Journal of Development Effectiveness
Vol. 3, No. 3, September 2011, 340–370

Routledge
Taylor & Francis Group

What prevents child diarrhoea? The impacts of water supply, toilets, and hand-washing in rural India

Victoria Yue-May Fan[a*] and Ajay Mahal[b]

[a]*Harvard School of Public Health, Global Health and Population, 665 Huntington Avenue, Boston, MA 02115, USA;* [b]*Monash University, School of Public Health and Preventive Medicine, Alfred Centre, 99 Commercial Road, Melbourne, Australia*

The authors apply three matching methods to estimate the impacts of water supply, toilet, and hand-washing interventions on child diarrhoea in rural India. Although propensity-score matching generally retains sample size, it can be associated with imbalance in the variables used to estimate propensity scores between treated and control groups. In contrast, exact matching is balanced over observables between treated and control units, but can result in considerable loss of observations. The authors also apply a novel 'coarsened exact matching' method that can potentially address the problem of sample attrition when matching. Their main finding using each of these three methods is that hand-washing, after defecating or before eating, significantly reduces prevalence and duration of a measure of overall diarrhea as well as acute watery diarrhoea among children under age five but not acute dysentery. In contrast, there may also be an effect of piped water on acute dysentery but not acute watery diarrhoea. Effects of improved water supply or improved toilets on different diarrhoeal outcomes are not observed consistently across matching methods.

Keywords: diarrhoea; hand-washing; water; sanitation and hygiene; infrastructure; environmental health; India

JEL classification: H54; I12

1. Introduction

Diarrhoea is the second leading cause of child mortality in the world after pneumonia, and diarrhoea alone kills more children than AIDS, malaria, and measles combined (UNICEF/World Health Organisation [WHO] 2009). The most recent global estimates for morbidity and mortality related to diarrhoea are for 2004, when nearly 2.5 billion cases of diarrhoea occurred among children aged younger than five, of which 1.5 million died (UNICEF/WHO 2009), despite the disease being both preventable and treatable. Systematic reviews and meta-analyses of interventions focused on improved water quality and supply and improved sanitation facilities suggest that these can prevent diarrhoea (Clasen *et al.* 2007, 2010, Waddington *et al.* 2009). Clasen *et al.* (2007) examined 42 studies, mostly randomised trials, in Latin America, Asia, and Africa. They concluded that interventions to improve microbial quality of drinking water can reduce the risk of diarrhoea among children under age five by about 30 to 40 per cent, and that water-quality interventions in the household were more effective than interventions at the water source

*Corresponding author. Email: vfan@alum.mit.edu

ISSN 1943-9342 print/ISSN 1943-9407 online

http://dx.doi.org/10.1080/19439342.2011.596941
http://www.tandfonline.com

in preventing diarrhoea. Another review by Waddington *et al.* (2009) examined 65 impact evaluations in 35 countries, concluding that interventions aimed at improving water *quality* are significantly more effective in reducing the risk of diarrhoeal disease compared with interventions that improve water *supply*. A recent Cochrane systematic review by Clasen *et al.* (2010) also examined the impacts of improved disposal of human excreta on occurrence of diarrhoea. Based on an analysis of 13 studies, they detected protective effects ranging from 20 to 80 per cent compared with baseline risk.

These findings suggest that although improving water quality and sanitation are effective, improving the *availability* of water supply alone without considering the *quality* of water accessed by the household is unlikely to be sufficient to prevent diarrhoea. Consider, for example, piped water, which is categorised as an 'improved water' source (WHO 2009). The availability of a piped water facility may not prevent diarrhoea if the piped water is susceptible to frequent shortages at the source, forcing households to seek out alternate, potentially 'unimproved', drinking water sources such as unprotected springs and surface water. (Similarly, if a toilet no longer functions, then a person may resort to 'unimproved' sanitation, which includes defecating in a bucket or in the open.) Moreover, even if piped water supply is uninterrupted, the water quality ingested may be poor if household containers for storing or for drinking water become contaminated over time; this argument is strengthened by the reviews by Clasen *et al.* (2007) and Waddington *et al.* (2009), who found that point-of-use treatments are more effective than those at the source.

The transmission of diarrhoea is not entirely attributable to environmental transmission through water supply and sanitation facilities. An especially important intervention for preventing diarrhoea that can be categorised as *hygiene*, rather than as water and sanitation, is individual hand-washing behaviour. The transmission of acute watery diarrhoea relies primarily on the 'faecal–oral' route via microbial agents such as viruses and bacteria that can survive on hands and hard surfaces (American Public Health Association 2008). This biological feature reinforces the fact that diarrhoea is not only transmitted through contaminated drinking water. This feature is particularly relevant in cultures such as those in India where large segments of the population eat directly with their hands. It is possible for microbial concentration to be higher on a dry surface compared with contaminated water (Cash 2011). Indeed, to contract diarrhoea, a person would need to ingest large amounts of water to be exposed to a pathogen, or ingest water with extremely high microbial concentration (Cairncross *et al.* 2010).

A Cochrane systematic review by Ejemot *et al.* (2008) examined 18 randomised trials of hand-washing on diarrhoea and found that interventions promoting hand-washing resulted in a 32 per cent reduction in diarrhoeal episodes in children living in low-income or middle-income countries, a conclusion supported by other reviews. Waddington *et al.* (2009) assessed that interventions that promote hand-washing behaviour and improve sanitation are both effective, although they were concerned about changing compliance rates over time. A review by Zwane and Kremer (2007) also lends support to the protective effects of point-of-use water treatments, some evidence of the effectiveness of piped water and sanitation infrastructure, and little evidence of the impact of communal rural water infrastructure. Finally, a systematic review by Cairncross *et al.* (2010) examined the effects mainly on diarrhoeal morbidity of three interventions: hand-washing with soap, drinking-water quality, and excreta disposal. They estimated diarrhoea risk reductions of 48 per cent for hand-washing with soap, 17 per cent for water treatment, and 36 per cent for sanitation, although they judge the available evidence to be of poor quality in general.

Despite the large numbers of analyses included in these reviews, they suffer from important weaknesses that affect both their validity and reliability. Most of the experimental studies reported in these reviews of water quality, toilets, or hand-washing on diarrhoea are characterised by small sample sizes, a focus on short-term outcomes, and limited or no compliance information. As Cairncross *et al.* (2010, p. i194) write: ‘There is a strong temptation to conduct evaluations of the health impact of water supply, sanitation, and hygiene interventions, but the challenges also are many’. Given the practical difficulties involved in undertaking carefully designed randomised experiments relating to water, sanitation, and hygiene interventions, an alternative is to examine the impact of interventions using quasi-experimental approaches, such as matching methods, on cross-sectional household survey data (Imai *et al.* 2008). Such observational data exist in many countries with a sample size an order (or orders) of magnitude greater than any individual study reported in the Cochrane reviews. Surprisingly, Waddington *et al.* (2009) found four studies that used propensity-score matching to evaluate the impacts of water, sanitation, and hygiene (Pradhan and Rawlings 2002, Jalan and Ravallion 2003, Khanna 2008, Bose 2009). Pradhan and Rawlings (2002), Jalan and Ravallion (2003) and Khanna (2008) evaluated the impacts of *water supply* on child diarrhoea; the first two found a significant impact of water supply on child diarrhoea, whereas Khanna (2008) did not. Pradhan and Rawlings (2002) and Bose (2009) also evaluated the impacts of *sanitation* on child diarrhoea; the former did not find a significant effect, whereas the latter found that it significantly reduced child diarrhoea.

In this paper, we go beyond these earlier quasi-experimental studies by using three different matching methods – propensity-score matching (Rosenbaum and Rubin 1983), exact matching and a novel ‘coarsened exact matching’ method (Iacus *et al.* 2011a) – to estimate the impacts of water supply, toilet, and hand-washing interventions, on the prevalence of child diarrhoea in rural India. Choosing between exact matching and propensity-score matching reflects a bias-efficiency trade-off (Ho *et al.* 2007, Imai *et al.* 2008). Exact matching methods do not require balance checking, but often lose many observations when matching on many variables or dimensions. In contrast, propensity-score matching tends to retain sample size but can lead to biased results without balance checking. Coarsened exact matching seeks a middle ground between the two approaches (Iacus *et al.* 2011b).

Our reason for choosing India for investigation is that it is the country with the highest number of both child cases and deaths from diarrhoea, with an estimated 237,482 deaths from diarrhoea in 2008 (Black *et al*. 2010). Moreover, concerns about access to water and sanitation facilities persist (see Figure 1). In 1990 it was estimated that just 7 per cent of India’s rural population had access to an improved sanitation facility and 66 per cent had access to an improved water source. Fifteen years later, these shares have increased to 18 per cent for sanitation and 81 per cent for improved water, but with a population of 1.2 billion people several hundred million people in rural India are still unable to adequately access water and/or sanitation (UNICEF/WHO 2009).

India also has good data on this subject from multiple surveys. For this paper, we rely on a large observational rural household survey undertaken by the National Council for Applied Economic Research (NCAER) in India during 1994. The survey data solicited information on source of water supply, type of toilet, and hand-washing behaviours, whose effects we are interested in, as well as a variety of village and household socio-economic characteristics. In doing so, we are also able to contrast our approach to matching with an earlier study that used the same dataset but that used only propensity-score matching and studied only the effects of piped water (Jalan and Ravallion 2003). This study concluded that piped water significantly reduced prevalence of under-five child diarrhoea by

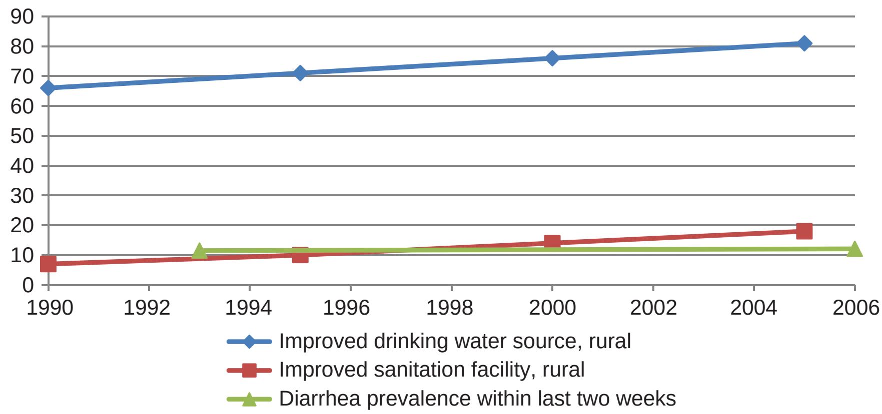


Figure 1. Trends in availability of improved water supply, sanitation facility and diarrhoea prevalence in India, 1990–2006.

Source: WHO (2011) and Macro International (2011).

21 per cent, which contrasts with findings from systematic reviews that tend to emphasise the role of water quality, hand-washing and sanitation. Apart from assessing the impacts of water, sanitation, and hygiene interventions in addition to piped water, our study builds on the Jalan and Ravallion (2003) framework in several ways. Two in particular are important. In our study, we conduct balance diagnostics according to best practice for propensity-score matching, at both individual and household levels by calculating a newly developed measure of imbalance, L1 (see Iacus *et al.* 2011a). We pursue multiple matching methods and thus assess and compare the robustness of results from propensity-score matching with those from exact matching and coarsened exact matching.

Our main conclusion is that hand-washing, after defecating or before eating, significantly reduces prevalence and duration of acute watery diarrhoea among children under age five but does not affect acute dysentery. There may be an effect of piped water on acute dysentery but not on acute watery diarrhoea, which is consistent with the study by Jalan and Ravallion (2003). No significant effect of improved water supply (including piped water) or improved toilets on different diarrhoeal outcomes is observed consistently across matching methods.

2. Matching methods and observational data

Our analysis relies on the 1994 Human Development Index survey, undertaken by India's NCAER. This survey collected information on health status, hand-washing behaviours, education, and other variables for 194,398 individuals including 25,117 children under age five living in rural areas. The survey collected data on socio-economic variables, source of water supply, and type of toilet for 33,230 households including 16,245 households with children under age five, and village-level variables such as the presence of village development (for example, schools, health facilities, road quality) for 1762 villages in 16 major states of India. The survey was conducted during January through June 1994 (Shariff 1999). The survey collected a rich set of data at the village, household and individual levels, which were essential for this study. By contrast, the recently collected National Family and Health Survey (2005/06) data did not include detailed village-level information or information on hand-washing behaviours (although it collected information on handling a child's stools). Periodically collected National Sample Survey data on health status and health-seeking behaviour also do not include this information. The NCAER data were collected using a multistage sampling design, which involved the construction of region-level strata by agricultural income and female literacy rates, followed by selection of districts and villages, from which the households were sampled.

In observational data of the kind collected in the NCAER survey, assignment of treatment is not randomised. Thus, the treated and untreated are likely to differ in characteristics other than the treatment; that is, there may be imbalance in characteristics (both observed and unobserved) between the treated and untreated groups. Matching is an attempt to construct a sufficiently plausible counterfactual using information on observables and thereby address the problem of selection bias or confoundedness. We use three matching methods – propensity-score matching (PSM), exact matching, and coarsened exact matching (CEM) – to estimate the effects of improved water supply, improved toilets, and hand-washing on prevalence and duration of child diarrhoea in rural India. These methods are applied to assess six 'binary' treatments: piped water; tube-well; improved water supply; improved toilet; hand-washing done after defecation or handling stools, which we call 'primary hand-washing'; and hand-washing done before eating, feeding, or preparing food, which we call 'secondary hand-washing'. We used the WHO definition of 'improved water supply' and 'improved toilet' (WHO 2009) and the categorisation of primary and secondary hand-washing as described by Curtis *et al.* (2000).

2.1. Matching methods

PSM involves a cyclical two-stage process: estimating propensity scores for treated units and control units (sometimes referred to as the 'first stage') with pre-treatment covariates in a logit (or probit) model, followed by matching treated and control units to each other with similar propensity scores (the 'second stage') using a matching algorithm such as 'nearest neighbour'. In contrast, exact matching and CEM match directly on pretreatment covariates and not on an intermediate score. In our study, we assume the placement of treatments of interest to be a function of village-level and household-level (rather than child-level) covariates that are plausible confounders of treatment, whereas the outcome of interest, a diarrhoea episode and its duration, is at the individual child level. One set of propensity scores were estimated at the household level and another set at the individual child level.[1] We use *only* the 16,245 households with children under age five as they are the population at risk of under-five child diarrhoea.

In the 'second stage', matching between the treatment and controls was undertaken at both the household level and the individual level, using the nearest-*n* neighbour matching algorithm without replacement. The nearest-*n* neighbour algorithm takes the average outcome measure of the closest *n* number of matched controls as the counterfactual for each treated unit. Conducting nearest-five neighbour matching, for example, depends on having at least five controls to one treated case.

A major challenge in implementing PSM is that it requires that the treated and control units matched on the (nearest) propensity score are balanced on pretreatment covariates. Current best practice in PSM requires the researcher to generate propensity scores, match observations on the scores, and then check balance, iteratively continuing until balance in the matched data is maximised, by pruning the data and thereby excluding observations for matching (Ho *et al.* 2007, Imai *et al.* 2008). In general, researchers can prune observations using the propensity score in at least two ways: by applying 'common support' and by using a calliper to match only neighbours within a certain 'distance'. The term 'common support' has been generally used to refer to the exclusion of control observations higher or lower than the range of scores for treated observations. In the case of a calliper, matches are only made within a certain distance to the propensity score. In PSM conducted in this paper, both common support and a calliper are applied for nearest-neighbour PSM.

However, pruning through either common support or a calliper does not guarantee a balanced matched sample under PSM, and thus the researcher must still check for balance after matching. One consequence of this practical challenge is that researchers commonly neglect to undertake or report balance diagnostics in applications of PSM (Ho *et al.* 2007, Imai *et al.* 2008). (This iterative, cyclical nature of PSM is masked when referring to it as a two-stage process.) Even when balance checks have been undertaken in the literature, they have often taken the form of hypothesis testing to compare means of pretreatment variables of the treated group and control group in the matched sample. However, Ho *et al.* (2007) and Imai *et al* (2008) argue that one should ideally compare the joint distribution of all covariates for the matched treatment and control units, and maximise balance as much as possible. In the past it has been relatively difficult to accurately measure balance under high dimensionality, that is, with many covariates to match on. Ho *et al.* (2007) suggested using lower-dimensional balance diagnostics, including calculating the differences in mean in the matched sample and relative improvement in balance over the full unmatched sample. This crude indicator of differences in means for a given variable or distance measure (for example, the propensity score) between the matched and full sample can still mask imbalances between matched pairs. Thus a recent paper by Iacus *et al* (2011a) proposed a new summary measure of multivariate imbalance, L1, and another paper by King *et al.* (2011) recommends comparing the effectiveness of different matching methods by choosing the matched sample with the smallest level of imbalance. We conduct these best practices in balance diagnostics by conducting multiple iterations in PSM specifications of pretreatment variables, varying the size of the calliper, and applying common support and in particular calculating the summary L1 measure (with further description of these diagnostics in the next section).

Two alternatives that do *not* require iterative balance checking are exact matching and CEM (Iacus *et al.* 2011b). Both of these methods match exactly on covariates and not on an estimated propensity score. Unlike PSM, exact matching methods have the advantage of *not* requiring any balance checking since units are matched exactly on all included covariates and not on a one-dimensional score. However, exact matching is handicapped by the smaller number of available units to match as the number of variables to match on increases. CEM is a variation on exact matching that seeks to address this concern; the procedure matches observations *exactly*, but only after 'coarsening'; that is, manually creating categories for continuous variables or creating coarser categories from finer categories, that are intuitive or natural (Iacus *et al.* 2011b). For example, one could consider the following 'natural' coarsening on education levels. Rather than using continuous years of education of a household member, one could use a level of education such as primary school, middle school, high school, and college or higher.

After applying each matching method, we take a sample weighted difference in outcome means to calculate the 'average treatment effect on the treated' in each case. Ho *et al.* (2007) argue that matching is a means to non-parametrically pre-process a dataset before parametric analysis, thereby making estimates less dependent on (parametric) modelling choices and specifications. Impact estimates should not vary much even when changing parametric modelling assumptions, and in many cases matching also reduces variance of the estimated causal effects. Thus, as a specification check after conducting PSM, we conducted parametric analysis by including potential confounding variables to account for any potentially remaining imbalances and compare those results with that from taking a simple difference of means.[2]

The key assumptions underlying any matching method are that treatment is uniform in implementation for all units and that there are no spillovers. It is not obvious that the treatment variables, especially those of water supply and sanitation facility, are uniform across households or villages. For instance, the quality and reliability of supply of water may vary across regions. As long as the variation in the interventions is not large, or if the variation is due to a characteristic that is irrelevant to the health outcome (for instance, depth of clean groundwater), this may not be a handicap. The risk of spillovers is expected to be small given the relatively small number of households sampled in each village (on average about 7 per cent of all households in each sampled village) and the small number of villages (about 1760 villages) randomly sampled from a total of nearly 565,000 Indian villages.

2.2. Measurement of treatment variables

Our water supply treatment variable is constructed from two questions posed to survey respondents: one on the main source of drinking water in the *non-summer* season, another on the main source of drinking water in the *summer* season. Respondents selected one of the following: ponds; dug well; running stream/canals; protected wells; tanker truck; piped outside the house; piped in the house; hand pump; others. We define piped water as any piped water in any season.[3] Using the WHO definition of improved water (WHO 2009), we construct 'improved water supply' as a binary variable that takes the value one for households using protected wells, tanker truck, piped water (inside or outside), and hand pump; and 'not improved' (or zero) if the main source was a pond, dug well, running stream, or others.[4] Our measure of sanitation relied on the type of toilet in the household (manual disposal, septic tank, drainage, water-sealed pits, or others). Using the definition of improved sanitation in WHO (2009), we define an 'improved toilet' as one with a septic tank, drainage, or water-sealed pits. Finally, measures of individual hand-washing behaviours are based on seven distinct questions on the sequence of hand-washing in relation to key events affecting disease transmission: after using the latrine or defecating; after cleaning a child's stools; after disposing child's stools; before cooking; before feeding food; before serving food; and before eating food. We use measures of hand-washing as a primary barrier or as a secondary barrier to diarrhoea depending on the sequencing of hand-washing (Curtis *et al.* 2000). 'Primary hand-washing' is defined as hand-washing done after contact with contaminated material from defecation or stools, and 'secondary hand-washing' is hand-washing done before activities related to preparing, serving, or consuming food. Although the hand-washing treatments are reported at the individual level, it is reasonable to assume that hand-washing behaviours of children are a function of household-level pre-treatment variables if we assume that hand-washing by (or of) children is a behaviour learned within the household (see, for example, Luby *et al.* 2004, which promoted hand-washing with soap for children as young as 30 months old). An alternative interpretation of this treatment is that it may also reflect the hand-washing behaviours of the survey respondent (such as the child's mother).

Table 1 reports the distribution of households by treatment status. Nearly one-quarter (23.4 per cent) of the sampled households with children under age five (that is, the households of interest) had access to piped water in any season. Whereas 72.9 per cent of households of interest had access to improved water, only 11.0 per cent of households had access to an improved toilet. Self-reported primary hand-washing and secondary

Table 1. Distribution of households or individuals by treatment status.

Treatment	All households	Household subset: With children under age five	Household subset: Without children under age five
Number of households	33,216	16,245	16,971
Number of individuals	194,398	25,117	169,281
Household water supply			
Piped water	25.8	23.4	27.8
Improved water supply	73.4	72.9	73.9
Tube-well (asset)	9.1	10.0	8.4
Household sanitation facility			
Improved toilet	12.3	11.0	13.3
Individual hand-washing			
Primary hand-washing	69.3	21.5	75.9
Secondary hand-washing	69.8	27.5	75.6
Household income quintile			
Lowest	20.0	24.8	16.0
Second	20.0	23.4	17.1
Third	20.0	19.9	20.1
Fourth	20.0	17.8	21.8
Highest	20.0	14.1	24.9
Total	100.0	100.0	100.0
Highest education of female in household			
Illiterate	47.0	48.0	46.1
Primary	29.4	30.3	28.6
Matriculation at most	19.8	18.2	21.2
Higher secondary or more	3.8	3.5	4.0
Total	100.0	100.0	100.0

Note: Of the 33,230 households in the household dataset, 14 households could not be matched to a village, hence reducing the total number of all households to 33,216. Household weights were used here and throughout the paper. Piped water and improved water supply both refer to that in any season, not all seasons. The hand-washing variables (primary or secondary) here are individual, rather than household, prevalence.

hand-washing were lower among those under age five (21.5 per cent and 27.5 per cent, respectively) compared with primary and secondary hand-washing among those aged five and higher (75.9 per cent and 75.6 per cent, respectively).[5]

2.3. Measurement of diarrhoea

The survey included several questions to elicit short-duration diarrhoeal morbidity experienced by household members in the past month and its duration. Diarrhoea was categorised in six forms: acute watery diarrhoea, acute dysentery, persistent diarrhoea, food poisoning, parenteral diarrhoea, and chronic diarrhoea. We construct three diarrhoea variables: 'all diarrhoea' (which includes any one of the six forms reported), acute watery diarrhoea, and acute dysentery. Acute watery diarrhoea was defined as three or more loose motions without blood in stools, lasting for 14 days or less. Acute dysentery was defined as three or more loose motions with blood mixed with stools. We use only the first reported episode of the illness in the reference period. That is, a child was defined as having the illness if the child had a first episode of diarrhoea regardless of whether there was a second episode. Given the reference period of one month (for reporting illnesses), the number of second (or more) episodes was small and would not affect our conclusions.

Table 2. Crude prevalence or duration of diarrhoea among children by treatment status.

Treatment type	Treatment status	Prevalence by type of diarrhoea: All diarrhoea	Acute watery diarrhoea	Acute dysentery	Duration of diarrhoea (days)
Overall sample		0.1086	0.0821	0.0095	0.5873
Piped water	Untreated	0.1124	0.0855	0.0110	0.6102
	Treated	0.0965	0.0714	0.0049	0.5136
Improved water supply	Untreated	0.1127	0.0844	0.0105	0.6044
	Treated	0.1071	0.0813	0.0092	0.5809
Tube-well	Untreated	0.1087	0.0823	0.0090	0.5845
	Treated	0.1083	0.0812	0.0137	0.6101
Improved toilet	Untreated	0.1117	0.0848	0.0098	0.5957
	Treated	0.0829	0.0605	0.0070	0.5181
Primary hand-washing	Untreated	0.1189	0.0928	0.0096	0.6418
	Treated	0.0710	0.0431	0.0091	0.3876
Secondary hand-washing	Untreated	0.1233	0.0972	0.0093	0.6699
	Treated	0.0700	0.0424	0.0102	0.3699

Note: Piped water and improved water supply refer to those treatments in any season.

Table 2 reports the crude prevalence of diarrhoeal disease, conditional on each of the water supply, toilet, or hand-washing treatments. Of these three categories of diarrhoea, acute watery diarrhoea was the most prevalent (8.2 per cent). The prevalence of acute dysentery was 1 per cent. Prevalence of all diarrhoea was 10.9 per cent. Among children under five in households with piped water, prevalence of all diarrhoea was 9.7 per cent compared with 11.2 per cent among those without piped water. By contrast, the difference in the crude prevalence and duration between those with improved toilet and those without was more accentuated. Among children with an improved toilet, prevalence of all diarrhoea was 8.3 per cent compared with 11.2 per cent among those without. Children with an improved toilet also had a shorter duration than those without (0.519 days compared with 0.642 days). Even larger crude differences in outcomes are observed for hand-washing. Among those who practiced secondary hand-washing, prevalence of all diarrhoea was 7.0 per cent compared with 12.3 per cent among those who did not. Duration of diarrhoea among those with secondary hand-washing was also much shorter compared with those without (0.370 days compared with 0.670 days).

These estimates are almost identical to those reported in the official NCAER *Human Development Report* based on the same dataset that we use (Shariff 1999) and are also comparable with data from a separate report based on the National Family Health Survey, 1992/93 (Macro International 2011). Shariff (1999) reports a prevalence of under-five child diarrhoea of 10.8 per cent, whereas our estimate of overall prevalence is 10.86 per cent. Similarly, Macro International (2011) estimated a diarrhoea prevalence of 11.5 per cent among children under age five in the past two weeks over 1992/93. Jalan and Ravallion (2003) in Table 3 reported the prevalence of diarrhoea to be 1.08 per cent among children under five living in households with piped water. It is possible that Jalan and Ravallion (2003) made a simple error by shifting the decimal place to the left throughout their results, or alternatively are referring to acute dysentery which indeed has a prevalence

Table 3. Summary of balance for treatment of piped water.

Method	Model	Model 1	Model 2	Model 3	Model 4
Exact matching	Untreated	16,816	10,566	1170	665
	Treated	6871	5420	938	528
CEM	Untreated	17,040	11,146	2042	1173
	Treated	6871	5462	1477	815
	L1	0.964	0.951	0.715	0.428
	%BI	99.6	99.6	99.4	99.7
PSM with narrowing callipers					
PSM, calliper 0.25	Untreated	6302	6332	6337	6281
	Treated	6302	6332	6337	6281
	L1	0.968	0.969	0.963	0.966
	%BI	88.6	87.1	86.4	86.8
PSM, calliper 0.1	Untreated	5977	5953	5995	5965
	Treated	5977	5953	5995	5965
	L1	0.966	0.966	0.965	0.967
	%BI	97.2	95.9	95.6	95.5
PSM, calliper 0.001	Untreated	5587	5006	4934	4955
	Treated	5587	5006	4934	4955
	L1	0.967	0.957	0.958	0.96
	%BI	100.0	100.0	100.0	100.0
PSM, calliper 0.0001	Untreated	5560	4257	2222	2156
	Treated	5560	4257	2222	2156
	L1	0.964	0.95	0.856	0.848
	%BI	100.0	100.0	100.0	100.0
PSM, calliper 0.00001	Untreated	5560	4115	919	650
	Treated	5560	4115	919	650
	L1	0.963	0.947	0.589	0.368
	%BI	100.0	100.0	100.0	100.0

Note: L1 here refers to multivariate L1, with zero indicating perfect balance and one indicating perfect imbalance. In contrast, %BI refers to 'percentage balance improvement' in distance measure and approaches 100% to indicate better balance. Model 1 includes state, household asset index (quintile), village infrastructure index (quintile). Model 2 includes household income (quintile) and highest education of female in addition to Model 1 variables. Model 3 includes scheduled caste, scheduled tribe, Muslim household, household size; household head is self-employed, household head is married in addition to Model 2 variables. Model 4 includes electricity, improved stove, and chimney in addition to Model 3 variables.

of 1 per cent in the under-five population, close to their estimate. (See the next section on balance diagnostics for description of our replication of the results by Jalan and Ravallion 2003.)

2.4. Covariates for matching

There were several potential pretreatment variables for the first-stage PSM model available in the dataset.[6] We briefly describe two important kinds of sets of variables used for matching – household wealth and village-level infrastructure – as they are likely to be important predictors of the treatment variables. An index of household assets was constructed using principal component analysis for the full sample of households (Filmer and Pritchett 2001). Separately, we constructed an index of village infrastructure by using information on the different facilities available in each village. In previous work, village wealth indices have mostly been constructed by averaging the wealth of households living in a village (Pritchett 2010), presumably because of the lack of detailed data on village infrastructure (which are available in the NCAER dataset). Village-level facility data include whether the village has an *Aaganwadi* (daycare) centre, a primary school and other schools, library, bank, market, and so forth. The construction of the household asset index and the village infrastructure index as summary measures of several variables also made exact and coarsened exact matching feasible. The highest education of the female in the household variable was coarsened to four categories (Table 1). Variables that were plausibly pre-treatment in the causal pathway were included, whereas variables that were potentially post-treatment were excluded in order to avoid post-treatment bias.[7] (See Tables 3 and 4 in the next section for list of covariates used in matching specifications.)

3. Balance diagnostics

In this section we consider the trade-off between balance and sample size in pretreatment variables in the matched samples from the three matching methods. In the context of the PSM model, this means following an iterative process of propensity-score estimation, matching and balance checking, as suggested by Gary King and his colleagues (Ho *et al.* 2007, Iacus *et al.* 2011a, 2011b, King *et al.* 2011). Because balance on characteristics can potentially worsen in the sample matched on the propensity score, the process should ideally be repeated until the matching procedure suggests maximum balance, at the cost of pruning observations. By contrast, exact matching and CEM are always balanced, but sample attrition can be large with many matching variables, particularly continuous variables.

The main measure of imbalance presented in this paper is the multivariate imbalance L1 measure, which offers a relatively intuitive interpretation '[For any given set of bins], if the two empirical distributions are completely separated, then $L1 = 1$; if the distributions exactly coincide, then $L1 = 0$' (see Iacus *et al.* 2011a, p. 352). This measure can be considered a common support restriction along a multi-dimensional histogram of multiple variables, not just a single variable or dimension such as the aggregate propensity score. We also constructed a measure, 'percent balance improvement', which is defined as $100(|A| - |B|)/|A|$, where (A) is the distance between the means of the treated and untreated groups in the unmatched sample and (B) is the distance between the means in the matched sample (Ho *et al.* 2007). In addition we considered percentage improvement in the propensity score for PSM and the L1 measure for CEM.

The three matching methods for each treatment were executed using MatchIt (Ho *et al.* 2011) and CEM (Iacus *et al.* 2011a) in R 2.10. For PSM we applied common support and started with a generous calliper of 0.25 standard deviations in propensity score. Nearest-one neighbour PSM was conducted for every treatment; nearest-five neighbour PSM was not possible for most treatments given a ratio of control to treated units of less than five (see Table 1).

In Table 3, we use piped water as an example to present a summary of balance in matched sample size across the three matching methods (exact matching, CEM, and PSM), which are shown in the first three sets of rows in the table. PSM retained a larger sample size than CEM, which in turn retained a larger sample size than exact matching. As one moves down a given column across rows in Table 3, we apply an increasingly narrow calliper for PSM as a sensitivity analysis. When using PSM, one should prune the data manually; for example, through 'common support' or applying a calliper or other procedure. In contrast, exact matching methods automatically prune the data by virtue of matching exactly and discarding those not matched. The pruning from general common support in PSM is generally limited. Table 3 shows that, by narrowing the calliper used for PSM, fewer units are matched while balance improves. (The robustness of impact estimates to calliper tightening and data pruning are shown in Section 6.)

In addition, Table 3 also presents summary measures of balance when increasing the number of matching variables (moving right in a given row from Models 1 to 4). Exact matching and CEM led to reductions in matched sample size when adding more variables to match on. In contrast, PSM retained sample size but did not balance as well on the matching variables. The final covariate specification presented in this paper was the result of successively adding four sets of variables that were plausibly pretreatment in the causal pathway. The four sets of variables used were: village development index (quintile), household asset index (quintile), and state; household income per capita and highest education of a female member in the household; demographic characteristics of scheduled caste, scheduled tribe, Muslim household, household size, whether the household head is married or not, and whether the household head is self-employed or not; and household 'infrastructure' that is, electricity, clean stove, and available chimney – especially as electricity is highly correlated with piped water, as shown in Appendix 2). Although impact estimates presented in the next section use Model 4, estimates from Models 1–3 were also calculated. As a sensitivity check, we also included in some specifications (results not presented) piped water, tube-well and improved toilet as matching variables for hand-washing given that water supply may affect whether one chooses to wash one's hands.

For the six treatments, we present balance diagnostics in Table 4 for the fourth and final covariate specifications using PSM and CEM. The table suggests that samples matched on hand-washing interventions had better balance than samples matched on other water and sanitation interventions. Moreover, for every intervention, CEM produced better balance than PSM.

In addition to our main matching specifications and analyses, we replicated the Jalan and Ravallion (2003) results using their pretreatment variables as well as their PSM specification of matching on the odds ratio of the propensity score within a calliper of 0.001. The first-stage model coefficients in our replication were highly comparable with their study, which used all 33,000 households including those *without* children (see Appendices 2 and 3). For our main analyses and for the second-stage of this replication (when matching units by the generated propensity score), we restricted the sample to only households with children who thus were at risk of the disease. After matching, imbalance remained on several variables between treated and control groups (see Appendix 4). As shown in Appendix 5, we estimated an effect of –0.019 of piped water on all diarrhoea, –0.019 on acute watery diarrhoea, and –0.006 on acute dysentery, all highly significant at $p < 0.05$. In comparison, Jalan and Ravallion (2003) found a piped water effect of –0.023 (if we assume a uniform shift in the decimal place, and thus their diarrhoea prevalence estimate of 1.08 per cent would correspond to the correct estimate of 10.8 per cent). Thus although we are able to replicate their results quite closely in terms of the outcome prevalence and the significance level, as we show next, the effects on every form of diarrhoea are not robust.

4. Main results and impact estimates

Impact estimates at the individual child level for water supply treatments on diarrhoeal outcomes are reported in Table 5 and estimates for toilet and hand-washing treatments in Table 6. Given limitations of space, the results presented use the fourth and final model; results with the three matching specifications were consistent (see Tables 3 and 4 for variables in the four covariate specifications). Impact estimates for exact matching and CEM are the simple difference in the outcomes between treated and controls. After PSM, no apparent differences in estimates were observed between simple differences (shown in table) or after parametric adjustment (data not shown), suggesting that estimates are 'doubly robust' from matching or from parametric adjustment (Ho *et al.* 2007).

Table 4. Summary of balance on matching variables.

	PSM						CEM					
Measure	Piped water	Improved water	Tube-well	Improved toilet	Primary hand-washing	Secondary hand-washing	Piped water	Improved water	Tube-well	Improved toilet	Primary hand-washing	Secondary hand-washing
Control	6281	6329	2562	2221	4779	6154	1173	1314	557	362	4258	4740
Treated	6281	6329	2562	2221	4779	6154	815	1458	429	238	2940	3587
Multivariate L1	0.968	0.944	0.971	0.987	0.804	0.756	0.428	0.419	0.485	0.412	0.130	0.128
Univariate L1												
State	0.039	0.082	0.030	0.022	0.016	0.016	0.020	0.006	0.016	0.017	0.006	0.002
H. asset index	0.006	0.023	0.020	0.038	0.009	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Village index	0.013	0.061	0.007	0.009	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000
Income quintile	0.011	0.038	0.005	0.012	0.002	0.006	0.000	0.000	0.000	0.000	0.000	0.000
Female education	0.021	0.053	0.002	0.022	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Scheduled tribe	0.002	0.040	0.003	0.002	0.004	0.009	0.000	0.000	0.000	0.000	0.000	0.000
Scheduled caste	0.001	0.006	0.009	0.014	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Muslim	0.006	0.020	0.016	0.025	0.004	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Household size	0.042	0.045	0.053	0.054	0.059	0.042	0.080	0.047	0.114	0.059	0.018	0.013
Household head: self-employed	0.001	0.012	0.007	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Household head: married	0.001	0.000	0.007	0.001	0.005	0.003	0.000	0.000	0.000	0.000	0.000	0.000
Electricity	0.027	0.067	0.008	0.032	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Improved stove	0.016	0.018	0.004	0.018	0.001	0.009	0.000	0.000	0.000	0.000	0.000	0.000
Chimney	0.000	0.050	0.000	0.014	0.004	0.004	0.000	0.000	0.000	0.000	0.000	0.000

Note: L1 here refers to multivariate L1, with zero indicating perfect balance and one indicating perfect imbalance.

The results in Table 5 indicate that there may be an effect of piped water on acute dysentery but not other diarrhoeal outcomes. Using PSM and CEM, there was a significant effect at $p < 0.05$ of piped water on acute dysentery and using exact matching at $0.05 \leq p < 0.1$. As noted earlier, in replicating the analysis by Jalan and Ravallion (2003), we found that piped water results in a negative effect on acute dysentery at the 5 per cent level of significance. In considering improved water supply and tube-well as a treatment, no significant effect on a given outcome is observed consistently across matching methods. When using only CEM (and not other methods), improved water supply reduces prevalence of acute dysentery by 0.009 ($p = 0.039$); that is, 0.9 percentage points compared with 1.8 per cent in controls in this matched sample.

We next consider treatments of improved toilets and hand-washing in Table 6. An effect of improved toilet on acute dysentery was detected only at $0.05 \leq p < 0.1$ using CEM but not other methods. When using exact matching, improved toilets appear to significantly reduce prevalence of all diarrhoea and its duration, but the sample suffered from high attrition.

In sharp contrast, hand-washing appears to have a large, negative, and highly significant effect ($p < 0.0001$) as estimated from every matching method on 'all diarrhoea', acute watery diarrhoea, and diarrhoeal duration. This contrasts quite starkly with the detected effects of water supply and sanitation for which the significance is both small and not robust across methods. Moreover, sample size reductions were lower across all methods when matching for hand-washing, compared with that for water and sanitation treatments. The effect sizes detected are consistent with the experimental literature. Using PSM, primary hand-washing reduced acute watery diarrhoea by 0.058 and was highly significant ($p < 0.0001$) or a 58 per cent reduction. Using PSM, secondary hand-washing reduced all diarrhoea by 0.059 ($p < 0.0001$). Although the effect sizes under exact matching and CEM are somewhat larger than those when using PSM, exact matching and CEM had higher sample attrition and retained just a fraction of the treated units, whereas PSM retained most treated units.

Although hand-washing appeared to reduce prevalence of acute watery diarrhoea and diarrhoeal duration, no effect of hand-washing was detected on acute dysentery using any matching method. Diarrhoeal duration was reduced by hand-washing by between one-third and more than one-half of a day, depending on the matching method used. Primary hand-washing reduces duration of diarrhoea by more than one-half a day (0.516, $p < 0.0001$) when using exact matching, whereas it reduces duration by 0.317 days when using PSM.

These results are consistent with the crude correlations observed in Table 2, which suggested a negative correlation between improved toilet on all diarrhoea and diarrhoeal duration as well as a negative correlation between hand-washing and different diarrhoeal outcomes. We also conducted analyses of household-level treatments at the household level; results are consistent with those at the individual child level (results not presented).

5. Sensitivity analyses and subgroup analyses

In this section we describe and present some of the sensitivity analyses conducted. There is always the chance that the main analyses failed to detect a significant effect when a true effect exists or detected a false positive. As mentioned earlier, a simple sensitivity check is to compare estimates between using simple differences and using parametric adjustment after matching; we found our estimates 'doubly robust' (Ho *et al.* 2007). Another sensitivity analysis relates to the analyses in Table 5, which analysed piped water relative

Table 5. Impact estimates of water supply on diarrhoeal outcomes in children.

		Exact		CEM		PSM	
Treatment	Outcome	Const.	Impact	Const.	Impact	Const.	Impact
Piped water	All diarrhoea	0.131	−0.011	0.142	−0.019	0.099	0.000
		(0.013)	(0.019)	(0.010)	(0.016)	(0.004)	(0.005)
	Acute watery diarrhoea	0.105	−0.013	0.107	−0.013	0.074	−0.001
		(0.012)	(0.018)	(0.009)	(0.014)	(0.003)	(0.005)
	Acute dysentery	0.014	−0.011	0.019	−0.015	0.009	−0.004
		(0.004)	(0.006)†	(0.003)	(0.005)*	(0.001)	(0.001)*
	Diarrhoeal duration	0.710	−0.167	0.714	−0.139	0.552	−0.027
		(0.081)	(0.123)	(0.058)	(0.093)	(0.028)	(0.040)
	Control *n* (of 18,006)	665		1173		6281	
	Treated *n* (of 6905)	528		815		6281	
	L1	n.a.		0.428		0.968	
	% BI	n.a.		99.67		86.6	
Improved water supply	All diarrhoea	0.148	−0.008	0.157	−0.021	0.114	−0.004
		(0.012)	(0.017)	(0.010)	(0.013)	(0.004)	(0.006)
	Acute watery diarrhoea	0.117	−0.004	0.118	−0.011	0.085	0.000
		(0.011)	(0.015)	(0.009)	(0.012)	(0.003)	(0.005)
	Acute dysentery	0.018	−0.004	0.018	−0.009	0.011	−0.003
		(0.004)	(0.006)	(0.003)	(0.004)*	(0.001)	(0.002)
	Diarrhoeal duration	0.747	0.089	0.847	−0.092	0.607	0.008
		(0.102)	(0.140)	(0.078)	(0.106)	(0.029)	(0.041)
	Control *n* (of 18,384)	871		1314		6329	
	Treated *n* (of 6527)	931		1458		6329	
	L1	n.a.		0.419		0.944	
	% BI	n.a.		99.94		68.77	
Tube-well	All diarrhoea	0.148	0.010	0.138	0.007	0.109	0.004
		(0.020)	(0.031)	(0.014)	(0.023)	(0.006)	(0.009)
	Acute watery diarrhoea	0.116	0.002	0.099	0.016	0.084	0.001
		(0.018)	(0.027)	(0.013)	(0.020)	(0.006)	(0.008)
	Acute dysentery	0.005	0.013	0.004	0.008	0.011	0.004
		(0.006)	(0.009)	(0.003)	(0.005)	(0.002)	(0.003)
	Diarrhoeal duration	0.667	0.235	0.724	0.141	0.652	−0.017
		(0.119)	(0.183)	(0.100)	(0.156)	(0.053)	(0.074)
	Control *n* (of 21,913)	313		557		2562	
	Treated *n* (of 2998)	247		429		2562	
	L1	n.a.		0.485		0.971	
	% BI	n.a.		99.90		94.47	

Note: Standard errors are listed below estimates. Sampling weights applied. L1 refers to the multivariate imbalance measure; % BI refers to percentage balance improvement.
*Significance at $p < 0.05$, †$0.05 \leq p < 0.1$. CEM refers to coarsened exact matching. PSM refers to nearest-one neighbour propensity score matching. The number of *matched* treated and control units are listed below estimates for each matching method.

to *any* water supply. As a sensitivity analysis, we analysed the effect of piped water relative to an *unimproved* water supply (Table 7). The results suggest that significance declines somewhat, although an effect of piped water on acute dysentery remains significant at $p < 0.05$ using CEM.

There is also a concern that the results from PSM are sensitive to data pruning with callipers. Whereas the main analyses above used a calliper of 0.25 standard deviations, in

Table 6. Impact estimates of improved toilet or hand-washing on diarrhoeal outcomes in children.

		Exact		CEM		PSM	
Treatment	Outcome	Const.	Impact	Const.	Impact	Const.	Impact
Improved toilet	All diarrhoea	0.165	−0.085	0.131	−0.017	0.089	0.006
		(0.023)	(0.035)*	(0.017)	(0.028)	(0.006)	(0.009)
	Acute watery diarrhoea	0.121	−0.057	0.088	−0.029	0.068	0.002
		(0.020)	(0.031)†	(0.014)	(0.022)	(0.005)	(0.008)
	Acute dysentery	0.014	−0.014	0.021	−0.017	0.004	0.005
		(0.006)	(0.009)	(0.006)	(0.010)†	(0.002)	(0.002)
	Diarrhoeal duration	0.924	−0.622	0.716	−0.218	0.421	0.175
		(0.137)	(0.208)*	(0.099)	(0.159)	(0.050)	(0.070)
	Control *n* (of 22,090)	213		362		2221	
	Treated *n* (of 2821)	155		238		2221	
	L1	n.a.		0.412		0.987	
	% BI	n.a.		99.65		96.53	
Primary hand-washing	All diarrhoea	0.151	−0.093	0.151	−0.087	0.128	−0.058
		(0.005)	(0.008)*	(0.005)	(0.008)*	(0.004)	(0.006)*
	Acute watery diarrhoea	0.120	−0.087	0.120	−0.083	0.100	−0.058
		(0.005)	(0.007)*	(0.004)	(0.007)*	(0.004)	(0.005)*
	Acute dysentery	0.013	−0.003	0.014	−0.004	0.012	−0.002
		(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)
	Diarrhoeal duration	0.836	−0.516	0.814	−0.468	0.702	−0.317
		(0.040)	(0.061)*	(0.036)	(0.057)*	(0.032)	(0.045)*
	Control *n* (of 20,131)	3639		4258		4779	
	Treated *n* (of 4780)	2730		2940		4779	
	L1	n.a.		0.130		0.804	
	% BI	n.a.		99.78		93.25	
Secondary hand-washing	All diarrhoea	0.155	−0.100	0.154	−0.093	0.125	−0.056
		(0.005)	(0.007)*	(0.005)	(0.007)*	(0.004)	(0.005)*
	Acute watery diarrhoea	0.127	−0.097	0.127	−0.092	0.101	−0.059
		(0.004)	(0.006)*	(0.004)	(0.006)*	(0.003)	(0.005)*
	Acute dysentery	0.012	−0.002	0.012	−0.002	0.010	0.000
		(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
	Diarrhoeal duration	0.823	−0.526	0.803	−0.471	0.646	−0.278
		(0.037)	(0.055)*	(0.033)	(0.051)*	(0.025)	(0.034)*
	Control *n* (of 18,756)	4079		4740		6154	
	Treated *n* (of 6155)	3320		3587		6154	
	L1	n.a.		0.128		0.756	
	% BI	n.a.		99.23		95.15	

Note: Standard errors are listed below estimates. Sampling weights applied. L1 refers to the multivariate imbalance measure; % BI refers to percentage balance improvement.
*Significance at $p < 0.05$, †$0.05 \leq p < 0.1$. CEM refers to coarsened exact matching. PSM refers to nearest-one neighbour propensity score matching. The number of *matched* treated and control units are listed below estimates for each matching method.

a sensitivity analysis we used progressively narrower callipers of 0.1, 0.001, 0.0001, and 0.00001 (with balance for these callipers reported in Table 3). Results with progressively narrower callipers are shown in Table 8. The results indicate that the effects of piped water on acute dysentery, and not other outcomes, are robust.

Another concern pertains to the coarsening of variables for CEM. In this study we manually coarsened several variables for CEM, which includes the construction of sample

Table 7. Sensitivity analysis I: impact estimates of piped water on diarrhoeal outcomes in children relative to control households with an unimproved water supply.

	Exact		CEM		PSM	
Outcome	Const.	Impact	Const.	Impact	Const.	Impact
All diarrhoea	0.168	−0.010	0.189	−0.049	0.114	0.011
	(0.022)	(0.032)	(0.017)	(0.025)	(0.005)	(0.007)
Acute watery diarrhoea	0.123	−0.002	0.137	−0.034	0.089	0.004
	(0.020)	(0.029)	(0.015)	(0.022)	(0.004)	(0.006)
Acute dysentery	0.012	−0.008	0.027	−0.022	0.008	−0.003
	(0.005)	(0.006)†	(0.005)	(0.007)*	(0.002)	(0.002)†
Diarrhoeal duration	0.829	−0.096	0.915	−0.260	0.639	0.063
	(0.134)	(0.196)	(0.102)	(0.153)†	(0.038)	(0.054)
Control *n* (of 18,006)	282		482		4148	
Treated *n* (of 6905)	247		386		4148	
Multivariate imbalance L1	n.a.		0.424		0.974	
% balance improv.	n.a.		99.8		81.3	

Note: Standard errors are listed below estimates.
*Significance at $p < 0.05$, †$0.05 \leq p < 0.1$. Sampling weights applied.

Table 8. Sensitivity analysis II: impact estimates of piped water on diarrhoeal outcomes in children on pruned samples.

	PSM calliper 0.1		PSM calliper 0.001		PSM calliper 0.0001		PSM calliper 0.00001	
Outcome	Const.	Impact	Const.	Impact	Const.	Impact	Const.	Impact
All diarrhoea	0.097	0.004	0.100	0.008	0.114	0.000	0.128	−0.014
	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)	(0.010)	(0.013)	(0.018)
Acute watery diarrhoea	0.069	0.007	0.075	0.005	0.084	0.005	0.094	−0.003
	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)	(0.009)	(0.011)	(0.016)
Acute dysentery	0.009	−0.004	0.010	−0.005	0.010	−0.004	0.019	−0.015
	(0.001)	(0.002)*	(0.001)	(0.002)*	(0.002)	(0.003)	(0.004)	(0.006)*
Diarrhoeal duration	0.526	0.008	0.513	0.053	0.574	0.024	0.732	−0.184
	(0.027)	(0.038)	(0.030)	(0.042)	(0.044)	(0.062)	(0.085)	(0.119)
Control *n* (of 18,006)	5965		4955		2156		650	
Treated *n* (of 6905)	5965		4955		2156		650	
Multivariate imbalance L1	0.967		0.960		0.848		0.368	
% balance improv.	95.5		99.9		100.0		100.0	

Note: Standard errors are listed below estimates.
*Significance at $p < 0.05$, †$0.05 \leq p < 0.1$. Sampling weights applied.

quintiles of the village development index and asset index (see Section 2). As a separate robustness check, we included the continuous index in place of quintiles and let CEM automatically coarsen these variables. This automated coarsening of continuous variables had a much smaller sample size than using the quintiles, but the results with this CEM variant (data not presented) were consistent with CEM with quintiles. A related concern to coarsening variables is whether impact estimates are sensitive to the specification of covariates for matching. As one coarsens to the extreme, a variable is effectively excluded and dropped (King *et al.* 2011). As noted in the results section, the results for all three matching methods are consistent (data not presented) with results from the fourth and final model. Given the large significance of the hand-washing results, as a robustness check we added additional covariates of piped water, tube-well, and improved toilet, and found robust effects of hand-washing (both primary and secondary).

Finally, we conducted subgroup analyses to estimate the impacts of primary hand-washing and secondary hand-washing on subgroups of interest, namely: households in the top three quintiles of assets compared with the bottom two quintiles; households in the top three quintiles of village development compared with the bottom two quintiles; and households where the highest education of a female member was secondary schooling matriculation or higher compared with those with primary schooling or lower education (Table 9). These subgroup analyses do not suggest major differences between subgroups. High-asset-holding households may benefit more from hand-washing than low-asset-holding households. In contrast, however, in low-developed villages, households benefit more from hand-washing than high-developed villages. Similarly, hand-washing may matter more in households with a low level of its highest educated female compared with a high level.

6. Discussion and concluding remarks

Our quasi-experimental study assessed the impacts of water supply, sanitation and hand-washing interventions for rural India on diarrhoeal disease using observational data. The matching methods used in this paper are consistent with best practices in matching methods, particularly in the use of balance diagnostics, and apply a novel Coarsened Exact Matching method. Our results are broadly consistent with the epidemiologic and clinical literature that suggests a robust and large significant effect of hand-washing. We also find smaller effects of piped water and improved toilets, although less robust to alternative matching methods. Obviously, our results do not test the effect of water *quality*, which is known to reduce diarrhoea. Moreover unlike epidemiologic studies that often lack information on detailed household-level and village-level indicators, our study controlled for confounding of household education, wealth, asset ownership, and village development. Our results on the effect of piped water on diarrhoea are consistent with a previous study by Jalan and Ravallion (2003) that used the same dataset, which found a significantly large effect of piped water in reducing diarrhoea but only if the outcome measure used is acute dysentery.

The paper contributes to the literature on the effects of various interventions on both acute watery diarrhoea and acute dysentery. Rarely do studies measure both outcomes of watery diarrhoea and dysentery; the handful of studies that measure both outcomes do not definitively suggest larger effects on either outcome (see Fewtrell *et al.* 2005, Cairncross *et al.* 2010). Similarly, there is no conclusive evidence suggesting that improved water supply reduces one of these outcomes more than the other. Biologically, it may be that pathogens spreading watery diarrhoea are less ‘sticky’ than those pathogens spreading dysentery. If this is true then incidence of watery diarrhoea may be more amenable than dysentery to hand-washing.

Furthermore, it is believed that dysentery (such as shigellosis) is more commonly associated in areas with insufficient water supply and inadequate sanitation, including wells, stagnant water sources, or sources of water that are contaminated by faecal matter

Table 9. Subgroup analyses: impact estimates of hand-washing on diarrhoeal outcomes in children, CEM.

	Asset Index				Village Development Index				Female Education			
	Low		High		Low		High		Low		High	
Outcome	Const.	Impact	Const.	Impact	Const.	Impact	Const.	Impact	Const.	Impact	Const.	Impact
Treatment: primary hand-washing												
All diarrhoea	0.162	−0.095	0.179	−0.120	0.211	−0.123	0.134	−0.097	0.181	−0.115	0.136	−0.088
	(0.008)	(0.012)*	(0.007)	(0.010)*	(0.008)	(0.012)*	(0.006)	(0.010)*	(0.006)	(0.009)*	(0.011)	(0.017)*
Acute watery	0.138	−0.096	0.146	−0.108	0.187	−0.131	0.100	−0.076	0.154	−0.109	0.097	−0.075
diarrhoea	(0.007)	(0.011)*	(0.006)	(0.009)*	(0.007)	(0.011)*	(0.006)	(0.009)*	(0.005)	(0.008)*	(0.009)	(0.014)*
Acute dysentery	0.007	0.004	0.009	0.001	0.009	0.007	0.006	−0.002	0.007	0.003	0.010	−0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)
Diarrhoeal duration	0.850	−0.468	0.938	−0.640	1.037	−0.614	0.771	−0.528	0.948	−0.594	0.707	−0.471
	(0.056)	(0.087)*	(0.044)	(0.069)*	(0.046)	(0.073)*	(0.051)	(0.080)*	(0.039)	(0.061)*	(0.070)	(0.110)*
Control *n* (of 20,131)	1855		2433		2295		1993		3547		741	
Treated *n* (of 4780)	1182		1747		1448		1481		2347		555	
Treatment: secondary hand-washing												
All diarrhoea	0.179	−0.109	0.181	−0.126	0.235	−0.147	0.132	−0.094	0.195	−0.128	0.124	−0.083
	(0.007)	(0.011)*	(0.006)	(0.010)*	(0.008)	(0.012)*	(0.006)	(0.009)*	(0.006)	(0.008)*	(0.009)	(0.014)*
Acute watery	0.158	−0.116	0.151	−0.114	0.208	−0.153	0.105	−0.081	0.170	−0.126	0.092	−0.072
diarrhoea	(0.007)	(0.010)*	(0.006)	(0.009)*	(0.007)	(0.011)*	(0.005)	(0.008)*	(0.005)	(0.008)*	(0.008)	(0.012)*
Acute dysentery	0.006	0.005	0.007	0.002	0.007	0.009	0.006	−0.002	0.005	0.006	0.012	−0.006
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.005)
Diarrhoeal duration	0.874	−0.475	0.927	−0.637	1.122	−0.702	0.712	−0.452	0.978	−0.612	0.625	−0.409
	(0.049)	(0.074)*	(0.042)	(0.063)*	(0.046)	(0.069)*	(0.044)	(0.067)*	(0.037)	(0.056)*	(0.059)	(0.090)*
Control *n* (of 18,756)	2497		3188		2752		2933		4559		1126	
Treated *n* (of 6155)	1058		1566		1307		1319		2118		508	

Note: Standard errors are listed below estimates. Sampling weights applied.
*Significance at $p < 0.05$. CEM refers to coarsened exact matching. PSM refers to nearest-one neighbour propensity score matching. The number of *matched* treated and control units are listed below estimates for each treatment.

(Lichnevski 1996; DuPont 2010). Thus if piped water yields larger volumes of water, is less stagnant and has less faecal contamination than other forms of water supply (for example, tube-wells), then it might be expected that piped water reduces dysentery more than tube-wells reduce dysentery. It is believed that dysentery is transmitted in a small concentration of bacteria relative to watery diarrhoea. Therefore watery diarrhoea, which is known to require larger concentrations of pathogen than dysentery, would be less amenable to large increases in water supply than dysentery.

The results from this study of rural India should be interpreted cautiously for policy purposes. Where sample size was reduced by a large measure from exact or coarsened exact matching, as in the case of water and sanitation, the results are likely to be less externally valid, although internally valid for the smaller matched sample. In contrast, the results of hand-washing are robust across matching methods, at levels of very high significance, and do not suffer from significantly reduced samples from matching. Yet with any matching method there is the concern that the estimated effect may be biased if there are unobserved variables predicting treatment. In the case of hand-washing, the effects may be over-estimated if the outcomes are driven by unobserved knowledge and behaviour other than hand-washing.

Our results may help to shed light on the long-term effects of policies relating to interventions on water supply and sanitation facilities, many of which are known to deteriorate in quality over time or suffer from low 'compliance'. Several reports have described the increasingly complex challenges of water supply in everyday Indian life. A recent World Bank report by Briscoe and Malik (2008) argued that India's water sector has been facing a major financing gap, which has led to crumbling or deteriorating infrastructure and the decline in quantity and quality of public irrigation and water supply services. In response to the decline, people have devised ingenious ways or 'coping strategies' around weak water supply systems; for example, individuals drilling tube-wells or relying on others' tube-wells. It is estimated that some 20 million individual tube-wells are now installed. Similarly, the urban middle class, in coping with irregular, unpredictable and often polluted and piped public water services, have devised household storage and purification systems to improve water quality, or private well access to tap groundwater. The challenges in infrastructure and service delivery are likely to be compounded by corruption (for example, Davis 2004).

These common challenges of water supply were also described in a 2007 report produced by the Planning Commission of the Government of India, which suggests that ground water is 'open access common property natural resource and anyone can bore a well and pump out water without limit' (Planning Commission, Government of India 2007). The variability in public provision of water has led to a National Rural Drinking Water Program, for which a 'good framework should consider different drinking water sources accessible in different situations and different points of time', and that drinking water should not rely on a single source but rely on multiple sources to ensure water security including groundwater, surface-water, and rainwater harvesting (Rajiv Gandhi National Drinking Water Mission 2009). It is notable that neither of the two reports cited above made a distinction between 'piped water' and other sources of water supply, such as boreholes or tube-wells. These reports have focused primarily on socio-economic implications of India's 'private, self-provision with groundwater', and place a minor emphasis, if any, on the health implications of water supply. It is not obvious how increasing diversity of household water sources would affect health outcomes.

The state of sanitation in India is dire. There has been an increasingly prominent focus of policy attention on the situation of sanitation in the country (see Mara *et al.* 2010, World Bank 2011). Mara *et al.* (2010) cited a few reports that found that toilets in India are often not used for their intended purpose of defecation, but instead used for other purposes such as firewood stores or storage sheds. This lack of 'compliance' may help to explain why our study did not detect a consistently significant effect of improved toilets on diarrhoea. Our study does not dispute the fact that improved sanitation, when sustained and maintained over time, can prevent major health risks as suggested in the limited epidemiologic literature. The costs of unimproved sanitation to Indians are large. The World Bank (2011) recently reported that 'inadequate sanitation' caused an annual economic loss of US$53.8 billion in 2006, as measured by premature mortality and other health-related impacts, as well as productive time lost, and impacts related to drinking water that are primarily shouldered by children younger than age five.

Hand-washing has been one of many potential interventions suggested by UNICEF/WHO (2009), including a number of other (what they call) 'primary preventive measures' such as improved drinking water supply, community-wide sanitation, and vaccination for rotavirus and measles. Despite this emphasis in that UNICEF/WHO report, hand-washing prevalence is not a routinely reported indicator in its WHO Statistics Database, unlike the indicators of percentage of individuals with improved water or with improved sanitation (or even other self-reported indicators of immunisation coverage or use of various family planning measures). Hand-washing is also not mentioned in a WHO (2009) report.[8] Some aspects of hand-washing have been measured in the Demographic and Health Surveys (which canvass detailed self-reported family planning behaviours) for multiple surveys over 1999–2005 in sub-Saharan Africa (see Macro International 2011). Thus, a key policy recommendation from the results of this paper would be that prevalence of hand-washing behaviour, even self-reported, is an important indicator for routine health surveillance, and is arguably as important as measures of water supply, sanitation, and other self-reported health-seeking behaviours.

The policy drivers of hand-washing behaviour in a population are not well understood. Of course hand-washing also requires access to water, though not necessarily an 'improved' water supply. There are studies that suggest that educational interventions can change a population's behaviour at least in the short run and can reduce diarrhoea incidence. A study by Stanton and Clemens (1987) tailored an intervention to a population's existing behaviours to provide an eight-week community-based health education programme with local experienced trainers. The intervention had three main messages: proper hand-washing before food preparation, defecating away from the house at a proper site, and suitable disposal of waste and faeces. They relied on public demonstrations, small-group discussions and community-wide meetings as well as posters, games and stories to illustrate that messages may be effective. As noted by Luby (2010), behaviour-change campaigns 'often require substantial resources (especially trained personnel, community organization, and funding)'. Temporal features of these interventions (that is, how frequently the intervention is offered and over what length of time) may determine whether the intervention produces long-lasting behaviour changes in the community. Changing norms through community leaders, social networks, and local community organisation is also likely to be an important feature of long-lasting interventions.

As we noted earlier, progress in this sector of water, sanitation, and hygiene has been slow in India. We have no evidence on whether hand-washing behaviours have improved in the past two decades. We hope that the results from this study can spur needed attention to improving water, sanitation, and hygiene, but particularly hygiene and hand-washing as well as sanitation in order to improve the health of India's children.

Notes

1. Estimating propensity scores at either household or individual level in the first-stage model followed by matching to the individual level will lead to equivalent matching results, if the same pretreatment covariates village-level and household-level covariates are used. If the propensity score is estimated at individual level, then individual children from the same village and household will be assigned the same propensity score if they come from the same household. If the score is estimated at household level, then a single household will be assigned a single propensity score and therefore children within the same household would receive the same score. Clustering standard errors in the first-stage logit model will not affect the distribution of propensity scores, only their standard errors.
2. Because exact matching is balanced on observables, estimates using parametric analyses are the same as estimates taking a simple difference. See Ho *et al.* (2007).
3. We chose the definition of piped water that appeared to be closest to that used in Jalan and Ravallion (2003). See Appendix 1.
4. A tube-well (or bore-well) relies on a mechanical or motorised pump, whereas a hand pump is similar to a tube-well but relies on a hand pump. Piped water refers to water distributed by pipeline, but does not refer to a kind of extraction method.
5. These estimates are consistent with a recent study of 288 households by Biran *et al.* (2009), who found a prevalence of *observed* hand-washing with both hands on any occasion of 28–31 per cent. A separate study by Biran *et al.* (2008) found that self-reported hand-washing with soap underestimated observed hand-washing with soap, whereas self-reported hand-washing generally overestimated observed hand-washing with water.
6. Household-level variables include demographic characteristics, whether the household belongs to a scheduled tribe, a scheduled caste, household religion, household size, occupation of household head, radio and television listening behaviours of household members, proportion of household members who are elderly, proportion of adults who are female, proportion of children who are male, whether household head is male, whether household head is single, whether household head is married, education level of household head, highest education level of female in household, total household income (in quintiles), and asset index (in quintiles) constructed from asset ownership (house, other property, bicycle, sewing machine, thresher, winnower, bullock cart, radio, television, fan, livestock, nature of house, condition of house, number of rooms in house, electricity, clean stove, chimney for cooking, ventilated kitchen, separate kitchen, use of landholding for cultivation, landholding size, gross cropped area, and gross irrigated area). Village-level infrastructure variables including proportion of gross cropped area that is irrigated, whether the village has a daycare centre, a primary school, a middle school, a high school, type of road approaching the village, whether the village has a bus stop, a railway station, a post office, a telephone facility, a community television centre, a library, a bank, and a market as well as village demographic variables of village population, the student–teacher ratio, male–female student ratio, and male–female minority student ratio.
7. Three pretreatment variables used by Jalan and Ravallion (2003) – the student–teacher ratio, the male–female student ratio, and the male–female minority student ratio – are not necessarily pretreatment and were thus excluded. Moreover, they had large numbers of missing observations.
8. Hand-washing is not mentioned in a WHO report on global health risks (WHO 2009). The report defines the category of ‘unsafe water, sanitation, and hygiene’ as ‘improved water’ and ‘improved sanitation’ with specific types of water sources (for example, piped water, borehole, and so forth) and sanitation facilities (for example, septic tank, pit latrine, and so forth) without mention of any ‘hygiene’-related behaviours such as hand-washing.

References

American Public Health Association, 2008. *Control of communicable diseases manual: an official report of the American Public Health Association*, ed. D.L. Heymann. Washington, DC: American Public Health Association.

Biran, A., *et al*., 2008. Comparing the performance of indicators of hand-washing practices in rural Indian households. *Tropical medicine & international health*, 13 (2), 278–285.

Biran, A., *et al*., 2009. The effect of a soap promotion and hygiene education campaign on handwashing behaviour in rural India: a cluster randomised trial. *Tropical medicine & international health*, 14 (10), 1303–1314.
Black, *et al*., 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The lancet*, 375, 1969–1987.
Bose, R., 2009. *The impact of water supply and sanitation interventions on child health: evidence from DHS surveys*. Mimeo. New Delhi: International Initiative for Impact Evaluation.
Briscoe, J. and Malik, R.P.S., 2008. *India's water economy: bracing for a turbulent future*. New Delhi: World Bank. Available from: http://go.worldbank.org/R09M773280 [Accessed 3 January 2011]
Cairncross, S., *et al*., 2010. Water, sanitation, and hygiene for the prevention of diarrhoea. *International journal of epidemiology*, 39, i193–i205.
Cash, R., 2011. Email. Personal communication, 7 June.
Clasen, T., *et al*., 2007. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *British medical journal*, 334 (7597), 782.
Clasen, T.F., *et al*., 2010. Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane database of systematic reviews*, 6, CD007180.
Curtis, V., Cairncross, S. and Yonli, R., 2000. Domestic hygiene and diarrhoea – pinpointing the problem. *Tropical medicine & international health*, 5 (1), 22–32.
Davis, J., 2004. Corruption in public service delivery: experience from South Asia's water and sanitation sector. *World development*, 31 (1), 53–71.
Department of Drinking Water Supply, undated. Rajiv Gandhi National Drinking Water Mission, Ministry of Rural Development, Government of India. Available from: http://www.ddws.nic.in/
DuPont, H.L., 2010. Shigella species (bacillary dysentery). *In*: G.L. Mandell, J.E. Bennett and R. Dolin, eds. *Mandell, Douglas, and Bennett's principles and practice of infectious disease*. 7th ed. Philadelphia: Churchill Livingston (Elsevier), 2905–2910.
Ejemot, R.I., *et al*., 2008. Hand washing for preventing diarrhoea. *Cochrane database of systematic reviews*, 1, CD004265.
Fewtrell, L., *et al*., 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet infectious diseases*, 5 (1), 42–52.
Filmer, D. and Pritchett, L.H., 2001. Estimating wealth effects without expenditure data-or tears: an application to educational enrolments in states of India. *Demography*, 38 (1), 115–132.
Ho, D.E., *et al*., 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15, 199–236.
Ho, D.E., *et al*., 2011. MatchIt: nonparametric pre-processing for parametric causal inference. *Journal of statistical software*, 42.
Iacus, S.M., King, G. and Porro, G., 2011a. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106 (493), 346–361.
Iacus, S.M., King, G. and Porro, G., 2011b. Causal inference without balance checking: coarsened exact matching. *Political analysis*, forthcoming.
Imai, K., King, G. and Stuart, E.A., 2008. Misunderstandings between experimentalists and observationalists about inference. *Journal of the Royal Statistical Society, Series A (statistics in society)*, 171, 481–502.
Jalan, J. and Ravallion, M., 2003. Does piped water reduce diarrhoea for children in rural India? *Journal of econometrics*, 112, 153–173.
Khanna, G., 2008. *The impact on child health from access to water and sanitation and other socioeconomic factors*. Geneva: The Graduate Institute of International and Development Studies, HEI Working Paper No. 02/2008.
King, G., *et al*., 2011. *Comparative effectiveness of matching methods for causal inference*. Working paper. Cambridge, MA: Harvard Unviersity.
Lichnevski, M., 1996. Shigella dysentery and shigella infections. *Eastern Mediterranean health journal*, 2 (1), 102–106. Available from: http://www.emro.who.int/publications/emhj/0201/14.htm [Accessed 1 March 2011].
Luby, S.P., *et al*., 2004. Effect of intensive handwashing promotion on childhood diarrhoea in high-risk communities in Pakistan: a randomised controlled trial. *Journal of the American Medical Association*, 291 (21), 2547–2554.

Luby, S.P., *et al*., 2010. A community-randomised controlled trial promoting waterless hand sanitizer and handwashing with soap, Dhaka, Bangladesh. *Tropical medicine & international health*, 15 (12), 1508–1516.

Macro International Inc, 2011. *MEASURE DHS STATcompiler*. Available from: http://www.measuredhs.com [Accessed 7 January 2011].

Mara, D., *et al*., 2010. Sanitation and health. *PLoS medicine*, 7 (11), e1000363.

Planning Commission, Government of India, 2007. *Report of the expert group on ground water management and ownership*. New Delhi: Government of India, iii. Available from: http://planningcommission.nic.in/reports/genrep/rep_grndwat.pdf

Pradhan, M. and Rawlings, L.B., 2002. The impact and targeting of social infrastructure investments: lessons from the Nicaraguan Social Fund. *World Bank economic review*, 16 (2), 275–295.

Prichett, L., 2010. Email. Personal communication, 31 March.

Rajiv Gandhi National Drinking Water Mission, 2009. *National rural drinking water programme – movement towards ensuring people's drinking water security in rural India: framework for implementation, 2009–2012*. New Delhi: Department of Drinking Water Supply, Rajiv Gandhi National Drinking Water Mission, Ministry of Rural Development, Government of India. Available from: http://ddws.gov.in/popups/RuralDrinkingWater_2ndApril.pdf

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Shariff, A., 1999. *India: human development report*. *a profile of Indian states in the 1990s*. New Delhi: National Council of Applied Economic Research and Oxford University Press, 300.

Stanton, B.F. and Clemens, J.D., 1987. An educational intervention for altering water-sanitation behaviours to reduce childhood diarrhoea in urban Bangladeshi: a randomized trial to assess the impact of the intervention on hygienic behaviours and rates of diarrhoea. *American journal of epidemiology*, 125 (2), 292–301.

UNICEF/WHO, 2009. *Diarrhoea: why children are still dying and what can be done*. Geneva: WHO and New York: UNICEF. Available from: http://www.childinfo.org/files/diarrhoea_hires.pdf [Accessed 25 March 2010].

Waddington, H., *et al*., 2009. *Water, sanitation, and hygiene interventions to combat childhood diarrhoea in developing countries*. New Delhi: The International Initiative for Impact Evaluation (3ie), 3ie Synthetic Reviews. Available from: http://www.3ieimpact.org/admin/pdfs2/17.pdf [Accessed 8 January 2011].

WHO, 2009. *Global health risks: mortality and burden of disease attributable to selected major risks*. Geneva: WHO, 23 and 38. Available from: http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf [Accessed 3 January 2011].

WHO, 2011. *WHO Statistical Information System (WHOSIS)*. Available from: http://www.who.int/gho/en/ [Accessed 1 May 2011].

World Bank, 2011. *The economic impacts of inadequate sanitation in India: inadequate sanitation costs India Rs. 2.4 trillion (US$53.8 billion)*. Water and Sanitation Program (World Bank), Asian Development Bank, AusAID, and UKAID. Available from: http://www.wsp.org/wsp/sites/wsp.org/files/publications/wsp-esi-india.pdf [Accessed 3 January 2001].

Zwane, A.P. and Kremer, M., 2007. What works in fighting diarrhoeal diseases in developing countries? A critical review. *World Bank Research Observer*, 22 (1), 1–24.

Appendix 1. Household access to piped water

Household access to piped water by income quintile and by highest education of female member.

Piped water treatment variable (dataset/s used)	Income quintiles	Number of households	Households with piped water	Households with piped water stratified by highest education of female members				Percentage of people with piped water
				Illiterate	Primary	Matriculation	Secondary or higher	
Piped water in any season (household dataset only)	Lowest	6584	1789	789	686	278	36	27.17
	Second	6509	1653	693	623	301	36	25.40
	Third	6542	1765	680	605	414	66	26.98
	Fourth	6694	1984	685	646	542	111	29.64
	Highest	6901	2320	688	630	766	236	33.62
	Full sample	*33,230*	*9511*	*3535*	*3190*	*2301*	*485*	*28.62*
Piped water in any season (merged household and village dataset)	Lowest	6579	1786	787	685	278	36	27.15
	Second	6506	1653	693	623	301	36	25.41
	Third	6539	1765	680	605	414	66	26.99
	Fourth	6692	1983	685	646	541	111	29.63
	Highest	6900	2319	688	629	766	236	33.61
	Full sample	*33,216*	*9506*	*3533*	*3188*	*2300*	*485*	*28.62*
Piped water in all seasons (merged household and village dataset)	Lowest	6579	1508	672	582	224	30	22.92
	Second	6506	1376	585	514	246	31	21.15
	Third	6539	1489	579	508	346	56	22.77
	Fourth	6692	1665	593	535	445	92	24.88
	Highest	6900	2025	598	553	667	207	29.35
	Full sample	*33,216*	*8063*	*3027*	*2692*	*1928*	*416*	*24.27*
Jalan and Ravallion (2003)	Lowest	6581	1707	768	655	251	33	27.18
	Second	6508	1567	674	590	274	29	25.40
	Third	6543	1658	667	560	371	60	26.96
	Fourth	6694	1814	660	602	462	90	29.62
	Highest	6904	2081	665	593	638	185	33.63
	Full sample	*33,230*	*8827*	*3434*	*3000*	*1996*	*397*	*28.62*

Notes: Piped water in any season refers to piped water use in either summer or non-summer seasons; piped water in all seasons refers to piped water use in both summer and non-summer seasons. After merging the household dataset to the village dataset, the total number of available household observations decreases from 33,230 to 33,216. (Jalan and Ravallion also used only 33,216 observations in the first-stage model [Table 2].) Note that the number of households with access to 'piped water' in Jalan and Ravallion (2003) was between these two numbers. We think that Jalan and Ravallion were more likely to have used piped water in any season, rather than in all seasons, as the '% of people with piped water' match exactly at 28.62 per cent, even though the total number of households with piped water differs (9511 or 9506 compared with Jalan and Ravallion's total of 8827).

Appendix 2. Replication of Table 2 in the paper by Jalan and Ravallion (2003)

	(1)		(2)		(3)		J&R	
Variable	Coefficient	*z*-stat	Coefficient	*z*-stat	Coefficient	*z*-stat	Coefficient	*t*-stat
Constant	−1.150	−4.24	−1.588	−5.50	−2.740	−7.66	−1.495	−5.40
Household variables								
tribe	−0.259	−4.73	−0.231	−4.13	−0.219	−3.57	−0.213	−4.20
caste	−0.028	−0.69	−0.017	−0.42	−0.010	−0.23	−0.010	−0.29
hindu	−0.237	−1.50	−0.254	−1.57	−0.306	−1.88	−0.242	−1.71
muslim	−0.206	−1.22	−0.207	−1.20	−0.272	−1.55	−0.216	−1.43
christian	0.552	2.98	0.349	1.84	0.573	2.88	0.404	2.43
sikh	−0.874	−4.12	−0.927	−4.29	−0.842	−3.81	−0.866	−4.53
hhsize	0.010	1.62	0.010	1.63	0.013	1.83	0.003	0.57
cultland	−0.154	−2.94	−0.055	−1.03	−0.037	−0.63	0.171	1.91
houseown	−0.189	−2.55	−0.172	−2.29	−0.153	−1.88	−0.190	−2.85
propown	0.041	0.93	0.001	0.02	0.050	1.04	0.002	0.04
bicyown	−0.246	−7.04	−0.254	−7.15	−0.238	−6.20	−0.265	−8.24
sewmown	0.025	0.50	0.010	0.19	−0.015	−0.26	0.012	0.25
thresown	−0.105	−1.01	−0.090	−0.85	−0.175	−1.47	−0.058	−0.58
winnown	0.217	1.71	0.174	1.34	0.150	1.06	0.218	1.82
cartown	−0.259	−5.05	−0.262	−5.01	−0.282	−5.06	−0.259	−5.43
radioown	−0.030	−0.69	−0.035	−0.80	−0.040	−0.85	0.010	0.25
tvown	0.128	2.33	0.106	1.88	0.057	0.94	0.081	1.34
fanown	0.022	0.48	0.018	0.38	−0.033	−0.65	0.013	0.32
livstown	−0.131	−3.55	−0.099	−2.64	−0.088	−2.15	−0.078	−2.34
hsnat_kuccha	−0.123	−3.12	−0.109	−2.71	−0.173	−4.00	−0.100	−2.78
hsnat_pucca	0.106	2.19	0.103	2.09	0.135	2.53	0.120	2.71
hscond_good	0.036	0.52	0.030	0.42	0.010	0.13	0.002	0.04
hscond_liv	0.096	1.64	0.097	1.64	0.049	0.76	−0.108	−1.37
rooms_1	−0.150	−1.76	−0.121	−1.40	0.027	0.28	0.068	−1.37
rooms_2	0.025	0.32	0.042	0.53	0.146	1.64	0.075	0.95
rooms_35	0.081	1.11	0.075	1.01	0.142	1.69	−0.020	1.11
kitchen	0.006	0.15	−0.008	−0.19	−0.003	−0.06	0.081	−0.53
vent	0.091	2.27	0.084	2.06	0.099	2.24	0.406	2.21
elec	0.471	11.80	0.433	10.69	0.435	9.97	−0.024	11.22
occ1	−0.082	−1.31	−0.050	−0.79	0.008	0.11	0.024	−0.48
occ2	−0.062	−0.87	−0.086	−1.20	−0.082	−1.06	0.146	0.43
occ3	0.135	1.73	0.095	1.20	−0.007	−0.08	−0.069	2.25
occ4	0.031	0.41	−0.043	−0.56	−0.002	−0.03	0.201	−0.96
radiom	0.081	1.31	0.089	1.42	0.071	1.05	−0.124	3.48
radiof	0.049	0.79	0.045	0.72	0.068	1.01	0.094	−2.18
tvm	0.173	2.80	0.122	1.95	0.186	2.76	0.039	1.29
tvf	−0.037	−0.57	−0.021	−0.31	−0.069	−0.98	0.090	0.49
newsm	0.011	0.24	−0.006	−0.12	−0.047	−0.93	−0.041	1.81
newsf	0.125	2.39	0.092	1.74	0.130	2.27	−0.114	−0.63
propold	−0.082	−0.50	−0.151	−0.92	−0.077	−0.43	0.046	−1.07
propfemad	−0.044	−0.34	−0.023	−0.18	−0.075	−0.53	0.084	0.33
propmalec	0.019	0.44	0.022	0.51	0.038	0.80	0.055	0.78
malehead	−0.223	−2.30	−0.246	−2.50	−0.332	−3.11	−0.180	−2.32
single	−0.185	−0.91	−0.182	−0.89	−0.165	−0.77	−0.167	−1.27
married	−0.040	−0.56	−0.018	−0.24	−0.002	−0.02	−0.026	−0.42
hhead_ed1	−0.212	−2.18	−0.219	−2.22	−0.243	−2.26	−0.130	−1.45
hhead_ed2	−0.100	−1.04	−0.123	−1.26	−0.154	−1.46	−0.037	−0.42
hhead_ed3	−0.059	−0.63	−0.076	−0.79	−0.091	−0.87	−0.034	−0.39
hhead_ed4	−0.111	−0.89	−0.135	−1.07	−0.111	−0.80	−0.055	−0.48

(*Continued*)

Appendix 2. (*Continued*)

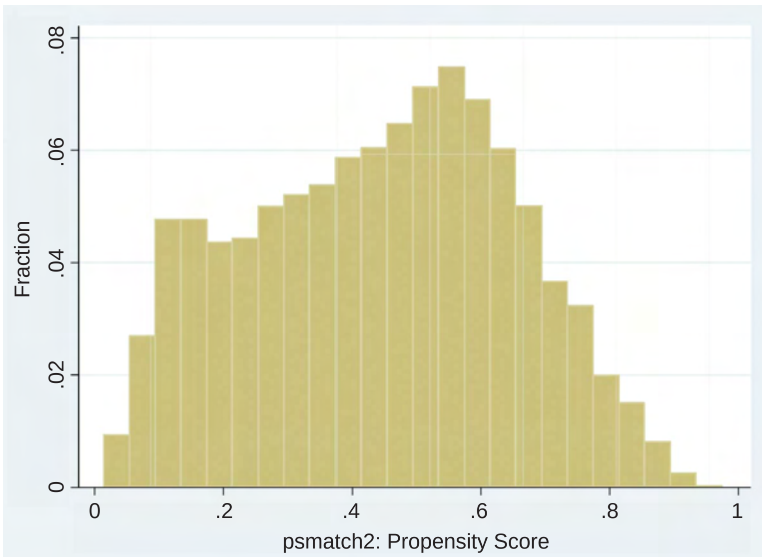
	(1)		(2)		(3)		J&R	
Variable	Coefficient	*z*-stat	Coefficient	*z*-stat	Coefficient	*z*-stat	Coefficient	*t*-stat
croparea	0.000	−0.17	0.000	−0.45	0.000	−1.07	0.000	−0.67
irrarea	0.000	−0.61	0.000	−0.64	0.000	−0.46	−0.001	−1.34
landless	−0.150	−3.01	−0.187	−3.70	−0.267	−4.93	−0.328	−4.00
landmarg	−0.187	−3.05	−0.219	−3.52	−0.325	−4.86	−0.311	−3.99
landsmall	−0.153	−2.33	−0.167	−2.49	−0.245	−3.42	−0.221	−2.92
Village-level variables								
vsize					0.117	4.43	0.082	4.27
pirrig_75			−0.102	−2.52	−0.202	−4.49	−0.048	−1.19
pirrig_50_75			0.108	2.08	0.145	2.56	0.194	4.18
angan			−0.084	−2.35	−0.126	−3.22	−0.072	−2.23
prim			−0.070	−1.27	0.223	2.30	−0.081	−1.43
mid			0.035	0.94	0.114	2.77	−0.090	−2.58
high			0.210	4.81	0.144	3.07	0.265	7.41
fmvill					−0.005	−3.16	0.106	3.01
fmmin							−0.077	−2.11
puccaroad			0.183	3.07	0.259	3.86	0.194	3.64
kuccharoad			−0.044	−0.80	0.033	0.53	−0.002	−0.03
busin			0.266	6.32	0.287	6.37	0.114	2.95
railin			−0.115	−1.34	−0.141	−1.54	0.009	0.18
postin			0.032	0.76	−0.043	−0.91	0.022	0.55
phonein			0.319	7.74	0.246	5.55	0.331	9.66
tvin			−0.003	−0.08	0.068	1.37	0.099	2.66
libin			−0.023	−0.47	−0.048	−0.93	−0.042	−1.12
bankin			0.252	5.59	0.165	3.43	0.191	4.66
mktin			0.361	6.25	0.468	7.29	0.317	6.09
stratio							0.002	5.30
n	26,876		26,867		22,396		33,216	

Notes: The regressions above attempt to replicate the first-stage logit model by Jalan and Ravallion (J&R) (2003). The regressions presented are done for all households including those without children. Regression (1) uses only household-level variables, (2) uses both household-level and village-level variables after dropping four village-level variables, and (3) adds in two village-level variables. The number of observations in Regression (1) is 26,876, although the overall dataset has 33,216 households, because when one includes the variable 'proportion of males among children' into the regression, households without children of any age are automatically dropped and not all households have children of any ages. Regression (2) does not include four of the village-level variables (village size, female to male students in the village, female to male students for minority groups, and student–teacher ratio in the village) that Jalan and Ravallion (2003) used. There were only 26,653 observations with village size, 22,555 observations with female–male student ratio in village, and 14,190 observations with female–male minority (SC, ST) ratio in village, and 13,903 observations with student–teacher ratio in village. Regression (3) does not include two of the village-level variables that Jalan and Ravallion (2003) used; these variables were not included because of the large number of missing observations and if included, these variables would reduce the total number of observations used in the regression. Except for the village size, these three variables were marked as missing if the denominator was zero (that is, if the village did not have any male students, did not have minority male students, or did not have any teachers). Village size and female-to-male ratio in village was included in Regression (3) as it lost 4471 observations (as opposed to the other two variables, which would have resulted in a loss of 12,964 observations). Jalan and Ravallion (2003) do not include the two stratifying variables, household income and highest educational attainment by female in household, in the logit model, although they estimate the subgroup effects separately.

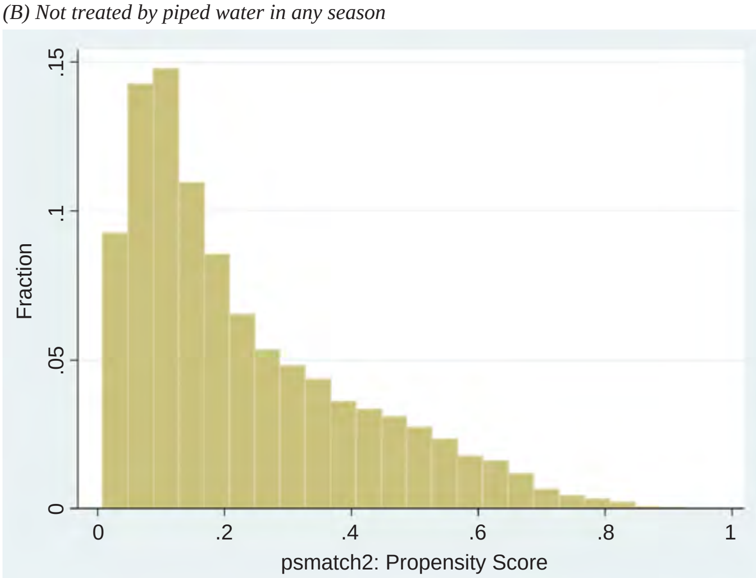
Appendix 3. Histograms of estimated propensity scores

Histograms of estimated propensity scores treated by piped water in any season (A) or without piped water in any season (B) at household level

(A) Treated by piped water in any season



(B) Not treated by piped water in any season



Note: These propensity scores correspond to the regression in column (3) in Appendix 2.

Appendix 4. Summary of balance diagnostics using the specification by Jalan and Ravallion (2003)

	Full unmatched data				Matched data				% BI
Variable	Means Treated	Means Control	SD Control	Mean Diff	Means Treated	Means Control	SD Control	Mean Diff	Mean Diff
distance	0.38	0.23	0.15	0.14	0.34	0.34	0.16	0.00	98.70
pirrig_75	0.23	0.23	0.42	0.01	0.25	0.25	0.43	0.00	47.62
pirrig_50_75	0.11	0.11	0.32	0.00	0.12	0.10	0.29	0.02	−2496.71
angan	0.63	0.46	0.50	0.17	0.61	0.64	0.48	−0.04	79.41
prim	0.90	0.87	0.33	0.02	0.90	0.90	0.30	0.00	97.71
mid	0.49	0.36	0.48	0.13	0.47	0.46	0.50	0.01	92.71
high	0.34	0.19	0.39	0.16	0.31	0.28	0.45	0.03	82.77
puccaroad	0.48	0.33	0.47	0.15	0.46	0.41	0.49	0.04	71.72
kuccharoad	0.42	0.57	0.50	−0.15	0.46	0.48	0.50	−0.02	87.22
busin	0.59	0.36	0.48	0.24	0.56	0.56	0.50	0.01	96.92
railin	0.04	0.03	0.17	0.02	0.04	0.04	0.19	0.00	96.43
postin	0.56	0.38	0.49	0.18	0.53	0.50	0.50	0.03	84.06
phonein	0.50	0.26	0.44	0.23	0.45	0.44	0.50	0.01	94.75
tvin	0.31	0.20	0.40	0.11	0.30	0.29	0.45	0.01	92.90
libin	0.20	0.12	0.33	0.08	0.19	0.16	0.37	0.02	69.65
bankin	0.24	0.17	0.38	0.07	0.23	0.20	0.40	0.03	60.26
mktin	0.11	0.08	0.27	0.03	0.10	0.07	0.26	0.03	0.27
tribe	0.09	0.15	0.36	−0.06	0.10	0.10	0.31	−0.01	89.80
caste	0.23	0.24	0.43	−0.01	0.24	0.24	0.43	0.00	73.79
hindu	0.85	0.81	0.39	0.04	0.83	0.85	0.36	−0.02	53.73
muslim	0.09	0.12	0.33	−0.03	0.10	0.09	0.29	0.01	69.42
christian	0.03	0.02	0.15	0.01	0.03	0.02	0.14	0.01	−59.00
sikh	0.02	0.03	0.18	−0.01	0.02	0.03	0.16	0.00	69.34
hhsize	7.06	6.93	3.01	0.13	7.00	6.98	3.06	0.02	84.81
cultland	0.61	0.69	0.46	−0.07	0.61	0.64	0.48	−0.02	71.03
houseown	0.94	0.97	0.18	−0.03	0.95	0.95	0.21	0.00	97.86
propown	0.17	0.12	0.33	0.05	0.15	0.15	0.36	0.00	98.75
bicyown	0.49	0.58	0.49	−0.09	0.52	0.51	0.50	0.01	87.46
sewmown	0.20	0.13	0.33	0.08	0.17	0.18	0.39	−0.01	83.66
thresown	0.03	0.04	0.18	−0.01	0.03	0.03	0.16	0.00	31.33
winnown	0.02	0.01	0.10	0.01	0.02	0.02	0.13	0.00	92.49
cartown	0.12	0.15	0.35	−0.03	0.13	0.15	0.35	−0.01	48.22
radioown	0.44	0.37	0.48	0.07	0.42	0.41	0.49	0.02	78.70
tvown	0.22	0.12	0.32	0.10	0.19	0.17	0.37	0.03	75.38
fanown	0.35	0.21	0.41	0.14	0.32	0.31	0.46	0.01	92.20
livstown	0.64	0.69	0.46	−0.05	0.64	0.68	0.47	−0.04	22.64
hsnat_kuccha	0.43	0.57	0.49	−0.14	0.45	0.48	0.50	−0.02	82.87
hsnat_pucca	0.23	0.15	0.36	0.08	0.23	0.20	0.40	0.02	69.89
hscond_good	0.27	0.21	0.41	0.06	0.26	0.24	0.43	0.02	61.20
hscond_liv	0.65	0.68	0.47	−0.03	0.65	0.66	0.47	−0.01	62.59
rooms_1	0.26	0.26	0.44	−0.01	0.26	0.30	0.46	−0.04	−566.26
rooms_2	0.37	0.37	0.48	0.00	0.37	0.37	0.48	−0.01	−102.06
rooms_35	0.31	0.32	0.47	0.00	0.32	0.28	0.45	0.04	−1560.46
kitchen	0.54	0.40	0.49	0.14	0.50	0.49	0.50	0.01	94.32
vent	0.46	0.32	0.47	0.14	0.41	0.42	0.49	0.00	96.80
elec	0.67	0.41	0.49	0.26	0.61	0.62	0.49	−0.01	96.90
occ1	0.55	0.60	0.49	−0.05	0.55	0.58	0.49	−0.03	32.46
occ2	0.16	0.17	0.38	−0.01	0.17	0.17	0.37	0.01	45.31
occ3	0.10	0.08	0.27	0.02	0.10	0.09	0.28	0.01	36.32

(*Continued*)

Appendix 4. (*Continued*)

	Full unmatched data				Matched data				% BI
Variable	Means Treated	Means Control	SD Control	Mean Diff	Means Treated	Means Control	SD Control	Mean Diff	Mean Diff
occ4	0.10	0.08	0.28	0.01	0.10	0.08	0.28	0.01	10.41
radiom	0.62	0.52	0.50	0.10	0.61	0.58	0.49	0.03	70.52
radiof	0.55	0.45	0.50	0.10	0.54	0.51	0.50	0.03	71.32
tvm	0.46	0.32	0.47	0.14	0.44	0.40	0.49	0.04	74.11
tvf	0.40	0.26	0.44	0.14	0.38	0.35	0.48	0.03	77.74
newsm	0.39	0.28	0.45	0.11	0.36	0.34	0.47	0.02	79.51
newsf	0.20	0.14	0.35	0.06	0.20	0.17	0.37	0.03	46.24
propold	0.06	0.05	0.09	0.01	0.06	0.06	0.10	0.00	76.89
malehead	0.95	0.97	0.18	−0.01	0.96	0.96	0.20	0.00	91.52
single	0.00	0.01	0.07	0.00	0.00	0.00	0.07	0.00	47.20
married	0.92	0.93	0.26	−0.01	0.92	0.92	0.28	0.00	100.00
hhead_ed1	0.43	0.52	0.50	−0.09	0.46	0.48	0.50	−0.02	77.13
hhead_ed2	0.29	0.24	0.43	0.05	0.28	0.29	0.45	−0.01	86.97
hhead_ed3	0.21	0.18	0.39	0.03	0.20	0.18	0.38	0.02	19.41
hhead_ed4	0.03	0.03	0.16	0.00	0.03	0.03	0.16	0.00	85.03
croparea	43.80	43.24	80.20	0.56	44.70	43.76	82.58	0.94	−69.78
irrarea	20.31	21.83	59.40	−1.52	21.99	21.16	54.78	0.84	44.76
landless	0.47	0.47	0.50	0.00	0.45	0.44	0.50	0.01	−218.81
landmarg	0.16	0.20	0.40	−0.04	0.16	0.17	0.38	−0.01	74.29
landsmall	0.10	0.10	0.31	−0.01	0.10	0.11	0.31	−0.01	−3.96
statecode	7.78	9.18	4.55	−1.40	8.12	8.12	4.23	0.00	100.00

Notes: %BI refers to percentage balance improvement. Common support was applied and a calliper of 0.001 propensity score (0.0425 standard deviation [SD]) that Jalan and Ravallion (2003) used. In total 5538 treated units (of 6905) and 12,800 control units (of 18,006) were matched. Although overall distance in propensity score improved, balance decreased in the following variables: village size, proportion of irrigation land, presence of bank in village, presence of market in village, caste, Christian, household size, ownership of cart or livestock, number of rooms, occupation, land ownership, and education of household head.

Appendix 5. Replication of Jalan and Ravallion (2003) with PSM

	J&R specification				F&M specification			
	Individual		Household		Individual		Household	
Outcome	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
(1) Nearest-five matching on OR (PS)								
All diarrhoea	−0.015	(0.005)*	−0.019	(0.007)*	−0.016	(0.005)*	−0.020	(0.007)*
Acute watery diarrhoea	−0.015	(0.005)*	−0.019	(0.007)*	−0.017	(0.004)*	−0.021	(0.006)*
Acute dysentery	−0.004	(0.001)*	−0.006	(0.002)*	−0.005	(0.001)*	−0.007	(0.002)*
Diarrhoeal duration	−0.046	(0.037)	..	..	−0.056	(0.036)	..	..
N_T	5666		3609		6655		4222	
N_C	17,462		11,399		17,854		11,652	
(2) Nearest-five matching on PS								
All diarrhoea	−0.010	(0.005)*	−0.012	(0.007)	−0.013	(0.005)*	−0.016	(0.007)*
Acute watery diarrhoea	−0.010	(0.004)*	−0.012	(0.006)*	−0.014	(0.004)*	−0.017	(0.006)*
Acute dysentery	−0.003	(0.001)*	−0.004	(0.002)*	−0.004	(0.001)*	−0.006	(0.002)*
Diarrhoeal duration	−0.025	(0.059)	..	..	−0.042	(0.035)	..	..
N_T	6743		4272		6873		4350	
N_C	17,857		11,648		17,960		11,721	

Notes: For household-level analyses, the outcome is the probability of having at least one case of child diarrhoea in the household. J&R, Jalan and Ravallion (2003); F&M, Fan and Mahal, present article. *Significance at $p < 0.05$. Analyses in this table used Stata SE 11, pscore program. Nearest five matching was not possible given insufficient numbers of control units. A calliper of 0.001 was used here.