Is poor sanitation killing more children in rural Zimbabwe? Results of propensity score matching method

Marshall Makate and Clifton Makate

Department of Economics, State University of New York at Albany, Albany, New York, USA., UNEP Tongji Institute of Environment for Sustainable Development , Tongji University , Shanghai, China

10 June 2016

Is poor sanitation killing more children in rural Zimbabwe? Results of propensity score matching method

Marshall Makate[1] and Makate Clifton[2]

Abstract: While studies in developing countries have examined the role of maternal and socio-demographic factors on child mortality, the role of poor sanitation (open defecation) on child mortality outcomes in rural communities of sub-Saharan Africa has received less attention. This study sought to examine the link between poor sanitation and child mortality outcomes in rural Zimbabwe. The analysis uses data from four rounds of the nationally representative Demographic and Health Survey for Zimbabwe conducted in 1994, 1999, 2005/06, and 2010/11. Using propensity score matching, we find that children living in households with no toilet facilities are 2.43 percentage points more liable to be observed dead by the survey date, 1.3, and 2.24 percentage points more likely to die before reaching the age of one and five years respectively. We also examined the possible differences in survival among female and male children. Our results indicate that male children are more liable to be observed dead by the survey date than female children. Also, female children have a slight survival advantage over boys during the under-five period. Our results suggest the need for more investments in basic sanitary facilities in Zimbabwe's rural areas to mitigate the potential devastating impacts of poor sanitation on child survival.

Key words: Poor sanitation; propensity score matching; child mortality outcomes; Zimbabwe

[1] Ph.D. Candidate, Department of Economics, State University of New York at Albany, Albany, New York, USA. For correspondence: email: mmakate@albany.edu

[2] Graduate student, UNEP Tongji Institute of Environment for Sustainable Development , Tongji University , Shanghai, China.

1. Introduction

The survival and health of many children within the first five years of life in developing regions is to a large extent compromised by amongst other factors, exposure to poor sanitation, especially open defecation (Hammer & Spears, 2016). Globally, nearly more than one billion people still practice open defecation which generates associated health costs of approximately US$260 billion annually (UNDP, 2016). Recent research has linked open defecation to adverse child health outcomes (Geruso & Spears, 2015; Hammer & Spears, 2016; Spears, 2013). However, very few studies in developing country contexts have been able to identify causal evidence of the impact of poor sanitation on child health outcomes (Gamper-Rabindran et al., 2010; Hammer & Spears, 2016). Understanding the link between poor sanitation and child health outcomes is important particularly for countries like Zimbabwe still experiencing high infant and child mortality (WHO, 2015). Besides, the public good nature of sanitation suggests important public health policy implications, hence the need to understand any causal connection between exposure to poor sanitation on infant and child health outcomes (Cutler & Miller, 2005).

In developed countries, improved sanitary facilities have been linked to improvements in child health outcomes (Cutler & Miller, 2005). However, there is a dearth in empirical evidence in the context of developing countries. Understanding the link between child health outcomes and poor sanitation has significant implications for public health policy particularly in countries such as Zimbabwe still struggling to curb the high infant mortality rates. For instance, in 2013, 55 children per every 1,000 live births died before reaching the age of one year while 88.5 children per 1,000 live births died before celebrating their first birthday (WHO, 2015). These mortality rates are unacceptably high given the SDGs global targets of reducing neonatal and under-five mortality to 12 and 25 deaths per 1,000 live births respectively by the year 2030.

A quantification of the impact of poor sanitation on child mortality is essential for public policy for at least two reasons. First, it helps policy makers to ascertain the appropriate amount of resources to the numerous other interventions competing for the same financial resources (Kumar & Vollmer, 2013). Second, it enriches our understanding of the importance of factors permitting other families in certain socioeconomic spheres to experience the worst consequences from poor sanitation than others. Zimbabwe is an ideal country to examine the association between poor sanitation and child mortality outcomes since it has made some progress in improving the sanitary situation of its rural citizens yet child mortality rates still wreck the country.

The primary objective in this study is to examine the impact of open defecation on infant and child mortality in Zimbabwe. Worldwide, ending open defecation has become a top priority amongst governments, non-governmental organizations, and other private stakeholders. The newly formulated Sustainable Development Goals (SDGs) have invigorated a new commitment towards the eradication of open defecation by 2030 (United Nations, 2015). Over the years, Zimbabwe has made efforts to improve the water and sanitation situation of its rural population through the introduction of the Integrated Rural Water Supply and Sanitation Programme (IRWSSP) in 1985 (Robinson, 2002). Despite registering early successes, nearly 48% of Zimbabwe's rural population still lack access to basic toilet facilities and hence resort to using the bush toilet system (Unicef, 2013). Moreover, the water situation in Zimbabwe's rural areas is still a cause for concern since nearly 29.5% of the households still lack improved sources of drinking water with 28% requiring travel time of about 30 minutes or more to fetch clean water (ZIMSTAT, 2012).

The challenge faced by the empirical literature in establishing a causal effect is that of potential endogeneity created by selection bias in poor sanitation. For instance, we do not know the reasons why some families in certain rural communities lack access to improved sanitary facilities. It's plausible that households with no access to sanitary facilities are also less forward-looking and possess inferior health knowledge and behavior. Thus, unobserved household-level characteristics might explain the differences in the child mortality outcomes across households and not the actual treatment itself.

To minimize the impact of selection bias in the use of poor sanitary facilities, this study uses propensity score matching (PSM) to assess the impact of poor sanitation (open defecation) on infant and child mortality in Zimbabwe. We exploit the richness of the nationally representative Zimbabwe Demographic and Health Survey (ZDHS) to create a suitable comparison group using the propensity score method to address selectivity bias. As a robustness check, a weighted least squares (WLS) regression model is estimated with the propensity score used as weights (Hirano & Imbens, 2001). Additionally, we conduct the bounds procedure to test the sensitivity of our estimates (Rosenbaum, 2002).

2. Related literature

The commitment to ending open defecation has increasingly gained attention amongst researchers, particularly in developing countries. Many of these studies have focused on India (Duflo et al., 2015; Kumar & Vollmer, 2013). In developed countries, studies have established an important link between public health infrastructure developments on child health outcomes. For instance, a U.S-based study by Watson (2006) revealed that a 10 percentage point increase in the share of American-Indian families receiving improvements in sanitary facilities was associated with an approximate 2.5 percent decrease in the infant mortality rate among US-Indian children. In a related study, Cutler and Miller (2005) examined the causal impact of clean water technologies in the form of filtration and chlorination on child mortality in a sample of major U.S. cities at the start of the 20^{th} century. They found that clean water technologies contributed to a reduction in nearly half of all the total mortality, three-quarters of the infant mortality and two-thirds of the child mortality in major U.S. cities.

While the developed country studies provide valuable insights into the link between improvements in sanitary facilities on child health outcomes, very little is known about this causal relationship in the context of developing countries. In sub-Saharan Africa, studies providing credible estimates on the causal link between poor sanitation and child health are scarce. This study fills an important gap in the literature by examining the effect of poor sanitation (open defecation) on infant and child mortality in Zimbabwe's rural communities.

Hammer and Spears (2016), examines this relationship using data from a cluster randomized control trial in rural Maharashtra, a village in India. They found that improvements in sanitation in Ahmednagar district were associated with a 0.3-0.4 standard deviation increase in children's height-for-age z-scores. Still in India, Geruso and Spears (2015) examined the connection between neighborhood sanitation and infant mortality. Exploiting the exogenous differences in latrines usage between Muslim and Hindu households induced by the religious compositions within these neighborhoods as an instrumental variable for sanitation, they found colossal infant mortality externalities of open defecation. In another related study for India, Kumar and Vollmer (2013), analyzed the association between improved sanitation and diarrhea among children living in rural India. Using the propensity score matching technique, they found that children from households with improved sanitary facilities had a lower chance of having diarrhea compared to their counterparts from families with non-improved clean facilities. In India, poor sanitation has also been linked to height deficiency in children under the age of five years (Spears, 2013).

Djimeu (2014), assesses the effect of improvements in sanitary facilities on child health and nutrition outcomes in Angola using propensity score, fixed effects, and propensity-score weighted regression techniques. Using cross-sectional household-level data collected between February 2000 and February 2001, he finds that improvements in sanitation were associated with a 0.335 standard deviation increase in height-for-age z-scores among under-five children. Kremer et al. (2011), used data from a randomized control trial in Kenya to assess the role of spring water protection on child health outcomes. Their results indicated that spring water infrastructure investments were associated with approximately 25% decline in child diarrhea. In a related study in Kenya, Kariuki et al. (2012) finds a significant reduction in child diarrhea associated with improved sanitation facilities. Improved sanitation not only enhance individual health but also has positive externalities on social and economic development in developing countries (Mara et al., 2010). In Brazil, Gamper-Rabindran et al. (2010) examined the link between the provision of piped water and infant mortality using a quantile treatment effect econometric approach with panel data. Their analysis suggests that improved water provision services in Brazil were associated with higher declines in infant mortality at the higher conditional quantiles compared to the lower conditional quantiles of the infant mortality distribution. Bose (2009), examines data from the Demographic and Health Survey (DHS) for Nepal to assess the impact of improved sanitation on child diarrhea. Relying on PSM, the empirical analysis suggests that improvements in sanitary facilities were associated with an approximate 5% reduction in diarrheal prevalence. Using data from rural Bangladesh and the Philippines, Lee et al. (1997) show that improvements in nutrition, sanitation and water quality can help lower mortality in children in developing countries. The analysis in this study builds on the above literature to assess the effect of open defecation on infant and child mortality in Zimbabwe – a country struggling to contain the infant and under-five mortality rates (WHO, 2015).

3. Methods

3.1. Data source

The analysis in this study uses data from four rounds of the nationally representative Zimbabwe Demographic and Health Survey (ZDHS) conducted in 1994, 1999, 2005/06 and 2010/11. The ZDHS is a cross-sectional household survey primarily funded by the United States Agency for International Development and implemented by Macro International in partnership with the government of Zimbabwe. The ZDHS is part of the global DHS program currently conducted in more than 40 developing countries. This survey collects detailed fertility and health information for women of reproductive ages 15-49 together with their children, for the five years preceding each survey. The survey uses a stratified two-stage cluster sample design based on the Zimbabwe population censuses of 1992 and 2002. At the first stage, enumeration areas are randomly sampled followed by a random sampling of households at the second stage. The DHS data also contains rich information on parental, household and community level characteristics. The individual response rates in the 1994, 1999, 2005/06, and 2010/11 ZDHS were high, 96%, 95.2%, 90%, and 93%, respectively.

All the child-level information is collected from the birth recode component of the ZDHS, which contains the birth histories (including the date of death for deceased children) of each interviewed woman for every birth they ever had. Household and maternal-level information is collected from the individual recode component of the ZDHS. After excluding all the observations from urban areas, 17,037 (24.43%) and with missing observations on poor sanitation, 1,002 (1.9%) we are left with an analytical sample of 51,690 child-level observations.

3.1.1. Outcome variables

This study collects child mortality information from the ZDHS birth histories component file containing detailed fertility history for each interviewed woman aged 15-49 years at the survey date. The information recorded in the birth history file includes the date of birth for all the births of each woman as well as the time of death for any deceased children. We construct three binary indicators to measure child mortality. First, an indicator variable is created that takes 1 if the child is observed to be dead by the survey date and 0 otherwise. Second, a binary indicator that equals 1 if the child died before reaching age 1 is created and 0 otherwise. As in Arulampalam and Bhalotra (2006) and Dancer et al. (2008), we exclude the children who were still alive at the survey date but were below the age of one and thus had not had a full year's exposure. Finally, we created a binary indicator taking 1 if the child died before reaching the age of five and 0 otherwise. Children who were alive at the survey date but had not reached the age of five were also excluded from the analysis since they had not had a full five-year's exposure (Grépin & Bharadwaj, 2015).

3.1.2. Treatment and explanatory variables

To assess the causal impact of poor sanitation on child mortality, this study uses a binary indicator variable that takes 1 if the household uses the bush toilet system (poor sanitation) and 0 otherwise. This definition is inspired by the United Nations Children' s Emergency Fund (UNICEF) Goal number 6 of the SDGs that seeks to achieve universal and equitable sanitation and hygiene for every world citizen as well as put to an end open defecation by 2030 (United Nations, 2015).

The observed variables used in the computation of the propensity score include the following individual, household and child-level variables: maternal age, years of schooling, ability to read and write, employment status, marital status, frequency of reading newspapers, household size, number of children under the age of five, number of boys and girls in the household, a dummy indicator for child's gender, indicators for household religious beliefs, indicators for household wealth, survey fixed effects, and region fixed effects.

3.2. Econometric framework

The empirical analysis in this study seeks to estimate the causal impact of poor sanitation facilities on child mortality, indicated by whether the child was observed dead by the survey date, before reaching age one, or before reaching age five. However, estimating the causal effect of poor sanitation on child mortality poses some econometric challenges since we cannot simultaneously observe the outcomes for the same children in the treatment and control groups (James J. Heckman, 1996; James J Heckman & Robb, 1985). For instance, in the present analysis, we can only observe outcomes for children from households using the bush toilet system or improved sanitary facilities, but we cannot see the outcomes for the same children in both states simultaneously. The ideal solution to this problem will be to implement a randomized controlled experiment in which counterfactuals are derived from the pool of eligible children from eligible households. However, randomly requiring individual families to have no access to toilets (or poor sanitation) or other public infrastructures such as schools, roads, electricity, hospitals and clinics is ethically infeasible.

The standard practice in the empirical literature in the absence of experimental data is to implement non-experimental methods such as the PSM to estimate the average treatment effects. The analysis in this study uses the PSM to assess the causal impact of poor sanitation on the likelihood that a child is observed to be dead by the survey date, before age one, or before age five. The PSM creates comparable treatment and control groups that are similar in observable characteristics (Rosenbaum, 2002; Rosenbaum & Rubin, 1985). Recently, the number of studies using matching methods in evaluating the causal impacts of public policy interventions or other public health-related interventions has significantly increased (Becker & Ichino, 2002; Djimeu, 2014; Kumar & Vollmer, 2013; McCrory & Layte, 2011).

The PSM is a two-step procedure that involves estimating either a probit or logit regression at the first stage to generate the probability (propensity score) $p(X)$ that a child i lives in a treated household (practices open defecation) conditional on some observed or background characteristics vector X. The second stage involves matching children from households with poor sanitation to the children with improved sanitation. A comparison of the average health outcome variables of the treatment and control groups can then be attributed to the impact of the treatment that selection into the program is sorely based on the vector X of observed covariates.

3.2.1. The average treatment effect on the treated

To minimize the potential bias created by the endogeneity nature of the treatment variable $psanit$ (poor sanitation), this study uses the PSM technique as briefly explained above. Let Y_{1i} and Y_{0i} denote the outcome variables for children residing in treated and control households, respectively. Define $psanit \in \{0,1\}$ as the binary indicator for the treatment. Following Rosenbaum and Rubin (1983), the propensity score $p(X)$ can be specified as follows:

$$p(X) = prob(psanit = 1|X) = E(psanit|X) \tag{1}$$

where X is a vector of observed characteristics believed to influence poor sanitation. Using the propensity score $p(X)$ calculated in equation (1), the average treatment effect on the treated $\left(\widehat{ATT}\right)$ can thus be specified as follows:

$$\begin{aligned}\widehat{ATT} &= E\{Y_{1i} - Y_{0i}|psanit_i = 1\} \\ &= E[E\{Y_{1i} - Y_{0i}|psanit_i = 1,\ p(X)\}] \\ &= E[E\{Y_{1i}|psanit_i = 1,\ p(X)\} - E\{Y_{0i}|psanit_i = 0,\ p(X)\}|psanit_i = 1]\end{aligned} \tag{2}$$

Equation (2) presents the average effect of the treatment under the conditional independence (CIA) and overlap assumption. The CIA assumes random program participation given the observed covariates X, and can be written as $(Y_1, Y_0 \perp psanit|X)$. The overlap assumption implies that for each X, there are both control and treatment observations, that is, $\{0 < Prob[psanit = 1|X] < 1\}$.

3.2.2. *Nearest-neighbor matching method*

The second stage in PSM technique involves matching treatment observations to control units. This paper uses the widely employed matching algorithm – one-to-one nearest-neighbor (NN-1) matching technique with replacement and within a caliper. The main idea in this matching method is that the propensity score of each treatment unit is matched to the closest control case with a propensity score closest in value. The NN matching estimator with replacement and within a caliper can formally be specified as follows:

$$\widehat{ATT} = \frac{1}{N1}\sum_{i=1}\{Y_i - Y_j\} \qquad (3)$$

For a given caliper of size $\varphi > 0$, j is selected such that,

$$\varphi > |p(X_i) - p(X_j)| = min_{k \in I}\{|p(X_i) - p(X_j)|\}$$

Any control observation j outside the caliper φ radius of the treated observation i will be left unmatched and thus excluded from the analysis. The analysis in this study uses the nearest-neighbor observation within the $\varphi = 0.01$ radius to create the counterfactual for each treatment unit i. STATA's $psmatch2$ function was used for the PSM analysis with bootstrapped standard errors using 500 replications.

3.2.3. *Propensity-based weighted regression*

In the program evaluation literature, another commonly used method to calculate the ATT is the multivariate regression model that uses the propensity score as a sampling probability weight. Many studies support the suitability of this approach to producing more efficient estimates (Hirano & Imbens, 2001; Hirano et al., 2003; Rosenbaum, 1987). The main idea in the propensity-based weighted regression is to use the propensity score $\hat{p}(X)$ to adjust the treatment and control groups and achieve covariate balance across the two groups. In this case, the weight is the inverse of the propensity score specified as $(1/\hat{p}(X))$ for children from treated households and $(1/(1-\hat{p}(X)))$. As a robustness check, we estimated the following regression model with the propensity score used as the sampling weight:

$$Y_{ijp} = \alpha_0 + \alpha_1 psanit_{jc} + \delta X_{jc} + \gamma_{jc} + \varepsilon_{ijc} \qquad (4)$$

where Y_{ijc} represents the mortality outcome for a child i living in household j in community c. The variable $psanit_{jc}$ is the indicator for poor sanitary facilities (no toilets); the vector of observed covariates X_{jc} includes child and household-level characteristics, γ_{jc} indicates community fixed effects and ε_{ijc} is the error term. All the analysis in this study was conducted in STATA SE version 13.0 (Stata, 2013).

4. Results

4.1. Descriptive statistics

The left panel in Fig. 1 shows the average trends of poor sanitation for all the rural households in our sample by year of DHS survey. The left panel shows that open defecation has persistently been an issue in rural Zimbabwe. In 1994, nearly 51% of the interviewed rural households had no toilet facilities. Over the years, the percentage of people with no toilet facilities has remained high averaging around 40%. The right panel in Fig. 1 shows the percentage of households with no toilet facilities by household wealth category. It is clear that among all the rural households, the relatively poor families appear to be the most disadvantaged and this trend has persisted over time.

[Insert Fig. 1 here]

Fig. 2 shows the time trends in overall child mortality by child's year of birth for the sample of children born in Zimbabwe's rural areas. According to Fig.1, infant and child death rates dropped significantly between 1980 and 1988. The observed decline in child mortality rates might be due to improving economic conditions as it coincides with the period just after Zimbabwe's independence in 1980. Mortality rates rose again reaching a peak in 1990 before showing a rather unstable pattern from 1990 to 1998. The period 2000 to 2009 marked the recessionary period as Zimbabwe's inflation reached unprecedented levels (Pindiriri, 2012). The average child mortality rates from 2005 appear to be on the rise again.

[Insert Fig. 2 here]

Table 1 provides the pre-matching descriptive statistics of the variables used in the estimation of the propensity score. The first three columns (1, 2, 3) in Table 1 show the means of the covariates for all the households, homes with poor sanitary facilities and improved health services, respectively. Column (4) tests the differences in means between houses with inadequate health services and those with improved health facilities. The majority (19%) of the women in our sample are aged 20-24 with women from the treatment group (poor sanitation) being relatively older than their counterparts from the control group (improved sanitation). Women from households with poor sanitary facilities have lower levels of education, literacy rates, media access rates and employment rates. The average household size is higher (6.23 vs. 6.13) in families with poor sanitary facilities. The percentage of homes with poor sanitation varies by province from 8.9% in Mashonaland East to 18.7% in Matabeleland North. The t-test for the differences between households with no toilet facilities and those with improved facilities reveal significant differences between the two groups. These observed significant differences motivate the use of PSM method to estimate the treatment effect of poor sanitation on infant and child mortality. The substantial differences between the two groups suggest that matching would purge the associated bias from the covariates and thus improve the precision of the estimates. Also presented in Table 1 are the results from the logit regression model (columns (5) and (6)) predicting the probability that a household practices open defecation or has no toilet facilities. The results indicate that most of the variables included in the regression model are significant predictors of poor sanitation.

[Insert Table 1 here]

4.2. Quality of the matching

Table 2 shows the differences in the mean values of the treatment and control groups. The matching process should result in a significant reduction in the bias to make the two groups comparable ensuring the overall balance of the covariates. The results in Table 2 indicate that the matching process resulted in a significant reduction in bias. Even though there is a substantial and significant decrease in the bias, the difference in the means of some of the covariates is still statistically significant between the two groups.

[Insert Table 2 here]

One of the necessary conditions required for a successful matching process is the fulfillment of the overlap condition. There needs to be a sufficient overlap of the propensity scores across the treatment and control groups (Rosenbaum, 2002). The visual analysis of the density distribution of the propensity score in Fig.3 indicates sufficient overlap between the two groups and thus satisfies the overlap condition of the PSM method.

[Insert Fig. 3 here]

4.3. Poor sanitation and child mortality

Table 3 presents the ATT for the effects of poor sanitation on child mortality outcomes. The ATT measures the difference in the average mortality rates for children living in households practicing open defecation and children living in households with improved sanitary facilities. The results indicate that children from families practicing open defecation are more liable to be observed dead by the survey date, before reaching the age of one, and before reaching the age of five. Specifically, we find that children living in households with no toilet facilities are 2.43 percentage points more likely to be observed dead before the survey date than their counterparts from households with better sanitation. Children living in households with poor sanitary facilities are 1.3 and 2.24 percentage points more liable to die before celebrating their first and fifth birthdays, respectively than those from households with improved facilities. Given that the average mortality rate for children living in homes with no toilet facilities is 9.4%, the 2.43 percentage point increase in child mortality represents an approximate 25.85% $(2.43 * 100/9.4)$ increase in child mortality. Similarly, given that the average infant mortality rate for children living in families with no toilet facilities, a 1.3 percentage point increase in infant mortality represents an approximate 23.63% $(1.3 * 100/5.5)$ overall increase in the probability that children die before reaching age one. Finally, the 2.24 percentage increase in under-five mortality represents an approximate 26.67% $(2.24 * 100/8.4)$ increase in mortality of children before reaching the age of five.

[Insert Table 3 here]

For comparison, we also estimated a WLS regression using equation (4). The results are presented in Table 4. The top panel in Table 4 shows the results from a linear regression model that ignores potential selectivity bias in the poor sanitation variable. The bottom panel in Table 4 shows the results of the WLS regression model using the propensity score as weights. The coefficients we obtained are consistent with the main results we found in Table 3. To summarize, the coefficient estimates from the LPM model are 0.022, 0.008, and 0.017 for overall child mortality, child is dead before age one, and child is dead before age five, respectively. For the WLS model, the estimates are 0.020, 0.008, and 0.017, for overall child mortality, child is dead before age one, and child is dead before age five, respectively. It is important to note that the results from the LPM and WLS cannot be interpreted as causal since they do not control for potential selection bias. The rest of the study focuses on the results from the PSM.

[Insert Table 4 here]

4.4. Heterogeneous treatment effects

As noted in the studies by Kumar and Vollmer (2013) and Djimeu (2014), it is possible that the effects of poor sanitation on child mortality might be different for boys than girls. In Vietnam, Le Pham et al. (2013) has shown that mortality in children might differ by the child's gender. To explore this possibility in the Zimbabwean context, we estimated the ATT by child's gender. The results are presented in Table 5. The results indicate that boys from households with no toilet facilities are 2.01 percentage points more liable to be observed dead by the survey date compared to their counterparts from homes with improved sanitation and this effect is statistically significant at 99% confidence level. Also, boys from families with no toilet facilities are 2.19 and 2.34 percentage points likely to die before reaching age one and five respectively. These results are statistically significant at the 95% and 99% confidence levels, respectively. For girls, we find a treatment effect of 1.15, 1.26, and 1.42 percentage points for mortality before the survey date, during infancy, and before reaching age five, respectively. The results for girls are statistically significant at the 90%, 99%, and 95% confidence levels, respectively. The treatment effect for boys is higher than for girls for the outcomes mortality by survey date and before age five. For infant mortality, our results show a higher treatment effect for girls than for boys.

[Insert Table 5 here]

4.5. Hidden bias and sensitivity checks

The biases arising due to the observed household-level characteristics are corrected by the PSM method. However, there could still be some other unobserved bias. For instance, it is possible that households with no toilet facilities are less forward-looking or illiterate and thus less liable to make investments in better sanitary facilities (Kumar & Vollmer, 2013). This behavioral aspect of the household is not observable and might impact the estimated treatment effects from PSM technique. If the treated and control groups do differ on unobservable characteristics, there might be hidden bias. We employed the bounds procedure suggested by Rosenbaum (2002) to test the extent to which the unobserved variables have to be changed to make the estimated ATT invalid.

The results for the sensitivity analysis are presented in Table 6. The results indicate that the estimated ATT with PSM does not remain significant even in the presence of large significant bias. Thus, one should take our estimated ATT with caution. The results indicating negative selection bias suggests that the PSM results are insensitive to selection bias arising from unobserved household characteristics. Here, we focus on the Q_{mh}^{-} and p_{mh}^{-} statistics. However, for the case of positive unobserved selection bias, the results are not always insensitive to the bias due to the unobserved characteristics. The relevant statistics to focus on are the Q_{mh}^{+} and p_{mh}^{+}. The fact that our PSM results are likely to be influenced by hidden bias, we interpret these results with caution.

[Insert Table 6 here]

5. Discussion

This study examines the vital link between the poor sanitation and child mortality in Zimbabwe – a country currently experiencing unfavorable child outcomes. We use PSM method to quantify the likely consequences of open defecation on mortality outcomes for children. Taken together, our results indicate that, at least in rural Zimbabwe from 1990 to 2011, children living in households with no toilet facilities had an elevated risk of dying before reaching the age of one or five compared to their counterparts from homes with improved sanitary facilities. Specifically, children from households that use the bush toilet system are 2.43, 1.30, and 2.24 percentage points more liable to be observed dead by the survey date, before the age of one year and the age of five respectively. This result might be explained by a combination of factors including the cholera outbreaks that have ravaged the country nearly every year since 1998. In 2008, approximately 4,282 deaths were reported to be as a result of cholera (Ahmed et al., 2011). The effects of cholera are even dire in communities with no toilet facilities. Also, families with poor sanitary facilities are more liable to have very high diarrhea prevalence rates and consequently higher mortality rates (ZIMSTAT, 2012). The findings corroborate the results from previous other studies (Gamper-Rabindran et al., 2010; Geruso & Spears, 2015; Spears, 2013).

Our results also indicate slightly larger treatment effects for boys than girls for mortality observed at the survey date and before age five. This finding is consistent with previous studies in demography that established a higher infant mortality rate among boys than boys (Pongou, 2013). It has been long-observed that females have better survival chances than do boys owing to their superior genetics and biological makeup, while boys are more vulnerable to diseases compared to their female counterparts (Naeye et al., 1971; Waldron, 1983). Other studies attribute the female survival advantage to the preconceived or prenatal environmental factors impacting the probabilities of giving birth to either a boy or girl. It's plausible that these same environmental factors explain the observed sex differentials in child mortality (Pongou, 2013). Our results are at odds with a study by Kumar and Vollmer (2013) for India who found an ATT of improved sanitation on diarrheal prevalence to be in favor of boys than girls.

While poor sanitation might be a plausible explanation for the high child mortality rates observed in Zimbabwe today, other factors might be at play here as well. We certainly did not control for other plausible and important factors such as the devastating impacts of HIV/AIDS among others. Though Zimbabwe has made important progress in the quest to improve the livelihoods of its citizens, approximately 48% of its rural population still practice open defecation (Unicef, 2013).

This study is not without limitations. First, due to data restrictions, we did not explore potential channels through which poor sanitation might result in child deaths. For example, the effects of poor sanitation might be worsened or reduced because of the household's socioeconomic status. Second, we did not control for other important factors that might have affected child mortality outcomes such as the HIV/AIDS prevalence rates. However, since HIV prevalence rates might differ by region or province, controlling for regional variation as we do might account for some of these effects. Despite the limitations, our study provides important insights on the extent and consequences of open defecation on child survival in Zimbabwe's rural communities. Given that the mortality rates for children remain high in the country (WHO, 2015), the findings in this study suggest the need for public policy makers to focus on addressing the inequalities in improved sanitation facilities access particularly in Zimbabwe's rural areas. Improving the sanitary facilities in rural Zimbabwe might be an effective strategy to reduce child mortality in Zimbabwe.

References

Ahmed, S., Bardhan, P.K., Iqbal, A., Mazumder, R.N., Khan, A.I., Islam, M.S., et al. (2011). The 2008 Cholera Epidemic in Zimbabwe: Experience of the icddr, b Team in the Field. *Journal of Health, Population and Nutrition*, 541-546.

Arulampalam, W., & Bhalotra, S. (2006). Sibling death clustering in India: state dependence versus unobserved heterogeneity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 829-848.

Becker, S.O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal*, 2, 358-377.

Bose, R. (2009). The impact of water supply and sanitation interventions on child health: evidence from DHS surveys. *International Initiative for Impact Evaluation: New Delhi*.

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

Dancer, D., Rammohan, A., & Smith, M.D. (2008). Infant Mortality and Child Nutrition in Bangladesh. *HEALTH ECONOMICS*, 17, 1015-1035.

Djimeu, E.W. (2014). The impact of social action funds on child health in a conflict affected country: Evidence from Angola. *Social Science & Medicine*, 106, 35-42.

Duflo, E., Greenstone, M., Guiteras, R., & Clasen, T. (2015). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation. National Bureau of Economic Research.

Gamper-Rabindran, S., Khan, S., & Timmins, C. (2010). The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics*, 92, 188-200.

Geruso, M., & Spears, D. (2015). Neighborhood sanitation and infant mortality. National Bureau of Economic Research.

Grépin, K.A., & Bharadwaj, P. (2015). Maternal education and child mortality in Zimbabwe. *Journal of Health Economics*, 44, 97-117.

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, 135-148.

Heckman, J.J. (1996). Randomization as an Instrumental Variable. (p. 336): MIT Press.

Heckman, J.J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30, 239-267.

Hirano, K., & Imbens, G.W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology,* 2, 259-278.

Hirano, K., Imbens, G.W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica,* 71, 1161-1189.

Kariuki, J.G., Magambo, K.J., Njeruh, M.F., Muchiri, E.M., Nzioka, S.M., & Kariuki, S. (2012). Effects of Hygiene and Sanitation Interventions on Reducing Diarrhoea Prevalence Among Children in Resource Constrained Communities: Case Study of Turkana District, Kenya. *Journal of Community Health,* 37, 1178-1184.

Kremer, M., Leino, J., Miguel, E., & Zwane, A.P. (2011). Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions. *Quarterly Journal of Economics,* 126, 145-205.

Kumar, S., & Vollmer, S. (2013). Does Access to Improved Sanitation Reduce Childhood Diarrhea in Rural India? *HEALTH ECONOMICS,* 22, 410-427.

Le Pham, T., Kooreman, P., Koning, R.H., & Wiersma, D. (2013). Gender patterns in Vietnam's child mortality. *Journal of Population Economics,* 26, 303-322.

Lee, L.-f., Rosenzweig, M.R., & Pitt, M.M. (1997). The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. *Journal of Econometrics,* 209.

Mara, D., Lane, J., Scott, B., & Trouba, D. (2010). Sanitation and Health. *PLoS Medicine,* 7, 1-7.

McCrory, C., & Layte, R. (2011). The effect of breastfeeding on children's educational test scores at nine years of age: Results of an Irish cohort study. *Social Science & Medicine,* 72, 1515-1521.

Naeye, R.L., Burt, L.S., Wright, D.L., Blanc, W.A., & Tatter, D. (1971). Neonatal mortality, the male disadvantage. *Pediatrics,* 48, 902-906.

Pongou, R. (2013). Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins. *Demography,* 50, 421-444.

Robinson, A. (2002). The Zimbabwe Experience: Lessons From a Review of 15 years of the Zimbabwe Integrated Rural Water Supply and Sanitation Programme. *World Bank Water and Sanitation Program—Africa region, Nairobi, Kenya, Field Note,* 26549.

Rosenbaum, P.R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika,* 74, 13-26.

Rosenbaum, P.R. (2002). *Observational studies*: New York : Springer, [2002] 2nd ed.

Rosenbaum, P.R., & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. (p. 41): Biometrika Trust.

Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. (p. 33): American Statistical Association.

Spears, D. (2013). How Much International Variation in Child Height Can Sanitation Explain?

Stata, S. (2013). Release 13. *Statistical Software. College Station. Texas, USA: StataCorp LP*.

UNDP, U.N.D.P. (2016). UNDP Support to the Implementation of Sustainable Development Goal 6.

Unicef, U.N.C.s.E.F. (2013). Supporting Zimbabwe's Rural WASH Sector. Rural WASH.

United Nations. (2015). Transforming Our World: The 2030 Agenda for Sustainable Development.

Waldron, I. (1983). Sex differences in human mortality: the role of genetic factors. *Social Science & Medicine*, 17, 321-333.

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, 1537-1560.

WHO. (2015). World Health Statistics 2015. World Health Organization.

ZIMSTAT. (2012). Zimbabwe Demographic and Health Survey 2010-11. Calverton, Maryland.

Figure Captions

Fig. 1 Prevalence of poor sanitation in rural Zimbabwe

Fig. 2 Trends in infant and child mortality by year of child's birth

Fig. 3 Distribution of the propensity score

Table 1: Descriptive statistics of household characteristics before propensity score matching

Household characteristics	All Households	Poor Sanitation	Improved sanitation	t-test (p-value)	Propensity score logit: Coefficients	Propensity score logit: Std. error
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal-related characteristics						
Age 20-24	0.193	0.188	0.198	0.108	-0.222**	(0.071)
Age 25-29	0.163	0.167	0.159	0.157	-0.181*	(0.081)
Age 30-34	0.132	0.134	0.130	0.456	-0.259**	(0.091)
Age 35-39	0.109	0.116	0.104	0.007	-0.323**	(0.103)
Age 40-44	0.088	0.092	0.085	0.083	-0.394***	(0.120)
Age 45-49	0.077	0.077	0.077	0.985	-0.620***	(0.139)
Education less than primary	0.185	0.209	0.167	0.000	0.079	(0.062)
Education more than primary	0.457	0.355	0.540	0.000	-0.154**	(0.056)
Able to read and write	0.875	0.834	0.907	0.000	0.023	(0.075)
Employed	0.388	0.341	0.426	0.000	-0.200***	(0.045)
Married	0.743	0.770	0.721	0.000	0.148*	(0.064)
Read newspapers frequently	0.299	0.214	0.366	0.000	-0.089	(0.048)
Household size	6.177	6.230	6.134	0.026	-0.079***	(0.008)
Number of children under age 5	1.168	1.289	1.070	0.000	0.136***	(0.025)
Number of girls	1.308	1.446	1.197	0.000	0.043*	(0.020)
Number of boys	1.337	1.481	1.221	0.000	0.062**	(0.019)
Household religious beliefs						
Catholic	0.286	0.269	0.300	0.000	-0.481***	(0.115)
Protestant	0.210	0.185	0.230	0.000	-0.524***	(0.123)
Pentecostal	0.102	0.092	0.110	0.000	-0.280*	(0.130)
Apostolic section	0.236	0.243	0.230	0.029	-0.364**	(0.126)
Other Christian	0.071	0.091	0.054	0.000	-0.107	(0.144)
Muslim/others	0.055	0.064	0.047	0.619	-0.416**	(0.146)
Household wealth						
Quintile 1 (poorest)	0.290	0.556	0.077	0.000	5.937***	(0.393)
Quintile 2	0.272	0.287	0.260	0.000	4.158***	(0.391)
Quintile 3	0.263	0.133	0.367	0.000	3.080***	(0.391)
Quintile 4 (richest)	0.144	0.023	0.242	0.000	1.631***	(0.397)
Survey fixed effects						
Year 1994	0.220	0.252	0.194	0.000	1.041***	(0.078)
Year 1999	0.206	0.189	0.220	0.000	0.037	(0.074)
Year 2005/06	0.286	0.296	0.279	0.010	0.413***	(0.059)
Provinces						
Mashonaland central	0.130	0.098	0.156	0.000	0.202*	(0.088)
Mashonaland east	0.115	0.080	0.144	0.000	0.726***	(0.081)
Mashonaland west	0.106	0.114	0.099	0.001	1.117***	(0.081)
Matabeleland north	0.121	0.187	0.068	0.000	1.199***	(0.083)
Matabeleland south	0.120	0.119	0.121	0.663	0.850***	(0.077)
Midlands	0.133	0.148	0.122	0.000	0.954***	(0.077)
Masvingo	0.138	0.171	0.112	0.000	1.196***	(0.076)

Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level. Robust standard errors are presented. Baseline categories are: mother's age 15-19; mother's education is less than primary school; traditional religious beliefs; household wealth quintile 5 (richest); survey year 2010/11; Manicaland province.

Table 2: Covariate balance - individual *t*-test

Household characteristics	Poor sanitation (1)	Improved sanitation (2)	% reduction in bias (3)	Differences (1)-(2) t-values (4)
Age 20-24	0.18656	0.18699	80.9	-0.12
Age 25-29	0.16017	0.16248	-229.5[3]	-0.72
Age 30-34	0.13255	0.13185	43.4	0.23
Age 35-39	0.11873	0.1015	-92	6.28
Age 40-44	0.09296	0.08945	32.1	1.39
Age 45-49	0.07426	0.08892	-143.4[1]	-6.11
Education less than primary	0.21965	0.21338	-9193.8[1]	1.74
Education more than primary	0.42553	0.432	32.5	-1.49
Able to read and write	0.72648	0.8839	-48	-41.25
Employed	0.3956	0.47913	25.7	-19.26
Married	0.9823	0.95271	-588.4[1]	19.09
Reads newspapers frequently	0.16088	0.3243	-41.4	-44.26
Household size	6.5433	6.0263	-231.7[1]	19.96
Number of children under age 5	1.4152	1.2171	1.4	21.04
Number of girls	1.348	1.3734	-283.8[1]	-1.86
Number of boys	1.3819	1.3521	32	2.14
Household religious beliefs				
Catholic	0.37097	0.35493	-1060.7[1]	3.8
Protestant	0.21896	0.20165	-71.1	4.84
Pentecostal	0.09076	0.08649	-1232.5[1]	1.71
Apostolic section	0.17406	0.1963	-423	-6.53
Other Christian	0.06775	0.07064	-591.1[1]	-1.3
Muslim/others	0.03386	0.04317	-145.7[1]	-5.52
Household wealth				
Quintile 1 (poorest)	0.29495	0.3033	-32.7	-2.08
Quintile 2	0.27763	0.28179	46.4	-1.06
Quintile 3	0.26286	0.25667	29.8	1.61
Quintile 4	0.14005	0.13535	63.7	1.55
Survey fixed effects				
Year 1994	0.31565	0.29172	-25.4	5.93
Year 1999	0.28675	0.26201	-432.9[1]	6.32
Year 2005/06	0.09634	0.11204	-89.8	-5.86
Provinces				
Mashonaland central	0.10819	0.1377	-6.3	-10.25
Mashonaland east	0.12597	0.10734	0.8	6.62
Mashonaland west	0.13532	0.09396	1.8	14.83
Matabeleland north	0.13539	0.13239	75.5	1
Matabeleland south	0.13258	0.123	-1226.2[1]	3.27
Midlands	0.11931	0.13028	-1509.8[1]	-3.78
Masvingo	0.12189	0.13882	-24.9	-5.73

Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level. Robust standard errors are presented. Baseline categories are: mother's age 15-19; mother's education is less than primary school; traditional religious beliefs; household wealth quintile 5 (richest); survey year 2010/11; Manicaland province.

[1] The percent bias reduction exceeds 100% because the difference that open defecation makes is far greater in the matched sample than the unmatched sample. This might be as a result of selection bias or imbalance in the propensity score that failed to meet the statistical threshold for imbalance according to Stata's propensity score balancing property.

Table 3: Average treatment effect of poor sanitation on child mortality outcomes

	Child dead before survey date		Child dead before age 1		Child dead before age 5	
Variables	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Poor sanitation	0.0243***	(0.00509)	0.0130***	(0.00413)	0.0224***	(0.00545)
Observations	51,683		48,694		38,272	
Average poor sanitation	0.094		0.055		0.084	
Average improved sanitation	0.084		0.051		0.073	

Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level. Standard errors for the ATT shown in parentheses are calculated using bootstrapping with 500 replications. One-to-one nearest neighbor matching was used to create the counterfactual.

Table 4: Poor sanitation and child mortality - robustness

	Child is dead before survey date		Child is dead before age 1		Child is dead before age 5	
Linear Probability Model (LPM)						
Poor sanitation	0.022***	(0.004)	0.008**	(0.003)	0.017***	(0.003)
Number of children	38207		38207		38207	
R-square	0.0234		0.0105		0.0172	
Weighted Least Squares (WLS)						
Poor sanitation	0.020**	(0.006)	0.008	(0.004)	0.017**	(0.006)
Number of children	38207		38207		38207	
R-square	0.0375		0.0201		0.0312	

Notes: *** significant at 1%, **significant at 5%, and * significant at 10%. Standard errors are clustered at the primary sampling unit. All the regression include controls for: maternal age, education, number of children under age five, breastfeeding status, birth order, child gender, household size, wealth status, child's year of birth, survey fixed effects, and region fixed effects.

Table 5: Heterogeneous impacts of poor sanitation on child mortality

	Child dead by survey date		Child dead by age 1		Child dead by age 5	
Variables	ATT	Std. error	ATT	Std. error	ATT	Std. error
Stratified by gender of children						
Boys	0.0201***	(0.00581)	0.0119**	(0.00497)	0.0234***	(0.00705)
Girls	0.0115*	(0.00599)	0.0126***	(0.00461)	0.0142**	(0.00643)

Notes: ***Significant at 1% level; **significant at 5% level; *significant at 10% level. Standard errors for the ATT shown in parentheses are calculated using bootstrapping with 500 replications. One-to-one nearest neighbor matching was used to create the counterfactual.

Table 6: Sensitivity analysis: Rosenbaum bounds

	Mantel-Haenszel (1959) bounds for child mortality variables											
	Child is dead before the survey				Child is dead before age one				Child is dead before age five			
Gamma	Q_{mh}^{+}	Q_{mh}^{-}	p_{mh}^{+}	p_{mh}^{-}	Q_{mh}^{+}	Q_{mh}^{-}	p_{mh}^{+}	p_{mh}^{-}	Q_{mh}^{+}	Q_{mh}^{-}	p_{mh}^{+}	p_{mh}^{-}
1	5.3606	5.3606	0.0000	0.0000	2.7076	2.7076	0.0034	0.0034	5.0445	5.0445	0.0000	0.0000
1.1	3.9066	6.8255	0.0000	0.0000	1.5700	3.8510	0.0582	0.0001	3.8763	6.2232	0.0001	0.0000
1.2	2.5862	8.1745	0.0049	0.0000	0.5345	4.9017	0.2965	0.0000	2.8167	7.3100	0.0024	0.0000
1.3	1.3754	9.4272	0.0845	0.0000	0.3321	5.8757	0.3699	0.0000	1.8463	8.3203	0.0324	0.0000
1.4	0.2563	10.5983	0.3988	0.0000	1.2115	6.7851	0.1128	0.0000	0.9504	9.2659	0.1710	0.0000
1.5	0.7177	11.6995	0.2365	0.0000	2.0317	7.6391	0.0211	0.0000	0.1174	10.1559	0.4533	0.0000
1.6	1.6912	12.7399	0.0454	0.0000	2.8009	8.4453	0.0025	0.0000	0.5772	10.9976	0.2819	0.0000
1.7	2.6072	13.7269	0.0046	0.0000	3.5260	9.2096	0.0002	0.0000	1.3077	11.7968	0.0955	0.0000
1.8	3.4728	14.6668	0.0003	0.0000	4.2126	9.9369	0.0000	0.0000	1.9974	12.5584	0.0229	0.0000
1.9	4.2941	15.5645	0.0000	0.0000	4.8651	10.6313	0.0000	0.0000	2.6513	13.2864	0.0040	0.0000
2	5.0760	16.4243	0.0000	0.0000	5.4873	11.2962	0.0000	0.0000	3.2734	13.9841	0.0005	0.0000

Gamma : Odds of differential assignment due to unobserved factors

Q_{mh}^{+} : Mantel-Haenszel statistic (assumption: overestimation of treatment effect)

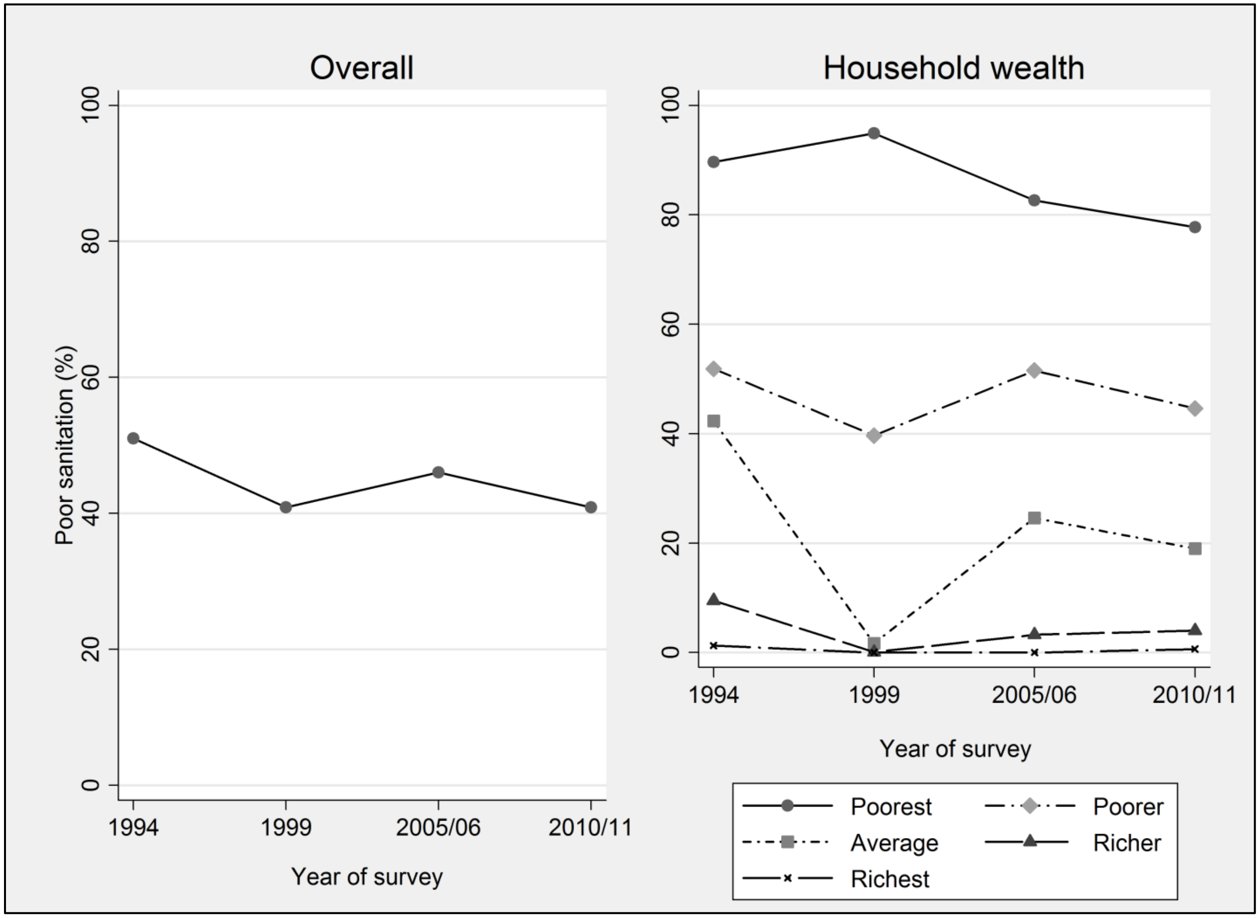
Q_{mh}^{-} : Mantel-Haenszel statistic (assumption: underestimation of treatment effect)

p_{mh}^{+} : significance level (assumption: overestimation of treatment effect)

p_{mh}^{-} : significance level (assumption: underestimation of treatment effect)

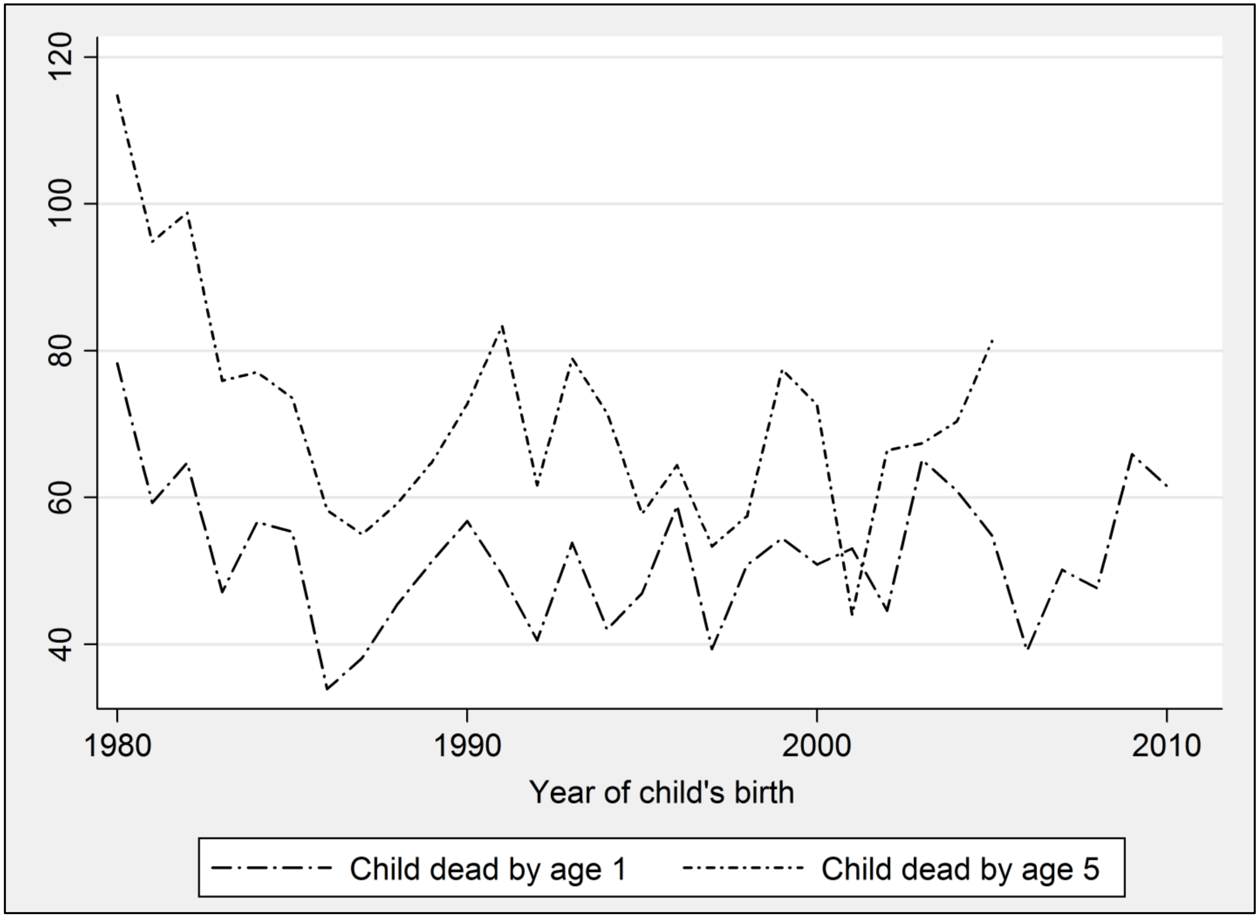
Source: MH bounds using STATA SE version 13.0

Fig. 1



Overall
Household wealth
Poor sanitation (%)
0
20
40
60
80
100
1994
1999
2005/06
2010/11
Year of survey
Poorest
Poorer
Average
Richer
Richest

Fig. 2



120
100
80
60
40
1980
1990
2000
2010
Year of child's birth
Child dead by age 1
Child dead by age 5

Fig. 3