LUND UNIVERSITY
School of Economics and Management

Master's Programme in Economic Development and Economic Growth (MEDEG)

Take the Poo to the Loo[1]

A cross-sectional analysis of stunting and open defecation in India

by

Marco Lovato

ma8311lo-s@student.lu.se

Abstract: The present study addresses the relationship between stunting and open defecation in India, 2015. The lack of proper sanitation allows pathogens to spread and generate fecal transmissible illnesses (FTIs), hence hampering the correct absorption of nutrients in the body and the consequent child's development. The expectation is that a higher open defecation rate is associated with a higher stunting prevalence. Moreover, being open defecation practices embedded within Hinduism, the Muslim population is expected to be negatively correlated with stunting prevalence. The results confirm the positive relationship between stunting and open defecation, whereas being Muslim does not seem to be significantly related with stunting prevalence. Increasing resources should be invested to introduce adequate sanitation in the country, thus ensuring the proper development of its human capital.

Keywords: *stunting, open defecation, India*

[1] Name of a UNICEF campaign launched to end open defecation in India. For more information:
https://www.unicef.org/india/campaigns/take-poo-loo
https://www.youtube.com/watch?v=Ll0GCPFpNQs

EKHS42
Master's Thesis (15 credits ECTS)
June 2020
Supervisor: Dr. Finn Hedefalk
Examiner: Dr. Luciana Quaranta
Word Count: 12,955

Table of Contents

List of Tables

List of Figures

1. Introduction

820 million people globally are undernourished (World Bank, 2020). 22% of worldwide children below 5 years are stunted, i.e. ‘having a height-for-age z-score (HAZ) that is more than two standard deviations below the age-sex median for a well-nourished reference population’ (Hoddinott et al., 2013). The pervasiveness of the phenomenon has rightfully led to the inclusion of the eradication of hunger among the Sustainable Development Goals (SDGs) proposed by United Nations member states in 2014. Undernutrition affects world regions differently, with the greatest burden being faced by developing countries: approx. 23% and 15% of the population is undernourished[2] in Sub-Saharan Africa (SSA) and South Asia, respectively, against much lower values in more developed countries. The prevalence of the condition in South Asia, in particular, has attracted increasing attention within the academic world: in contrast with SSA, where economic growth has been low and discontinuous, the region has been experiencing persistent positive economic growth rates in the past two decades. The positive economic performance, however, seems to have failed to improve health indicators to the same degree, raising many eyebrows among the academics, leading some to talk about an ‘Asian enigma’ (Ramalingaswami et al., 1996).

Within South Asia, India has probably been the most emblematic case: between 1995 and 2005, GDP growth per annum averaged 6.14%, with the poverty headcount ratio declining consistently, suggesting a widespread improvement of living conditions. Yet, as of 2005, undernutrition levels in the country were higher than for most countries in Sub-Saharan Africa, with Indian children being shorter, on average, than their African counterparts (Spears, 2013). For years, the literature has struggled to find an explanation. More recent studies, however, seem to have find the blind spot inherent to past research: sanitation. Typically, the studies attempting to explain the ‘Asian enigma’ has focused on food intake, micronutrients and food security for the poor (Chambers and von Medeazza, 2014), effectively neglecting the role of sanitation. In a country like India, where one out of three persons practiced open defecation (hence, not using any kind of toilet facility) as of 2015, to neglect sanitation as a determinant of malnutrition is to overlook the crucial role of fecal transmissible illnesses (FTIs) on growth faltering. FTIs, which range from hookworms to diarrhea, negatively affect the absorption of nutrients in the body, hence potentially canceling out any benefit originating from dietary improvements (Chambers and von Medeazza, 2014).

Quantitative research (Spears, 2013, 2014; Quattri and Smets, 2014) has found a significant relationship between stunting and open defecation prevalence, hence confirming the crucial role of sanitation on health indicators. While open defecation practices in SSA are typically the consequence of a lack of sanitary infrastructures, in the case of India, home to 60% of the world population practicing open defecation, the habit has also cultural and religious roots. Hinduism considers emptying latrines as a both physically and ritually dirty task, which is typically assigned to dalits, the lower caste. The religious determinant of open defecation practices in India has been also investigated, unveiling a significant heterogeneity of the practices between religious groups, with Hindu being the group exhibiting the highest levels (Vyas and Spears, 2018).

[2] Defined as a caloric intake insufficient to meet the minimum energy requirements necessary for a given individual.

The deeply rooted open defecation prevalence in India offers the starting point for this thesis, which aims to provide further evidence of the relationship between open defecation practices and stunting prevalence. The main research question that this study addresses is:

What are the implications of open defecation practices on stunting prevalence in India, 2015?

In the last 10 years, the Indian government has launched a series of campaigns to introduce the use of latrines and end open defecation across the country, obtaining a discrete success (Gupta et al., 2019). Whether the positive relationship highlighted by previous literature applies also to more recent times represents one of the key contributions of this research. Moreover, with regard to religion, the higher prevalence of open defecation practices among Hindus relative to Muslims (Vyas and Spears, 2018) has been found to translate into a consistent life expectancy advantage of the latter on the former (Geruso and Spears, 2018). Yet, whether the advantage persists with regard to stunting has, to the best of the author's knowledge, yet to be addressed by the literature.

2. Literature Review

The present section defines stunting and reviews the previous literature on the condition, while also describing the economic costs it entails. Next, the determinants of children undernutrition are defined, with a particular attention to the role of sanitation. The prevalence of stunting and open defecation is discussed within the context of South Asia, and India in particular. Previous studies over the relationship between the two variables are reviewed, finally presenting the research question of this research.

2.1 Background and definition of stunting

Stunting is defined as 'having a height-for-age z-score (HAZ) that is more than two standard deviations below the age-sex median for a well-nourished reference population' (Hoddinott et al., 2013). Typically, the condition prevails among children, with the strongest effect being observed in the first 24 months of life (Victora et al., 2010). The severe undernutrition affecting children in what is probably one of the most delicate phases of their growth has a permanent damage on both their physical and cognitive development.

Developing countries are faced with the majority of worldwide stunting cases: in 2011 165 million children were estimated to be stunting globally (Black et al., 2013), with Africa and Asia alone representing 35% and 27% of documented cases. Stunting prevalence, on the other hand, has significantly reduced in Asia in the past decades, whereas it largely stagnated around 40% in Africa (Hoddinott et al., 2013). The decrease experienced by the continent, however, has been largely driven by the consistent reduction occurred in China, Thailand and Vietnam, whereas other countries (e.g. Pakistan, India, Cambodia) have experienced little or no change (World Bank, 2020). The physical and cognitive impairments that originate from stunting represent a fundamental obstacle to the improvement of living conditions in developing countries, leading to what can be rightfully defined a consistent loss of human capital potential.

Next section reviews the research on consequences of stunting on physical and cognitive development, together with the economic costs they entail.

2.2. The consequences of stunting

2.2.1 The development of physical capabilities

A vast body of literature has investigated the consequences of stunting on the physical body. Stunting typically leads to an underdevelopment of an individual's stature, thus negatively affecting the height and muscular strength attained once adult (Lundborg et al., 2014; Stein et al., 2010). While a lower muscular strength alone may already negatively affect an individual's living conditions, the variable has been also found associated with increasing risk of mortality (Metter et al., 2002). On a similar matter, Crimmins and Finch (2006) found how inflammations and infections that delay the height growth of a person also contribute to cardiovascular diseases at older ages. Reduced muscular strength and height have a significant impact on labour productivity of physically demanding jobs: Dinda et al. (2006), for instance, find how Indian coalminers of above average height earn between 9% and 17% more than their shorter counterparts.

Other studies have further confirmed a significant positive effect of height on wage gains in Ghana and Brazil (Thomas and Strauss, 1997; Schultz, 2002). Increased height, while leading to a higher wage, has been found to have also consistent effects on the economic growth experienced by a country: on this matter, Fogel (2004) goes as far as claiming that height gains may have accounted for as much as 30% of British economic growth in the past two centuries.

Developing countries seem to be the ones that are bound to benefit the most from height gains, for at least two reasons. First, these countries are typically faced with a much poorer health environment than their more developed counterparts, which implies that even comparatively small intervention may yield high returns. Second, given their lower level of development, automation is not as pervasive as in more developed countries, with physically demanding jobs still representing a consistent proportion of the total workforce. An overall healthier and stronger workforce would then lead to consistent productivity gains across a number of economic activities, hence fostering economic growth.

2.2.2 The development of cognitive capabilities

In addition to the physical development, the first years in life represent a crucial phase for the cognitive development as well. Previous research has highlighted how undernutrition in early life has a profound effect on the following development of the central nervous system and the hippocampus, reducing synapses and synaptic neurotransmitters (Levitsky and Strupp, 1995; Huang et al., 2003; Kar et al., 2008). Kar et al. (2008) conducted a study on the impact of malnutrition on the cognitive skills of a sample of Indian children, finding how malnourished children exhibited a poorer performance in most of neuropsychological tests involving working memory, attention, visual perception, and verbal comprehension.It should thus come as no surprise that height has been found to be positively correlated with schooling (Thomas and Strauss, 1997) and that stunting between 12 and 36 months of life was linked to lower overall school achievement (Grantham-McGregor et al., 2007).

Schooling , in turn, has a significant positive effect on economic growth, with studies documenting how an increase of 1 percentage point in the enrollment ratio is expected to raise the growth rate by 0.025 percentage points for primary education alone (Barro et al., 1991). If the effect of schooling on economic growth may seem low, – 'the private rate of return to an extra year has been estimated to oscillate between 5 and 15 percent'(p.81, Temple, 2001).

Given the documented positive impact of schooling on human capital formation and economic growth, and the profound impairment stunting entails on cognitive capabilities, stunting reduction presents a clear economic rationale.

2.2.3 The development of social capabilities

While the physical and cognitive effects of height gains occupy the vast majority of the discourse around stunting, other studies point at the impact growth faltering may have on non-cognitive skills. In a social context, being taller may infuse a certain authority or dominance, which in turn may shape the development of specific personal traits (e.g. leadership, motivation, persistence). On this matter, Frieze et al. (1990) suggest how taller people tend to be more successful due to reasons such as interpersonal dominance and self-confidence. Persico et al. (2004) go further, arguing how being taller may foster a teenager's participation in social activities (e.g. athletics, school clubs, etc.), thus, leading to a higher social adaptability. Personal traits that are likely to develop with social activities (e.g. leadership, charm, motivation, persistence) are undoubtedly valued in the workplace and, according to Lundborg et al. (2014), can explain as much as one-fifth of the total height premium. Hence, while perhaps more difficult to grasp and define, social skills might also be affected by stunting, adding a further channel through which the condition impacts economic success and labor productivity.

Stunting has been found to have important repercussions on the development of physical, cognitive and social skills. This, in turn, affects the living conditions and labor productivity of an individual, impeding a country to fully exploit its human capital potential and generate economic growth. Developing countries are bound to benefit the most from a reduction in stunting, given the prevalence of physically demanding jobs and the higher marginal gains that health improvements would entail in these settings.

Having extensively discussed the different channels through which stunting affects the development of children and the following human capital formation, we now turn to its major determinants, focusing particularly on the role of diseases and infections stemming from poor sanitation environments.

2.3 The determinants of stunting

Adult height is affected by a number of different determinants, which can be distinguished as genetic and environmental. With regard to the former, the debate is still open in the academic world, with some anthropologists arguing for the importance of genetics in the determining the height of a population and others downplaying its role (Koepke and Baten, 2005). On this matter, Vella et al. (1994) argue how environmental factors tend to play a much more important role than genetics. Among the environmental factors, the literature on stunting distinguishes three major determinants: nutrition, health environment, and prenatal health (Voth-Gaeddert et al., 2018).

Of the three, nutrition has typically received most of the attention. Bharvaga (2001) argues how ‘poor-quality diet supplies inadequate quantities of protein and micronutrients such as iron’ which are deemed essential for a child’s development. Since proteins and iron are typically found in animal products, the author stresses the budget constraint of poor households as a crucial factor perpetuating the high prevalence of undernourished people in developing countries. The impact of undernutrition has been found to persist for as long as three generations (Victora et al., 2008), indicating the severe long-term consequences of malnutrition.

The health environment that characterizes an individual’s early life, however, is also important. Adair and Guilkey (1997) observed the growth of Filipino children between birth to 24 months of age, finding how diarrhea and febrile respiratory infections significantly increased the prevalence of stunting across the population. In fact, diarrhea entails a consistent loss of nutrients and fluids, which are not absorbed by the body, considerably worsening undernutrition conditions. The relationship, however, is not merely unidirectional: stunting has been found to significantly increase the likelihood of infections such as diarrhea, pneumonia, and measles (Black et al., 2013), hence leading to the continuous reinforcement of a vicious cycle. With the risk of future stunting being the highest in the first 29 months (Vella et al., 1994) and diarrhea being most common among 6- and 23-month old children (Arabi et al., 2012), it is understandable that diarrheal diseases represent the second leading cause of death in children under 5 years old (WHO, 2013).

Extensive research on the timing of growth faltering has unveiled how the condition does not develop solely after birth, but is significantly affected also from prenatal health. Harmful behaviors (e.g. smoking, drinking) or inadequate nutrition can in fact damage the fetus and reduce the growth potential after birth. In Indonesia, newborn length was estimated to be the strongest determinant of length-for-age at 12 months (Schmidt et al., 2002), while in Sub-Saharan Africa, maternity care (through, for instance, improved iron supplementation) has been found positively correlated with stunting reduction (Buisman et al., 2019). This, in turn, implies that maternal health is also a major determinant of stunting: if a pregnant mother is exposed to pathogens and an overall poor health environment, that would have repercussions on her health and, consequently, on the fetus development. While the proportion of growth failure for which prenatal health is responsible has yet to be fully understood, previous research has estimated it predicts as much as 20% of stunting prevalence (Christian et al., 2013).

The three factors discussed do not act alone, but mutually reinforce themselves: diarrheal diseases impede the correct absorption of nutrients by the body, thus, hampering growth, while poor nutrition and health environment have also dire consequences on maternal health before and during pregnancy, thus affecting the prenatal health of the child.

Figure 1 summarizes the causes and consequences of stunting highlighted by previous literature and discussed in the study. The figure represents the theoretical framework guiding the construction of the model and the subsequent analysis carried out in this paper.

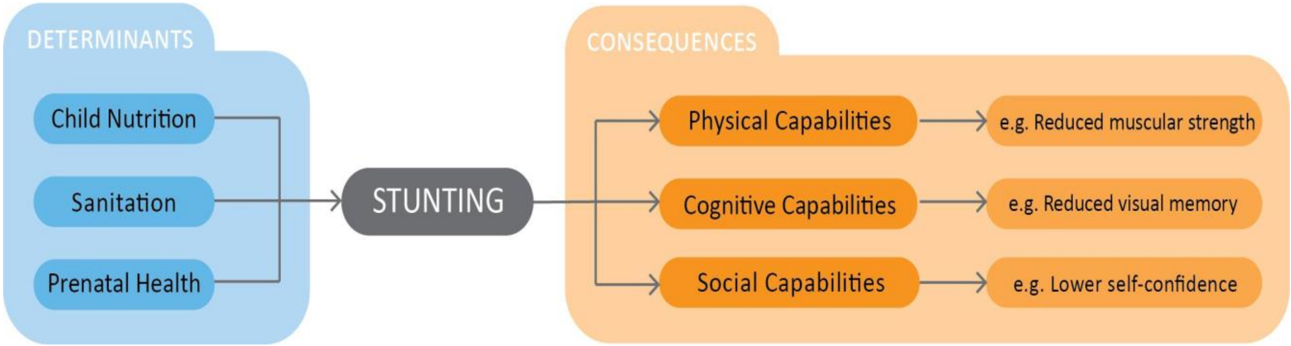

Figure 1: Theoretical framework of the study

The present section has addressed the major determinants of stunting, which have been found to lie in an scarce nutrition, poor health environment, and inadequate prenatal health. The next section will address the prevalence of stunting in India, discussing past and present trends of stunting prevalence and the most common causes described by the literature.

2.4 Stunting prevalence in India: the 'Asian enigma'

Although worldwide stunting prevalence has experienced a decrease in the last decades, stunting currently affects as many as 144 million children under 5 worldwide, with the highest number of cases being registered in Asia. Figure 2 reports the country percentages of the world total number of children (less than 5 years old) with a negative height-for-age z-score: Asian countries figure among the ones with the highest values, with India alone presenting as much as 32% of worldwide malnourished children. The country has a long history of undernutrition, with stunting prevalence remaining virtually unchanged (approx. 46%) between 1995 and 2005.

Economic indicators, on the contrary, showed consistent improvements: during the same period GDP growth per annum averaged 6.14, while the poverty headcount ratio (1.90$ poverty line) declined from 45.9% to 38.2% (World Bank, 2020). Yet, as of 2005, undernutrition levels in India were higher than for most countries in Sub-Saharan Africa, with Indian children being shorter, on average, than their African counterparts (Spears, 2013). It is worth noting that, starting in 2005, India did experience a decline in stunting prevalence, which reached 38.4% in 2015, suggesting the beginning of a positive trend. Nevertheless, the improvement in the past decade has barely allowed the country to match the level of Nepal and Bangladesh, both countries significantly poorer than India.

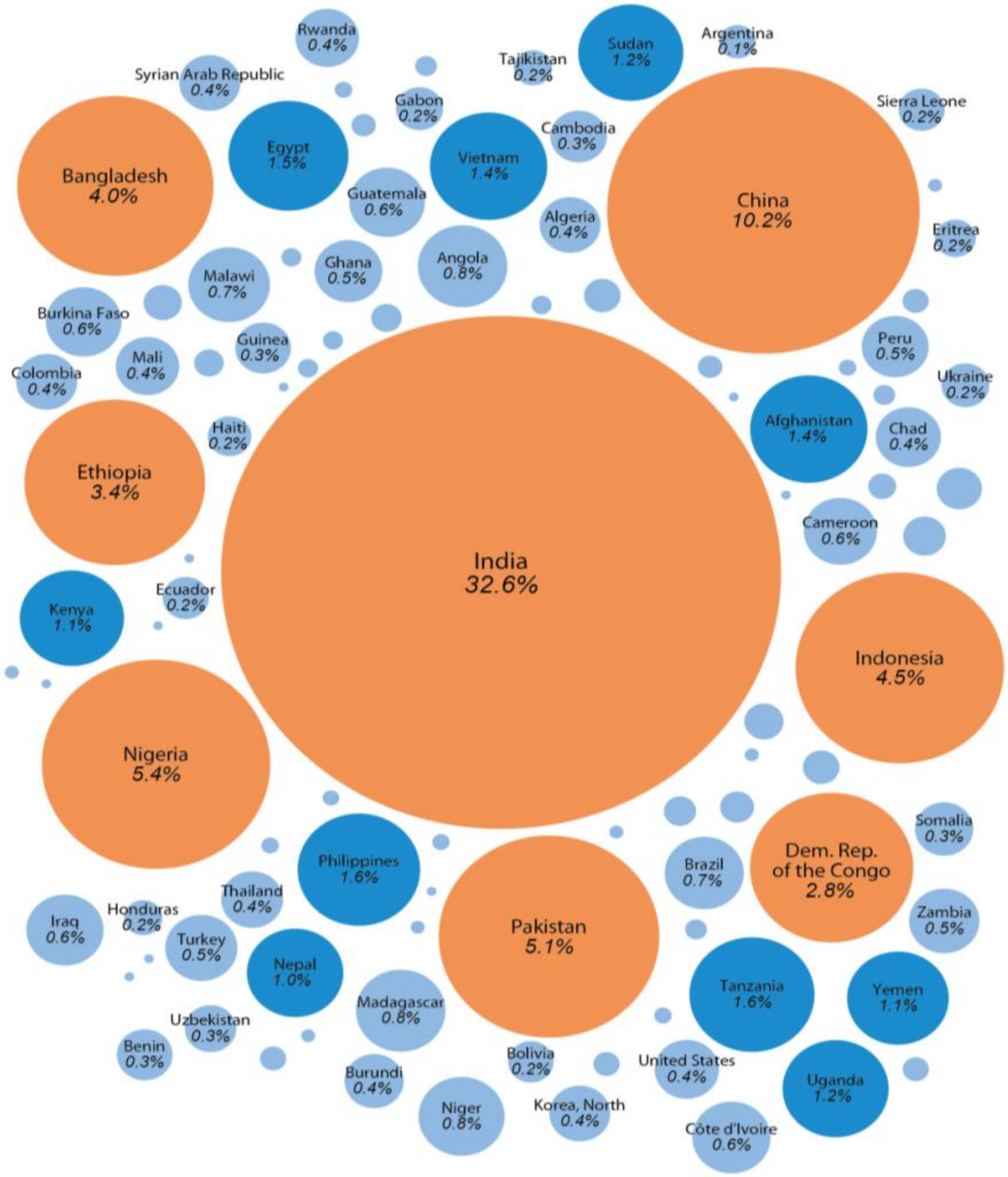

Figure 2: Malnourished children (z-score height-for-age < 0), country percentages of world total

Source: Chambers and von Medeazza, 2014

The contrasting trend between stunting prevalence and economic performance in the country has attracted increasing attention within the academic world, with studies (Deaton and Dreze, 2009) observing how 'the nutrition situation in India is full of puzzles' and that economic growth seems to have largely failed to reduce malnutrition, with average calories intake declining over the years. In light of the consistent dietary improvements that followed India's positive economic performance, the high prevalence of malnutrition and overall poor performance on health indicators has been defined by some as the 'Asian enigma' (Ramalingaswami et al., 1996).

According to Chambers and von Medeazza (2014), however, the literature attempting to address the so-called ‘enigma’ suffered from a significant flaw, i.e. it was mostly related to food intake, food security for the poor, and feeding programmes, completely neglecting sanitation.

To neglect the sanitary determinants of stunting in a country where more than 72% of rural people practice open defecation (also OD in this study) (Kumar, 2017) means to ignore the enormous burden fecal transmissible illnesses (henceforth FTIs) have on growth faltering. While the impact of diarrheal episodes has already been discussed, FTIs comprise a much wider range of diseases (e.g. environmental enteropathy[3] and hookworms) which entail profound consequences in terms of nutrients absorption and anemia.

One of the first studies attempting to shed light on this otherwise overlooked channel was by Spears (2013). By investigating the relationship between open defecation and height across 65 developing countries, the author estimates how the former explains a consistent part of the latter. In particular, when restricting the sample to South Asia and Sub-Saharan Africa, open defecation reduces the height gap between the children in the two regions by 30 percent.

The present section has underlined the stark contrast between India’s economic performance and stunting prevalence. To this date, India exhibits stunting levels comparable only to less developed countries. Open defecation practices have been found to be a major factor in explaining what had been previously described as the ‘Asian enigma’.

The following section begins by discussing the prevalence of open defecation practices in South Asia, explaining the main reason underlying its persistence in the region. It then goes on by reviewing the previous literature over the relationship between open defecation and children’s health indicators, with a particular attention towards stunting prevalence, finally presenting the research question of the study.

2.5 Stunting and Open Defecation

South Asia is home to over 60% of the people practicing open defecation worldwide (UNICEF, 2020). In India alone, approx. one in three people practiced open defecation as of 2015. Figure 3 illustrates how more than half of the world population practicing open defecation can be found in India, with Indonesia, Pakistan, Nigeria, and Ethiopia following suit, but presenting much lower values. If the current level of development of SSA may justify the lower diffusion of adequate sanitation in the region, the same cannot be said about many Asian countries.

[3] Infection caused by the ingestion of faecal bacteria.

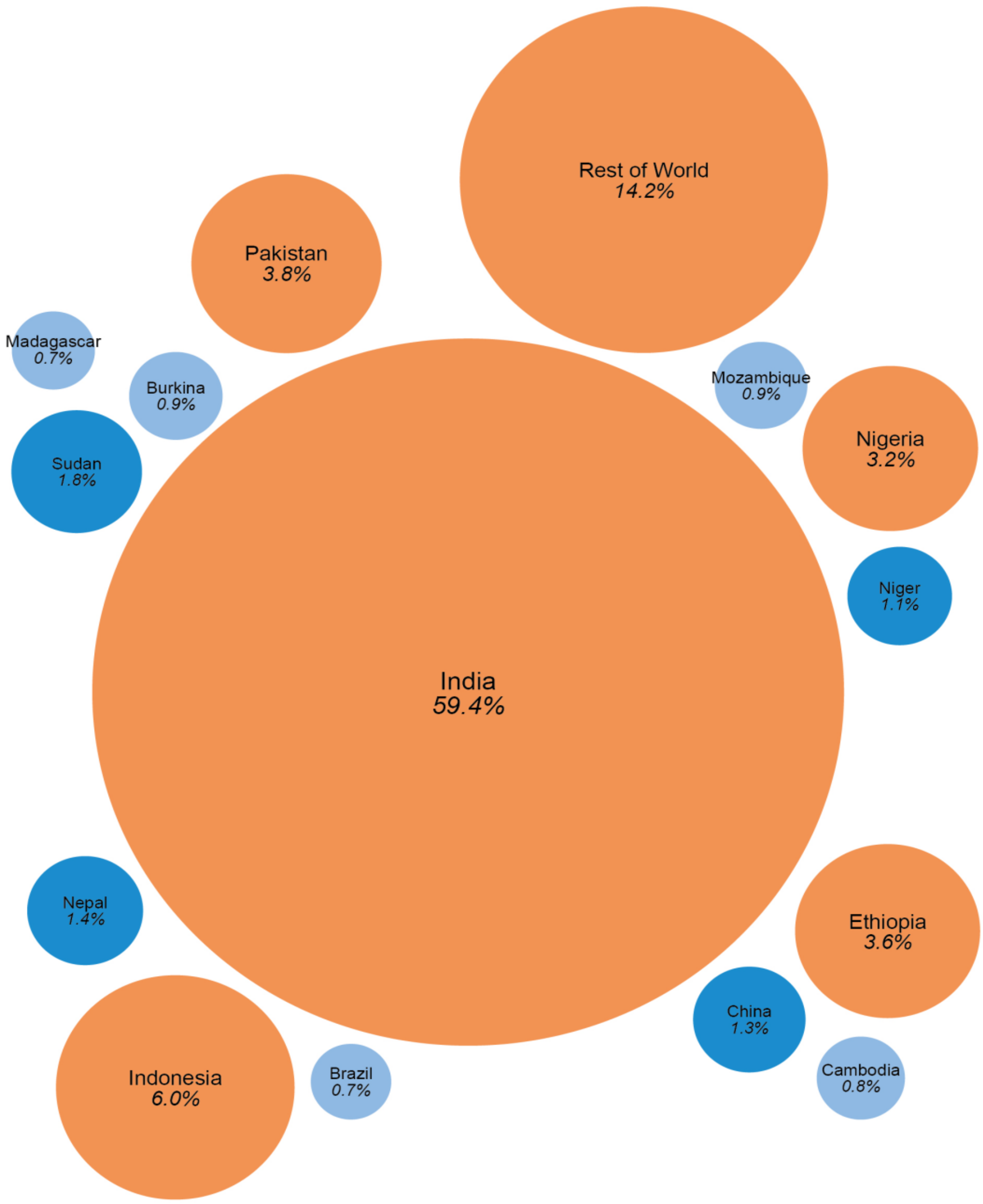

Figure 3: Open defecation, country percentages of world total

Source: Chambers and von Medeazza, 2014

Table 1 shows the recent trend of open defecation across a subgroup of South Asian countries. The prevalence of OD practices in these countries has experienced a sharp decline over the last years, with Bangladesh exhibiting a value as low as 1.2% as of 2015. What is probably the most interesting information in the table, however, is the significant gap between the countries already in 2000, with India and Nepal facing a much higher OD prevalence than their peers, a factor that clearly influenced the current level. What is the reason of this striking difference? Why does India present such a high level, in contrast to, for instance, Bangladesh, which shared the same history as recently as 1948? What is the underlying determinant of open defecation practices?

Table 1: People practicing open defecation (% of population) for a subgroup of South Asian countries, 2000-2015

Country/Year	2000	2005	2010	2015
India	72,6	58.5	44.6	31
Nepal	66,8	53	39.6	26.6
Bangladesh	17,6	11.8	6.3	1.2
Pakistan	40,2	31.1	22.3	13.7

Source: World Bank, 2020

According to Coffey et al. (2017), what is 'perhaps the most important barrier to the adoption of affordable latrines in rural India is the unique history of untouchability and its continuing practice'. What allowed Bangladesh to reduce its open defecation prevalence, in fact, was the construction of inexpensive soak pit latrines. These latrines, however, need to be emptied manually, a task that is considered physically and ritually dirty according to Hinduism. Traditionally, this task has been carried out by dalits (i.e. the untouchables), since any individual belonging to a different caste would refuse to perform it. In recent years, however, dalits have begun to move away from these socially degrading jobs, hence making it more difficult or expensive for a higher caste individual to find someone to empty the pit (LoPalo et al., 2019).

As a response to the changing social environment, the rural rich have turned towards more expensive latrines with large pits or cemented tanks that can be emptied by a truck, while the poor has continued to practice open defecation (Coffey et al., 2014). This has led to what Coffey et al. (2014) defined 'missing middle rungs' in the sanitation ladder, i.e. the scarcity of intermediate sanitation measures between the extremes of open defecation and flush toilets.

The influence of religious and cultural beliefs on the practice of open defecation would provide a good explanation as for why Nepal and India are faced with such a high prevalence, whereas Bangladesh, characterized by a prevalently Muslim population, not only improved considerably over the years, but was benefiting from a much lower value to start with. In recent years, the Indian government has invested increasing resources to reduce the open defecation rate, launching sanitation campaigns and promoting the construction of latrines in rural areas. These efforts, while not succeeding in eliminating open defecation, have contributed to reduce its prevalence across the country (Gupta et al., 2019).

The absence of the 'middle rungs' described by Coffey et al. (2014) has dire consequences on health indicators. The association between OD, stunting and overall poor children health has been documented by a number of studies (Quattri and Smets, 2014, Hathi et al., 2017; Spears et al., 2014; Spears, 2018), while high quality sanitation has been found to reduce the odds of child diarrhea, stunting and mortality by 13%, 27% and 23%, respectively (Fink et al., 2011).

Given the persistently high open defecation and stunting rates exhibited by India, research on the topic has consistently focused on the country. Spears et al. (2014) conduct a study on 112 Indian districts, making use of four different data sources for stunting, open defecation, infant mortality, and calorie consumption, respectively. The study documents that a 10 percent increase in open defecation is associated with a 0.7 percentage increase in stunting.

Further literature on the topic has added measures of population density to the model. On this matter, Spears (2014) observes how 'even if the number of people defecating in the open decreases, average exposure to open defecation density can increase if the population is increasingly arranged such that people live near to people who defecate in the open'. By making use of 2001 and 2011 census data, the author performs an analysis of India both at the state- and district-level. The findings suggest how the average exposure to density of OD has increased between 2001 and 2011, hence effectively undoing any effort to improve sanitation across the country. Other studies have further confirmed the negative interplay between population density, OD, and stunting at the international level (Hathi et al., 2017, Spears, 2018). In particular, although rural areas are typically characterized by a higher OD prevalence, densely populated urban settings have also been found to have a consistent positive effect on stunting rates.

The relationship between Hinduism and open defecation practices, on the other hand, has recently come under scrutiny by a number of quantitative studies. Vyas and Spears (2018) investigate the relationship between open defecation and religion by using a large dataset comprising Bangladesh, Nepal and India. The authors find a high correlation between open defecation rate and Hindu population. Moreover, the study sheds light on a spillover effect between households identifying with different religion: Hindu households are found to alter their behavior in areas that are predominantly non-Hindu and vice versa. This 'neighborhood effect' persists also with regard to children health: Spears and Geruso (2018) document how the different prevalence of open defecation practices allows to explain a consistent gap in infant and neonatal mortality rates between Muslim and Hindu households. Muslim households benefit from both a lower OD rate and the vicinity with other Muslim households, which together contribute to create a less favorable environment to the diffusion of FTIs. The 'neighborhood effect' has thus important implications both on the OD rate and the health of its population, covering an important role in fostering or hampering children full development. Whether the relationship persists also with regard to stunting prevalence, however, has, to the best of the author's knowledge, yet to be addressed.

The studies discussed have investigated the relationship between open defecation and children health by accounting for population density and religion. To this date, however, a more comprehensive research including all of the factors mentioned seems to be missing from the literature. Furthermore, most of the literature made use of data no more recent than 2011 (when the last census took place). As previously discussed, India has been experiencing a discrete success in driving down both the stunting prevalence and open defecation rate in recent years. Whether the positive relationship between the indicators persists despite the recent improvements represents one of the key contributions of the study. Moreover, the different open defecation prevalence among Hindus and Muslims has been addressed only with regard to infant and neonatal mortality rate. It is unclear whether the relationship persists also with regard to morbidity (i.e. stunting). The main research question that this study addresses is:

What are the implications of open defecation practices on stunting prevalence in India, 2015?

The main independent variables of the model are open defecation rate, open defecation density, and Muslim faith. Given the negative externalities of open defecation and the ‘neighborhood effect’ of different religious groups, the variables are included both at the household and neighborhood level.

3. Data & Methodology

The present section describes the data and methods used for the present analysis. First, the data source is presented, and the variables employed are listed and discussed. Second, the methodology is defined through the model outline and the main regression equation. Finally, the sensitivity analysis that will follow the main results is discussed.

3.1 The Data: The National Family and Health Survey (NFHS)

This study addresses the relationship between stunting, open defecation rate, open defecation density, and religion in India, unifying the approaches taken by previous research.

In order to conduct the analysis, comprehensive data on health and socio-economic indicators should be used. In the case of India, this can be gathered from two major data sources: census data and the National Family and Health Survey (henceforth NFHS), India’s national version of the Demographic and Health Survey (DHS).

Unfortunately, the last census data for India was collected in 2011, and thus, could possibly provide an inaccurate representation of the country’s present situation. As previously argued, India has been experiencing a discrete improvement with regard to both open defecation and stunting prevalence in the last decade. To use 2011 data would thus neglect the latest trend exhibited by the country. For this reason, the latest NFHS-4, conducted in 2015-6, was preferred for the current analysis. The nationally representative survey was conducted across each of the country’s 36 states and 640 districts. The households interviewed were selected through a two-stage sample design, ensuring proper randomization and weighted representation of each area. The NFHS-4 provides extensive data over socio-economic and health indicators at the household and individual (e.g. man, woman, child) level, reporting information such as height-for-age, birth history, education, and asset ownership.

The DHS Program offers different versions of the data collected in order to make the analysis easier. For the present study, the children dataset was chosen, which includes relevant health information on the child and its mother. The dataset is comprised of 259,627 observations of children below 5 years.

3.2 Dependent and independent variables

The research is carried out at the individual level including PSU- and district-level variables. PSUs represent the clusters chosen in the first stage of the survey. In rural settings, PSUs may be entire villages (of 100-200 households), whereas they are census enumeration blocks in urban areas (Geruso and Spears, 2018). In line with previous studies (Spears, 2013; Hathi et al., 2017; Geruso and Spears, 2018) PSU-level variables will be used to test for the ‘neighborhood effect’ caused by OD practices. The following sections present and discuss the variables employed in this thesis. The variables construction (when applicable) and possible limitations are also described.

3.2.1 Dependent variable: stunting rate

Stunting rate represents the dependent variable of the study, and is obtained, as most of the variables, from the NFHS-4. Height is measured for each child <5 years (or 60 months). Children with less than 24 months are measured by using the recumbent length, whereas older children are measured standing. The variable used to assess the stunting rate is height-for-age, i.e. ‘the difference between the height of the observed child and the average height of healthy children, scaled by the standard deviation of child height in the healthy population’ (Coffey and Spears, 2018). A child with a negative height-for-age z-score is shorter than the average child in a healthy population of reference, with stunting being defined as having a height-for-age z-score lower than -2.

Stunting prevalence has been widely used as a measure of a child’s health in a number of studies. Coffey and Spears (2018), however, warn against the use of a dichotomized variable, since it does not contain any useful information on the overall distribution, focusing only on the lower tail. When analyzing time trends, it might be the case, for instance, that we observe a reduction in stunting cases, but an increase in the number of children with a negative height-for-age. By using stunting prevalence, we would thus overlook an overall worsening of children health, incorrectly defining it as an improvement. Although the present analysis is cross-sectional, stunting rate will be replaced by height-for-age as part of a sensitivity analysis, in order to check the robustness of the results. This in turn allows to look at the impact of our variables of interest on overall undernutrition, and not only on the extreme case of stunting.

3.2.2 Main independent variables

This section presents the three main independent variables of interest: open defecation rate, open defecation density, and Muslim religion.

Open defecation rate. Both NFHS and census data contain information on the household toilet facility, and thus, could be used for the open defecation rate variable. As previously discussed, NFHS data is, however, more recent, hence providing a more accurate picture of the current situation in India. Moreover, the two questions are also slightly different: the NFHS refers to toilet usage, whereas census data asks about toilet ownership. Although one may think it is not a relevant difference, Spears (2013) notes that many households owning a toilet do not use it. As a consequence, to be using census data for the variable would possibly lead to an underestimation of the open defecation rate. This precaution, however, does not completely solve the possible mismeasurement issue. Household-level questions are affected by social desirability bias. Studies (Vyas et al., 2019) comparing individual- and household-level questions on open defecation found a 20 percentage points difference in the open defecation rates between the two samples. Unfortunately, there is no practical way to correct this bias, which represents a clear limitation of this research.

Following Spears and Geruso (2018) a household is defined as practicing OD if it reports using no facility or a bush. Since open defecation does not affect solely those who practice it, but also the households around them, a neighborhood-level variable indicating the proportion of open defecators within a certain PSU is also added to the analysis, in line with previous research (Spears, 2013; Hathi et al., 2017).

Open defecation density. Open defecation density is constructed following Spears (2013, 2014). The scholar constructs a measure of ‘open defecators per square km’ by multiplying population density per square kilometer with the fraction of people reporting open defecation. The variable is then transformed in a logarithm, due to the wide variation in population density. The formula to obtain the variable is the following:

$$\ln\left(\frac{number\ of\ open\ defecators}{total\ count\ of\ households} \times \frac{number\ of\ people}{area\ (in\ square\ km)} + 1\right)$$

Since the NFHS does not provide information on population density, the data is retrieved from the 2011 Census at the district level. It is important to note how, by matching the census data to the 2015-6 NFHS, one is effectively assuming that population size at the district level did not change significantly or changed approx. the same way between 2011 and 2015. This is necessary due to the absence of population size data closer to the year the survey was taken. According to Spears (2018), such assumption is reasonable given the relatively low tendency of households to change their living place: the 2012 Indian Human Development Survey (IHDS) documents how less than 1 percent of the whole sample reported to have moved in the previous 5 years.

Another implicit assumption lies in not considering the different population growth rates experienced by each district within those years. If a certain district had experienced a higher population growth, then the following adverse effects stemming from increased open defecation density would certainly have an impact on the newborns, hence affecting the stunting rate. There seems to be no practical way to control for this bias, which is therefore added to the limitations of the study. Nevertheless, being the variable included in its logarithmic form, and thus, being its variance reduced, the bias originating from this assumption is likely to be small.

Religion. Following Geruso and Spears (2018), this study includes a variable representing whether a certain household is Muslim. As previously argued, however, the negative impact of OD does not depend only from the single individual, but it is associated with the practices kept within the neighborhood. On the matter, Vyas and Spears (2018) have found a lower prevalence of open defecation practices among Muslims and Hindus living in prevalently Muslim neighborhoods. Other studies (Geruso and Spears, 2018) suggest that the lower open defecation rate that characterizes Muslim neighborhoods is at the root of an otherwise inexplicable life expectancy advantage of Muslims compared to Hindus. In order to test whether this channel holds also with respect to stunting prevalence, the present study will include the proportion of Muslims living within a certain PSU. PSUs is an approximation of the neighborhoods in the model, thus allowing to account for the ‘neighborhood effect’ of the Muslim population.

Given the minority status of Muslims in India, the positive spillover effect their presence in a neighborhood can cause on Hindus’ open defecation practices may be limited, at least at lower levels. If the relationship was nonlinear, the *muslim* variable at the PSU level might fail to capture the effect intended. This possibility is addressed in Section 4.3 by replacing the proportion of Muslims within a certain PSU with binary variables indicating whether the PSU is characterized by a certain percentage of Muslims (e.g. 60%, 70% 80%).

3.2.3 Control variables

After the three variables of interest and their neighborhood-level counterparts, a vector of controls is added to the regression. The model includes four types of control variables: socio-economic determinants, birth demography, prenatal health, and child nutrition.

Socio-economic determinants. While questions referring explicitly to the income level of the household are absent from the NFHS, the survey reports a series of questions over the ownership of durable goods (e.g. car, truck, fridge, etc.). From these and other questions related to asset ownership, the DHS Program has constructed a Wealth Index, an indicator that attempts to measure the economic condition of a certain household. The index divides the sample into quintiles, ranging from the poorest to the richest. It is important to note that, among the factors taken into consideration by the index, there is the toilet facility used by the household, which is also present among the key independent variables. For this reason, Spears and Geruso (2018) choose to construct a new index excluding the toilet facility question. The presence of multicollinearity issues arising from the joint inclusion of the variables in the model will be controlled for in Section 4.1. Further controls referring to socioeconomic status also include an urban indicator, whether the household is Hindu, total number of household members, and the mother's total years of education attained. Husband's total years of education was also considered as an additional indicator, but the information was collected for far fewer households. Its inclusion would have thus consistently reduced the sample size, possibly undermining the robustness of the results. When included, its coefficient was insignificant across all the specifications, hence suggesting it had no predictive power.

Birth demography. Following Hathi et al. (2017), birth demography is also taken into account by adding mother's age at first birth, number of children ever born, sex of the child, month of birth, year of birth, indicators for first born child and multiple births. The occurrence of multiple births or the order of birth can significantly alter the care given to the child, thus, possibly resulting in phenomena such as growth faltering. Moreover, an interaction term between sex of the child and its age in month was further added to the model. Previous studies (Panagariya, 2012) have questioned the appropriateness of WHO height-for-age reference population for Indian children, deeming them inaccurate. If that was the case, our dependent variable would be clearly biased. Spears (2013) addresses the issue by including an interaction term between age-in-months and sex, a choice implemented also in the current study. Being sex and age-in-months the factors used in the calculation of the height-for-age z-scores, by including their interaction in the model one is effectively accounting for the possible inaccuracy of the WHO reference population with respect to Indian children. Finally, since the survey was conducted in different places in different times of the year, different children would be eligible in different areas (Spears and Geruso, 2018). For this reason, month-of-birth indicators and year of birth (linearly) were also included among the birth demography controls.

The literature has distinguished three major determinants of stunting, i.e. poor health environment, inadequate nutrition and insufficient prenatal health (see Figure 1 for reference). While the health environment is accounted for by the OD variable, the other determinants are yet to be added to the analysis. To exclude these variables from the regression would clearly lead to an omitted variable bias, affecting the reliability of the results. For this reason, two groups of controls, referring respectively to children nutrition and prenatal health, will be added to the regression.

Prenatal health. With regard to prenatal health, in line with previous literature (Hathi et al., 2017; Spears, 2018) the analysis includes the mother's Body Mass Index (henceforth BMI), height-for-age, anemia level, whether the mother has a health insurance and whether she smokes. As previously argued, malnutrition persists for as long as three generations, and can be thus observed in the mother's height-for-age and BMI. Moreover, unhealthy habits such as smoking can negatively affect the fetus formation during the pregnancy. Finally, the presence of a health insurance is an adequate approximation of the maternal care a woman receives during her pregnancy.

Child nutrition. With regard to child's nutrition, the NFHS does not collect information over calories intake. For this reason, the present study makes use of variables indicating whether the child was given protein-rich food in the past 24 hours (e.g. chicken, fortified baby food, legumes, eggs, and milk products). It is important to stress that to use these variables effectively implies assuming they are representative of the usual diet followed by the child, which may not reflect the truth. Finally, together with the variables related to child nutrition, an indicator referring to whether the child is being given iron pills is also included. Table 2 presents a list of the controls employed in the model together with the type of variable and a short definition.

In this section, the variables employed in the analysis were presented. Open defecation and Muslim faith, both at the household and PSU level, together with district-level open defecation density, represent the key independent variables. Controls are also included and divided in four main subgroups: socio-economic status, birth demography, child nutrition, and prenatal health (see Table 2 for an overview). Next section proceeds to describe the methodology implemented in the study, presenting the model and regression equation used for the analysis. At last, a subsection is dedicated to the sensitivity checks that will be carried out following the main results.

Table 2: Controls employed in the models

Group of controls	Variables	Type of variable	Definition
Socio-economic status (SES)	Urban area	Binary	Whether the household is situated in an urban area
	Hindu	Binary	Whether the household is Hindu
	Count of household members	Continuous	Total number of household members
	Wealth Index	Categorical	Quintile distribution of wealth in the sample
	Mother's education	Continuous	Total n° of education years of the mother
Birth demography	Multiple birth	Binary	Whether the child was part of a multiple birth

	Child's sex	Binary	Whether the child is female
	Age of mother at first birth	Continuous	Age of mother at first birth
	Number of children	Continuous	Total children ever born
	Month of birth	Binary	Binary variable for each month of birth
	Year of birth	Continuous	Year of birth
	First born	Binary	Whether the child is the firstborn
	Interaction between age-in-months and sex of the child	Interaction term	Interaction between age-in-months and sex of the child
Prenatal health	Anemia	Binary	Whether the mother presents moderate or severe levels of anemia
	Health insurance	Binary	Whether the mother has a health insurance
	Height-for-age	Continuous	Mother's height-for-age z-score
	Body Mass Index (BMI)	Continuous	Mother's BMI
	Smoking	Binary	Whether the mother smokes
Child nutrition	Iron pills	Binary	Whether the child was given iron pills
	Dairy products	Binary	Whether the child was given this food in the last 24h
	Fortified baby food	Binary	Whether the child was given this food in the last 24h
	Eggs	Binary	Whether the child was given this food in the last 24h
	Legumes	Binary	Whether the child was given this food in the last 24h
	Birds	Binary	Whether the child was given this food in the last 24h

3.3 Methodology

3.3.1 Model Outline

This study aims to address the relationship between stunting prevalence and open defecation in India in 2015. The analysis is conducted at the individual level. Following previous studies, open defecation is included at the household, PSU, and district ('open defecators per square km') level. Furthermore, Muslim faith is added at both the household and PSU level, having the literature highlighted the lower open defecation prevalence within those communities (Vyas and Spears, 2018).

The main results are reported through nine different models. Models 1-4 include the key variables of interest progressively: Model 1 regresses stunting prevalence on household open defecation rate, Model 2 includes household Muslim faith, Model 3 adds the PSU level variables, and finally open defecation density is included in Model 4. Models 5-9 include the groups of control variables in the following order: prenatal health (5), child nutrition (6), socio-economic determinants (7), birth demography (8), and state dummies (9).

The model includes robust standard errors to control for heteroskedasticity, i.e. non-constant variance of the error terms across observations. Given the several different levels at which heteroskedasticity may occur (e.g. PSU, district, or state), no cluster is specified. In order to account for unobservable features, district dummies were also considered. Being the analysis cross-sectional and the open defecation density variable at the district level, however, the inclusion of district fixed effects would have caused multicollinearity issues. For this reason, state dummies were included instead. Although not allowing to account for the same bias, the inclusion of state indicators allows to reduce the confounding factors affecting the results.

Since the outcome variable is binary, the relationship between the dependent variable and the regressors is nonlinear, thus making the linear Ordinary Least Squares (OLS) model not suitable for the analysis (Carter Hill et al., 2011). If one were to use it regardless, the model would return predicted probabilities outside the interval [0,1]. A more appropriate model to conduct the analysis is the Logit model, which is employed in this study. To employ a Logit model significantly affects the interpretation of the outcomes: the coefficients no longer represent the marginal effects of the independent variables on the dependent variable (as in the OLS) and are not immediately interpretable. In order to circumvent the issue, the table reports the odds ratios. Odds ratios reflect the predicted change in the odds given a one unit increase in the predictor. The nominator is given by the probability that the outcome variable will be 1, whereas the denominator represents the probability that the outcome variable will be 0. For instance, a continuous variable's odds ratio of 2 would imply that a one unit increase of the predictor doubles the odds that the outcome variable will be 1.

The equation employed in the study is the following:

$$
\begin{aligned}
stunting_i = \beta_0 &+ \beta_1 open\ defecation_i + \beta_2 open\ defecation_p + \beta_3 muslim_i + \beta_4 muslim_p \\
&+ \beta_5 open\ defecation\ density_d + \beta_6 prenatal\ health_m + \beta_7 nutrition_i + \beta_8 SES_i \\
&+ \beta_9 birth\ demography_i + \gamma_s + \varepsilon_i
\end{aligned}
$$

Where *i* represents individual children below 5 years, *p* stands for PSU, *d* are districts, *m* refers to mothers of the children, and *s* refers to states. The dependent variable, $stunting_i$, is a binary variable equal to 1 if the height-for-age z-score of the child is lower than -2 and 0 otherwise. $open\ defecation_i$ is a binary variable equal to 1 if the household (and thus, child) practices open defecation, and 0 otherwise, whereas

$open\ defecation_p$ is the proportion of households surveyed that practices open defecation within a certain PSU. In the same way, $muslim_i$ is a binary variable equal to 1 if the household (and thus, child) is Muslim and 0 otherwise, while $muslim_p$ represents the proportion of Muslims within a certain PSU. PSUs are the clusters used in the first randomization stage of the survey. According to Geruso and Spears (2018) PSUs can be approximated to entire villages in rural settings and census enumeration blocks in urban areas. In this study, PSUs approximate neighborhoods. Finally, $open\ defecation\ density_d$ is a variable indicating the 'density of open defecators per square km' within a certain district *d*, constructed following Spears (2018). As previously discussed, a number of controls is also included in the regression. In line with the literature, the controls have been divided according to the major determinants of stunting (see Figure 1), i.e. prenatal health and child nutrition. The remaining controls have been grouped in birth demography and socio-economic determinant groups, respectively, to facilitate the interpretation of the changes in the estimates once the variables are added to the model. Finally, γ_s represents a vector of 36 state dummies, and ε_i the error term.

The dataset comprises of 259,627 observations of children below 5 years. Some of the variables included in the model, however, are available for far fewer observations. When these missing values were randomly distributed, this would not represent a problem, and specifications involving different samples would still be comparable. On the other hand, when the distribution was non-random, specifications referring to different data samples could not be compared. To rule out any issue arising from this possibility, the 'exceeding' observations were dropped from the first specifications, reducing the sample size to 124,205 observations across all models.

Finally, a further aspect to take into consideration is the geographical nature of the dataset in use. Observations may be nonindependent over space, with observations closer to each other being more similar to those further apart. This may be the result of common underlying factors or a spatial spillover effect between the observations (Fischer and Griffith, 2008). In the presence of unaccounted spatial autocorrelation, the model would be misspecified, and the consequent estimates unreliable. Being a spatial regression out of the scope of the present study, this possibility has not been investigated further, hence representing a limitation of this thesis.

3.3.2 Sensitivity Analysis

In order to control for the robustness of the main results, a series of sensitivity analyses is also included in the study. Severe stunting prevalence and height-for-age z-score replace stunting prevalence in Model 1 and 2, respectively. Model 3 replaces the proportion of Muslims within a PSU with indicator variables representing a number of Muslim population's share (60%, 70%, 80%, and 90%). Finally, Models 4-6 are carried out at the PSU, district, and state level, respectively.

First, the dependent variable is substituted with severe stunting prevalence. Severe stunting is defined as a height-for-age z-score below -3, i.e. one point lower than a standard stunting condition. The specification allows to control how does the relationship with the key independent variables changes when focusing on a subgroup of even lower height-for-age z-scores.

Second, stunting prevalence is substituted with height-for-age. As previously argued (see 3.2.1) the dichotomous nature of the stunting indicator restricts the analysis to the lower tail of the distribution, effectively neglecting the relationship of the independent variables with comparatively healthier children. Since height-for-age is a continuous variable, the OLS model is used for this specification. The variable is normally distributed, thus satisfying OLS assumptions. Third, in order to account for the possible nonlinear relationship between the Muslim population and open defecation practices (and thus, stunting prevalence), the proportion of Muslims within a certain PSU is substituted with indicator variables representing different Muslim shares. The joint inclusion of those variables in the sensitivity analysis allows to detect a significant relationship at a more precise level (e.g. between 80% and 90% Muslim share).

Further sensitivity analyses are also carried out at the PSU, district, and state level, respectively. Whether the relationship is confirmed at these higher aggregation levels will depend on the quantity of information that are lost by collapsing the data and the strength of the relationship between dependent and independent variables. By using PSU-, district-, and state-averages of the variables the heterogeneity existing within each of those units is lost, with possible significant consequences on the estimates.

Only the most complete specification (Model 9 in the previous section) is reported for each sensitivity analysis. At the district and state level, however, a number of control variables are dropped. First, at these aggregation levels, certain variables (e.g. month and year a child was born) can be assumed to be randomly distributed, and thus, become irrelevant. Second, given the far fewer observations available for those models (640 and 36, respectively), to include a high number of variables can lead to overspecification issues. Third, a number of variables reported a VIF above the value of 10, hence suggesting the existence of multicollinearity. The wealth index figured among the third group of variables. Given the importance of a wealth indicator as a proxy of socio-economic status, the variable was substituted by 5 asset ownership indicators referring to electricity, radio, bicycle, motorcycle, and car.

State dummies were included in all models with the exception of the last one, since they would clearly lead to perfect multicollinearity, being the analysis cross-sectional. A detailed list of the variables included in Model 5 (district) and 6 (state) is available in Table 1 in the Appendix, whereas Table 2 and 3 report the VIF values for the variables causing multicollinearity in the district and state model, respectively.

4. Results

This section will present the results of the study. It will be divided into three main subsections. The first subsection will discuss the descriptive statistics of the variables employed, making use of tables and figures. The second subsection will present and briefly describe the main results. Finally, the third subsection will be dedicated to the results of the sensitivity analysis.

4.1 Descriptive Results

Table 3 reports the descriptive statistics of the variables employed in the study. Controls are excluded, since their high number would have negatively affected the legibility of the table. Table 4 and 5 in the Appendix provide the descriptive statistics for those variables. Stunting, open defecation rate and Muslim faith are binary variables, which allows to interpret their means as the proportion of stunting children, households practicing open defecation and Muslim households, respectively. In order to provide a feel of the overall distribution of height-for-age z-scores among Indian children below 5 years, the continuous variable height-for-age is also included.

The average height-for-age z-score for Indian children below 5 years is -1.29 standard deviations lower than the reference value for a healthy population, with a standard deviation of 1.61. Stunting prevalence is as high as 32 percent, with half the sample being below the -1.35 z-score threshold. The statistics confirm that stunting represents a widespread burden for the country, affecting one out of three children below 5 years. Further descriptive statistics show that almost one out of two Indian households in the sample practices open defecation. This number, while high, is likely to be an underestimation of the true value, given the social desirability bias that affects household-level questions (Vyas et al., 2019). Finally, the binary variable for the Muslim population confirms that Hinduism is the predominant religion in India: only 17 percent of the households in the sample are Muslim, with the average PSU exhibiting the same percentage. Table 3 also includes the measure of open defecation density, as defined in the previous section. The expression with which the variable has been calculated impedes any significant interpretation of the coefficient, but the rather high standard deviation hints at the even greater variance of the variable before the logarithmic transformation.

Table 3: Descriptive statistics of the main variables

Variables	Mean	Standard Deviation	Median (50^{th} percentile)	Observations (N)
Height-for-age	-1.29	1.61	-1.35	124,205
Stunting	0.32	0.47	0	124,205
Open defecation	0.46	0.50	0	124,205
Open defecation (PSU)	0.47	0.36	0.5	124,205
Muslim	0.17	0.38	0	124,205
Muslim (PSU)	0.17	0.30	0	124,205
Open defecation density	5.15	1.34	5.22	124,205

Note: The variables stunting, open defecation and Muslim are binary, hence, the mean values represent the percentage of children under 5 being stunted, the percentage of households practicing open defecation and the percentage of households being Muslim within the dataset, respectively.

The dichotomous nature of the indicators in use impedes a graphic visualization of the relationship between the variables. To circumvent the issue, Table 4 reports the pairwise correlation coefficients. Stunting is positively correlated with open defecation rate, both at the household and the PSU level. The *muslim* variable, on the other hand, presents an extremely low positive correlation coefficient (0.006 and 0.005 at the household and PSU level, respectively), posing doubts over the expected negative relationship between the values. In line with the literature previously discussed, the Muslim indicator is negatively correlated with open defecation both at the household and PSU level. Particularly, the correlation between the proportion of Muslims living in a PSU and the open defecation rate within that PSU is equal to -0.207, suggesting the existence of the 'neighborhood effect' highlighted by Geruso and Spears (2018). Muslims tend to practice less open defecation than other communities, but the lower rate does not seem to translate into a lower stunting prevalence. Finally, the variable indicating open defecation density is positively correlate with stunting prevalence, although by a rather low coefficient. The poor correlation may be the result of the higher level of aggregation at which the variable has been constructed (i.e. district level), which in turn might not give an accurate representation of the open defecation density within each PSU.

Table 4: Correlation table of stunting and key variables of interest

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) Stunting	1.000					
(2) Open defecation	0.144	1.000				
(3) Open defecation (PSU)	0.138	0.760	1.000			
(4) Muslim	0.006	-0.148	-0.174	1.000		
(5) Muslim (PSU)	0.005	-0.155	-0.207	0.845	1.000	
(6) Open defecation density	0.098	0.366	0.486	0.032	0.038	1.000

Further evidence of the relationship between the dependent variable and the variables of interest is provided through the use of ArcGIS. The software allows to create maps reflecting the spatial distribution of the variables across the Indian territory. Figure 4 shows the spatial patterns of stunting prevalence, open defecation rate, Muslim population, and open defecation density. Natural breaks were chosen to represent the distribution of the variables across space, meaning that class ranges were formed following similar characteristics between observations, hence minimizing the average standard deviation from a class mean.

The maps seem to confirm the relationships suggested by Table 4: districts characterized by a higher stunting prevalence typically exhibit also a higher open defecation rate. With regard to the Muslim population, the relationship seems less clear: districts characterized by a higher Muslim population may be affected by a high or low stunting rate, in line with the extremely low correlation coefficient discussed before. The Muslim population appears mostly concentrated along the borders with Pakistan and Bangladesh, reflecting the partition implemented by the British in 1947. Finally, with regard to the open defecation density variable, the map highlights a significant 'belt' of northern and north-eastern districts characterized by high values, corresponding approximately to the states of Uttar Pradesh, Bihar, Jharkhand and West Bengal. Those districts are also characterized by a high stunting prevalence, probably being among the observations increasing the positive correlation coefficient exhibited by the variable in Table 4.

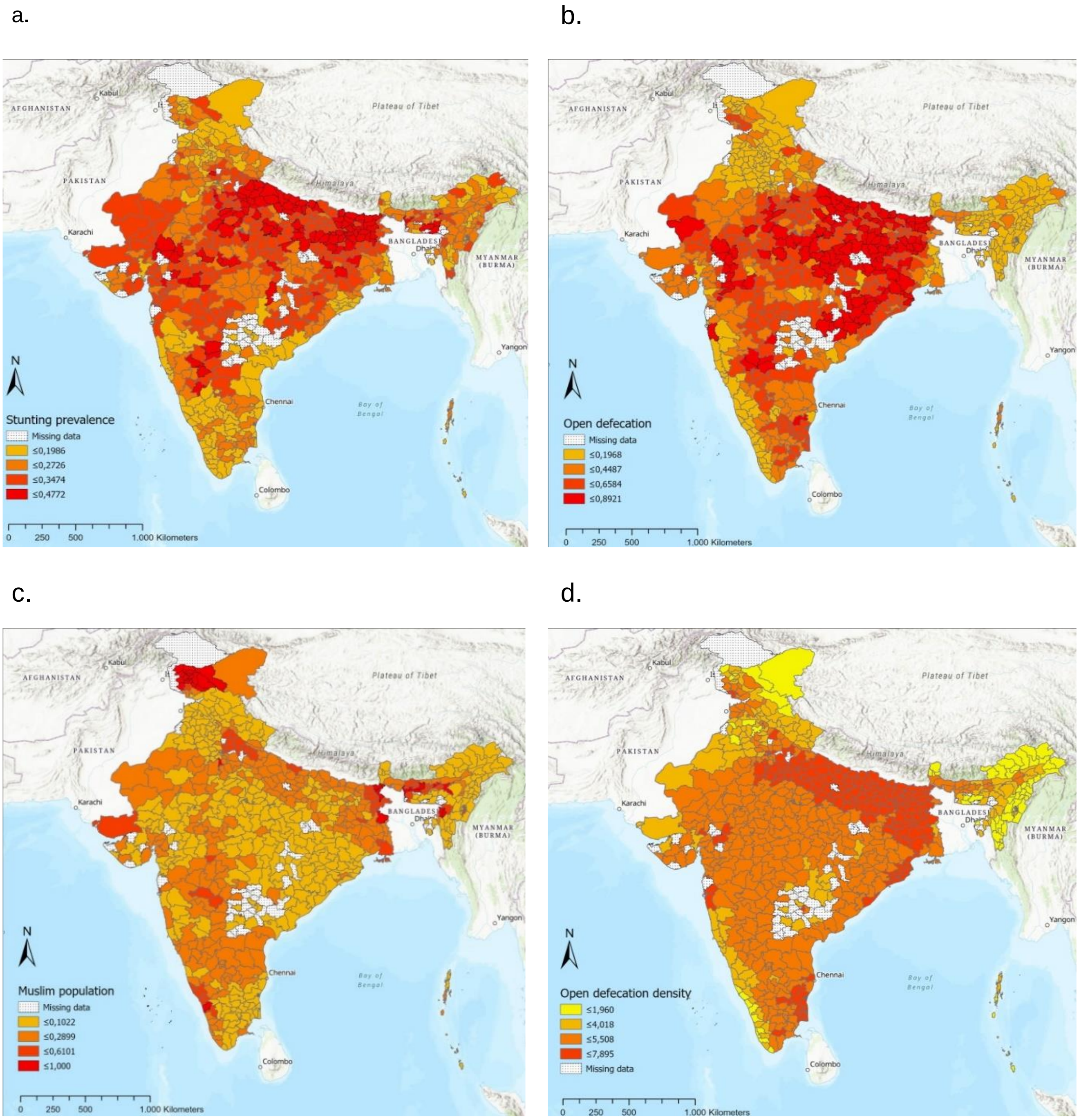

Figure 4: Spatial patterns of a) stunting prevalence, b) open defecation, c) Muslim population, d) open defecation density (district-level)

The variables *open defecation* and *muslim* are included both at the individual and PSU level. The rationale for such decision has been already extensively explained in the previous section (see 3.2.2). However, multicollinearity issues may arise when including variables that are highly correlated between each other. In fact, Table 4 reports high pairwise correlation coefficients for *open defecation* (0.760) and *muslim* (0.845) between the individual and PSU level.

In order to rule out any issues stemming from such a high correlation, the Variance Inflation Factor (VIF) was calculated (see Table 6 in the Appendix). Following the discussion on the possible multicollinearity between the wealth index and open defecation rate (see 3.2.3), the VIF for the former was also included. A rule of thumb at a value of 10 is commonly applied to determine whether there is any issue arising from the correlation between the independent variables. None of the VIFs exceeds 3.5, well below the value of 10, suggesting multicollinearity should not be a problem in this case.

The present section has described the data source of the study, while also discussing the construction, advantages, and limitations of the variables used. Different methods (i.e. tables and maps) have been employed to show the correlation between the key variables of interest and the dependent variable: stunting and open defecation show a positive correlation, whereas the correlation coefficients are rather small for open defecation density and close to zero for Muslim faith.

4.2 Empirical Results

The results of the main regression are reported in Table 5. All standard errors reported are robust. The significance level chosen is 5%. The coefficients reported are the odds ratios. An odds ratio of 1 implies that the regressor does not significantly predict the outcome variable, since the probabilities that the dependent variable assumes a value of 1 or 0 given the predictor is the same (50%).

Table 5: Results from the main model

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Open defecation	1.899***	1.943***	1.520***	1.519***	1.388***	1.384***	1.160***	1.105***	1.081***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Muslim		1.209***	1.023	1.023	1.101**	1.102**	1.072	1.040	0.978
		(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)	(0.07)	(0.07)
Open defecation (PSU)			1.617***	1.490***	1.307***	1.312***	1.148***	1.204***	1.173***
			(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Muslim (PSU)			1.361***	1.330***	1.233***	1.228***	1.123**	1.123**	1.159***
			(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
OD density				1.061***	1.023***	1.024***	1.005	1.003	0.982
				(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	124,205	124,205	124,205	124,205	124,205	124,205	124,205	124,205	124,205
Prenatal Health					Yes	Yes	Yes	Yes	Yes
Child Nutrition						Yes	Yes	Yes	Yes
SES Determinants							Yes	Yes	Yes
Birth Demography								Yes	Yes
State Fixed Effects									Yes

Note: dependent variable is stunting prevalence (height-for-age z-scores below -2). The estimates reported *are the odds ratios. An odds ratio of e.g. 2 means that the odds of a child being stunted double when the independent variable increases by one. Table 7 in the Appendix reports the extended results for each control employed.*

Open defecation is significant at 5% confidence level across all specifications, both at the household and PSU level. Initially, the odds ratios exhibited by the variables are rather high, with the odds of a child being stunted increasing by half if the household practices open defecation (column 4). As the various groups of controls are added to the model, however, the odds ratios decline to 1.081 and 1.173 in the most complete specification, meaning that, all else equal, whether a household practices open defecation (or a 100% increase in the open defecation rate within a certain PSU) increases only slightly the odds of a higher stunting prevalence (8% and 17%, respectively). The inclusion of socio-economic determinants seems to play a significant role in driving down the values, suggesting the existence of possible underlying channels.

The *muslim* variable at the household level, on the other hand, is mostly insignificant across the different models. Unexpectedly, the Muslim population within a certain PSU has a positive and significant relationship across all specifications. This seems to suggest that a higher Muslim population tends to increase the odds of stunting prevalence among children, in contrast with the findings of Geruso and Spears (2018) with regard to the relationship between Muslim neighborhoods and infant and neonatal mortality rates. Finally, open defecation density does not seem to entertain any significant relationship with the dependent variable: while initially positive and significant, the odds ratio turns insignificant with the inclusion of socio-economic determinants. The result seems to relegate the only district-level variable of the model to an insignificant role in explaining stunting prevalence. This may be a consequence of the higher aggregation level at which the variable is constructed, which might capture less information than variables at the PSU level.

4.3 Sensitivity Results

In order to control for the robustness of the results obtained, this section presents a series of sensitivity analyses. Table 6 presents the most complete specification for each sensitivity analysis. Model 1 replaces stunting prevalence (height-for-age z-score below -2) with severe stunting prevalence (height-for-age z-score below -3). Model 2 introduces the height-for-age z-score as the dependent variable. Model 3 substitutes the *muslim* variable at the PSU level with indicators for different shares of Muslims within a certain neighborhood. Model 4, 5, and 6 show the estimates when the data is collapsed at the PSU, district and state level. As previously discussed, some control variables were dropped in the last two models to avoid over-specification and multicollinearity issues (see Table 1 in the Appendix). Given the much lower number of observations available in Model 5 and 6, the estimates obtained in those specifications are the least reliable.

Table 6: Results from the sensitivity analysis

VARIABLES	(1) Severe Stunting	(2) Height-for-age	(3) Stunting	(4) PSU	(5) District	(6) State

Open defecation	1.023	-0.045**	1.083***	0.022***	0.059***	0.123**
	(0.04)	(0.02)	(0.03)	(0.01)	(0.02)	(0.05)
Muslim	0.865	-0.011	1.028	0.024***	0.012	0.024
	(0.09)	(0.05)	(0.07)	(0.01)	(0.02)	(0.05)
Open defecation (PSU)	1.330***	-0.064**	1.165***			
	(0.08)	(0.03)	(0.05)			
Muslim (PSU)	1.147*	-0.061*				
	(0.08)	(0.04)				
OD density	0.998	0.004	0.982	0.002	-0.000	-0.003
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
Muslim (PSU) >60%			1.119*			
			(0.07)			
Muslim (PSU)>70%			0.929			
			(0.08)			
Muslim (PSU)>80%			0.999			
			(0.09)			
Muslim (PSU)>90%			1.037			
			(0.08)			
Observations	124,205	124,205	124,205	26,356	640	36
R-squared		0.198		0.226	0.796	0.887
Prenatal Health	Yes	Yes	Yes	Yes	Yes	Yes
Child Nutrition	Yes	Yes	Yes	Yes	Yes	Yes
SES Determinants	Yes	Yes	Yes	Yes	Yes	Yes
Birth Demography	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	

Note: odds ratios are reported for the Logit model employed in Column 1 and 3. Robust standard errors are included in each model. The dependent variable in Column 4, 5 and 6 is the proportion of stunting children within a certain PSU, district and state, respectively.

The results from the sensitivity analysis seem to confirm a significant negative relationship between open defecation and children undernutrition. The relationship appears to be stronger with regard to severe stunting rather than overall height-for-age. In particular, with severe stunting as the dependent variable, open defecation exhibits a positive and significant coefficient only at the PSU level. The *muslim* variable, on the other hand, exhibits a significant coefficient only in Model 4. When the proportion of Muslims within a certain neighborhood is replaced by indicator variables in Model 3, the estimates are insignificant. Finally, open defecation density is insignificant across all specifications, in line with the estimates presented in Table 5.

5. Discussion

The present study investigated the relationship between open defecation and stunting prevalence. The channel at the root of the positive relationship between the indicators lies in the pathogens diffused in the air by feces, which in turn cause FTIs (e.g. diarrhea, hookworms) and impede a correct absorption of nutrients in the body of the children. The positive and significant coefficient exhibited by open defecation prevalence at the PSU level seems to confirm the underlying channel, in line with previous literature (Spears, 2013; Hathi et al., 2017). The magnitude of the estimate, on the other hand, is rather low, suggesting a limited relationship between the variables. A possible explanation for this difference may come from the data used for the present analysis. While past studies have typically focused on multiple survey rounds (typically as recent as 2011), this research focused exclusively on the latest NFHS, which took place in 2015-6. As previously discussed (see Section 2.5), open defecation rate in India has been experiencing a consistent decline in the last 15 years. It is therefore possible that the comparatively smaller coefficients are a consequence of the overall lower open defecation rate within the country.

In contrast with the expectations, open defecation at the household level was also found to be positively correlated with stunting prevalence, with a coefficient only slightly lower than the PSU-level variable. If the indicator suffered from spatial autocorrelation, i.e. if households practicing open defecation would tend to be closer to one another, then the variable would indirectly pick up a part of the neighborhood effect, thus explaining the positive and significant coefficient. Interestingly, the coefficient is insignificant when severe stunting (height-for-age z-score below -3) is used as the dependent variable. A possible explanation is therefore that whether a household practices open defecation does have an effect on stunting prevalence, but it is insignificant for more extreme malnutrition cases, where only the proportion of household practicing open defecation within a certain area affects the stunting rate.

With regard to the religion variables, the findings fail to confirm the expectations. Whether a household is Muslim does not significantly predict stunting prevalence. The small pairwise correlation coefficients reported in Section 4.1 were already suggesting the possibility of an insignificant relationship. The household level variable presented a significant and positive coefficient only in Model 5 and 6, which becomes insignificant once socio-economic indicators are added to the model. The positive relationship between the variables was thus likely to be a consequence of the poorer socio-economic status of Muslims in India, rather than any feature related to cultural and religious beliefs.

The proportion of Muslims within a certain PSU, on the other hand, exhibits a positive and significant coefficient across all specifications. As previously discussed, the positive spillover effect Muslims have on Hindus' open defecation practices may be insignificant until much higher levels. Model 3 in Table 6 explored the existence of a nonlinear relationship between the variables. When Muslims make up 60% of a certain neighborhood, the coefficient is positive and significant, although only at the 10% level[4], to then turn insignificant at higher rates. A possible explanation lies in the fact that, being Muslims a minority in India, even when a neighborhood is composed by a 60% of Muslim population, the spillover effect works from Hindus towards Muslims rather than the other way around. The overall open defecation rate would then increase, hence worsening stunting prevalence. When the proportion of Muslims raises further the relationship becomes insignificant. This seems to suggest that the comparatively lower open defecation rate within Muslim communities is not strong enough to translate into a significantly lower open defecation rate among Hindus within the same neighborhood, nor into a lower overall stunting prevalence. Previous studies focusing on OD practices, Muslim population, and health indicators have either addressed exclusively the link between the first two (Vyas and Spears, 2018) or chosen different health indicators (i.e. infant and neonatal mortality rates) (Spears and Geruso, 2018). The lower open defecation rate that characterizes Muslim communities may present a significant negative relationship with mortality indicators but not be strong enough to persist even with regard to morbidity indicators (i.e. stunting and overall malnutrition).

Finally, open defecation density fails to significantly predict stunting prevalence. The higher aggregation level at which the variable was constructed may have affected the estimates. In fact, the variable varies only across districts, hence exhibiting the same value for PSUs within a certain district characterized by different open defecation rates. Yet, if that was the case, the sensitivity analysis carried out at the district level should report an insignificant coefficient for the open defecation rate, which is positive and significant instead. A possible explanation may thus lie in the way the variable was created: 2011 census data were used to obtain population size and area in square kilometers, which were then matched to the district division in place in 2015. In between those years, however, the district division of India has changed. The differences between the 2011 and 2015 Indian district division may have therefore affected the calculation of the variable, and thus, the estimates obtained.

These results should be interpreted with caution, due to the several limitations of this research. First, the existence of spatial autocorrelation was not accounted for. When the data were spatially autocorrelated, the estimates would be consequently affected. Second, child nutrition was proxied by the food given to the child in the last 24 hours. This approximation may be inaccurate and thus, omit a fundamental determinant of stunting prevalence. Third, previous research (Vyas et al., 2019) has unveiled the existence of a social desirability bias that leads to consistent underestimation of the actual open defecation rate when household level questions are used. This may in turn have led to downward biased estimates.

6. Conclusion

What was previously described as the ‘Asian enigma’, i.e. the fundamental mismatch between the economic prosperity enjoyed by South Asian countries (India, in particular) and the poor performance with regard to health indicators, has recently been explained through the lack of adequate sanitation. Sanitation, together with child nutrition and prenatal health are considered the most important determinants of stunting prevalence (Voth-Gaeddert et al., 2018). The high rate of open defecation in India, deep-rooted in religious

[4] And thus, insignificant in this study, where a 5% significance level is chosen.

and cultural beliefs proper of Hinduism, represents a key challenge to the diffusion of adequate sanitation infrastructure within the country.

A number of studies have documented the significant positive relationship between open defecation and stunting prevalence within the country (Spears, 2013; Hathi et al., 2017; Gosha et al., 2014; Spears et al., 2014). Moreover, the high population density characterizing the Indian territory has been found to further exacerbate the negative impact on stunting prevalence (Hathi et al., 2017; Spears, 2018). Muslim communities, on the other hand, typically exhibit lower open defecation rates than Hindu ones (Vyas and Spears, 2018), which translates into a positive impact on neonatal and infant mortality rates (Geruso and Spears, 2018).

The aim of this study was to contribute to the understanding of the relationship between open defecation and stunting. The research attempted to answer the following research question:

What are the implications of open defecation practices on stunting prevalence in India, 2015?

The key variables of interest were open defecation rate, Muslim population, and open defecation density ('open defecators per square kilometer'). The study made two fundamental contributions to the literature. First, it investigated whether the positive relationship between open defecation and stunting persists despite the recent discrete success obtained by the Indian government promoting the construction of toilet facilities (Gupta et al., 2019). Second, it investigated whether the positive effect of the lower open defecation rate exhibited by Muslim communities extended to morbidity indicators (i.e. stunting).

The results show a significant and positive relationship between open defecation and stunting prevalence, both at the household and neighborhood level. The relationship remains significant when controlling for socio-economic determinants, child nutrition, prenatal health and birth demography. The estimates are robust also to sensitivity analyses replacing stunting with severe stunting and height-for-age, respectively, and at the neighborhood, district, and state level. Open defecation at the household level, in particular, is significant across all models and sensitivity analyses with the exception of the one where stunting prevalence is substituted by severe stunting (height-for-age z-score below -3). The finding seems to suggest that whether a household practices open defecation has a positive impact on comparatively less severe malnutrition conditions. For more extreme cases, only the proportion of households practicing open defecation within a certain neighborhood seems to entertain a significant positive relationship with stunting prevalence.

Open defecation density and Muslim faith, instead, did not present the expected positive and negative relationship, respectively, with the dependent variable. Several explanations were offered to account for this event. With regard to open defecation density, the higher aggregation level might have affected the adequacy of the indicator. Moreover, the imperfect match between the district division in place in 2011 and 2015 may have led to inaccuracies in the construction of the variable.

With regard to Muslim faith, on the other hand, a possible explanation is the existence of a nonlinear relationship between this variable and stunting prevalence. The relationship may be positive for lower percentage of Muslims within a certain area, since those would adopt the same practices as their Hindu neighbors, with the consequent adverse impact on stunting prevalence. As the percentage of Muslims raises, however, the estimate remains insignificant. The lower open defecation rate within Muslim communities may thus fail to translate in significant advantages in terms of stunting prevalence.

Previous literature (Geruso and Spears, 2018) has documented a positive effect of lower open defecation practices on mortality indicators (e.g. infant mortality rate). The findings of this study, however, seem to suggest that the positive relationship does not extend towards morbidity indicators (i.e. stunting).

The study presents a number of limitations that call for caution in the interpretation of its results. First, the geographical nature of the data used may entail risks of spatial autocorrelation that were not accounted for. Second, child nutrition was approximated by the food given in the last 24 hours, which can be inaccurate and thus, omit an important determinant of stunting from the model. Third, the open defecation rate obtained through household level survey questions has been found to be an underestimation of the actual value (Vyas et al., 2019) due to social desirability.

Future research on the topic should be directed at providing further evidence with regard to the relationship between stunting prevalence and open defecation density. For instance, the construction of variables able to approximate population density at lower aggregation levels (e.g. PSUs) should be considered. Moreover, the possible positive health externalities of the lower open defecation rate exhibited by Muslims offer the starting point for several spatial analyses involving different sets of morbidity indicators.

Open defecation, although being the target of several government sanitation campaigns, remains a great obstacle to the improvement of the living conditions of many Indian citizens, affecting the development of physical, cognitive and social skills of the future generations. The investment of increasing resources to tackle the issue would have important consequences on both economic development and growth, considerably increasing the country’s human capital.

References

Adair, L. S., & Guilkey, D. K. (1997). Age-specific determinants of stunting in Filipino children. *The Journal of nutrition*, *127*(2), 314-320.

Arabi, M., Frongillo, E. A., Avula, R., & Mangasaryan, N. (2012). Infant and young child feeding in developing countries. *Child development*, *83*(1), 32-45.

Barro, R. J. (1991) Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics*, 106, 2, 407–443.

Bhalotra, S., Valente, C., & Van Soest, A. (2010). The puzzle of Muslim advantage in child survival in India. *Journal of Health Economics*, *29*(2), 191-204.

Bhargava, A. (2001). Nutrition, health, and economic development: Some policy priorities. *Food and Nutrition Bulletin*, *22*(2), 173-177.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., ... & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The lancet*, *382*(9890), 427-451.

Buisman, L. R., Van de Poel, E., O'Donnell, O., & van Doorslaer, E. K. (2019). What explains the fall in child stunting in Sub-Saharan Africa?. *SSM-population health*, 8, 100384.

Carter Hill, R., Griffiths, W. E., & Lim, G. C. (2011). Principles of Econometrics.

Chambers, R., & von Medeazza, G. (2014). Reframing undernutrition: faecally-transmitted infections and the 5 As.

Christian, P., Lee, S. E., Donahue Angel, M., Adair, L. S., Arifeen, S. E., Ashorn, P., ... & Hu, G. (2013). Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low-and middle-income countries. *International Journal of Epidemiology*, 42(5), 1340-1355.

Crimmins, E. M., & Finch, C. E. (2006). Infection, inflammation, height, and longevity. *Proceedings of the National Academy of Sciences*, *103*(2), 498-503.

Deaton, A., & Drèze, J. (2009). Food and nutrition in India: facts and interpretations. *Economic and political weekly*, 42-65.

Dinda, S., Gangopadhyay, P. K., Chattopadhyay, B. P., Saiyed, H. N., Pal, M., & Bharati, P. (2006). Height, weight and earnings among coalminers in India. *Economics & Human Biology*, *4*(3), 342-350.

Fischer, M. M., & Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. *Journal of Regional Science*, 48(5), 969-989.

Fogel, R. W. (2004). *The escape from hunger and premature death, 1700-2100: Europe, America, and the Third World* (Vol. 38). Cambridge University Press.

Geruso, M., & Spears, D. (2018). Neighborhood sanitation and infant mortality. *American Economic Journal: Applied Economics*, *10*(2), 125-62.

Gupta, A., Khalid, N., Desphande, D., Hathi, P., Kapur, A., Srivastav, N., ... & Coffey, D. (2019). Changes in open defecation in rural north India: 2014-2018.

Hoddinott, J., Alderman, H., Behrman, J. R., Haddad, L., & Horton, S. (2013). The economic rationale for investing in stunting reduction. *Maternal & child nutrition*, *9*, 69-82.

Koepke, N., & Baten, J. (2005). The biological standard of living in Europe during the last two millennia. *European review of economic history*, *9*(1), 61-95.

Kumar, A. (2017). Beyond toilets and targets: sanitation mission in India. *Development in Practice*, *27*(3), 408-413.

Levitsky, D. A., & Strupp, B. J. (1995). Malnutrition and the brain: changing concepts, changing concerns. *The Journal of nutrition*, *125*(suppl_8), 2212S-2220S.

Lundborg, P., Nystedt, P., & Rooth, D. O. (2014). Height and earnings: The role of cognitive and noncognitive skills. *Journal of Human Resources*, *49*(1), 141-166

Mara, D., & Evans, B. (2018). The sanitation and hygiene targets of the sustainable development goals: scope and challenges. *Journal of Water, Sanitation and Hygiene for Development*, *8*(1), 1-16.

Persico, N., Postlewaite, A., & Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy*, *112*(5), 1019-1053.

Quattri, M., & Smets, S. (2014). Lack of community-level improved sanitation causes stunting in rural villages of Lao PDR and Vietnam.

Ramalingaswami, V., Jonsson, U., & Rohde, J. (1996). The Asian enigma. *Progress of nations*, 10-17.

Schmidt, M. K., Muslimatun, S., West, C. E., Schultink, W., Gross, R., & Hautvast, J. G. (2002). Nutritional status and linear growth of Indonesian infants in West Java are determined more by prenatal environment than by postnatal factors. *The Journal of Nutrition*, 132(8), 2202-2207.

Schultz, T. P. (2002). Wage gains associated with height as a form of health human capital. *American Economic Review*, *92*(2), 349-353.

Spears, D. (2013). *How much international variation in child height can sanitation explain?*. The World Bank.

Spears, D. (2014). Increasing average exposure to open defecation in India, 2001–2011. Rice Institute

Stein, A. D., Wang, M., Martorell, R., Norris, S. A., Adair, L. S., Bas, I., ... & Victora, C. G. (2010). Growth patterns in early childhood and final attained stature: data from five birth cohorts from low- and middle- income countries. *American Journal of Human Biology*, *22*(3), 353-359.

Tarozzi, A., & Mahajan, A. (2007). Child nutrition in India in the nineties. *Economic Development and Cultural Change*, *55*(3), 441-486.

Temple, Jonathan. 2001. Growth effects of education and social capital in the OECD countries. OECD Economic Studies No.33.

Thomas, D., & Strauss, J. (1997). Health and wages: Evidence on men and women in urban Brazil. *Journal of econometrics*, *77*(1), 159-185.

Vella, V., Tomkins, A., Ndiku, J., Marshal, T., & Cortinovis, I. (1994). Anthropometry as a predictor for mortality among Ugandan children, allowing for socio-economic variables. *European Journal of Clinical Nutrition*, *48*(3), 189-197.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., ... & Maternal and Child Undernutrition Study Group. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *The lancet*, *371*(9609), 340-357.

Voth-Gaeddert, L. E., Stoker, M., Cornell, D., & Oerther, D. B. (2018). What causes childhood stunting among children of San Vicente, Guatemala: Employing complimentary, system-analysis approaches. *International journal of hygiene and environmental health*, 221(3), 391-399.

Vyas, S., & Spears, D. (2018). Sanitation and religion in South Asia: what accounts for differences across countries?. *The journal of development studies*, 54(11), 2119-2135.

Vyas, S., Srivastav, N., Mary, D., Goel, N., Srinivasan, S., Tannirkulam, A., ... & Coffey, D. (2019). Measuring open defecation in India using survey questions: evidence from a randomised survey experiment. *BMJ open*, 9(9), e030152.

WHO (2013) ‘Diarrhoeal disease Fact sheet N°330’, Geneva: World Health Organization, www.who.int/mediacentre/factsheets/fs330/en/ (accessed 1 March 2020).

Appendix

Table 1: Control variables included in the district- and state-level sensitivity analysis (Models 5 and 6 of Table 6)

Group of controls	Variables		Model 5 (district)	Model 6 (state)
Socio-economic status (SES)	Urban area		X	X
	Hindu		X	X
	Count of household members			
	Wealth Index was replaced by:	Household has electricity	X	X
		Household has radio	X	X
		Household has bicycle	X	X
		Household has motorcycle/scooter	X	X
		Household has car/truck	X	X
	Mother's education		X	X
Birth demography	Multiple birth		X	
	Child's sex		X	
	Age of mother at first birth		X	X
	Number of children		X	X
	Month of birth			
	Year of birth			
	First born			
	Interaction between age-in-months and sex of the child			
Prenatal health	Anemia		X	X
	Health insurance		X	X
	Height-for-age		X	X
	Body Mass Index (BMI)		X	

	Smoking	X	X
Child nutrition	Iron pills	X	X
	Dairy products	X	X
	Fortified baby food	X	X
	Eggs		
	Legumes	X	
	Birds	X	

Note: the sensitivity analyses presented in Model 5 & 6 of Table 6 are at the district- and state-level. The higher aggregation level necessarily reduces the observations available to 640 and 36, respectively. In order to avoid overspecification and multicollinearity issues, only a limited number of control variables was included. Both specifications presented a multicollinearity issue with the Wealth Index (see Table 5 and 6 below). The variable was then discarded and substituted by the variables reported above, indicating ownership of a number of assets. At the district level, comparatively more variables were preserved, since the multicollinearity issues that arose were less severe. At the state level, multicollinearity issues led to the exclusion of several variables within each subgroup.

Table 2: VIF values of control variables excluded at district level due to multicollinearity issues

Variables	VIF
Wealth Index	18.29
Eggs	12.91

Table 3: VIF values of control variables excluded at the state level due to multicollinearity issues

Variables	VIF
Wealth Index	124.03
Legumes	12.93
Birds	26.19
Eggs	13.6
BMI	33.74

Table 4: Descriptive statistics of the binary control variables employed in the study

Group of controls	Variables	Proportion over the whole sample	Standard Deviation

Socio-economic status (SES)	Household located in an urban area	28%	0.44
	Household is Hindu	78%	0.40
Birth demography	Multiple birth	2%	0.12
	Child is female	47%	0.49
	First born	38%	0.48
Prenatal health	Anemia	7%	0.26
	Health insurance	14%	0.35
	Smoking	6%	0.23
Child nutrition	Iron pills	25%	0.43
	Dairy products	7%	0.27
	Fortified baby food	11%	0.32
	Eggs	11%	0.32
	Legumes	11%	0.31
	Birds	4%	0.19

Note: the descriptive statistics were divided between binary and non-binary control variables in order to ease the interpretation of the mean for the former, which represents the proportion over the whole sample.

Table 5: Descriptive statistics of the non-binary control variables employed in the study

Group of controls	Variables	Mean	Standard Deviation	Min	Max
Socio-economic status (SES)	Count of household members	6.51	2.93	1	41
	Wealth Index	2.74	1.39	1	5
	Mother's education	6.38	5.19	0	20
Birth demography	Age of mother at first birth	20.97	3.53	5	47
	Number of children ever born	2.48	1.45	1	17
Prenatal health	Height-for-age	-2.02	0.99	-5.99	5.99
	Body Mass Index (BMI)	21.22	3.89	12.02	59.95

Table 6: VIF values of key independent variables at the individual and PSU level

Variables	VIF

Open defecation	2.61
Open defecation (PSU)	2.91
Muslim	3.35
Muslim (PSU)	3.42
Wealth Index	1.66

Table 7: Models 5-9 from the main results, coefficients reported for each control

VARIABLES	(5)	(6)	(7)	(8)	(9)
Open defecation	1.388***	1.384***	1.160***	1.105***	1.081***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Muslim	1.101**	1.102**	1.072	1.040	0.978
	(0.05)	(0.05)	(0.07)	(0.07)	(0.07)
Open defecation (PSU)	1.307***	1.312***	1.148***	1.204***	1.173***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Muslim (PSU)	1.233***	1.228***	1.123**	1.123**	1.159***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
OD density	1.023***	1.024***	1.005	1.003	0.982
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Height-for-age	0.996***	0.996***	0.996***	0.996***	0.996***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BMI	0.999***	0.999***	1.000***	1.000***	1.000***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mother is anemic	1.091***	1.091***	1.054**	1.049**	1.035
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Dairy products		1.053	1.082**	0.994	0.956
		(0.04)	(0.04)	(0.04)	(0.04)
Fortified baby food		0.863***	0.949	0.894***	0.934**
		(0.03)	(0.03)	(0.03)	(0.03)
Eggs		1.107***	1.117***	0.920**	0.974
		(0.04)	(0.04)	(0.04)	(0.04)
Legumes		1.159***	1.172***	0.970	0.993
		(0.04)	(0.04)	(0.03)	(0.03)
Birds		0.891**	0.866**	0.839***	0.846***
		(0.05)	(0.05)	(0.05)	(0.05)
Wealth Index			0.921***	0.911***	0.885***
			(0.01)	(0.01)	(0.01)
Urban area			1.078***	1.074**	1.074**
			(0.03)	(0.03)	(0.03)
Mother's education			0.955***	0.968***	0.970***
			(0.00)	(0.00)	(0.00)
Multiple birth				1.766***	1.826***
				(0.16)	(0.16)
Female				27.374***	27.889***
				(5.95)	(6.94)
Firstborn				0.886***	0.883***
				(0.02)	(0.02)
Age at first birth				0.995*	0.991***
				(0.00)	(0.00)
Children ever born				1.031***	1.019**

				(0.01)	(0.01)
Year of birth				1.042**	1.047
				(0.02)	(0.03)
Number of HH members			1.006*	1.003	0.999
			(0.00)	(0.00)	(0.00)
Hindu			1.055	1.031	0.964
			(0.05)	(0.05)	(0.06)
Mother does not smoke	0.767***	0.769***	0.886***	0.911***	0.901***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Health insurance	0.880***	0.870***	0.901***	0.909***	1.004
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Iron pills		1.074***	1.098***	1.042*	1.038
		(0.02)	(0.02)	(0.02)	(0.02)
Observations	124,205	124,205	124,205	124,205	124,205
Prenatal Health	Yes	Yes	Yes	Yes	Yes
Child Nutrition		Yes	Yes	Yes	Yes
SES Determinants			Yes	Yes	Yes
Birth Demography				Yes	Yes
State Fixed Effects					Yes

Note: the interaction term between age-in-month and sex was not included due to the high number of variables (120). Month dummies were also excluded.