Gender Composition of Children and Sanitation Behavior In India*

Deepak Saraswat†

September 2, 2019

Abstract

Open Defecation has been linked to various public health issues and has gained significant policy attention. Investing in adoption of better sanitation has also been advocated on the grounds of providing women with privacy and protection from potential harassment. Nonetheless, previous research has shown that due to son-biased preferences, households in India under-invest in outcomes for their female children. Linking the gender composition of children in Indian households to their sanitation behavior I find that, households reduce open defecation in presence of female children. Reduction in Open Defecation ranges from 7-14% in urban and 3-7% in rural areas, respectively. Various heterogeneity analysis suggest that the reduction in OD is driven by households facing a potentially high cost of harassment related to the female children. The findings in this paper provide a new first stage association between gender composition of children and sanitation behavior and also contribute to the economic literature on decision making in households belonging to developing countries.

JEL Codes: O10, O18, J16, J18
Keywords: Sanitation, Open Defecation, Gender Composition of Children, India

*I thank Dean Spears, Nishith Prakash, Stephen L. Ross, Nathan Fiala, Shrijata Chattopadhyay, Gaurav Chiplunkar and seminar participants at the University of Connecticut (Department of Economics, Department of Agricultural and Resource Economics), Southern Economic Association's 88th Annual Meeting, Jindal School of Government and Public Policy and Indian Institute of Management (Bangalore) for their helpful comments and suggestions. I declare that I have no relevant or material financial interests that relate to the research described in this paper. I am responsible for any errors that remain.

†Department of Economics, University of Connecticut. deepak.saraswat@uconn.edu.

1 Introduction

India accounts for 60% of world's open defecation. Open defecation, henceforth OD (defecating in open places, behind bushes, near roads, near railway tracks etc.), imposes a large cost on public health and has been a focus of policy attention for more than a decade in India. While Government of India, in its flagship "Swachh Bharat Abhiyaan (Clean Indian Campaign)" has planned and executed large budget outlays with an aim to make India free from Open Defecation[1], the progress and take-up of improved sanitation is slow[2], leaving improved sanitation in India an open question for researchers.

Understanding and reducing OD (thereby improving sanitation) has been a central question in economics and public health literature. While improvements in sanitation has been linked to gains in public health and human capital in developed world[3], poor sanitation in developing countries like India has been associated with significant negative health externalities[4]. Due to the negative health externalities of OD, the economic literature has largely looked at this issue through the lens of a "public good problem". Apart from the health externalities there are other negative consequences connected to OD such as gender based harassment[5] which, unlike a public good problem, would likely have direct implications on the agent making a choice (e.g. adopt a toilet or defecate in open). This potential association between gender and sanitation behavior has received relatively less attention in the literature.

In this paper, I focus on incentives of the households to switch away from open defecation due to such negative consequences associated with it. Defecating in the open may have

[1] As per Ministry of Information and Broadcasting, Government of India has spent about 83 percent of their advertising expenditure on this flagship program (see http://164.100.47.190/loksabhaquestions/annex/8/AU2287.pdf).

[2] See Spears & Coffey (2018) for detailed analysis using NFHS-4.

[3] Bleakley (2007), Cutler & Miller (2005), and Watson (2006) provide evidence about the role of sanitation in achieving better health and human capital in US.

[4] Duflo *et al.* (2015), Geruso & Spears (2018), Spears & Lamba (2016) and Coffey *et al.* (2017) provide evidence connecting the practice of open defecation to acute malnutrition, child mortality, poor child cognition and anemia.

[5] Jadhav *et al.* (2016) and JAGORI & UN-Women (2010) finds evidence related to sexual harassment faced by women who defecate in open.

significant costs to women and girls because, as compared to men and boys, they need more privacy. These costs include loss of dignity, lack of privacy, and the possibility of harassment when they defecate, urinate or attend to menstrual hygiene in the open. This paper attempts to answer a specific question - Do households adopt better sanitation (or stop defecating in the open) due to the presence of adult female children?

A common challenge while comparing the households with and without the presence of female children is to adequately account for unobserved heterogeneities in preferences of decision makers, which affect both the gender composition of children and other observable outcomes (e.g. sanitation behavior). This is particularly important in context of India where son preference is widely prevalent. In presence of son biased preferences, a correlation develops overtime between fertility, gender composition of children and observed household characteristics. In this paper, I exploit a plausibly exogenous variation related to the gender of first born child in a household which has been widely used in the economic literature (see Bhalotra & Cochrane.C (2010), Rosenblaum (2013), Portner (2010), Barcellos *et al.* (2014) and Jayachandran & Kuziemko (2011)). There is evidence to support that first pregnancy in India has a biologically normal sex ratio and that sex selective abortions are costly and prevalent at higher birth orders.[6] Nevertheless, for this paper, the direction of potential bias resulting due to issues such as sex-selective abortions is less of a concern. Son biased preferences begets a desire to have certain number of boys which results in girls living in larger families. With lower per capita resources in a larger family[7], the impact if at all (in observed reduction in OD), is likely to be *downwards*, thereby providing a lower bound on estimates.

I use National Family Health Survey (NFHS), a large representative household level sur-

[6]Portner (2010), Bhalotra & Cochrane.C (2010) and Jha *et al.* (2011) show that first pregnancy in India exhibit a biologically normal male:female sex ratio of 1.04-1.07. Hesketh & Xing. (2006) report that sex-selective abortions are common at later birth orders. Working in the research theme similar to this paper, Kishore & Spears (2014) and Anukriti *et al.* (2018) use gender of first born child as an indicator of presence of male and female children in a household.

[7]Tests conducted in this paper for differences in household characteristics with first born child as a girl versus boy provide support for this.

vey conducted in 2015-16 in India as the main data source. Results suggest that households reduce open defecation by 7-14% in urban areas and 3-7% in rural areas, if the first born child is a girl. These estimates are economically meaningful.[8] To understand the potential mechanisms underlying this reduction in OD, I conduct further analysis to gauge the timing of reduction in OD, and examine the heterogeneities in this behavior across various socio-economic dimensions. Related to timing, the reduction in OD is found to be associated with first born female children who are in the age range where they are of/ have crossed *pubescence*, and not with younger first born children. In addition to that, I also find that reduction in OD picks up larger magnitude and becomes statistically significant when the first born girl is reaching a marriageable age. There are other factors that further increase the cost to harassment or its likelihood. These include regions with crime against women, weaker socio-economic indicators etc. Heterogeneity analysis suggests that households in states where crime against women is higher show larger reduction in OD when first born child is female. Similarly, states with slow economic progress and lower literacy rates, as a proxy of weak socio-economic indicators, show a higher reduction in OD. Overall, I find suggestive evidence that reduction in OD in connection to first born female child is higher when the *costs* from a potential harassment and its likelihood are higher.

In Indian context where there is a preference for son, the private benefits for daughters are often ignored. Sen (2003), Jeffery *et al.* (1989) show that India has a widespread preference for male child. Barcellos *et al.* (2014), Jayachandran & Kuziemko (2011) and Deaton (2003) show that households in India selectively under-invest in private benefits such as nutrition, education, postnatal time and attention for a girl child (e.g. breastfeeding). Just like these private benefits, access to toilet also acts as a private good for girls, as presence/absence of that is related to private benefits/costs to them. Recent economic literature has linked access to toilet to benefits like girl's education. Adukia (2017) shows that school latrine

[8] Geruso & Spears (2018) find that a 10 percent point reduction in open defecation around the neighborhood is associated with a decline in infant mortality of 6 per 1,000, or about 8 percent of the population mean infant mortality rate. Stopnitzky (2017) finds a increase of 21% in toilet ownership (from a base level of 29%) in response to the "no toilet - no bride" campaign in Haryana.

construction program in India increases school enrollment of pubescent-age girls and much more when there is access to sex-specific toilets. The absence of toilets, on the other hand, has been linked to unpleasant social environment. Linking sanitation practices and sexual harassment, (Jadhav *et al.* (2016)) provide evidence from Indian context that women who openly defecate are twice as likely to suffer from non-partner sexual violence compared to those who do not. (JAGORI & UN-Women (2010)) provide evidence from urban slums in New Delhi (where OD is higher) that 66% women living in these slums report a verbal abuse, 46% report a visual abuse and 10% report a sexual assault.

Economic literature has demonstrated that households in India do not invest in private benefits of a girl child (education, health etc.). If access to toilet provides private benefits to female child, why should we expect households to invest in improved sanitation for them? While education and health are private benefits, unpleasant social environment (i.e. potential of harassment) is more likely to bring the costs to household as a whole. Religious doctrines in India have long established that in a *patriarchal* society, women are the responsibility of men, they are to be dependent on men and are subject to ostracization in various events of not following the doctrines[9]. It suggests that households may provide a sanitation facility if the costs of a female child being harassed accrues to everyone in the household.

The empirical findings in this paper align with a model of household decision making where the household *internalizes* the potential cost of female harassment and given the other factors (economic condition, social costs etc.) driving these costs, invests in a *private* good (sanitation facility) for female children. In this framework, the *non-private* nature of potential *costs* acts as the motivation for household to internalize these costs. The absence of these *non-private* costs on the other hand, would result in the underinvestment in female

[9] *Manusmriti* is one of the oldest religious doctrines for *Hindus*, dictating the duties of a hindu. According to *Manusmriti* a hindu woman must not be independent and should be under the custody of their father, brother, husband or son depending on stages of life and age (*Manusmriti 5/151 and 9/3*). It also states that in case a woman tears the membrane [hymen] of her Vagina, she shall instantly have her head shaved or two fingers cut off and made to ride on Donkey (Manusmriti 8/369). *Quran* lays out the duties for followers of *Islam* and states that men are the protectors and maintainers of women and should spend for and support women in their means (Quran 4:34)

children as observed in the literature. First contribution of this paper is to the literature analyzing decision making at a household level with differential outcomes for male and female children in presence of strong gender biases.

Closest to this paper is Stopnitzky (2017), which finds an increase in toilet adoption due to a social campaign "no toilet-no bride" in Indian state of *Haryana*[10], thereby highlighting the importance of bargaining in marriage market in adoption of *private* good for female members. My paper makes a contribution which is distinct as I analyze a different household structure in which the incentives for reduction in OD comes from a different female member of the household. In addition to that, the reduction in OD analyzed in this paper comes organically from differences in gender composition across household and not as a response to an external intervention such as the campaign. Nevertheless, the findings in this paper can also be viewed as complementary to Stopnitzky (2017) since the incentives for reduction in OD in both these studies are coming from the potential costs to female members which, owing to their *non-private* nature, have to be internalized by the household as a whole.

Second, this paper provides a unique new first stage result related to the gender of the first born child and its association with sanitation behavior of a household. This first stage result has a potential to contribute to further economic research like understanding peer-effects of sanitation behavior on neighbors or in social network of a household. Guiteras *et al.* (2015) show that rural households in Bangladesh adopt sanitation when they are in proximity to a household who received subsidy for building toilets, thereby highlighting the importance of social connections in sanitation behavior. However, economic literature still has room for studies analyzing spillovers and externalities in context of improved sanitation adoption. Identifying the reasons for lack of research in understanding the externalities of open defecation, Geruso & Spears (2018) highlight the lack of a strong first-stage in take-up of better sanitation in experimental studies as one of the key reasons. Difficulty in generating a large enough first-stage effect has been demonstrated by three recent field experiments in

[10]This campaign urged the families looking for the marital match for their female children to demand toilet adoption from potential suiters.

rural India (Hammer & Spears (2016); Clasen *et al.* (2014), Pl *et al.* (2015)). The result from this paper provides evidence of previously not known natural incentives for taking up better sanitation, and when coupled with potential spillovers in sanitation practices, it can help future experimental studies in generating larger first-stage effects.

Finally, empirical findings in this paper are also relevant for policy makers. It contributes to the understanding of who adopts toilet and who does not, which is likely to improve the targeting of resources towards improving sanitation.

Rest of the paper is organized as follows. Section 2 provides the background of the research problem in this paper. Section 3 organizes a conceptual framework looking at a household level decision making regarding sanitation practices and deriving testable implications. Section 4 describes the data and empirical strategy. Section 5 describes the Identification. Main results are discussed in Section 6. Section 7 presents falsification tests followed by robustness tests in Section 8. Section 9 explores heterogeneity of results and Section 10 concludes.

2 Background

A billion people worldwide defecate in the open and Indian alone accounts for 60% of them (UNICEF & WHO (2014)). These stark numbers along with well-known health consequences of open defecation makes it a high-priority policy concern. Apart from the effects on public health, defecating in open has negative externalities like the potential for harassment of girls and women who go out to defecate in the open. Women value toilets to a greater extent than men because they suffer disproportionately from male harassment when they defecate, urinate, or attend to menstrual hygiene in open (Stopnitzky (2017)).

Whether or not households practice open defecation also depends on critical factors like the region a household lives in (for example urban vs. rural) and how rich or poor they are. Households living in rural areas have access to large fields, open space, and more privacy

while defecating in open which contributes significantly to high open defecation rates in these areas[11]. Defecating in the open in fields far from their home does not pose any immediate cost of pollution and impurity near houses and hence does not result in a higher social cost. These factors may induce low enough social costs that the monetary cost of adopting a toilet is higher, and we can expect even richer households in rural areas to be practicing open defecation. In rural areas, where the community is more integrated and privacy concerns are lower for women, harassments while defecating in the open may also be a lower probability event. On the contrary, in urban areas, households live in constrained spaces which provide less access to open space and less privacy while defecating in the open. Since urban areas have better and modern infrastructure, the social costs (shame) of polluting it and spreading impurity (as perceived by residents) is also higher. These costs are higher in high-income areas and we can expect richer households to be adopting a toilet, irrespective of gender composition of their children. Households in poorer pockets of urban areas, have higher costs of defecating in the open due to significant space constraints, lack of privacy and a higher probability of harassment. In spite of being poor, we may expect the households with adult female children to be incentivised to invest in a toilet and not defecate in the open.

3 Conceptual Framework

3.1 Setup

This sub-sections sets-up a framework in which a household decided whether or not to adopt a toilet (or stop defecating in the open). This decision depends on the presence of a female child, the income of a household and the social costs of open defecation they face.

In a short term one-period framework, a household i is maximizing their utility over a

[11] Qualitative work (Coffey & Spears (2017)) to understand open defecation practices has documented some interesting features of sanitation practices in rural areas, like - a) Households in rural areas may also prefer to defecate in open since they like open environment and not being constrained by walls of a toilet, b) They have been doing so for generations and a behavioral change is harder for them, c) Women in rural areas also prefer to defecate in open since it gives them a chance to go out of house and meet friends.

bundle of consumption good X and adopting a toilet t (or a decision to not Openly Defecate), as follows:

$$\max_{\{X_i,t_i\}} V_i(X_i,t_i) = U_i(X_i,(\varphi_i|t_i)) - \omega(C_i(F_i)|t_i) \tag{1}$$

$$s.t. \quad X_i + Pt_i \leq I_i$$

Where, X_i is a bundle of all consumption goods, the price for which is normalized to 1. $t_i \in \{0,1\}$ is a decision to adopt a toilet, the price for which is P. I_i is the aggregate set of resources a household has.

The first term $U_i(.)$ represents the utility from consuming X and is increasing in this argument; dis-utility from φ_i which represents the social cost of open defecation and is decreasing in it. The factor φ_i is the social cost factor related to the region a household lives in (such as the cost of shame, polluting the neighborhood etc., when defecating in the open)[12]. φ_i is conditional on having a toilet in the household and is discussed later.

The second term, $C_i(F_i)$, represents the cost to the household of female children getting harassed (such as shame) when defecating in the open. $C_i(F_i) > 0$ if the indicator for presence of female children in the household, $F_i = 1$, and $C_i(F_i) = 0$ if $F_i = 0$. The factor ω represents the probability of harassment faced by female children while defecating in open[13].

Finally, in a short-term one period framework, I assume φ_i and $C_i(F_i)$ to be exogenous to the decision making problem. In other words, in a short-term one period frame, a household is assumed to be not choosing to relocate to new areas to change φ_i (which is a large investment) and neither they are able to change the gender composition of their children and hence $C_i(F_i)$.

[12] φ_i can range from very small values in sparsely populated remote rural areas to very high value in a posh urban residential society. For simplicity in mathematical proofs, I assume it to be continuous between arbitrarily defined lower and upper bounds and non-negative.

[13] The probability of harassment links more closely to the crime rate in the region a household lives in. For mathematical simplicity, I assume it is constant. Even if we let it vary by region (low for rural areas, higher for urban areas), the direction of results would not change but the mathematical arguments become cumbersome

Toilet adoption, $t_i \in \{0, 1\}$ relates to both the presence/absence of social costs φ_i, and the costs of female children facing harassment $C_i(F_i)$. Households do not face the social costs or the cost of harassment if they adopt a toilet ($t_i = 1$) i.e. $(\varphi_i|t_i = 1) = 0$ and $(C_i(F_i)|t_i = 1) = 0$. Both costs enter the utility as positive if $t_i = 0$.

I assume the term $U_i(.)$ in equation (1) to be additively separable in its arguments:

$$U_i(X_i, (\varphi_i|t_i)) = u(X_i) - (\varphi_i|t_i)$$

and accordingly, the value function becomes,

$$V_i(X_i, t_i) = u(X_i) - (\varphi_i|t_i) - \omega(C_i(F_i)|t_i)$$

In addition, to represent the role of toilet adoption decision, φ_i and $C_i(F_i)$ enter V as follows:

$$V_i(X_i, t_i) = u(X_i) - \varphi_i(1 - t_i) - \omega C_i(F_i)(1 - t_i) \tag{2}$$

Substituting the binding version of budget constraint and toilet adoption decision, $t_i \in \{0, 1\}$, a household chooses from following optimal $V_i^*(X_i^*, t_i^*)$:

$$V_i^* = \begin{cases} V_i(I_i, 0) = \quad u(I_i) - \varphi_i - \omega C_i(F_i, 0) & if \quad t_i = 0, \\ V_i(I_i - P, 1) = \quad u(I_i - P) & if \quad t_i = 1 \end{cases} \tag{3}$$

3.2 Propositions

Given the optimal choice of value based on toilet adoption decision, there are following proposition (Appendix A provides relevant proofs):

Proposition 1. *Households with low enough social costs od OD never adopt a toilet.*

This proposition states that there exists a level of social cost $\bar{\varphi}$, such that $\forall \quad \varphi_i < \bar{\varphi}, \quad u(I_i) - \varphi_i - \omega C_i(F_i) > u(I_i - P) \implies V_i(I_i, 0) > V_i(I_i - P, 1)$.

Proposition 2. *Households with high enough social costs and high enough income always adopt toilet.*

This proposition states that for $\forall \;\; \varphi_i \geq \bar{\varphi}$, there exists a high enough income level $\bar{I}$, such that, $\forall \;\; I_i \geq \bar{I}, \;\; u(I_i - P) \geq u(I_i) - \varphi_i - \omega C_i(F_i, 0) \implies V_i(I_i - P, 1) > V_i(I_i, 0)$.

Proposition 3. *In presence of high enough social costs and low enough income, household only adopt toilet if they have female children in the household.*

This proposition states that for $\forall \;\; \varphi_i \geq \bar{\varphi}$, there exists a low enough income level $\bar{I}$, such that, $\forall \;\; I_i \leq \bar{I}$:

$$u(I_i) - \varphi_i - \omega C_i(F_i) \geq u(I_i - P) \implies V_i(I_i - P, 1) \leq V_i(I_i, 0) \quad \text{if } F_i = 0, \text{ i.e. } C_i(.) = 0$$

$$u(I_i) - \varphi_i - \omega C_i(F_i) < u(I_i - P) \implies V_i(I_i - P, 1) > V_i(I_i, 0) \quad \text{if } F_i = 1, \text{ i.e. } C_i(.) > 0$$

Proposition 4. *Households with very tight budget constraint never adopt a toilet.*

This proposition states that $\forall \;\; \varphi_i$, there exists a very low income level $\tilde{I}$, such that, $\forall \;\; I_i \leq \tilde{I} < \bar{I}, \;\; u(I_i) - \varphi_i - \omega C_i(F_i) > u(I_i - P) \implies V_i(I_i, 0) > V_i(I_i - P, 1)$.

3.3 Testable Predictions

The proposition and the cases put forth in the previous section provides us cases where a representative household may or may not adopt a toilet depending on aggregate resources, the social costs they face and the presence of female children they have. These theoretical cases give us empirically testable hypothesis, that, given the *treatment* (presence of or some female children) status of a household, under what conditions they are likely to be incentivised to adopt better.

Taking into account the Indian context (as discussed in Section I), we can relate the social costs to sanitation choices of households. A large proportion of households in rural areas are likely to be living in areas where there are low regional costs to open defecation (low φ) and going by proposition 1 in the previous sub-section, irrespective of income and

gender composition of kids; they are likely to demonstrate lower toilet adoption. Going by proposition 2, richer households in urban areas face high enough social costs of open defecation (high φ) that, irrespective of the gender composition of children, they are likely to adopt a toilet. Households living in poor urban regions have a higher cost of defecating in the open but, at the same time, they are poorer. Their marginal utility gain from not spending in adopting a toilet is high enough, such that, a) In the absence of an adult female child, the utility gain outweighs the social cost, but b) In the presence of an adult female child, the total costs (social + potential cost of female harassment) outweighs the utility gain. Going by proposition 3, these households will only adopt a toilet if they have adult female children in a household. In that sense, they are *compliers*. However, some of these poor households will be soo poor that the budget constraint remains tight for them even when they have female children and going by proposition 4, they will not adopt toilet.

4 Data and Empirical Strategy

4.1 Data

The main dataset used in the analysis is National Family Health Survey (NFHS) of India, conducted in 2015-16. The NFHS (India's version of Demographic and Health Survey) is a large, nationally representative survey and is regarded as a very high-quality demographic survey. The respondents are women aged 15-49 and report birth histories and other information about their children. This survey also includes information on household and member's characteristics, assets, infrastructure and other health related reports. The main variables I use in the analysis come from the birth records of NFHS. These include birth order, gender, date of birth, whether the child is alive, and whether or not s/he continues to live in the household for each of the child ever born to the surveyed women. Apart from these, I use characteristics of the women surveyed, of the head of the household, the residence (rural or urban) and indicators of wealth in a household (data on categories of assets). Following the

recent studies in the area, I use the survey question where a household reports: *"What kind of toilet facility do members of your household usually use?"* to construct the main outcome of interest. I create an indicator OD equals to 1 if a household reports: "*Having no toilet facility, going to field/bush to relieve themselves*" and 0 otherwise. Table 1 provides summary statistics related to few key variables.

As discussed in previous sections, household wealth could be a key factor related to adoption of better sanitation facilities. Although, NFHS does not record income or consumption of surveyed households, it records the assets a household owns. Recent empirical studies use these asset ownerships as a proxy of household wealth (Geruso & Spears (2018)). I create an 'Asset Index' as a measure of the wealth of a household by summing over the presence of various assets in household and creating a standard normal index of it.

For main analysis, I place few restrictions on the data to be used. I use households where, 1) Either the mother or father of the first born child are household head, and 2) first born child is alive and is living in parents house. Another key factor which could relate to the adoption of sanitation facilities is the age of oldest girl child. As discussed before, households might be incentivized to adopt a toilet when the eldest girl child is entering (or near) puberty. A general age range of attaining Puberty in Girls is about 10-14 years in India (Khadgawat *et al.* (2016)). Other similar studies also document early puberty starting as early as eight years of age. Hence, in addition to the restrictions mentioned above, I apply a minimum age cutoff for first born to be eight years to the main analysis sample. It includes the usual starting range of 10 years for the onset of puberty and also two years before that to account for the earlier onset of puberty and/or an earlier recognition of a need for privacy by parents. This is refereed to as the *main analysis sample* in rest of the paper.

4.2 Empirical Strategy

For all results reported in section 6 to 9, I run a reduced form linear probability model as follows:

$$OD_{ir} = \beta FB_{ir}^{Girl} + Y_{ir} + \delta_{ir} + \epsilon_{ir} \tag{4}$$

where, i indicates the household and r is the region they lives in (rural or urban). Outcome is the indicator variable OD. FB^{Girl} is indicator for first born gender being female. Y is the asset index. δ is the vector of household level controls. Full list of control are - Asset index as a quadratic polynomial; total fertility (living children only) as a quadratic polynomial; indicators for age of first born child as quadratic polynomial; indicators for sources of water; indicators for construction material of floor, roof and wall; an indicator for households reporting Islam as their religion; indicators of caste; indicators of current age of the mother; indicators for years of education of mother; indicator for mother being pregnant and finally an indicator for nuclear households. District fixed effects are also used in the richest specification. Standard errors are clustered at the level of primary sampling unit (PSU)[14]. In further analysis, main results are also tested for robustness to PSU fixed effects and various interactions of district, caste and year of birth fixed effects.

Gender of first born has a direct bearing on average family size and hence per capital resources. In order to achieve desired number of sons, households with first born child as girl end up with higher fertility on average, and are likely to be poorer compared to households with first born boy. Empirical analysis in next section shows that households with a first born girl fare lower in asset index as compared to their counterparts. While gender of first child being female may create incentives to adopt a toilet when the child reaches puberty, the tighter resource constraint will push the incentives in the opposite direction. In this case, 'asset index' (as a proxy for wealth) becomes an indispensable control. Hence, the baseline specification in all analysis includes a control for asset index.

[14]The NFHS is a two-stage random sample, first sampling Primary Sampling Unit (PSU) and then households within sampled PSUs. PSU is that sense is a sampling cluster. It is usually city blocks in urban areas and villages in rural areas.

5 Identification

An ideal independent variable to address the research question in this paper is the presence/absence of female children in the household, which is not exogenous. An ideal (but hypothetical) comparison would be between households with and without an elder female child in which the presence of that child is randomly assigned. To get close to this ideal comparison, various studies in Indian context use gender of first child as a plausible random assignment (see Barcellos *et al.* (2014), Kishore & Spears (2014)). I use the same identification strategy in the main analysis of this paper. Gender of first born child is considered random in many economic studies (see Rosenblaum (2013), Bhalotra & Cochrane.C (2010)). A problem central to these studies is that, due to son-biased preferences, households in India practice fertility stopping rules. As a result, a correlation develops over time between household level outcomes and the gender of the first child. To get around this problem, these studies restrict the samples to households with a very young first born child.

In this paper, the sample of interest are the households in which the first born child is near or has reached puberty, hence the solution to restrict sample does not apply here. Households in which the first-born child is a girl are more likely to have a higher fertility rate to achieve the desired number of sons, and hence larger family on an average. Larger families have lower per capita resources which make them poorer on average, and hence more likely to practice open defecation (or not invest in a toilet). The bias induced due to the gender of the first child if at all, will only induce a downward push on any association I find between gender of the first child and the sanitation practices. In the empirical analysis below, using mean difference tests, I first check if household characteristics and other relevant outcomes are statistically similar across households with young first born children. Next, to check if relevant outcomes change in a direction which could potentially induce upward bias on the results, I run similar mean difference tests on the entire *main analysis sample*.

Finally, it remains to see if the gender of the first born child affects the presence of female children in my data. Figure 1 uses the Main Analysis Sample and plots the average number

of female kids vs. the average total number of kids for all the households in the sample. The association shown in the plot is separated by the gender of the first born child. As observed, households with the first born child as a girl have, on average, a higher number of female children compared to households with the first born child as a boy.

6 Main Results

6.1 How do households with first born girl compare to households with first born boy

Before the main analysis I run mean difference checks on various household characteristics and relevant outcomes. I use equation (3) with sanitation behavior of a household and other household characteristics as outcome variable and estimating it without controls.

The first set of mean difference check looks at the difference in relevant outcomes of households in main analysis sample with first born child of age 5 years or less. Table 2 reports the results, separately by rural and urban regions. Reassuringly, the main outcome of interest, open defecation is not statistically different in both regions across the comparison groups. Households with better infrastructure such as piped water, cement/concrete floor, walls, etc. may find it easier to adopt a toilet. As observed, households with first child as girl do not differ significantly from those with first child a boy in any of these categories. Households with the first born girl are also not observed to be different in education of mother. Households show evidence of gender-biased fertility stopping rules, as the mothers of the first-born girl are more likely to be pregnant at the time of survey as compared to mothers of first-born boys.

The second set of mean difference checks uses the *main analysis sample* (with the age restriction for first born child being 8 years or above). Table 3 reports these results. Looking at asset index, households with first born girl child are significantly poorer than households with the first born boy. As expected, first born girls live in larger households and with

more siblings. Households with the first born girl are also not observed to be having better construction (material for floor, walls, etc.) or more educated mothers.

Overall, these mean difference checks help support the identification strategy that there are no significant differences in the variables related to Open Defecation which may put an upward bias on main outcome of interest.

6.2 Do households reduce open defecation due to presence of female children?

Using equation (3), I analyze the main hypothesis that households might reduce open defecation (adopt a better sanitation facility) if they have a girl child; the presence of whom is identified by a the gender of first born child. Results are reported starting from the baseline version of equation (3) with only asset index as control and adding subsequent controls to arrive at the preferred specification with all controls.

Column 1-3 of Table 4 reports the results for urban areas. Starting with column 1, households with first born girl are observed to be reducing open defecation. This association between first born gender and open defecation retains its strength and statistical relevance with the full set of household level controls and district fixed effects. The preferred specification in column 3 shows a statistically significant (at the level of 1 %) reduction in open defecation of about 6.7% of the mean. Column 4-6 show analogous results for rural areas. Column 6 shows a reduction in open defecation of about 2.5%. This association retains its strength and significance with all controls and to inclusion of district fixed effects.

This analysis provides support for testable predictions in section 3. While both urban and rural regions show reduction in open defecation in association with the gender of first born child, it shows up stronger for urban as compared to rural areas. This indicates that households in urban areas potentially face higher costs of open defecations when they have girl child in households.

6.3 Who drives reduction in OD in presence of female children?

As observed in previous section, the gender of first born child relates to the sanitation behavior of a household. It gives empirical support to the research question and testable predictions from Section 3. As argued in section 3, this could be a result of interaction of both social and economic costs of reducing open defecation. It is conceivable that social and economic costs go in opposite directions. Richer households live in areas where they face higher social costs out of shame of defecating in open while poorer households live in areas where social costs are low because a majority of neighbors practice open defecation. In this set of analysis, I aim at restricting the potential economic costs and arrive at sub-samples who are driving the results observed in previous section.

To keep the economic costs constant, I divide the households into *deciles* of asset index, separately for rural and urban areas. Using the preferred specification of equation (3), Panel A and B of Table 5 report the coefficient of interest for urban and rural regions, respectively. Column 1-10 include households in *deciles* of asset index, going from poorest to richest households in both regions.

Looking first at the results for urban areas in Panel A, mean OD rate suggests that open defecation rate is falling rapidly as households get richer. Open defecation rate is high and ranges from 47% to 25% for the two poorest groups of asset index (Column 1 & 2). It starts to fall as households get richer and reduces to zero for richest households. Poorer households also have higher OD rates are most likely to be the ones to reduce OD if they have female children at home. Poorer households with higher OD rates signifies they are living in worser conditions of the urban area and might be facing the higher crime rate and anti-social elements. Also given a higher OD rate, they have much larger room for improvement.

Moving to rural areas in Panel B, it shows that open defecation rates are not only higher on average than urban areas, but they stay higher for even the richer groups of households. For example, while a set of *median* households (ranked by by asset index) in urban areas had OD rate of lower than 8%, its counterpart in rural areas has almost a 60% OD rate. In

contrast to the case in urban areas, this provides empirical evidence for lower social costs of open defecation in rural households where space is not constrained. As discussed in the conceptual framework, rural households may have low costs of defecating in the open and may have weaker incentives to switch from OD.

Looking at reduction in OD for urban areas (Panel A), households in poorest decile do not show a reduction even if their first born child is girl. This provides evidence of a very tight budget constraint outweighing the cost of girl child defecating in open, and is consistent with proposition 4 of the conceptual framework. As we move to richer households and budget constraint starts to relax, the reduction in OD starts to show up for households with first born girl child. This reduction is statistically significant upto 4th decile group, where the open defecation rate is above 10% and there is room to improve[15]. In terms of magnitude, this amounts to an average of 14% reduction in open defecation in households with *below median wealth* (except the poorest decile) in urban areas. Looking at rural areas in panel B, while the households with *below median wealth* show signs of reduction in OD, the magnitude is very small and statistically not different from zero. Richer households with first born girl show significant reduction in OD. This reduction, however, is small in magnitude as compared to the urban areas. It amounts to an average of 7% reduction in rural households in top half of the income distribution.

These results also contrasts well with potential costs from a threat of harassment of girl child. Table A-1, using Indian Human Development Survey (IHDS - 2011/12) shows the percentage of households reporting harassment of girls being a frequent event in their neighborhood[16]. These percentages are broken down by rural and urban areas and by eco-

[15] 8th decile group also shows a significant reduction. Given that the average open defecation rate for this group is just 1% and that this particular comparison will not satisfy multiple hypothesis correction, it cannot be statistically distinguished from a *false-positive*.

[16] IHDS 2011-12 asks a question where household reports about frequency of events where girls are harassed in their neighborhood. I code the response categories *sometimes* and *often* as 1 with *rarely/never* as 0. Column 2 and 3 are derived using the asset index which is the standardized value of sum of assets reported in the survey. Self-reports of crime are less likely to suffer from reporting bias unlike other sources such NCRB (National Crime Record Bureau), which provide information on crimes reported to police stations. In addition, it is easier to divide the IHDS self-reports in rural versus urban regions, which is much harder and inaccurate using NCRB (since the reporting is at police station level with no clear division of rural

nomic status of households. As shown, the urban areas report 34% higher harassment as compared to rural areas, which provides empirical support for potential costs due to harassment being higher in urban as compared to rural areas. The difference remains same when comparing urban poor (below median wealth) to rural rich (above median wealth). This helps explain the results observed so far; poorer urban households facing higher costs from potential harassment of girl child are incentivised to switch away from OD even in tighter budget constraints while the budget constraint has to relax much further for rural areas to change their sanitation behavior.

Overall, these results indicate that interaction of social costs and resource constraint is a key determinant in change in household's sanitation behavior in response to the gender composition of their children. Social costs of OD are potentially higher in urban areas and as a result, small improvements in economic status induces a household to switch from OD if they have first born girl (and thereby are exposed to a potential harassment cost). On the other hand, in rural areas, the improvement in economic status have to be much higher for a household to switch away from OD in presence of girl child.

Finally, since the analysis in this section makes use of multiple groups within the main analysis sample, they must be subjected to multiple hypothesis correction to determine which groups demonstrate a statistically valid reduction in OD. I subject the results to two correction methods - Bonferroni correction (see Dunn (1961)) and Benjamini-Hochberg correction (see Benjamini & Hockberg (1995)). The 2nd asset decile group in urban areas and 7th to 10th decile groups in rural areas pass these two tests[17]. They are referred to as *complier* groups when used in further analysis below.

6.4 What drives the need of a toilet?

Empirical evidence shown above indicates that households reduce OD in presence of female children, which is likely due to the costs associated with a possibility that they face harass-

versus urban areas.)

[17] A false discovery rate of 0.05 is used for Benjamini-Hochberg correction.

ment when going outdoors to relieve themselves. With strict budget constraints (as in urban compliers), it is likely that a toilet would be adopted when a potential harassment is likely to be the most costly. It is very hard to measure the cost of harassment and hence, in turn hard to establish when it will be highest. However, there are certain times when a household would like to protect the daughters from any potential harassment the most. One such event is when the daughter is of marriageable age.

Figure A-1 and A-2 show the proportion of first born children getting married, by age groups ranging from 13-22 years for urban and rural areas respectively. Proportion of married children are separated by the gender of the first born child. As observed, age 17 is associated with first born girls marrying in urban areas. The corresponding age in rural areas is 16 years.

Figure 2 and 3 plots the coefficients from the preferred specification on the households belonging to urban and rural complier groups. Each coefficient represents an age group ranging from 8 to 23 years (on x-axis). As observed in the Figure 2, gender of first born is not associated with reduction in OD until they are in age group of 14-17 years in urban areas. The reduction in OD in urban compliers shows up most strongly (and statistically significant) when the first born girl in close to the marriageable age. Figure 3 is for rural compliers in similar age groups. Unlike urban compliers, rural counterparts demonstrate reduction in OD in each first born age group and not concentrated around the marriageable age. This exercise indicates that urban compliers are under tight budget constraints and are only able to switch away from OD when it is likely to cost the most. In rural compliers, the budget constraints are relatively relaxed and the reduction in OD is spread across all age groups of first born child.

7 Falsification

1. Do households reduce OD in presence of younger first born girls?

As discussed in Section 3 and demonstrated in the conceptual framework, a central notion is that households will be incentivised to invest in better sanitation when the expected costs of female children practicing OD is higher. If the reduction in open defecation also shows up when the female kids in the household are below the age of puberty, it may indicate that the results are potentially driven by causal factors other than what I have considered so far. To test this, I run a falsification test using preferred specification of equation (3). While the data used for falsification analysis is same as the main analysis Sample, it differs in the age cutoff. Only the households with the first born child less than 8 years of age are included.

Table 6 reports the results on main analysis sample and table 7 on complier sample. The association between first born gender and OD is not in the same direction as the main results for most of the specifications for both urban and rural areas. Wherever it turns negative and in same direction, it is small and statistically insignificant. The falsification exercise suggests that association between first born female child and reduction in open defecation is valid only for elder female children and provides support for testable predictions from conceptual framework.

2. Do households reducing OD also increase other asset ownership?

As shown in table 2 and 3, households with first born girl child are relatively poorer than households with first born boy. This association between first born gender and economic status is likely driven by higher fertility and larger family size which comes with first born being girl. While the households with first born girl are poorer on average, they have to cater to a large family and will differ in asset composition as compared to households with first born boy. Table A-5 explores the asset composition of both urban and rural complier samples. Second column in both panels show the difference in asset ownership of households with first born girl versus those with first born boys. Urban compliers show higher ownership of few assets like mattress, pressure cookers and sewing machines. Rural compliers also show higher ownership of these assets along with very small differences in table and chair ownership. Most of these assets are not luxury items and the difference is likely driven by larger family size

in households with first born girl. On the other hand, household with first born girl report less ownership of more luxurious aseets like internet, computer, air conditioners, washing machine, motor vehicles etc. This difference in asset ownership suggests that households with first born girl have relatively inferior quality of asset ownership, and it is unlikely that a reduction in OD is driven by any omitted variables correlated with economic status and gender of first born child.

8 Robustness

1. Robustness to Controls

Administrative blocks (e.g. districts) may have differences in sanitation preferences, policies and priorities. These policies and priorities may also be focused on households from specific socio-economic background (such as caste). Also, these differential policies may appeal differently to households with different age composition of children (e.g. households with teenagers may respond more to sanitation policies). I subject the main results to various variables attempting to control for the differential sanitation preferences, policies and priorities. Table A-2 takes the results from table 4 and start by including fixed effects at a more granular level than district; at the level of PSU in column 1 and 4. Column 2,3,5 and 6 subject the main results to various interactions of districts with caste and age of first born child. Both urban and rural results appear to be robust and stable to inclusion of these additional controls.

2. Are results driven by female members other than first born girl?

If there are additional female members in the household other than daughters, it cannot be ruled out that the results shown in the previous section are driven by the additional female member and not by the female child. Table 3 shows that in households with first born girl, there are significantly higher number of women who are 15-49 years of age. Possible explanation of this could be the presence of first born girl herself, additional female kids the

household had after first child (in order to achieve desired number of sons) or systematic presence/entry of additional female members in these households. The latter one, if present is of a deep concern for validity of the results. For e.g. entry of new daughter in law could be a major confounder, in regard to which Stopnitzky (2017) provides causal empirical evidence of reduction in OD.

I analyze the difference in gender composition of households with first born child as male versus a female for complier sample in urban and rural regions (without the age restriction). Table A-3 reports the result of estimating equation (3) on gender composition of complier households. Households with first born girl have about one additional female member (Column 1) in both rural and urban areas. Column 2 shows that the number of additional female members in Column 1 can be entirely explained by presence of additional *daughters only*. Column 3 estimates the regression with all female members other than daughter of household head as outcome and it indicates lower presence of other female members in households with first born daughters. Looking specifically at the presence of daughter in law (Column 4), households with first born girl has lower number of daughters in law (potentially due to delay in marriage of son as son is younger). This exercise indicates that the results discussed in previous section are not driven by any other female member in household apart from additional daughter(s).

3. Is mother's education playing a role?

More educated mothers may understand the need for privacy and hygiene for their daughters better and may also have stronger bargaining power. If the association between gender of first born and reduction in OD is observed only for educated mothers, it might enter as a serious endogeneity concern. Table A-4 reports the results from interaction of first born gender with a variable for years of education for mother. Results are reported for both the main analysis sample and complier sample for urban and rural areas. As shown, while the coefficient on first born gender dummy stays negative and significant (as expected), the coefficient on interaction of gender and mother's years of education are in opposite direction,

small and insignificant. As shown by the coefficients for mother's years of education, while educated mothers live in households practicing less OD, they do not seem to influence the reduction in OD due to the first child being girl.

9 Heterogeneity

1. Crime Against Women and Sanitation Behavior

Hypothesis in section 3 and subsequent results establish that households respond to private demand of sanitation when the costs of female children defecating in open are likely to be higher. It is conceivable that households living in areas where there are higher crime against women would have higher incentives to reduce OD as their girl child reaches marriageable age as compared to their counterparts in lower crime regions. This is more likely to be the case for households for whom the budget constraint is tight and who will only invest in better sanitation when the costs of not doing so are higher. I look at this possibility using the Indian Human Development Survey (IHDS-2011/12) which asks a questions related to the frequency of events related to harassment of girls in the neighborhood[18]. I code the response of this question as 1 if reported *sometimes* or *often* and as 0 if reported *rarely/never*. Using this variable, I rank states and UTs in India from highest to lowest rate of harrasment against girls, separately for urban and rural areas. Complier sample for both regions is then divided into *High Crime* (top 10 ranking states) versus *Low Crime* (all other states).

Table 8 reports the result of estimating equation 3 on complier sample divided by high versus low crime. The open defecation rate is similar in two groups of high versus low crime for both regions, indicating that both groups have equal room to improve. As observed in urban regions, the entire reduction in OD is driven by states with high rates of harassment. While the remaining states show small reduction in OD, it is statistically insignificant. In

[18]Since this is self-reported, it is less likely to be affected by reporting bias as opposed to other sources like crimes reported to police (as recorded by National Crime Record Bureau). In addition, unlike NCRB and official reports from police station level, it is easier to generate separate crime rates for rural versus urban regions using self-reports in representative surveys like IHDS.

rural areas, both high and low groups show reduction in OD but high crime region show larger reduction as opposed to low crime regions. These results suggest that the possibility or extent of exposure of girl child to harassment is a key determinant in switching away from OD.

2. Sanitation behavior in High vs Low performing regions

Regions which lag behind on socio-economic indicators are poorer, less developed and usually have more patriarchal societal structure. Given male dominant household structure, the incidences and costs of girl child being harassed are likely to be higher[19], resulting in higher potential costs from harassment. In addition, these regions usually have weaker state machinery and institutions. The citizens have fewer avenues to seek protection against violations of law and are largely self-reliant. It is conceivable that these regions have stronger incentives to reduce OD in presence of girl child. I use two such indicators to classify states as weaker versus stronger in socio-economic indicators - first, BIMARU versus non-BIMARU states in India and second, literacy rate. Indian states of Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh are typically tagged as economically low performing states and are termed *BIMARU* states[20]. Population Census of India (2011) is used to rank states/UTs on literacy rate.

Table 9 shows the result of estimating the equation (3) on complier samples divided into BIMARU and Non-BIMARU states. Table 10 shows the analogous results for complier samples divided in group of 10 states with lowest literacy rates versus other states. Urban compliers in BIMARU and Non-BIMARU states have similar open defecation rates, hence similar room to improve. BIMARU states exhibit a much higher reduction in urban OD rate as compared to Non-BIMARU. Similarily, the reduction in urban OD rate seems to be driven by states with lower literacy as compared to high literacy. Rural compliers on the other hand show similar reduction in OD in BIMARU vs Non-BIMARU and Low versus

[19]Reporting and booking the male perpetrators is likely to be hard

[20]See https://www.financialexpress.com/india-news/bimaru-redux-niti-aayog-ceo-says-bihar-madhya-pradesh-uttar-pradesh-rajasthan-keeping-india-backward/1143709/

High literacy divisions. This heterogeneity analysis supports the presence of tight budget constraints in urban compliers where they would switch the behavior only when the costs of not doing so are higher.

10 Conclusion

This paper investigates a potential association of gender composition of children in households with their sanitation behavior. Using the gender of first born child as a plausible exogenous shock to the gender composition of children in a household, this paper finds that poor urban households and richer rural households reduce open defecation in presence of female children. This behavior is explained potentially by an interaction of costs associated with female children defecating in open thereby getting exposed to harassment, and budget constraints. After exploring various subsamples and heterogeneities, I find considerable support in favor of hypothesis related to interaction of costs mentioned before.

This paper contribute towards understanding the sanitation behavior and its drivers in India. Results provide a new insight into who adopts a toilet and who does not, which provides a new avenue for policymakers to target. In addition, these provide a new first stage association between gender composition of children and household sanitation behavior which has a potential to contribute to further economic research investigating social network and spillover effects on sanitation behavior.

References

Adukia, A. 2017. Sanitation and Education. *American Economic Journal: Applied Economics*, **9(2)**, 23–59.

Anukriti, S., Kwon, S., & Prakash, N. 2018. Household Savings and Marriage Payments: Evidence from Dowry in India. *Working papers 2018-09, University of Connecticut, Department of Economics*.

Barcellos, Silvia H., Carvalho, L., & Lleras-Muney, A. 2014. Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently? *American Economic Journal: Applied Economics*, **6**, 157–189.

Benjamini, Y, & Hockberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bhalotra, S., & Cochrane.C, T. 2010. Where Have All the Young Girls Gone? Identification of Sex Selection in India. *IZA Discussion Paper no. 5381, Institute for the Study of Labor, Bonn*.

Bleakley, H. 2007. Disease and Development: Evidence from Hookworm Eradication in the American South. *Quarterly Journal of Economics*, **122 (1)**, 73–117.

Clasen, T, Boisson, Sophie, Routray, Parimita, Torondel, Belen, Bell, Melissa, Cumming, Oliver, Ensink, Jeroen, & et al. 2014. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: A cluster-randomised trial. *Lancet Global Health*, **2(11)**, e645–53.

Coffey, D., & Spears, D. 2017. *Where India Goes: Abandoned Toilets, Stunted Development and Costs of Caste*. Harper Collins.

Coffey, D., Geruso, M., & Spears, D. 2017. Sanitation, Disease Externalities, and Anemia: Evidence From Nepal. *The Economic Journal*, **Accepted**.

Cutler, D., & Miller, G. 2005. The role of public health improvements in health advances: The twentieth-century United States. *Demography*, **42 (1)**, 1–22.

Deaton, A. 2003. Health, Inequality, and Economic Development. *Journal of Economic Literature*, **41, No.1**, 113–58.

Duflo, E., Greenstonea, M., Guiteras, R., & Clasen, T. 2015. Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation. *NBER WP 21521*.

Dunn, J. D. 1961. Multiple Comparions Among Means. *Journal of the American Statistical Association*, **56**, 52–64.

Geruso, M., & Spears, D. 2018. Neighborhood Sanitation and Infant Mortality. *American Economic Journal: Applied Economics*, **10**, 125–62.

Guiteras, R., Levinsohn, J., & Mobarak, AM. 2015. Sanitation subsidies. Encouraging sanitation investment in the developing world: a cluster-randomized trial. *Science*, **348(6237):903-6**.

Hammer, J., & Spears, D. 2016. Village sanitation and child health: Effects and external validity in a randomized field experiment in rural India. *Journal of Health Economics*, **48**, 135–48.

Hesketh, Therese, & Xing., Zhu Wei. 2006. Abnormal Sex Ratios in Human Populations: Causes and Consequences. *Proceedings of the National Academy of Sciences of the USA 103*, **36**, 13271–75.

Jadhav, A., Weitzman, A., & Smith-Greenway, E. 2016. Household sanitation facilities and women's risk of non-partner sexual violence in India. *BMC Public Health*.

JAGORI, & UN-Women. 2010. *Safe Cities Free of Violence Against Women and Girls Initiative: Baseline Survey in Delhi*. Report. Jagori and UN Women.

Jayachandran, S., & Kuziemko, I. 2011. Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India. *Quarterly Journal of Economics*, **126**, 1485–1538.

Jeffery, P., Jeffery, R., & Lyon, A. 1989. Labor Pains and Labour Power. *London:Zed*.

Jha, P., Kesler, M., Kumar, R., Ram, U., Aleksandowicz, L., Bassani, D., Chandra, S., & Banthia., J. 2011. Trends in Selective Abortions in India: Analysis of Nationally Representative Birth Histories from 1990 to 2005 and Census Data from 1991 to 2011. *Lancet*, **377**, 1921–28.

Khadgawat, R., Marwaha, RK., Mehan, N., Surana, V., Dabas, A., Sreenivas, V., Gaine, MA., & Gupta, N. 2016. Age of Onset of Puberty in Apparently Healthy School Girls from Northern India. *Indian Pediatrics*, **53 Issue 5**, 383–387.

Kishore, A., & Spears, D. 2014. Having a son promotes clean cooking fuel use in urban India: Women's status and son preference. *Economic Development and Cultural Change*, **62 Issue 4**, 673–699.

Pl, Gertler., Shah, Manisha, Alzua, Maria Laura, Cameron, Lisa, Martinez, Sebastian, & Patil., Sumeet. 2015. How Does Health Promotion Work? Evidence from the Dirty Business of Eliminating Open Defecation. *NBER WP 20997*.

Portner, C. C. 2010. Sex Selective Abortions, Fertility and Birth Spacing. *Working Paper no. 4, University of Washington.*

Rosenblaum, D. 2013. The Effect of Fertility Decisions on Excess Female Mortality in India. *Journal of Population Economics*, **26 No. 1**, 147–80.

Sen, A. 2003. Missing Women Revisited: Reduction in Female Mortality Has Been Counter balanced by Sex Selective Abortions. *British Medical Journal*, **327**, 1297–98.

Spears, D., & Coffey, D. 2018. Open defecation in rural India, 2015-2016: Levels and trends in the NFHS-4. *Working Paper*.

Spears, D., & Lamba, S. 2016. "Effects of Early-Life Exposure to Rural Sanitation on Childhood Cognitive Skills: Evidence from India's Total Sanitation Campaign. *Journal of Human Resources*, **51 (2)**, 298–327.

Stopnitzky, Y. 2017. No Toilet No Bride? Intrahousehold Bargaining in Male-Skewed Marriage Markets in India. *Journal of Development Economics*, **127**, 269–282.

UNICEF, & WHO. 2014. *Progress on Drinking Water and Sanitation 2014 Update*. Report. UNICEF.

Watson, T. 2006. Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations. *Journal of Public Economics*, **90 (8–9)**, 1537–60.

A: Mathematical Appendix

In addition to assumptions in Section III-A, I assume, $C(1)$ is given; the aggregate resources an household i, $I \in [I_L, I_H]$ where $I_L > 0$ and $I_H < \infty$.

Proposition 1.: $\forall\ I_i\ \exists\ \bar{\varphi}$, such that, for $\varphi_i < \bar{\varphi}$,

$$V_i(I_i, 0) > V_i(I_i - P, 1) \qquad \forall i \tag{5}$$

Proof: Consider equation (5),

$$\begin{aligned} & V_i(I_i, 0) > V_i(I_i - P, 1) \\ \implies & U(I_i, (\varphi_i | t_i = 0)) - \omega(C(F_i)|t_i = 0) > U(I_i - P) \\ \implies & u(I_i) - \varphi_i - \omega C(F_i) > U(I_i - P) \\ \implies & \hat{\varphi} \equiv U(I_i) - U(I_i - P) - \omega C(F_i) > \varphi_i \end{aligned} \tag{6}$$

In equation (6), we obtain the cutoff as a function of income and presence of children. We observe that, $\hat{\varphi}$ is decreasing in both, I_i[21] and F_i[22]. Hence, the lowest value of $\hat{\varphi}$ is the cutoff level below which all households do not adopt a toilet. Now, we obtain,

$$\bar{\varphi} \equiv \hat{\varphi} = U(I_H) - u(I_H - P) - \omega C(F_i) \tag{7}$$

■

Proposition 2: For all $\varphi_i > \bar{\varphi}$, $\exists\ \bar{I}$ such that for $I_i > \bar{I}$,

$$V_i(I_i - P, 1) > V_i(I_i, 0) \quad \forall i, \tag{8}$$

[21]The derivative of $\hat{\varphi}$ wrt I_i is $u'(I_i) - u'(I_i - P)$. Since $u(.)$ is concave in X, $I_i > I_i - P$ implies that $u'(I_i) < u'(I_i - P)$.

[22]The derivative of $\hat{\varphi}(I_i, F_i)$ wrt F_i is $-\omega C_F(F_i)$. Here, $C_F(F_i) > 0$ implies that $-\omega C_F(F_i) < 0$.

Proof: Equation (8) suggests that the household has a higher net utility from adopting a toilet:

$$U(I_i - P, (\varphi_i | t_i = 1)) > U(I_i, (\varphi_i | t_i = 0)) - \omega(C(F_i) | t_i = 0) \tag{9}$$

In this case, equation (9) should hold for all possible values of F_i. Since, the RHS is decreasing in F_i, if the inequality holds for $F_i = 0$, it also holds for $F_i = 1$. Therefore, the condition in equation (9) reduces to:

$$U(I_i - P, (\varphi_i | t_i = 1)) > U(I_i, (\varphi_i | t_i = 0)) \tag{10}$$

Note, that equation (10) does not depend on ω. From equation (10), we get,

$$U(I_i - P) > U(I_i) - \varphi_i \tag{11}$$

$$\implies \varphi_i > U(I_i) - U(I_i - P) \tag{12}$$

Note, that the LHS is constant for all I_i[23] and the RHS is decreasing in I_i[24]. Hence, corresponding to each level of $\varphi_i > \bar{\varphi}$, there exists an $\bar{I}$ such that for $I_i > \bar{I}$ households will always adopt a toilet. $\bar{I}$ is given by:

$$\varphi_i = u(\bar{I}) - u(\bar{I} - P) \tag{13}$$

■

[23]

[24]Derivative of the RHS wrt I_i is $u'(I_i) - u'(I_i - P)$. Since $u(.)$ is concave in X, $I_i > I_i - P$ implies that $u'(I_i) < u'(I_i - P)$

Proposition 3: For $\varphi_i > \bar{\varphi}$, and $I_i < \bar{I}$, such that $\forall I_i$,

$$U(I_i, (\varphi_i | t_i = 0)) > U(I_i - P, (\varphi_i | t_i = 1)), \qquad \text{if } F_i = 0 \tag{14}$$

$$U(I_i - P, (\varphi_i | t_i = 1)) > U(I_i, (\varphi_i | t_i = 0)) - \omega(C(F_i) | t_i = 0) \qquad \text{if } F_i = 1 \tag{15}$$

Conceptual Explanation: I will provide a conceptual explanation supporting the validity of proposition 3 above. Consider equation (14),

$$U(I_i, (\varphi_i | t_i = 0)) > U(I_i - P, (\varphi_i | t_i = 1)) \tag{16}$$

$$\implies U(I_i) - \varphi_i - \omega C(F_i) > U(I_i - P) \tag{17}$$

$$\implies U(I_i) - U(I_i - P) > \varphi_i \tag{18}$$

In this case, there are no girl child in household, I_i is sufficiently low, such that marginal utility out of income is higher and a gain in utility because of not adopting a toilet is higher than the social cost of open defecation. Hence, in this case a household will not adopt a toilet.

Similarily, consider equation (15),

$$U(I_i - P, (\varphi_i | t_i = 1)) > U(I_i, (\varphi_i | t_i = 0)) - \beta C(F_i) \tag{19}$$

$$\implies U(I_i) - U(I_i - P) < \varphi_i + \omega C(F_i) \tag{20}$$

In this case, there are adult girl child in household (hence, $C(.) > 0$), all other factors are same as equation (14). The gain in utility by not adopting a toilet in this case is outweighed by the social cost + cost associated with an adult female child in the household. Hence, in this case a household will adopt toilet only if $C(.) > 0$

■

Tables

Table 1: Summary Statistics: NFHS 2014-15

Outcome	Rural	Urban
Total population	72%	28%
Number of HH members	5.8	5.6
Number of Children	2.9	2.4
Mother's Age	36	35
Father's Age	47	46
First Born Gender: *Female*	43.4%	43.1%
First Born Age	6.8	4.3
Religion: *Hindu*	76.7%	73.4%
Religion:*Muslim*	10.03%	14.7%
Open Defecation	49%	12%

Notes: This table provides summary statistics of key variables related to the entire NFHS 2014-15 data.

Table 2: Mean Difference Check - First Born $\leq$ 5 years

	Urban			Rural		
Variable	Mean	Difference	Standard Error	Mean	Difference	Standard Error
Asset Index	0	-0.012	[.026]	0	-0.029	[.017]*
Open Defecation Rate	0.106	0.004	[.008]	0.485	0.005	[.008]
Water Source						
Piped into Dwelling	0.302	-0.02	[.011]*	0.092	-0.005	[.005]
Piped to yard/plot	0.185	-0.002	[.009]	0.096	0.001	[.005]
Public tab/standpipe	0.156	0.011	[.01]	0.137	0.001	[.006]
Tubewell or borehole	0.2	0.001	[.01]	0.475	-0.005	[.008]
Cooking Fuel						
LPG, Natural Gas	0.765	-0.001	[.01]	0.215	0	[.007]
Wood	0.13	0.007	[.009]	0.621	-0.006	[.008]
Floor Material						
Mud/Clay/Earth	0.094	-0.01	[.007]	0.437	-0.001	[.009]
Animal Dung	0.016	0.004	[.003]	0.094	0.001	[.005]
Cement	0.558	0.003	[.012]	0.323	0.008	[.008]
Roof Material						
Metal/GI	0.19	-0.006	[.011]	0.265	0.006	[.007]
RCC/RBC/Cement/Concrete	0.507	-0.005	[.013]	0.216	0	[.007]
Thatch/Palm Leaf	0.01	-0.005	[.003]*	0.073	0.001	[.004]
Burnt Brick	0.023	-0.004	[.004]	0.063	-0.002	[.004]
Wall Material						
Mud	0.045	-0.008	[.005]	0.213	0.009	[.007]
Cement/Concrete	0.478	-0.004	[.014]	0.214	0.005	[.007]
Burnt Brick	0.222	-0.009	[.011]	0.214	-0.002	[.007]
Household Characteristics						
Religion: Muslim	0.153	-0.003	[.009]	0.124	0.005	[.006]
Mother's current age	26.145	-0.096	[.121]	24.469	0.037	[.075]
HH Head's current age	30.959	-0.158	[.141]	29.167	0.018	[.108]
Is mother currently pregnant?	0.13	0.031	[.009]***	0.166	0.027	[.006]***
Number of HH members	3.616	0.007	[.031]	3.799	0.01	[.025]
Number of Children under 5 years	0.997	0.005	[.007]	0.999	0.007	[.005]
Number of women (15-49 years)	1.14	0.012	[.011]	1.164	-0.002	[.008]
Prenatal care taken for first birth?	0.768	-0.002	[.011]	0.582	0.002	[.009]
Is the first born child alive?	0.98	0.001	[.004]	0.967	0.006	[.003]*
Mother's education (years)	4.284	-0.049	[.044]	4.066	0.033	[.031]

Notes: This table reports the mean differences across various household characteristics for the households with the first born being young ($\leq$ 5 years of age). Coefficient β from equation (3) is reported (estimated without without any controls) along with the p-value. Standard errors are clustered at PSU level. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 3: Mean Difference Check - *Main Analysis Sample*

	Urban			Rural		
Variable	Mean	Difference	Standard Error	Mean	Difference	Standard Error
Asset Index	0.029	-0.052	[.009]***	0.013	-0.092	[.007]***
Water Source						
Piped into Dwelling	0.329	-0.011	[.004]***	0.1	-0.004	[.002]**
Piped to yard/plot	0.177	0.003	[.003]	0.113	0.001	[.002]
Public tab/standpipe	0.157	0.003	[.003]	0.146	-0.002	[.002]
Tubewell or borehole	0.199	0	[.003]	0.453	0.001	[.003]
Cooking Fuel						
LPG, Natural Gas	0.754	-0.005	[.004]	0.211	-0.001	[.002]
Wood	0.149	-0.001	[.003]	0.612	-0.004	[.003]
Floor Material						
Mud/Clay/Earth	0.079	0.003	[.002]	0.385	0.009	[.003]***
Animal Dung	0.018	0.001	[.001]	0.103	0	[.002]
Cement	0.527	-0.001	[.004]	0.342	-0.005	[.003]
Roof Material						
Metal/GI	0.149	0.004	[.003]	0.213	0.005	[.003]
RCC/RBC/Cement/Concrete	0.539	0	[.004]	0.258	-0.002	[.003]
Thatch/Palm Leaf	0.011	0	[.001]	0.062	0.004	[.002]***
Burnt Brick	0.027	0	[.001]	0.062	0.001	[.001]
Wall Material						
Mud	0.042	0.002	[.002]	0.191	0	[.002]
Cement/Concrete	0.478	0	[.005]	0.239	-0.008	[.003]***
Burnt Brick	0.236	-0.008	[.004]**	0.244	0.004	[.003]
Household Characteristics						
Religion: Muslim	0.171	0.001	[.003]	0.114	0.01	[.002]***
Mother's current age	36.609	-1.464	[.059]***	35.303	-1.97	[.048]***
HH Head's current age	41.547	-1.409	[.065]***	39.914	-2.099	[.054]***
Is mother currently pregnant?	0.017	0.006	[.001]***	0.028	0.015	[.001]***
Number of HH members	4.902	0.134	[.015]***	5.267	0.179	[.012]***
Number of Children under 5 years	0.435	0.078	[.005]***	0.623	0.148	[.006]***
Number of women (15-49 years)	1.485	0.354	[.008]***	1.434	0.241	[.006]***
Mother's education (years)	4.128	-0.011	[0.014]	3.791	0	[.011]

Notes: This table reports the mean differences across various household characteristics for the households with the first born ≤ 8 years old (in rural and urban regions, respectively). Coefficient β from equation (3) is reported (estimated without without any controls) along with the p-value. Standard errors are clustered at PSU level. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 4: Gender of first born child and reduction in OD

Dependent Variable:	Open Defecation Rate					
	Urban			Rural		
	(1)	(2)	(3)	(4)	(5)	(6)
First Born (Girl)	-0.009***	-0.008****	-0.008****	-0.014****	-0.015****	-0.013****
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	71,478	71,478	71,478	159,334	159,334	159,334
Mean OD	0.12	0.12	0.12	0.52	0.52	0.52
Controls	No	Yes	Yes	No	Yes	Yes
District FE	No	No	Yes	No	No	Yes

Notes: This Table reports the estimates from equation (3) on the *main analysis sample*. Outcome variable in all columns is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Results are reported separately by Urban and Rural areas. Column (3) and (6) are the richest specification. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 5: First Born Gender and Reduction in OD by Wealth Categories

Dependent Variable:	Open Defecation Rate									
	Deciles of Asset Index									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
	Panel A: Urban Areas									
First Born (Girl)	0.002	-0.040****	-0.020*	-0.018**	-0.009	-0.008	0.000	-0.007**	-0.001	0.001
	(0.012)	(0.010)	(0.010)	(0.008)	(0.007)	(0.006)	(0.005)	(0.003)	(0.002)	(0.001)
Observations	4,959	5,532	5,094	5,349	5,488	5,605	5,783	5,829	5,975	6,325
Mean OD	0.47	0.25	0.19	0.12	0.08	0.05	0.03	0.01	0	0
	Panel B: Rural Areas									
First Born (Girl)	-0.005	-0.001	-0.010	-0.010	-0.013*	-0.013*	-0.033****	-0.031****	-0.039****	-0.020****
	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.005)
Observations	11,422	11,618	11,764	11,827	11,962	12,059	12,121	14,440	10,244	12,967
Mean OD	0.84	0.79	0.74	0.64	0.6	0.49	0.41	0.32	0.21	0.08

Notes: This Table reports the estimates from richest specification of equation (3) on the *main analysis sample* divided into deciles of asset index. Outcome variable in all columns of both Panel A and B is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Results are reported in two panels of Urban and Rural areas. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 6: Gender of First Born Child and Reduction in OD - *Falsification*

Dependent Variable:	Open Defecation Rate					
	Urban			Rural		
	(1)	(2)	(3)	(4)	(5)	(6)
First Born (Girl)	0.000	-0.001	-0.002	0.007	-0.001	0.003
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
Observations	15,539	15,539	15,539	38,910	38,910	38,910
Mean OD	0.13	0.13	0.13	0.55	0.55	0.55
Controls	No	Yes	Yes	No	Yes	Yes
District FE	No	No	Yes	No	No	Yes

Notes: This Table reports the estimates from equation (3) on a *falsification* version version of *main analysis sample*, where the age of first born is restricted to be $\leq$ 8 years. Outcome variable in all columns is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Results are reported separately by Urban and Rural areas. Column (3) and (6) are the richest specification. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 7: First Born Gender and Reduction in OD by Wealth Categories-*Falsification*

Dependent Variable:	Open Defecation Rate									
	Deciles of Asset Index									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
	Panel A: Urban Areas									
First Born (Girl)	0.034	-0.013	0.004	0.001	-0.011	0.006	-0.007	-0.008	-0.002	0.000
	(0.029)	(0.020)	(0.023)	(0.019)	(0.021)	(0.018)	(0.010)	(0.005)	(0.008)	(-)
Observations	2,054	1,939	1,755	1,784	1,587	1,570	1,392	1,364	1,061	1,033
Mean OD	0.42	0.21	0.17	0.11	0.08	0.05	0.03	0.01	0	0
	Panel B: Rural Areas									
First Born (Girl)	-0.001	-0.002	0.023*	0.002	0.001	-0.020	0.024	0.004	0.024	-0.000
	(0.009)	(0.011)	(0.012)	(0.014)	(0.016)	(0.015)	(0.017)	(0.016)	(0.019)	(0.011)
Observations	4,825	4,302	4,145	4,041	3,972	3,752	3,753	4,235	2,910	2,975
Mean OD	0.84	0.77	0.72	0.64	0.6	0.52	0.41	0.35	0.23	0.09

Notes: This Table reports the estimates from equation (3) on a *falsification* version version of *main analysis sample*, (where the age of first born is restricted to be ≤ 8 years), divided into deciles of asset index. Outcome variable in all columns of both Panel A and B is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Results are reported in two panels of Urban and Rural areas. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 8: Crime Against Women and Change in Sanitation Behavior

Dependent Variable:	Open Defecation Rate			
	Urban		Rural	
	(1)	(2)	(3)	(4)
First Born (Girl)	-0.075****	-0.017	-0.039****	-0.026****
	(0.016)	(0.013)	(0.007)	(0.004)
Observations	2,301	3,231	12,100	37,672
Crime	High Crime	Low Crime	High Crime	Low Crime
Mean OD	0.28	0.24	0.24	0.26
Household Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table reports the estimates of equation (3) on *complier sample* for urban and rural regions divided by prevalence of crime against women. Columns are divided by crime reports across 10 states/UTs with highest incidences and remaining other states/UTs (using crime reports from Indian Human Development Survey). Column 1 & 3 report the results for 10 states/UTs with highest crime against women and column 2 & 4 report the results for all other states/UTs. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 9: State Performance and Change in Sanitation Behavior

Dependent Variable:	Open Defecation Rate			
	Urban		Rural	
	(1)	(2)	(3)	(4)
First Born (Girl)	-0.071***	-0.030***	-0.042****	-0.024****
	(0.022)	(0.011)	(0.008)	(0.004)
Observations	1,461	4,071	12,548	37,224
Region	BIMARU	Non-BIMARU	BIMARU	Non-BIMARU
Mean OD	0.26	0.25	0.39	0.21
Household Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table reports the estimates of equation (3) on *complier sample* for urban and rural regions divided by low and high performing states. Columns are divided by indicators of low versus high performing states. Column 1 & 3 report the results using *BIMARU* states (Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh). Column 2 & 4 report the results for*Non-BIMARU* states (all others). Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table 10: Literacy and Change in Sanitation Behavior

Dependent Variable:	Open Defecation Rate			
	Urban		Rural	
	(1)	(2)	(3)	(4)
First Born (Girl)	-0.071****	-0.011	-0.035****	-0.024****
	(0.017)	(0.012)	(0.006)	(0.004)
Observations	2,386	3,146	21,414	28,358
Literacy	Low	High	Low	High
Mean OD	0.31	0.21	0.34	0.2
Household Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table reports the estimates of equation (3) on *complier sample* for urban and rural regions divided by low and high literacy rates. Columns are divided by indicators of states with low versus high literacy. Column 1 & 3 report the results using 10 states/UTs with lowest literacy rates. Column 2 & 4 report the results for all other states. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

11 Figures

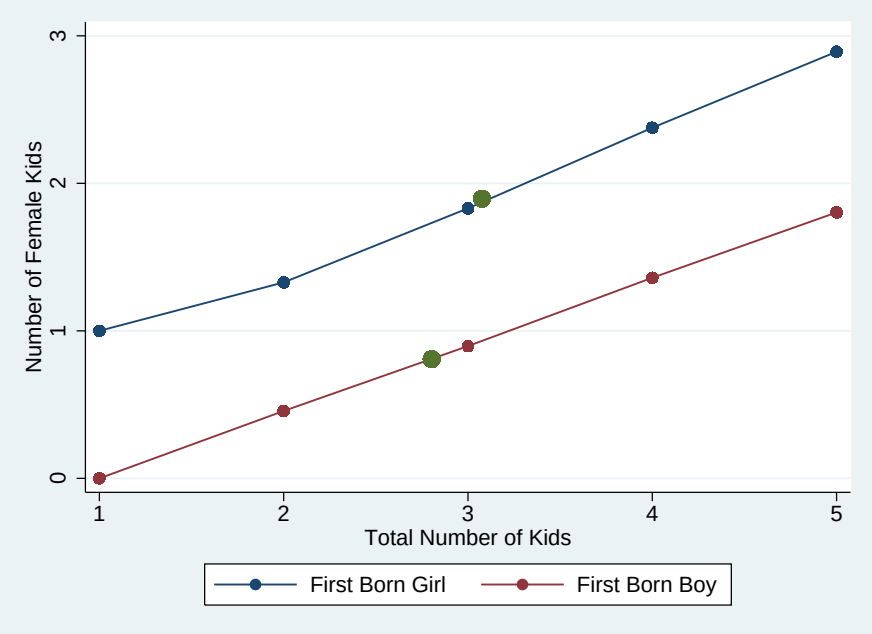

Figure 1: Sibling Size and Gender Composition by Gender of First Born Child

Notes: This Figure plots the association between total number of children and number of female children in a household, separated by the gender of first born child in the *main analysis sample*. The Green dots represent the interaction point of means from both axis.

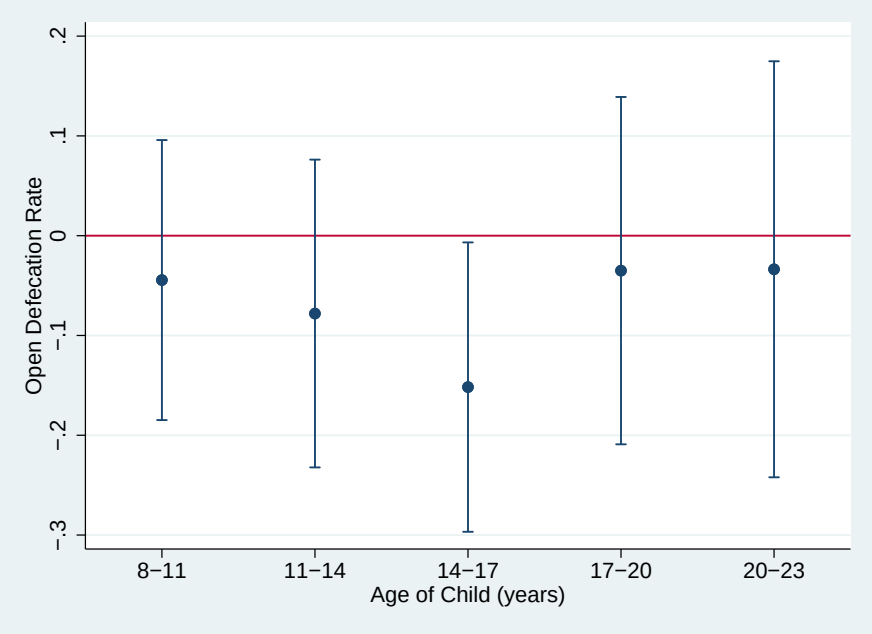

Figure 2: Change in Sanitation Behavior and Age of First Born- *Urban Areas*

Notes: This figure shows the coefficient from estimating the richest specification of equation (3) on urban *complier sample* divided into groups of age of first born child (ranging from 8-23 year old). Standard errors are clustered at level of PSU.

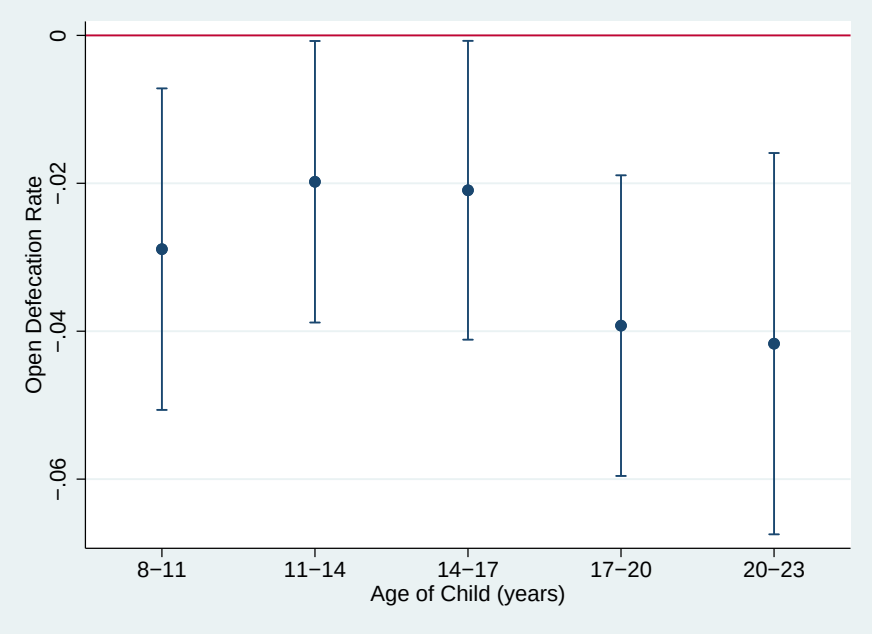

Figure 3: Change in Sanitation Behavior and Age of First Born- *Rural Areas*

Notes: This figure shows the coefficient from estimating the richest specification of equation (3) on rural *complier sample* divided into groups of age of first born child (ranging from 8-23 year old). Standard errors are clustered at level of PSU.

12 Appendix

Table A-1: Rate of Crime Against Women

	Overall	Asset Index $<$ median	Asset Index $\geq$ median
Urban	26.2%	26.3%	26.12%
Rural	19.5%	19.53%	19.48%

Notes: This table reports the rate of crime against women by urban and rural regions and by economic status of households within these regions. The crime rates are derived using self reports of frequency of incidences related to crime against women in Indian Human Development Survey (2011-12).

Table A-2: Robustness to Controls - *Main Results*

Dependent Variable:	Open Defecation Rate					
	Urban			Rural		
	(1)	(2)	(3)	(4)	(5)	(6)
First Born (Girl)	-0.006***	-0.006***	-0.007***	-0.010***	-0.013***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	71,478	71,478	71,478	159,334	159,334	159,334
Mean OD	0.12	0.12	0.12	0.52	0.52	0.52
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Distrtct FE	No	Yes	No	No	Yes	No
PSU FE	Yes	No	Yes	Yes	No	Yes
District*Caste FE	No	Yes	Yes	No	Yes	Yes
YOB*Caste FE	No	Yes	Yes	No	Yes	Yes
District*YOB FE	No	Yes	Yes	No	Yes	Yes

Notes: This Table reports the estimates from equation (3) on the *main analysis sample*. Outcome variable in all columns is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Results are reported separately by Urban and Rural areas. Column 1-3 and 4-6 demonstrate the robustness of coefficient of interest with respect to various sets of controls. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table A-3: Gender of First Born Child and Gender Composition of Household

	Gender Composition (Female Members)			
	(1)	(2)	(3)	(4)
Dependent Variable:	All Female Members	Only Daughters	Other than Daughters	Daughter in Law
	Panel A: Urban Areas			
First Born (Girl)	0.996***	1.066***	-0.070***	-0.051***
	(0.024)	(0.021)	(0.015)	(0.005)
Observations	7,471	7,471	7,471	7,471
	Panel B: Rural Areas			
First Born (Girl)	0.912***	1.067***	-0.155***	-0.143***
	(0.010)	(0.008)	(0.008)	(0.004)
Observations	63,645	63,645	63,645	63,645

Notes: This table reports the estimates of equation (3) on *complier sample* (including first born children of all ages) for Urban and Rural areas. The outcome variable here are total number of female members of household (Column 1), number of daughters of household head (Column 2), number of female members other than daughters of household head (Column 3) and, number of daughters in law of household head (Column 4). Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table A-4: Interaction of Mother's Education with Gender of First Born

Dependent Variable:	Open Defecation Rate			
	Rural		Urban	
Sample:	Complier	Main	Complier	Main
First Born (Girl)	-0.014***	-0.018***	-0.084**	-0.039***
	(0.005)	(0.003)	(0.043)	(0.007)
Years of Education (Mother)	-0.010***	-0.016***	-0.019*	-0.017***
	(0.001)	(0.001)	(0.010)	(0.001)
First Born Girl * Education	0.002	-0.000	0.008	0.003
	(0.001)	(0.001)	(0.013)	(0.002)
Observations	55,939	120,424	5,532	49,772
Household Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table reports the estimates from interaction of gender of the first born child and mothers years of education in both complier and main analysis sample. Standard errors (in parenthesis) are clustered at Primary Sampling Unit. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Table A-5: Mean Differences Check - *Assets in Complier Sample*

	Urban Compliers			Rural Compliers		
Asset	Mean	Difference	Standard Error	Mean	Difference	Standard Error
Electricity	0.987	-0.003	[.003]	0.994	0	[.001]
Radio	0.034	0.003	[.005]	0.137	0.006	[.004]*
Refrigerator	0.081	0.002	[.007]	0.406	0.009	[.005]*
Bicycle	0.381	-0.051	[.012]***	0.608	-0.016	[.005]***
Motorcycle/Scooter	0.149	-0.036	[.009]***	0.562	-0.044	[.005]***
Car/Truck	0.003	0.001	[.001]	0.09	-0.002	[.003]
Mattress	0.591	0.038	[.012]***	0.915	0.005	[.003]*
Pressure cooker	0.649	0.044	[.014]***	0.835	0.022	[.004]***
Chair	0.704	0.011	[.013]	0.958	0.005	[.002]***
Cot or Bed	0.787	-0.003	[.011]	0.982	0	[.001]
Table	0.46	0.021	[.014]	0.847	0.011	[.003]***
Electric fan	0.772	-0.039	[.012]***	0.936	0.001	[.002]
Colour Television	0.751	-0.006	[.012]	0.84	-0.005	[.003]
Sewing Machine	0.172	0.041	[.011]***	0.374	0.023	[.005]***
Internet	0.025	-0.003	[.004]	0.159	-0.043	[.004]***
Computer	0.007	0.001	[.002]	0.09	-0.012	[.003]***
Air conditioner/Cooler	0.085	0.002	[.007]	0.24	-0.024	[.004]***
Washing Machine	0.02	-0.002	[.004]	0.158	-0.007	[.004]*
Water Pump	0.058	-0.011	[.006]*	0.213	-0.009	[.004]**
Thresher	0.002	-0.001	[.001]	0.025	-0.003	[.001]**

Notes: This table reports the mean differences across various assets owned by the household by the gender of first born child, in *complier sample*. Standard errors are clustered at PSU level. Significance level is denoted by: * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

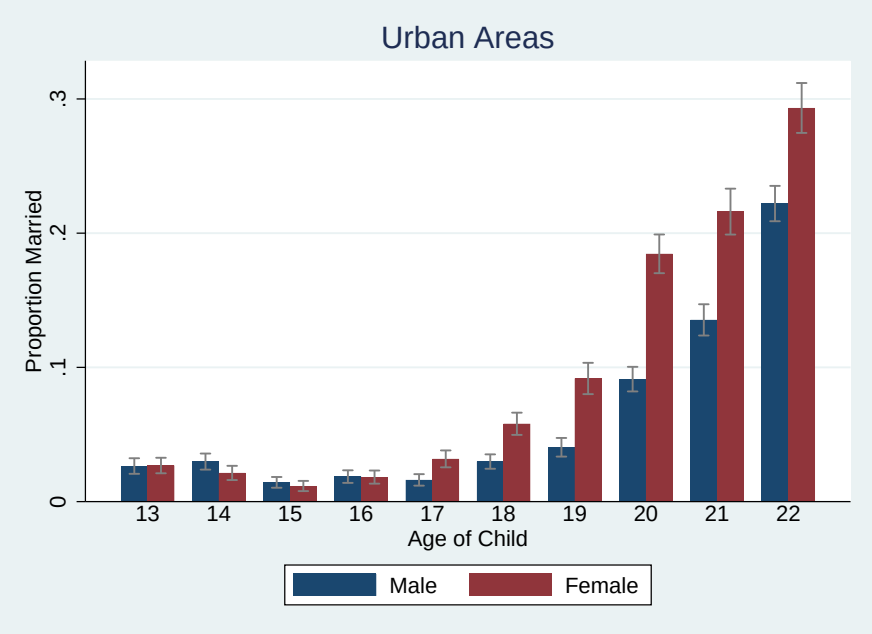

Figure A-1

Notes: This figure shows the age of first born child on x-axis and the proportion of them who are married on y-axis, using the *main analysis sample* for urban areas. The plots are separated by gender of the first born child and bars representing means include the associated confidence intervals.

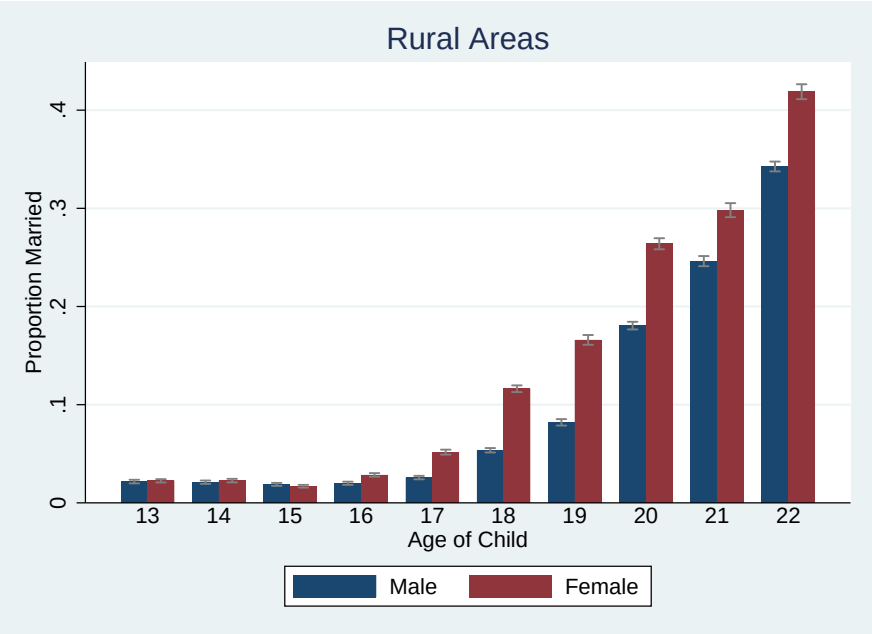

Figure A-2

Notes: This figure shows the age of first born child on x-axis and the proportion of them who are married on y-axis, using the *main analysis sample* for rural areas. The plots are separated by gender of the first born child and bars representing means include the associated confidence intervals.