evidence from South Asia." *Population and Development Review*, 41(3): 439–463.

Burgard, Sarah. 2002. "Does race matter? Children's height in Brazil and South Africa." *Demography*, 39(4): 763–790.

Burgess, Robin, Rohini Pande, and Grace Wong. 2005. "Banking for the poor: Evidence from India." *Journal of the European Economic Association*, 3(2-3): 268–278.

Case, Anne, and Christina Paxson. 2010. "Causes and consequences of early-life health." *Demography*, 47(1): S65–S85.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic status and health in childhood: The origins of the gradient." *American Economic Review*, 92(5): 1308–1334.

Coffey, Diane. 2015. "Prepregnancy body mass and weight gain during pregnancy in India and sub-Saharan Africa." *Proceedings of the National Academy of Sciences*, 112(11): 3302–3307.

Coffey, Diane, Aashish Gupta, Payal Hathi, Nidhi Khurana, Dean Spears, Nikhil Srivastav, and Sangita Vyas. 2014. "Revealed preference for open defecation." *Economic & Political Weekly*, 49(38): 43.

Cramer, James C. 1995. "Racial and ethnic differences in birthweight: The role of income and financial assistance." *Demography*, 32(2): 231–247.

Cummins, Steven, Sarah Curtis, Ana V Diez-Roux, and Sally Macintyre. 2007. "Understanding and representing 'place' in health research: a relational approach." *Social science & medicine*, 65(9): 1825–1838.

Currie, Janet. 2009. "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of economic Literature*, 47(1): 87–122.

Das Gupta, Monica. 1987. "Selective discrimination against female children in rural Punjab, India." *Population and Development Review*, 77–100.

Deaton, Angus. 2007. "Height, health and development." *Proceedings of the National Academy of the Sciences*, 104(33): 13232–13237.

Desai, Sonalde, and Amaresh Dubey. 2012. "Caste in 21st century India: competing narratives." *Economic and political weekly*, 46(11): 40.

Desai, Sonalde, and Gheda Temsah. 2014. "Muslim and Hindu women's public and private behaviors: Gender, family, and communalized politics in India." *Demography*, 51(6): 2307–2332.

Desai, Sonalde, and Soumya Alva. 1998. "Maternal education and child health: Is there a strong causal relationship?" *Demography*, 35(1): 71–81.

Desai, Sonalde, and Veena Kulkarni. 2008. "Changing educational inequalities in India in the context of affirmative action." *Demography*, 45(2): 245–270.

Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research. 2005. *India human development survey (IHDS), 2005.*

Deshingkar, Priya, and Edward Anderson. 2004. *People on the move: New policy challenges for increasingly mobile populations.* Overseas Development Institute.

Deshingkar, Priya, and John Farrington. 2009. "A framework for understanding circular migration." *Circular migration and multilocational livelihood strategies in rural India*, 1–36.

Deshpande, Ashwini. 2000. "Does caste still define disparity? A look at inequality in Kerala, India." *The American Economic Review*, 90(2): 322–325.

Deshpande, Ashwini. 2017. *The grammar of caste: Economic discrimination in contemporary India.* Oxford University Press.

Deshpande, Ashwini, and Rajesh Ramachandran. 2016. "The changing contours of intergroup disparities and the role of preferential policies in a globalizing world: Evidence from India." CDE WP No. 267.

Deshpande, Ashwini, and Thomas E Weisskopf. 2014. "Does affirmative action reduce productivity? A case study of the Indian railways." *World Development*, 64: 169–180.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5): 1001–1044.

Elo, Irma T. 2009. "Social class differentials in health and mortality: Patterns and explanations in comparative perspective." *Annual Review of Sociology*, 553–572.

Elo, Irma T, and Samuel H Preston. 1996. "Educational differentials in mortality: United States, 1979–1985." *Social science & medicine*, 42(1): 47–57.

Filmer, Deon, and Lant H Pritchett. 2001. "Estimating wealth effects without expenditure data – or tears: An application to educational enrollments in states of India." *Demography*, 38(1): 115–132.

Geruso, Michael. 2012. "Black-white disparities in life expectancy: How much can the standard SES variables explain?" *Demography*, 49(2): 553–574.

Geruso, Michael, and Dean Spears. 2018. "Neighborhood Sanitation and Infant Mortality." *American Economic Journal: Applied Economics*.

Gillespie, Stuart, Jody Harris, and Suneetha Kadiyala. 2012. "The Agriculture-Nutrition Disconnect in India: What Do We Know?" IFPRI discussion paper no. 01187.

Government of India. 2001. "Census 2001: Metadata."

Guillot, Michel, and Keera Allendorf. 2010. "Hindu-Muslim differentials in child mortality in India." *Genus*, 66(2): 43–68.

Hathi, Payal, Sabrina Haque, Lovey Pant, Diane Coffey, and Dean Spears. 2017. "Place and child health: The interaction of population density and sanitation in developing countries." *Demography*, 54(1): 337–360.

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh Paul. 2012. "Castes and labor mobility." *American Economic Journal: Applied Economics*, 4(2): 274–307.

Hummer, Robert A. 1993. "Racial differentials in infant mortality in the US: An examination of social and health determinants." *Social Forces*, 72(2): 529–554.

International Dalit Solidarity Network. 2009. "Caste-based discrimination in South Asia: Situational overview, responses and ways forward." International Dalit Solidarity Network. technical report.

International Institute for Population Studies(IIPS), and Macro International. 2007. "National Family Health Survey (NFHS-3), 2005-06: India: Volume I." Mumbai: IIPS.

Jeffrey, Craig. 2001. "'A fist is stronger than five fingers': caste and dominance in rural north India." *Transactions of the Institute of British Geographers*, 26(2): 217–236.

Jodhka, Surinder S, and Ghanshyam Shah. 2010. "Comparative contexts of discrimination: Caste and untouchability in South Asia." *Economic and Political Weekly*, 99–106.

Jones, Steve. 1978. "Tribal underdevelopment in India." *Development and Change*, 9(1): 41–70.

Kawachi, Ichiro, and Lisa F Berkman. 2003. *Neighborhoods and health.* Oxford University Press.

Kumar, Satendra. 2016. "Agrarian transformation and the new rurality in Western Uttar Pradesh." *Econ Pol Weekly*, 51(26): 61–71.

Lamba, Sneha, and Dean Spears. 2013. "Caste, cleanliness, and cash: Effects of caste-based political reservations in Rajasthan on a sanitation prize." *Journal of Development Studies*, 49(11): 1592–1606.

Lauderdale, Diane S. 2006. "Birth outcomes for Arabic-named women in California before and after September 11." *Demography*, 43(1): 185–201.

LaVeist, Thomas, Keshia Pollack, Roland Thorpe, Ruth Fesahazion, and Darrell Gaskin. 2011. "Place, not race: Disparities dissipate in southwest Baltimore when blacks and whites live under similar conditions." *Health Affairs*, 30(10): 1880–1887.

Lee, Barrett A, Sean F Reardon, Glenn Firebaugh, Chad R Farrell, Stephen A Matthews, and David O'Sullivan. 2008. "Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales." *American Sociological Review*, 73(5): 766–791.

Leung, ManChui, and David T Takeuchi. 2011. "Race, place, and health." In *Communities, Neighborhoods, and Health.* , ed. L.M. Burton, S.P. Kemp, M. Leung, S.A. Matthews and D.T. Takeuchi, 73–88. Springer.

Maharatna, Arup. 2000. "Fertility, mortality and gender bias among tribal population: An Indian perspective." *Social Science & Medicine*, 50(10): 1333–1351.

Massey, Douglas S. 1990. "American apartheid: Segregation and the making of the underclass." *American Journal of Sociology*, 96(2): 329–357.

Massey, Douglas S. 2004. "Segregation and stratification: A biosocial perspective." *Du Bois Review*, 1(01): 7–25.

Mitra, Aparna. 2008. "The status of women among the scheduled tribes in India." *The Journal of Socio-Economics*, 37(3): 1202–1217.

Mosse, David. 2005. *Cultivating development: An ethnography of aid policy and practice.* Pluto Press.

Murthi, Mamta, Anne-Catherine Guio, and Jean Dreze. 1995. "Mortality, fertility, and gender bias in India: A district-level analysis." *Population and Development Review*, 745–782.

Nathan, Dev, and Virginius Xaxa, ed. 2012. *Social exclusion and adverse inclusion: Development and deprivation of Adivasis in India.* Oxford University Press.

Office of Registrar General & Census Commissioner. 2001. "2001 Census of India."

Onis, Mercedes. 2006. "WHO Child Growth Standards based on length/height, weight and age." *Acta Paediatrica*, 95(S450): 76–85.

Osypuk, Theresa, Lisa Bates, and Dolores Acevedo-Garcia. 2010. "Another Mexican birthweight paradox? The role of residential enclaves and neighborhood poverty in the birthweight of Mexican-origin infants." *Social science & medicine*, 70: 550–560.

Palloni, Alberto. 2006. "Reproducing inequalities: Luck, wallets, and the enduring effects of childhood health." *Demography*, 43(4): 587–615.

Pande, Rohini P. 2003. "Selective gender differences in childhood nutrition and immunization in rural India: The role of siblings." *Demography*, 40(3): 395–418.

Reardon, Sean F, Stephen A Matthews, David OŠSullivan, Barrett A Lee, Glenn Firebaugh, Chad R Farrell, and Kendra Bischoff. 2008. "The geographic scale of metropolitan racial segregation." *Demography*, 45(3): 489–514.

Reichman, Nancy E, Julien O Teitler, and Erin R Hamilton. 2009. "Effects of neighborhood racial composition on birthweight." *Health & Place*, 15(3): 814–821.

Rogers, Richard G. 1992. "Living and dying in the USA: Sociodemographic determinants of death among blacks and whites." *Demography*, 29(2): 287–303.

Rosenzweig, Mark R, and Oded Stark. 1989. "Consumption smoothing, migration, and marriage: Evidence from rural India." *Journal of Political Economy*, 97(4): 905–926.

Roy, TK, Sumati Kulkarni, and Y Vaidehi. 2004. "Social inequalities in health and nutrition in selected states." *Economic and Political Weekly*, 677–683.

Sastry, Narayan. 1996. "Community characteristics, individual and household attributes, and child survival in Brazil." *Demography*, 33(2): 211–229.

Sastry, Narayan. 2004. "Trends in socioeconomic inequalities in mortality in developing countries: The case of child survival in São Paulo, Brazil." *Demography*, 41(3): 443–464.

Sastry, Narayan, and Jon M Hussey. 2003. "An investigation of racial and ethnic disparities in birth weight in Chicago neighborhoods." *Demography*, 40(4): 701–725.

Shah, Ghanshyam, Harsh Mander, Sukhadeo Thorat, Satish Deshpande, and Amita Baviskar. 2006. *Untouchability in rural India.* Sage.

Singh, Abhishek, Sandip Chakraborty, and Tarun K Roy. 2008. "Village Size in India: How relevant is it in the context of development?" *Asian Population Studies*, 4(2): 111–134.

Spears, Dean. 2012. "Height and cognitive achievement among Indian children." *Economics & Human Biology*, 10(2): 210–219.

Spears, Dean. 2013. "The nutritional value of toilets: How much international variation in child height can sanitation explain?" Princeton working paper.

Spears, Dean. 2016. "Caste and life satisfaction in rural north India." *Economic and Political Weekly*, 12–14.

Srinivas, Mysore Narasimhachar. 1976. *The remembered village.* Univ of California Press.

Srinivas, Mysore Narasimhachar, ed. 1955. *India's villages.* Asia Publishing House.

Thorat, Amit, and Omkar Joshi. 2015. "The Continuing Practice of Untouchability in India: Patterns and Mitigating Influences." *India Human Development Study Working Paper*, 3.

Thorat, Sukhadeo, and Katherine S Newman. 2012. *Blocked by caste: Economic discrimination in modern India.* Oxford University Press.

Thorat, Sukhdeo, and Joel Lee. 2005. "Caste discrimination and food security programmes." *Economic and Political Weekly*, 4198–4201.

Torche, Florencia. 2011. "The effect of maternal stress on birth outcomes: exploiting a natural experiment." *Demography*, 48(4): 1473–1491.

Valmiki, Om Prakash. 2003. *Joothan: A Dalit's life.* Columbia University Press.

Williams, David R. 1999. "Race, socioeconomic status, and health the added effects of racism and discrimination." *Annals of the New York Academy of Sciences*, 896(1): 173–188.

Williams, David R, and Chiquita Collins. 1995. "US socioeconomic and racial differences in health: Patterns and explanations." *Annual review of sociology*, 349–386.

Williams, David R, and Chiquita Collins. 2001. "Racial residential segregation: A fundamental cause of racial disparities in health." *Public health reports*, 116(5): 404.

Xie, Yu, and Margaret Gough. 2011. "Ethnic enclaves and the earnings of immigrants." *Demography*, 48: 1293–1315.

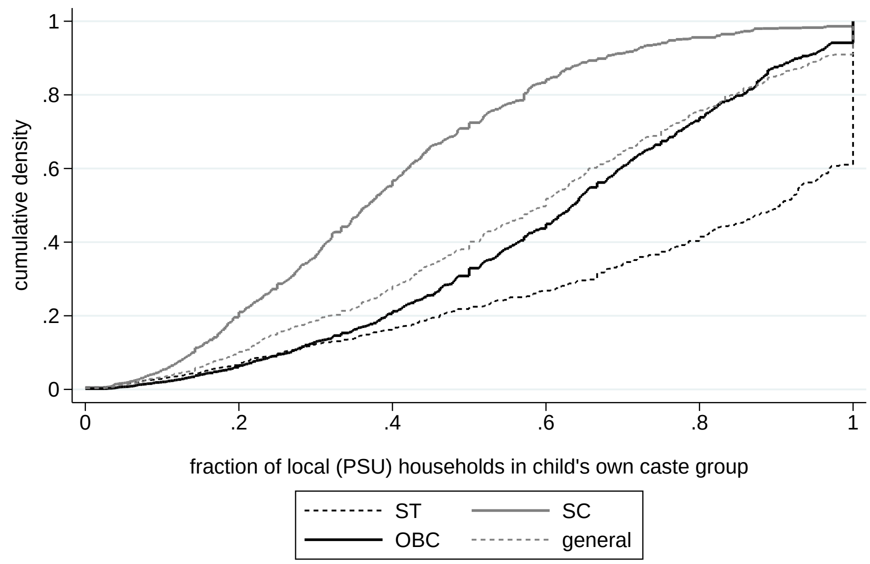


Figure 1: Cumulative distribution of the fraction of households in a child's primary sampling unit (PSU) from her own caste group

Observations are rural children whose heights were measured by the NFHS-3.

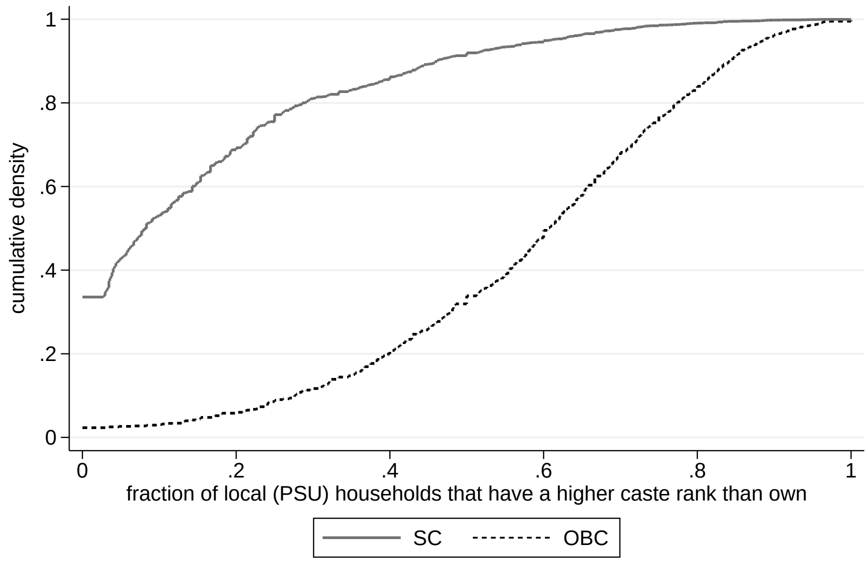


Figure 2: Cumulative distribution of the fraction of households in a child's primary sampling unit (PSU) that have a higher caste rank than her own household's

Observations are rural children whose heights were measured by the NFHS-3.

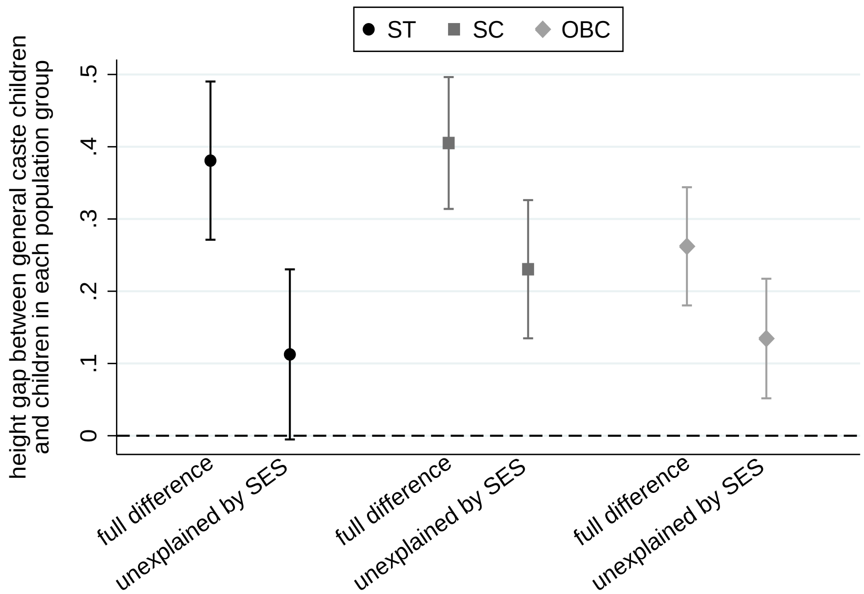


Figure 3: Decomposition results: Fraction of the height gap between general caste children and ST, SC, and OBC children unexplained by socioeconomic variables

Estimates represent reweighting decomposition results that describe height gaps between rural general caste children and rural children in other population groups. Confidence intervals represent 95% confidence, computed using clustered standard error to reflect the survey design of the DHS. The socioeconomic variables used in the decomposition are floor type (4 categories) and education level of her mother (6 categories), for a combined set of 24 bins. These variables and the reweighting decomposition method are described in section 4.

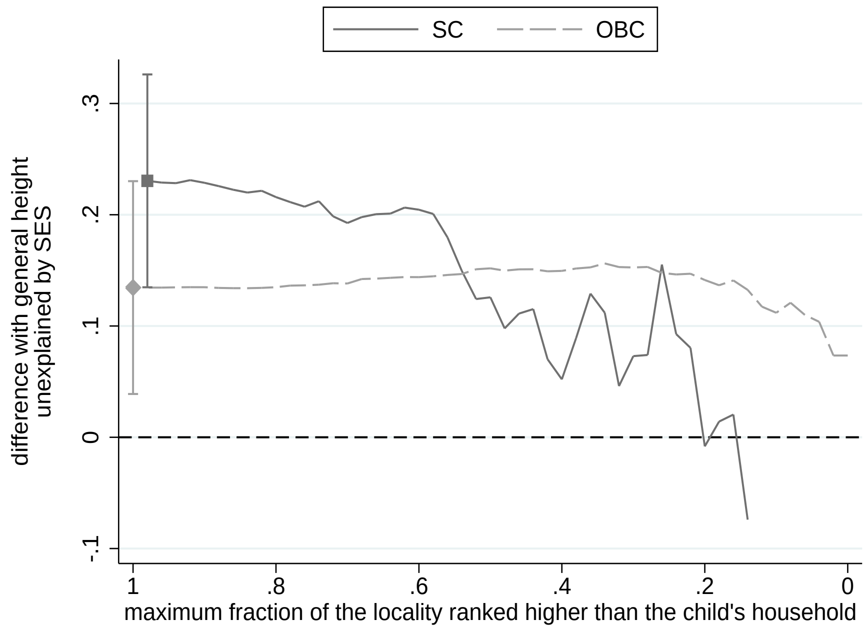


Figure 4: Decomposition results: SES explains more of the SC and OBC height gaps for children who are less locally outranked

The figure uses NFHS-3 data on rural India to estimate height gaps between general caste and OBC children, and general caste and SC children, computed using many replications of the reweighting decomposition described in section 4. The horizontal axis indicates which OBC or SC children are used in the reweighting decomposition. For example, at 0.4, an OBC or SC child is included if she lives in a locality in which 40% or fewer of households in the PSU outrank her household by caste. The figure proceeds horizontally in intervals of 0.02 and connects points with no smoothing. See Sections 4.2 and 5.2 for further details.

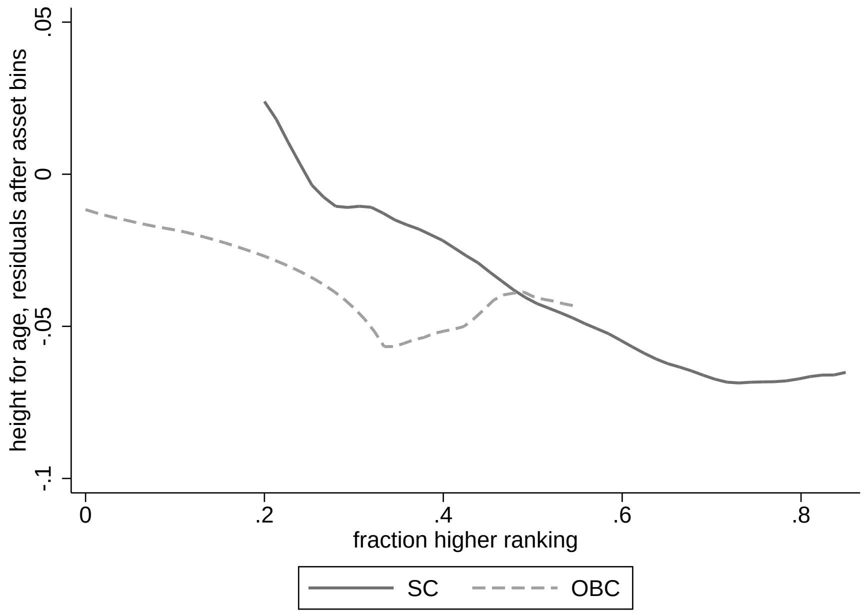


Figure 5: Child height, net of SES, and the fraction of households in a child's locality that outranks hers

Local kernel-weighted regression. Epanechnikov kernel, bandwidth = 0.15. Here the large set of regression SES controls are used. Observations are children included in column 4 of Table 3 for whom the fraction of households that outrank their own is between the 10th and 90th percentiles of this variable for their own caste group.

Table 1: Summary statistics

	ST	SC	OBC	general
height for age *z*-score				
25th percentile	-3.24	-3.22	-3.06	-2.76
50th percentile	-2.20	-2.21	-2.05	-1.77
75th percentile	-1.12	-1.18	-1.00	-0.74
SES variables used in reweighting decomposition				
dirt floor	0.80	0.72	0.61	0.55
rudimentary floor (brick, stone, etc.)	0.05	0.03	0.05	0.05
finished floor	0.11	0.18	0.25	0.32
other floor type[a]	0.05	0.07	0.09	0.09
mother has no education	0.70	0.64	0.58	0.33
mother did not complete primary	0.08	0.07	0.07	0.08
mother completed primary	0.04	0.07	0.08	0.07
mother did not complete secondary	0.15	0.19	0.23	0.36
mother completed secondary	0.01	0.01	0.02	0.05
mother has higher education	0.01	0.01	0.02	0.04
additional SES variables used in regression				
electricity	0.41	0.45	0.50	0.58
phone	0.01	0.04	0.07	0.12
radio	0.18	0.22	0.28	0.36
television	0.14	0.25	0.30	0.39
refrigerator	0.02	0.03	0.05	0.13
bicycle	0.44	0.54	0.60	0.57
motorcycle	0.06	0.06	0.13	0.19
car	0.01	0.00	0.01	0.02
uses toilet or latrine	0.10	0.14	0.16	0.40
owns land	0.67	0.44	0.66	0.63
local infrastructure controls				
fraction of households in PSU with electricity	0.44	0.49	0.48	0.51
fraction of births in PSU with prenatal care	0.71	0.72	0.69	0.76
local caste composition controls				
fraction of PSU that is SC	.	0.39	0.18	0.18
fraction of PSU that is OBC	.	0.35	0.63	0.25
fraction of PSU that is general	.	0.21	0.14	0.53
demographic controls				
number of household members	6.53	6.78	7.53	7.01
birth order	3.17	3.01	2.90	2.64
female	0.50	0.49	0.47	0.47
lives with paternal grandparents	0.19	0.23	0.32	0.32
n	4,730	5,134	8,613	6,354

Statistics presented are means unless otherwise stated. Observations are rural children whose heights were measured by the NFHS-3. Because averages are representative of children, they may differ from published India-wide summary statistics. [a]A floor type of "other" is also listed for children whose mothers were not interviewed in their permanent home. Most likely, these women were interviewed in their parents' home.

Table 2: The apparent similarity of the SC and ST height deficits

	(1)	(2)	(3)	(4)
	height-for-age *z*-score			
SC or ST		-0.397***	-0.381***	-0.405***
		(0.0424)	(0.0559)	(0.0465)
SC	-0.405***		-0.0244	
	(0.0465)		(0.0557)	
ST	-0.381***			0.0244
	(0.0559)			(0.0557)
OBC	-0.262***	-0.262***	-0.262***	-0.262***
	(0.0417)	(0.0417)	(0.0417)	(0.0417)
n	24,840	24,840	24,840	24,840
R^2	0.008	0.008	0.008	0.008

OLS regressions, weighted using sample weights. Observations are rural children whose heights were measured by the NFHS-3. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$. Standard errors clustered by primary sampling unit (PSU) in parentheses. ST = Scheduled Tribe, SC = Scheduled Caste, OBC = Other Backwards Classes.

Table 3: Explaining height gaps between general caste children and children from ST, SC, and OBC groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variable:	height for age z-score (general caste children are omitted category)						
ST	-0.381***	-0.0705					
	(0.0559)	(0.0653)					
SC	-0.405***	-0.153**	-0.154**	0.0383	-0.132*	-0.200**	0.0080
	(0.0465)	(0.0545)	(0.0551)	(0.118)	(0.0646)	(0.0622)	(0.159)
OBC	-0.262***	-0.123*	-0.134**	-0.0200	-0.142*	-0.143*	0.0045
	(0.0417)	(0.0485)	(0.0494)	(0.0774)	(0.0598)	(0.0590)	(0.104)
n	24840	23111	18141	18141	18141	18147	18140
R^2	0.008	0.201	0.222	0.364	0.344	0.229	0.369
own SES bins		✓	✓			✓	
own SES × caste rank bins				✓			✓
own SES × SES rank bins					✓		
demographic & neighborhood						✓	✓
composition controls						✓	✓
STs in the sample	yes	yes	no	no	no	no	no

OLS regressions, weighted using NFHS sample weights. Observations are rural children whose heights were measured by the NFHS-3. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors clustered by primary sampling unit (PSU) in parentheses. ST = Scheduled Tribe, SC = Scheduled Caste, OBC = Other Backwards Class. The construction of controls for "own SES bins," "own SES × caste rank bins," and "own SES × SES rank bins" are discussed in section 4. "demographic controls" include child birth order, child sex, and whether the child lives in a joint family with his/her grandparents. "neighborhood controls" include the fraction of households in a child's PSU with electricity, the fraction of births (last births to the mother) in the child's PSU that got prenatal care, the fraction of SC households, the fraction of OBC households, and the fraction of general caste households in a child's PSU.

Distribution of caste and tribal status among children of different religions.
Table A1 shows the distribution of caste and tribal status by religion among rural children whose heights were measured in the NFHS-3. The religion of the child's household head is used as the measure of religion.

CDFs of height by population group.
The summary statistics table (Table 1) in the main text described the 25th, 50th, and 75th percentiles of height-for-age by population group. Figure A1 plots the empirical CDFs of child height-for-age for the four population groups in rural India that we study. Note that the distribution for general caste children stochastically dominates (is everywhere to the right of) the distribution for OBC children, which stochastically dominates the distributions for SC and ST children.

Floor type predicts child height in rural India.
In the reweighting decomposition described in Section 4 of the main text, we use 4 categories of floor type as a measure of child SES. Table A2 uses OLS regression to show that floor type predicts child height in rural India. This is true whether or not we additionally control for mother's education, the second variable used in the reweighting decomposition.

Using father's education in the reweighting decomposition produces similar results as using mother's education.
In the main text, we perform reweighting decompositions that ask the question: what would be the average heights of rural ST, SC, and OBC children be if they had the same distribution of SES (as measured by flooring material and mother's education) as general caste children? Here, we perform the same analysis, but use father's education instead of mother's education. The results are similar: reweighting over the intersection of father's education and floor type entirely explains the ST–general height gap, but does not entirely explain the SC–general gap, nor the OBC–general gap. Figure A2 reproduces Figure 3 from the main text using father's education instead of mother's education.

Regression analysis: Robustness to omitting OBC children.

The caste composition of the OBC category varies at the state-level throughout India, as does the everyday experience of being someone from an OBC caste. In some contexts, some OBC households may not be considered socially lower ranking than some households classified here as general. However, SC households are unambiguously considered socially lower ranking than OBC and general caste households. To verify that no special property of OBC children or households is responsible for the pattern of our results, Table A3 replicates our main results excluding all OBC children from the sample. The same results are found: SES accounts for all of the ST height gap, but not the SC height gap, which is explained by SES variables and local caste rank.

OLS regression analysis, separately by boys and girls.

Many demographic process in India differ by sex; gender attitudes and practices have important consequences for health, nutrition, and mortality. Our demographic controls in the regression results include a control for child sex. Table A4 verifies that the general pattern of our results is robust to splitting the sample by sex. For both girls and boys, there are large height gaps between SC, ST, and OBC children, and general caste children. For ST children, the gap can be explained by SES controls.

Regression analysis: Robustness to controlling for local open defecation and other controls.

Recent research has demonstrated the relevance of the local gastro-intestinal disease environment to child height in India (for example, Hathi et al., 2017). Table A5 reproduces the main results (from Table 3) and adds a control for local open defecation. That is, we control for the fraction of households in a child's PSU that report practicing open defecation rather than toilet or latrine use. Consistent with the prior literature, we find that local open defecation predicts child height. However, the inclusion of this control variable does not change our main results: SES differences still explain height gaps between general caste and ST children, and a combination of SES differences and local caste rank still explain height gaps between general caste and SC children, and between general caste and OBC children.

Table A6 reproduces Table A5, and uses additional controls in columns 6 and 7: the number of siblings a child has at the time of the survey (no siblings is omitted category), the religion of the household (Hindu is the omitted category), and whether the households owns a mattress, a pressure cooker, a chair, a bed, a table, or an electric fan. These additional controls are added as dummy variables. The results in Table A6 show that the addition of these new controls does not change the interpretation of the main results.

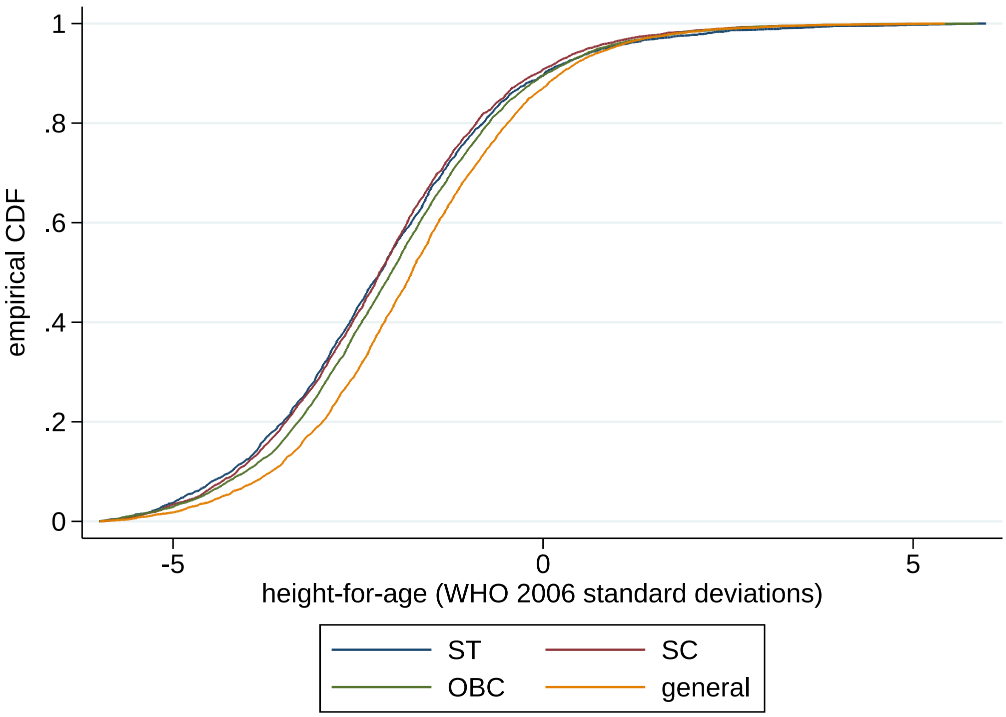
SC respondents are more likely to report open defecation near their houses. Table A5 controls for PSU-level open defecation, but does not control for individual disease burdens. Section 6 of the main text discusses the possibility that differences in sanitation externalities might partially explain why, conditional on SES, lower caste children are shorter than general caste children in localities where they are outranked, but not in homogeneous villages. Table A7 finds suggestive evidence that in mixed villages, SC disease burdens may be higher than SC disease burdens in homogeneous villages. Table A7 uses data from a sanitation survey collected in five states in north India in 2013-14 (see Coffey et al. (2014) for details of the survey) to show that SC respondents are more likely to report seeing someone defecate in the open near their houses than general caste respondents. This result is robust to controls for whether the respondent defecates in the open, for estimated village level open defecation, and for the respondent's sex and age. We note, however, that this is not true for OBC respondents.

The heights of Scheduled Caste children are similar in large and small villages. Section 6 of the main text discusses the possibility that the fraction of a child's village that outranks her household, a key variable in our analysis, may be correlated with village size. If lower caste children living in smaller villages are shorter than lower caste children living in larger villages, this would suggest that village size could be an important omitted variable in our analysis. Unfortunately, the NFHS-3 does not report village size. In order to investigate the correlation between village size and child height, we turn to the India Human Development Survey (IHDS), 2005 (Desai et al., 2005). The IHDS is a smaller, but nationally representative, sample of children for whom height, social category, and village size were collected. Figure A3 shows that at every age at which rural children were measured (ages zero to five and eight to eleven), Scheduled Caste children living in smaller villages (less than 500 households) are as tall as children living in larger villages (500+ households). Further, Table A8 shows the same conclusion using an OLS regression: it uses the same IHDS data on rural children ages zero to five and eight to eleven to show that, controlling for state fixed effects, household consumption per capita, and the education level of the most educated adult in the household, living in a large village does not predict height for rural SC children.

The fraction of households that are Scheduled Caste is similar in NFHS-3 PSUs and in 2001 Census villages.
Section 6 of the main text discusses the fact that the NFHS data allow us to observe local caste rank at the primary sampling unit (PSU) level, rather than at the village level. Figure A4 provides suggestive evidence that if we could observe a child's local caste rank at the village level, we would find similar results. It plots the cumulative distribution of the fraction of households in the NFHS PSUs that are Scheduled Caste, as well as the cumulative

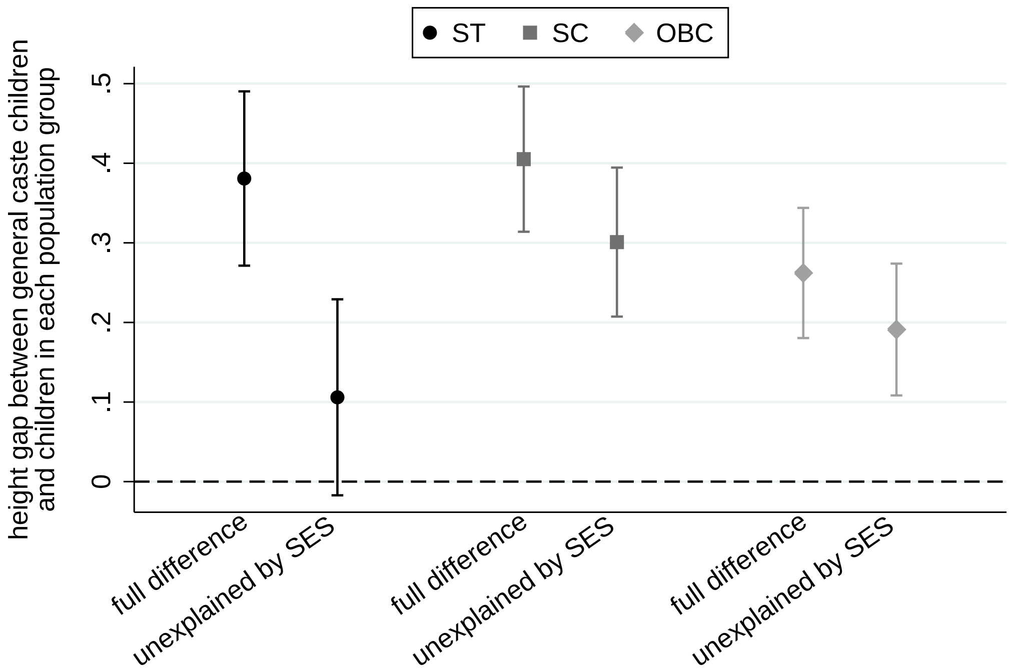
distribution of the fraction of households in 2001 Census villages that are Scheduled Caste. The two cumulative distributions are very similar.

Figure A1: Empirical CDF of height-for-age, by population group



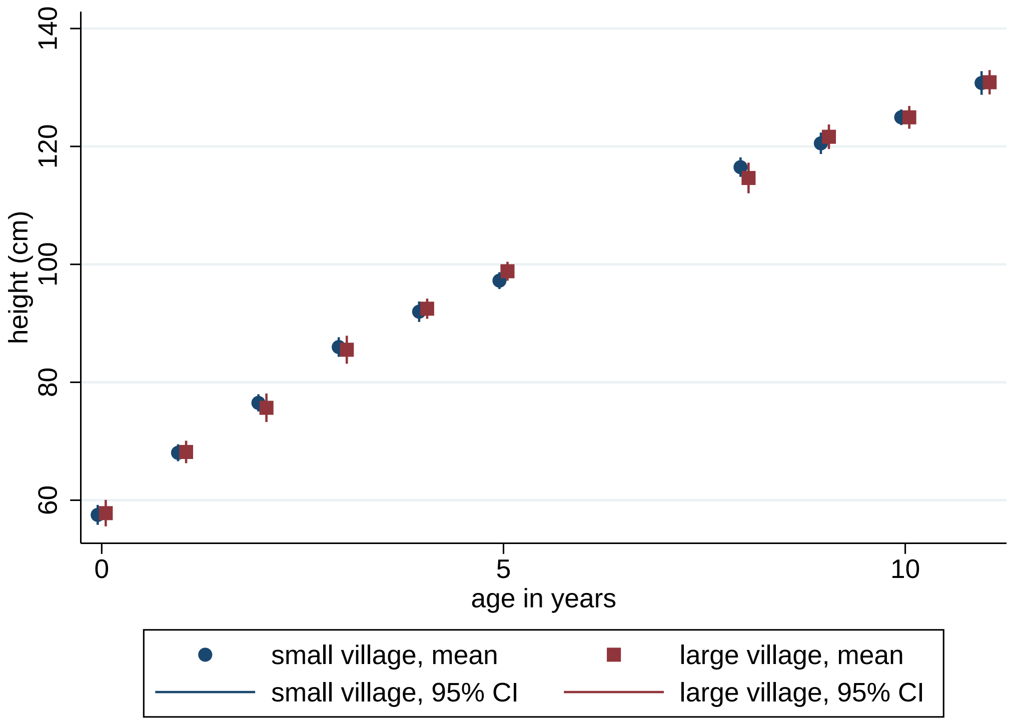
Observations are rural children whose heights were measured by the NFHS-3.

Figure A2: Decomposition results: Fraction of the height gap between general case children and ST, SC, and OBC children unexplained by socioeconomic variables



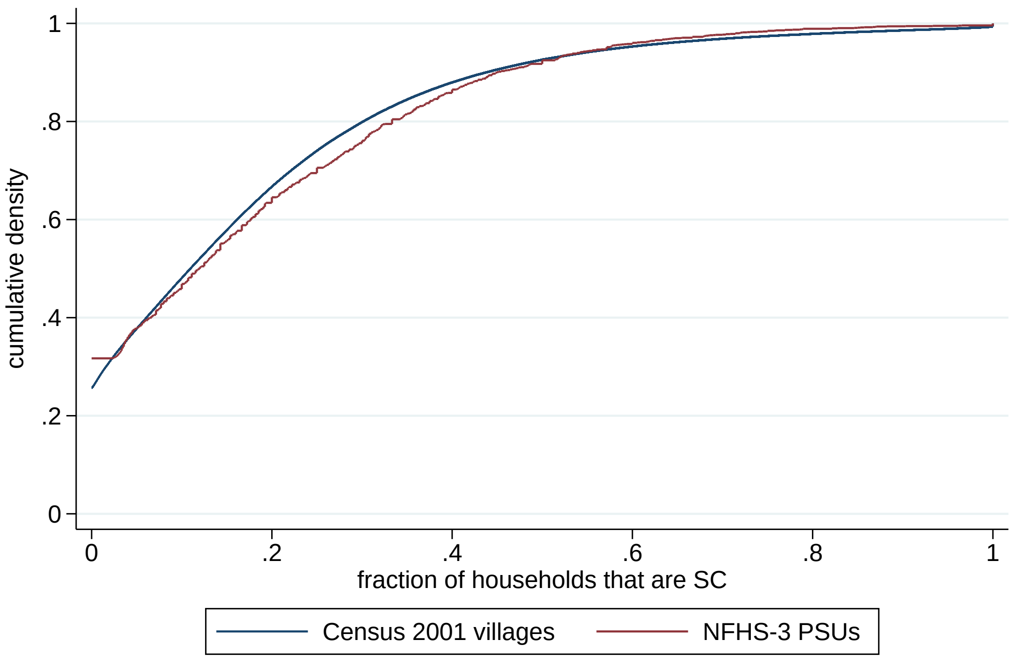
Estimates represent reweighting decomposition results that describe height gaps between general caste children and children in other population groups. Confidence intervals represent 95% confidence, computed using clustered standard error to reflect the survey design of the DHS. The socioeconomic variables used in the decomposition are floor type (4 categories) and education level of her father (6 categories), for a combined set of 24 bins. These variables and the reweighting decomposition method are described in section 3. The figure in the main text uses mother's education instead of father's education.

Figure A3: Heights of Scheduled Caste children are similar in large and small villages



Observations are rural SC children whose heights were measured by the India Human Development Survey, 2005. The graph shows average height and 95% confidence intervals for children of different ages. $n = 5,445$. Large villages have 500 or more households; small villages have less than 500 households.

Figure A4: The fraction of households that are Scheduled Caste is similar in NHFS-3 PSUs and in 2001 Census villages



Observations are rural primary sampling units in the NFHS data and villages in the 2001 Census data (Office of Registrar General & Census Commissioner, 2001). Cumulative distributions of fraction of the PSU or village that are SC are shown.

Table A1: Distribution of caste and tribal status among children of different religions

	religion of the child's household head		
	Hindu	Muslim	Other religion
ST	0.12	0.01	0.34
SC	0.24	0.02	0.28
OBC	0.45	0.32	0.10
General	0.18	0.58	0.23
social group data missing	0.01	0.17	0.05
n	18,485	3,587	3,993

This table shows the caste and tribal status of children by the religion of their household head. Observations are rural children whose heights were measured by the NFHS-3.

Table A2: Floor type predicts child height-for-age

	(1)	(2)
rudimentary floor	0.307***	0.176***
	(0.0504)	(0.0502)
finished floor	0.569***	0.346***
	(0.0249)	(0.0263)
other floor type	0.461***	0.327***
	(0.0382)	(0.0382)
mother did not complete primary		0.226***
		(0.0398)
mother completed primary		0.281***
		(0.0414)
mother did not complete secondary		0.490***
		(0.0260)
mother completed secondary		0.815***
		(0.0662)
mother has higher education		1.094***
		(0.0704)
constant	-2.158***	-2.290***
	(0.0129)	(0.0145)
n	24840	24840
R^2	0.023	0.046

Observations are rural children whose heights were measured by the NFHS-3. The omitted category of floor type is a dirt floor. The omitted category for mother's education is no education.

Table A3: Regression robustness: Similar results omitting OBC children

	(1)	(2)	(3)	(4)
dependent variable:		height-for-age z-score		
ST	-0.381***	-0.0979		
	(0.0559)	(0.0636)		
SC	-0.405***	-0.193***	-0.204***	-0.164
	(0.0465)	(0.0515)	(0.0526)	(0.207)
own SES bins		✓	✓	✓
local caste rank bins				✓
n	16,227	16,227	11,082	11,082
R^2	0.014	0.190	0.226	0.339

† two-sided $p < 0.10$; * $p < 0.05$; ** $p < 0.01$. Standard errors are clustered by survey PSU. These regressions fully exclude OBC children, but are otherwise identical to the same numbered columns in the main regression results in Table 3 of the main text.

Table A4: OLS regression results, by sex

	girls before SES	girls after SES	boys before SES	boys after SES
ST	-0.30***	0.03	-0.46***	-0.15
	(0.07)	(0.10)	(0.07)	(0.09)
	$n = 5,572$		$n = 5,925$	
SC	-0.36***	-0.17*	-0.44***	-0.24**
	(0.06)	(0.07)	(0.06)	(0.07)
	$n = 5,262$		$n = 5,822$	
OBC	-0.23***	-0.11	-0.29***	-0.13*
	(0.05)	(0.07)	(0.05)	(0.06)
	$n = 7,118$		$n = 7,849$	

Each estimate is from a separate comparison of children in the disadvantaged population group (ST, SC, or OBC) with general caste children. The standard errors are clustered by survey PSU. SES controls include an indicator for the household having electricity, a phone, a radio, a TV, a refrigerator, a bicycle, a motorcycle, a car, land, and whether the household uses a toilet or latrine. These are the same SES controls used in Table 3 of the main text. Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$.

Table A5: Main results (Table 3), adding a control for local open defecation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variable:	height for age z-score (general caste children are omitted category)						
ST	-0.381***	-0.0560					
	(0.0559)	(0.0648)					
SC	-0.405***	-0.147**	-0.148**	0.0636	-0.125+	-0.148**	0.0636
	(0.0465)	(0.0544)	(0.0551)	(0.119)	(0.0643)	(0.0551)	(0.119)
OBC	-0.262***	-0.112*	-0.122*	0.00217	-0.129*	-0.122*	0.00217
	(0.0417)	(0.0482)	(0.0493)	(0.0783)	(0.0595)	(0.0493)	(0.0783)
local open defecation		-0.213*	-0.235*	-0.231+	-0.237+	-0.235*	-0.231+
		(0.0949)	(0.103)	(0.134)	(0.138)	(0.103)	(0.134)
n	24840	23111	18141	18141	18141	18148	18141
R^2	0.008	0.201	0.222	0.364	0.344	0.222	0.364
own SES bins		✓	✓			✓	
own SES × caste rank bins				✓			✓
own SES × SES rank bins					✓		
demographic & neighborhood						✓	✓
composition controls						✓	✓
STs in the sample	yes	yes	no	no	no	no	no

OLS regressions, weighted using NFHS sample weights. Observations are children under five in the NFHS-3. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$. Standard errors clustered by primary sampling unit (PSU) in parentheses. ST = Scheduled Tribe, SC = Scheduled Caste, OBC = Other Backwards Class. The construction of controls for “own SES bins,” “own SES × caste rank bins,” and “own SES × SES rank bins” are discussed in Section 4 of the main text. “demographic controls” include child birth order, child sex, and whether the child lives in a joint family with his/her grandparents. “neighborhood controls” include the fraction of households in a child’s PSU with electricity, the fraction of births (last births to the mother) in the child’s PSU that got prenatal care, the fraction of SC households, the fraction of OBC households, and the fraction of general caste households in a child’s PSU. “local open defecation” is the fraction of households sampled in a child’s primary sampling unit that defecate in the open.

Table A6: Main results (Table 3), with additional controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variable:	height for age z-score (general caste children are omitted category)						
ST	-0.381***	-0.0705					
	(0.0559)	(0.0653)					
SC	-0.405***	-0.153**	-0.154**	0.0383	-0.132*	-0.219***	-0.0345
	(0.0465)	(0.0545)	(0.0551)	(0.118)	(0.0646)	(0.0618)	(0.161)
OBC	-0.262***	-0.123*	-0.134**	-0.0200	-0.142*	-0.152**	-0.0171
	(0.0417)	(0.0485)	(0.0494)	(0.0774)	(0.0598)	(0.0588)	(0.104)
local open defecation						-0.168	-0.122
						(0.109)	(0.138)
one sibling						-0.322***	-0.327***
						(0.0639)	(0.0823)
two siblings						-0.562***	-0.602***
						(0.0861)	(0.110)
three siblings						-0.812***	-0.853***
						(0.116)	(0.145)
four or more siblings						-0.998***	-1.095***
						(0.168)	(0.205)
Muslim						-0.0284	-0.0371
						(0.0664)	(0.0798)
neither Hindu nor Muslim						0.0897	0.281*
						(0.0859)	(0.125)
n	24840	23111	18141	18141	18141	18089	18082
R^2	0.008	0.201	0.222	0.364	0.344	0.235	0.375
own SES bins		✓	✓			✓	
own SES × caste rank bins				✓			✓
own SES × SES rank bins					✓		
additional controls						✓	✓
demographic & neighborhood						✓	✓
composition controls						✓	✓
STs in the sample	yes	yes	no	no	no	no	no

The same note as for table A5 applies here. This table uses additional controls in columns 6 and 7: the number of siblings a child has at the time of the survey (no siblings is omitted category), the religion of the household (Hindu is the omitted category), and whether the households owns a mattress, a pressure cooker, a chair, a bed, a table, or an electric fan. These additional controls are added as dummy variables.

Table A7: SC villagers are more likely to see others defecate in the open near their home

	(1)	(2)
dependent variable: has seen OD near home		
SC household	0.0588*	0.0412†
	(0.0229)	(0.0235)
OBC household	0.0123	-0.00384
	(0.0180)	(0.0182)
village OD		0.284**
		(0.0332)
respondent ODs		-0.0222
		(0.0197)
respondent is female		0.00261
		(0.0161)
respondent's age		0.000780
		(0.000612)
n (rural households)	2,947	2,914

Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$. OD = open defecation; village OD is the fraction 0-1 of households in the village that defecate in the open. Data are from the SQUAT survey; see Coffey et al. (2014) for more information about the survey.

Table A8: Heights of Scheduled Caste children are similar in large and small villages: OLS results

dependent variable: height in cm	
large village (> 500 HH)	-0.05
	(0.31)
age 1	9.73***
	(0.94)
age 2	18.84***
	(0.62)
age 3	27.92***
	(0.57)
age 4	33.84***
	(0.66)
age 5	40.81***
	(0.69)
age 8	58.51***
	(0.43)
age 9	64.17***
	(0.42)
age 10	67.91***
	(0.70)
age 11	72.80***
	(0.73)
state fixed-effects	✓
fixed effects for years of education of the most educated adult in HH	✓
consumption per capita	0.00139***
	(0.000334)
constant	47.96***
	(3.14)
n	5434
R^2	0.81

Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$. Observations are rural SC children whose heights were measured by the India Human Development Survey, 2005. The table shows coefficients from an OLS regression of height in centimeters on village size, age in years, and controls. Large villages have 500 or more households; small villages have less than 500 households.