RESEARCH ARTICLE

Prevalence and clustering of diarrhoea within households in India: some evidence from NFHS-4, 2015–16

Bevin Vijayan and Mala Ramanathan*

Achutha Menon Centre for Health Science Studies, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, Kerala, India
*Corresponding author. Email: mala@sctimst.ac.in

Abstract

Diarrhoeal disease is one of the major causes of morbidity and mortality in children and is usually measured at individual level. Shared household attributes, such as improved water supply and sanitation, expose those living in the same household to these same risk factors for diarrhoea. The occurrence of diarrhoea in two or more children in the same household is termed 'diarrhoea clustering'. The aim of this study was to examine the role of improved water supply and sanitation in the occurrence of diarrhoea, and the clustering of diarrhoea in households, among under-five children in India. Data were taken from the fourth round of the National Family and Health Survey (NFHS-4), a nationally representative survey which interviewed 699,686 women from 601,509 households in the country. If any child was reported to have diarrhoea in a household in the 2 weeks preceding the survey, the household was designated a diarrhoeal household. Household clustering of diarrhoea was defined the occurrence of diarrhoea in more than one child in households with two or more children. The analysis was done at the household level separately for diarrhoeal households and clustering of diarrhoea in households. The presence of clustering was tested using a chi-squared test. The overall prevalences of diarrhoea and clustering of diarrhoea were examined using exogenous variables. Odds ratios, standardized to allow comparison across categories, were computed. The household prevalence of diarrhoea was 12% and that of clustering of diarrhoea was 2.4%. About 6.5% of households contributed 12.6% of the total diarrhoeal cases. Access to safe water and sanitation was shown to have a great impact on reducing diarrhoeal prevalence and clustering across different household groups. Safe water alone had a greater impact on reducing the prevalence in the absence of improved sanitation when compared with the presence of improved sanitation. It may be possible to reduce the prevalence of diarrhoea in households by targeting those households with more than one child in the under-five age group with the provision of safe water and improved sanitation.

Keywords: Diarrhoea; Sanitation; Clustering

Introduction

Globally, 3 out of 1000 children under the age of five died due to diarrhoea in 2016, accounting for 8% of all under-five deaths. The situation in India was similar, with an under-five mortality rate due to diarrhoea of 4 per 1000 live births in 2016, representing 9% of total deaths among under-five children (UNICEF, 2018). The majority of these deaths occurred in the post-neonatal period, indicating a preponderance of exogenous causes of mortality. Most of these deaths were avoidable and were usually related to poor household conditions, including sanitation and water supply, and also poor access to health care facilities (Saha & van Soest, 2011). Improved health care access would mitigate this, but prevention would contribute a significant contribution towards reducing deaths due to diarrhoea.

Deaths in under-five children show significant heterogeneity in risk across households (Arulampalam & Bhalotra, 2006; Das Gupta, 1990; Ronsmans, 1995; Saha & van Soest, 2011; Shabani *et al.*, 2010). In particular, Das Gupta (1990) and Ronsmans (1995) noticed that a small proportion of households in India and rural Senegal contributed to a significant proportion of under-five children mortality, which is otherwise termed as 'death clustering'. This clustering or heterogeneity in the mortality experience across households could be due to many exogenous factors that are common to communities or households. Such factors could be household socio-economic status and the household's ability to provide for the survival of children through access to sanitation and safe drinking water (Edvinsson & Janssens, 2012) or the maternal capability of autonomous and effective functioning through education (Das Gupta, 1990; Gupta, 1997). Alternatively, the heterogeneity could be due to the effect of the death of a previous child affecting the subsequent child's mortality. This effect has been called 'state dependency' or 'scarring'. This process may be operationalized through death of a previous child resulting in shorter birth intervals for the succeeding infant or through maternal depression due to the previous child's death affecting maternal functioning and therefore care for the subsequent child. Alternatively, death of an infant may also result in a household learning to reduce similar deaths, resulting in a reduction in the scarring effect. A similar clustering effect has been noticed for diarrhoea in pre-school children (Katz *et al.*, 1993).

For acute conditions such as diarrhoea, the determination of the potential sources of heterogeneity in terms of within-household factors that are common to all children in the household (inter-household heterogeneity) and those due to individual characteristics of the child with diarrhoea (intra-household heterogeneity) may not be possible. Also, it may not possible to chronologically sequester the primary case of diarrhoea within the household to make a determination of its potential transmission to other children, and thus to be able to identify the original case of diarrhoea in a household with multiple cases. Moreover, it has been established that improved water supply and improved sanitary conditions have decisive effects on the prevalence of diarrhoea at the individual level (Cairncross *et al.*, 2010). Therefore, a focus on these exogenous factors with respect to diarrhoeal prevalence is still relevant.

Diarrhoeal diseases and their correlates are often conceptualized and measured at the individual level, whereas improved water supply and improved sanitation are shared household attributes that impact the health and well-being of all residents of a household. Children living in the same home environment are exposed to similar risks. This can be explained using the example of a child in a household suffering from diarrhoea. For this child it is possible that the other children in the household are also exposed to similar risks of getting diarrhoea. This child may or may not be a sibling sharing the same genetic factors. Apart from this, there are also the factors of caste, socio-economic status and place of residence, which are exogenous and have an effect on the occurrence of diarrhoea in the household and its clustering. An understanding of diarrhoea prevalence and its tendency to cluster within households, and the changes likely to be wrought in it through modifiable factors such as safe water and sanitation, will help to provide appropriate, targeted policies. This study aimed to examine the role of improved water supply and sanitation in the occurrence of diarrhoea and the clustering of diarrhoea in households among under-five children in India.

Methods

Data were from the fourth round of the National Family Health Survey (NFHS-4) conducted in India in 2015–2016 (IIPS & ICF, 2017). The NFHS-4 is a nationally representative multistage survey designed to provide estimates of vital indicators at the district, state and national level. A total of 699,686 women were interviewed from 601,509 households. The survey collected data on sanitation and improved source of drinking water using a household questionnaire. Information on diarrhoeal disease in under-five children was obtained using a women's questionnaire.

Households were classified as having 'improved sanitation' or an 'improved source of drinking water' using definitions taken from the Joint Monitoring Program (JMP) (WHO & UNICEF, 2018). An added refinement was that water was designated as 'safe' if some form of treatment was done before consumption. Having an improved source of drinking water alone will not guarantee protection from infection as it depends on the quality of water and the way it is stored. Treating water before drinking takes care of the quality. This improved drinking water variable has been labelled 'safe water'.

The information about the occurrence of diarrhoea in a child was matched with the household characteristics. If any child belonging to the same household had reported diarrhoea during the reference period then that household was designated a 'diarrhoeal household'. The number of children in the household that reported diarrhoea in the 14-day reference period prior to the date of the survey was recorded. Computation of the diarrhoea prevalence for clustering was carried out considering only eligible households, i.e. those with at least two children. This led to the exclusion of households with fewer than two children under the age of five years. Household clustering of diarrhoea was defined as the occurrence of diarrhoea in more than one child in households with two or more children.

The presence of clustering of diarrhoea within households was assessed by examining whether the prevalence of diarrhoea was conditioned by the number of children within a household. In the absence of clustering, the distribution of households with diarrhoea would be binomial and the probability of observing k cases of diarrhoea in a household with n children ($0 \leq k \leq n$) (Ronsmans, 1995) is given by:

$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \text{ for } 0 \leq k \leq n$$

The average number of children with diarrhoea in households with n children is np, and the variance $np(1-p)$. The value of 'p', is estimated from the observed proportion of households with diarrhoea in the sample for each category of number of children in households. The excess of the observed number of households with diarrhoea over the expected number of households would indicate clustering. The difference between the observed and expected number of households with clustering was tested using a χ^2 test to establish the statistical significance of the departure from expected number of children with diarrhoea within households. This helps determine the presence of clustering without ascertaining the exact cause for the departure.

The overall prevalence of diarrhoea in the household and prevalence of diarrhoea clustering were examined by the exogenous variables caste, wealth quintile and the place of residence.

The variables 'improved sanitation' and 'safe water' were combined to create four categories to manage potential interaction effects: Both present, Only sanitation, Only safe water and Both absent. Adjusted odds ratios for each household characteristic category, i.e. caste, wealth quintile, place of residence and presence of sanitation and safe water, were computed by dividing the odds of diarrhoea in each of the categories by the overall odds of diarrhoea. Similarly, adjusted odds ratios for clustering were computed in the same manner. Standardizing the odds in this way enabled a comparison across the different groups within each of the selected variables. This enabled comparisons across characteristics in terms of the increase or decrease in the odds for a characteristic against the same odds without the characteristic. This standardization procedure is similar to the Prevalence Odds Ratio (POR), but does not lend itself to estimates of the statistical significance of the effect through p-values. The POR is estimated by the cross-product of a 2×2 contingency table, which represents the odds of disease in exposed as against the odds of disease in the unexposed. Because the standardization is not through such a cross-tabulation of each characteristic against the prevalence of diarrhoea or diarrhoea clustering but through the computation of an alternative ratio, it is difficult to discern its equivalence to PORs. However, a parallel can be drawn to make comparisons of the specific mean against the overall mean in ordinary least squares regression.

The use of the Prevalence Odds Ratio (POR) or Prevalence Ratio (PR) has been the subject of debate in epidemiology (Lee, 1994; Thompson *et al.*, 1998), mostly related to the appropriateness of the measure, statistical tools and the availability of software programmes (Lee, 1994; Zocchetti *et al.*, 1997). The POR is preferred for chronic diseases with long latent periods whereas the PR is used for acute illness. The ideal measure to use in the case of acute diarrhoea would be the PR, because when the prevalence of the outcome is low the POR and PR are almost the same. The choice of the measure used should be made on epidemiological reasoning (Zocchetti *et al.*, 1997), and as it was anticipated that the prevalence of diarrhoea would not be very high the POR was used in this analysis.

Multinomial logistic regression analysis was used to predict the effect of the combined categories 'safe water' and 'improved sanitation' on the occurrence of diarrhoea and clustering of diarrhoea within households. This analysis was done for households with two or more children where clustering was possible. In the first model (Model 1) the outcome variable was divided into three categories, namely 'no diarrhoea', 'one child with diarrhoea' and 'two or more children with diarrhoea', which was then examined with the predictor variable of availability of water supply and sanitation within the household. In the second model (Model 2), the exogenous variables caste, wealth quintile and place of residence were also included. The results are presented as Relative Risk Ratios (RRR) with 95% confidence intervals. Data were analysed using Stata (Version 15) (StataCorp, 2018) and graphs created using the *ggplot2* package in R software (Wickham, 2009; R Core Team, 2017).

Results

There were a total of 176,531 households with children under five years of age with information on the occurrence of diarrhoea and 59,103 households were eligible for analysis of clustering wherein there was more than one child available in the household. The household prevalence of diarrhoea was about 12%. Furthermore, the prevalence of clustering (meaning more than one child in the household with diarrhoea during the 2-week reference period before the survey) was 2.4%. Of the 22,500 children with diarrhoea, 12.6% came from just 6.5% of the total households with diarrhoea, clearly indicating the contribution of multiple cases within the same households.

The probability of clustering was evaluated by examining the null hypothesis that there was no difference in the probability of diarrhoeal occurrence by the number of children within a household. The probability of diarrhoea occurrence across households with varying number of children below five years ranging between 2 and 7 was computed and contrasted against the expected probability of diarrhoea based on the overall probability of diarrhoeal prevalence among household (i.e. $p=0.1202$). The observed and expected probabilities are shown in Fig. 1. The expected probability of diarrhoea was computed for each number of children in households using the binomial distribution given by the formula given in the Methods section for $2 \leq k \leq 7$. Using these expected probabilities, the expected number of households with diarrhoea clustering was estimated as a product of the specific expected probability of diarrhoea and the number of households with 2–7 children. This was contrasted with the observed number of households with diarrhoea clustering.

The observed and expected numbers of cases of diarrhoea within households with two or more children were tested using the chi-squared test of independence with five degrees of freedom (to allow for households with 2–7 children). The computed value of the χ^2 statistic was 82.13, which was significant at $p<0.001$. This served to reject the null hypothesis that there was no possible clustering of diarrhoea across households, indicating the existence of a statistically significant level of clustering. However, it does not necessarily provide causal reasoning for the same.

The overall prevalence of diarrhoea and the prevalence of clustering of diarrhoea in households by household characteristics are shown in Table 1. The overall household diarrhoeal prevalence
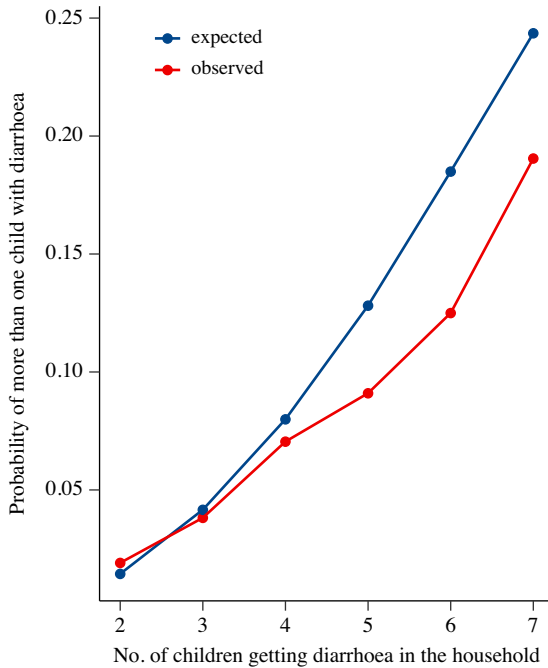
Figure 1. Expected and observed probabilities of diarrhoeal occurrence in children across households with 2 or more children, India, 2015–16.

increased as the number of children in the household increased. A similar trend was observed for clustering, with a higher prevalence occurring when the number of children in the household exceeded five.

The prevalences of overall diarrhoea and diarrhoea clustering in rural households were 12.7% and 2.5%, respectively, which were higher than those in urban areas. The 'Other' group representing the higher caste groups had a prevalence of 10.6%, which was low when compared with those of Schedule Caste, Schedule Tribe or Other Backward Caste groups. However, the prevalence of clustering was lowest for the Scheduled Tribe group. Scheduled Caste groups had the highest burden of overall diarrhoea (12.6%) and clustering (2.6%), which was higher than the overall prevalence of diarrhoea and prevalence of clustering. The prevalences of diarrhoea and clustering were higher in poorer households than in middle and higher quintile households and lowest in the richest quintile group (13.8% and 2.8%, respectively). This association was in the expected direction. Access to improved sanitation and safe water did have an impact on both the prevalence of diarrhoea and its clustering. When both improved sanitation and safe water were available, the prevalence of diarrhoea was 8.8% and the clustering of diarrhoea in households was 1.7% when compared with the absence of both (13.63 % and 2.7 %, respectively). When it came to clustering, safe water alone had more impact on reducing diarrhoea prevalence than sanitation alone.

The odds ratios (ORs) for overall prevalence of diarrhoea were computed across the different groups of sanitation and safe water, and were adjusted with the overall odds of diarrhoea to derive the Prevalence Odds Ratios (Table 2). A similar exercise was carried out for the household clustering of diarrhoea and the POR was computed against the overall odds of household clustering of diarrhoea (Table 3). The overall OR for diarrhoea in the household was 0.1366 and for household clustering of diarrhoea it was 0.0246. To enable an easy visual comparison, the POR for diarrhoea and diarrhoea clustering within households was represented for the two extreme conditions – when both safe drinking water and sanitation were present and when both were absent (Fig. 2).

The reduced risk in urban and rural areas due to the provisioning of sanitation and safe water was strongest in urban areas. However, the risk of diarrhoea was higher in rural areas when compared with urban areas when both were absent. When both sanitation and safe water were

Table 1. Overall prevalence of diarrhoea and prevalence of household clustering of diarrhoea in children aged 0–5 by household characteristics, India, 2015–16

Characteristic	Prevalence of diarrhoea (any) in households with at least one child n (%)[b]	Prevalence of household clustering of diarrhoea in households with more than one child[a] n (%)[b]	Total number of households
No. children in household			
1	11,250 (9.58)	—	117,428
2	7741 (15.97)	948 (2.0)	49,645
3	1585 (20.26)	299 (3.8)	7828
4	358 (29.14)	90 (7.0)	1276
5	88 (29.49)	25 (10.8)	275
6	19 (26.55)	7 (9.0)	56
7	8 (41.15)	4 (19.2)	21
8	0 (0.0)	0 (0.0)	1
9	0 (0.0)	0 (0.0)	1
Place of residence			
Urban	4868 (10.34)	289 (2.2)	44,354
Rural	16,181 (12.73)	1084 (2.5)	132,177
Caste			
Scheduled Caste	4347 (12.6)	302 (2.6)	33,095
Scheduled Tribe	3406 (10.95)	186 (1.9)	35,686
Other Backward Class	8690 (12.79)	562 (2.4)	67,025
Other	4606 (10.64)	314 (2.4)	40,725
Wealth quintile			
Poorest	5639 (13.81)	424 (2.8)	43,903
Poorer	5022 (12.7)	326 (2.5)	40,664
Middle	4264 (12.27)	257 (2.1)	35,417
Richer	3449 (10.87)	223 (2.3)	30,546
Richest	2675 (9.52)	143 (2.0)	26,001
Improved sanitation			
No	12,990 (13.23)	911 (2.6)	100,683
Yes	8059 (10.38)	462 (2.1)	75,848
Safe drinking water			
No	15,519 (12.94)	1091 (2.6)	119,716
Yes	5528 (9.87)	282 (1.8)	56,788
Safe water and sanitation			
Both present	3164 (8.83)	151 (1.7)	35,463
Only sanitation	4895 (11.6)	311 (2.3)	40,370
Only safe water	2364 (11.58)	131 (1.9)	21,325
Both absent	10,624 (13.63)	780 (2.7)	79,346
Total	21,049 (12.02)	1373 (2.4)	176,531

[a]The total number of households for the calculation of clustering of diarrhoea was 59,103.
[b]The numbers in brackets are weighted percentages.

Table 2. Prevalence and Prevalence Odds Ratios (PORs) of diarrhoea in households with at least one child by socioeconomic characteristics, India, 2015–16

Characteristic	Sanitation and safe water	Prevalence of diarrhoea (any) in households with at least one child	Odds of diarrhoea	Adjusted/Prevalence Odds ratios
Place of residence				
Urban	Both present	8.4	0.092	0.671
	Only sanitation	10.9	0.123	0.899
	Only safe water	10.8	0.122	0.891
	Both absent	12.3	0.141	1.031
Rural	Both present	9.4	0.104	0.760
	Only sanitation	12.0	0.136	0.998
	Only safe water	11.9	0.135	0.986
	Both absent	13.8	0.161	1.176
Caste				
Scheduled Caste	Both present	10.9	0.122	0.892
	Only sanitation	10.7	0.120	0.879
	Only safe water	12.8	0.147	1.075
	Both absent	13.6	0.158	1.153
Scheduled Tribe	Both present	9.3	0.102	0.749
	Only sanitation	9.8	0.109	0.798
	Only safe water	13.1	0.150	1.102
	Both absent	10.8	0.121	0.885
Other Backward Class	Both present	9.1	0.100	0.732
	Only sanitation	12.6	0.144	1.054
	Only safe water	11.0	0.124	0.909
	Both absent	14.8	0.173	1.268
Other	Both present	7.8	0.084	0.616
	Only sanitation	11.1	0.124	0.909
	Only safe water	10.2	0.114	0.833
	Both absent	12.9	0.148	1.083
Wealth quintile				
Poorest	Both present	8.3	0.091	0.663
	Only sanitation	10.5	0.118	0.861
	Only safe water	12.1	0.138	1.011
	Both absent	14.2	0.166	1.212
Poorer	Both present	9.6	0.106	0.778
	Only sanitation	11.9	0.136	0.993
	Only safe water	11.9	0.134	0.984
	Both absent	13.3	0.154	1.125

Table 2. (*Continued*)

Characteristic	Sanitation and safe water	Prevalence of diarrhoea (any) in households with at least one child	Odds of diarrhoea	Adjusted/Prevalence Odds ratios
Middle	Both present	11.1	0.125	0.917
	Only sanitation	12.2	0.139	1.019
	Only safe water	11.8	0.133	0.976
	Both absent	12.9	0.148	1.080
Richer	Both present	8.9	0.098	0.716
	Only sanitation	11.4	0.128	0.939
	Only safe water	10.6	0.119	0.871
	Both absent	13.3	0.153	1.120
Richest	Both present	8.1	0.088	0.645
	Only sanitation	11.3	0.127	0.933
	Only safe water	10.0	0.112	0.818
	Both absent	13.6	0.157	1.148

available to the household these tended to protect against the occurrence of diarrhoea. This reduction was observed in all caste groups except the Schedule Caste, where the reduction was only 11%. The reduction ranged from 25.1% in the Scheduled Tribes to 38.4 % in the Other (higher caste) category. However, in households with only sanitation, the risk of diarrhoea reduced by almost 10% in Schedule Caste and Other and 20% among Scheduled Tribes, but such reductions were not evident in Other Backward Class households. Having safe water reduced the risk in only the Other Backward Class and Other (higher caste) category households. Not having both increased the risk in all the groups except Scheduled Tribes.

The risk of household clustering of diarrhoea (Table 3) was reduced when there was provisioning of both improved sanitation and safe water and the reduction was almost 50% in rural areas. When both sanitation and safe water were absent the risk increased by 15% in rural areas compared with 5% in urban areas. In households with only safe water the risk reduction was greater in urban than in rural areas. The protection against clustering of episodes of diarrhoea was stronger in rural areas than urban areas, due to the presence of these two components of public health infrastructure.

The provisioning of safe water and sanitation was not effective among the Scheduled Caste group for the prevention of clustering episodes of diarrhoea. One of the possible reasons for this could be the heterogeneous nature of Scheduled Caste groups. In all other caste groups, the presence of sanitation and safe drinking water had a protective effect. The protective effect ranged from 73% in Scheduled Tribes to 23% in Other Backward Classes. In the Scheduled Tribe group, not having safe water and sanitation did not seem to increase the risk of clustering.

The presence of sanitation and safe water had a protective effect across all income groups with respect to clustering of diarrhoea in the same household. These benefits were to the tune of a decrease in adjusted/prevalence odds from 1.227 to 0.357 and from 1.115 to 0.303 among the poorest and the poorer groups on comparing the absence of sanitation and safe water with their joint presence within the household. A very significant improvement was noticed in the richest households, where the adjusted/prevalence odds declined from 2.077 to 0.674. However, it should be noted that the richest households accounted for 10% of the total households with diarrhoea clustering, and this could be the reason for the relatively larger decline in the odds of clustering depicted.
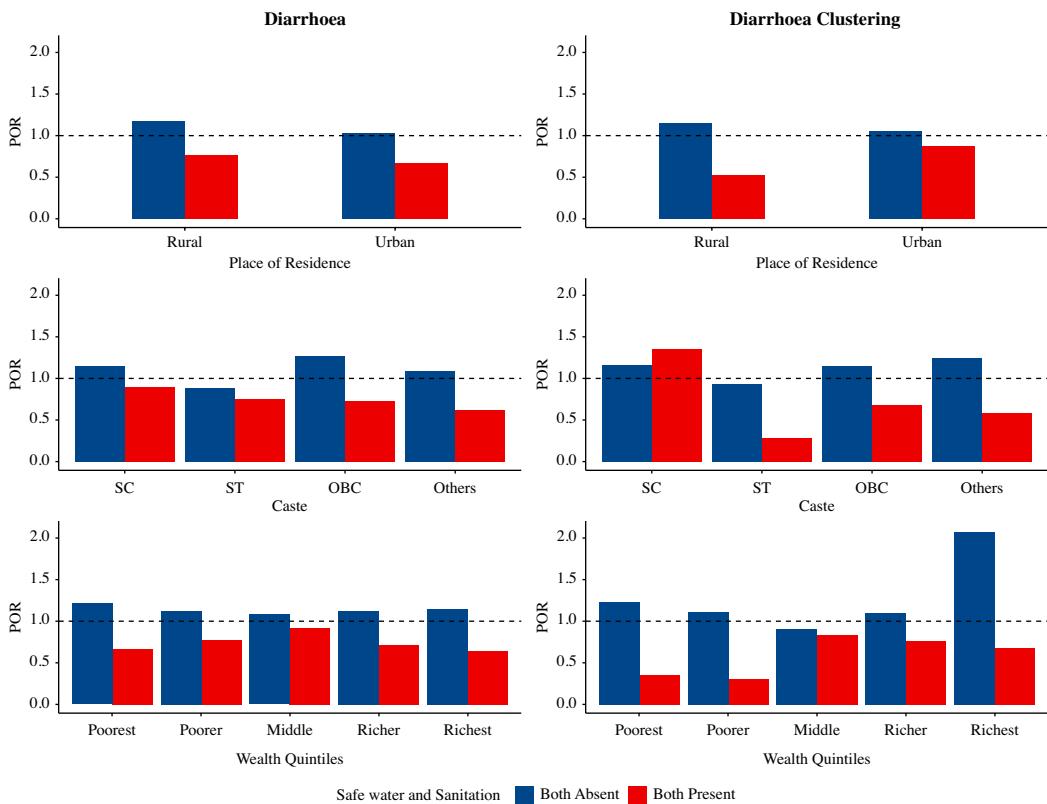
Table 3. Prevalence and Prevalence Odds Ratios (PORs) of household clustering of diarrhoea in households with at least one child by socioeconomic characteristics, India, 2015–16

Characteristic	Sanitation and safe water	Prevalence of diarrhoea (any) in households with at least one child	Odds of Diarrhoea	Adjusted/Prevalence Odds ratios
Place of residence				
Urban	Both present	2.1	0.021	0.872
	Only sanitation	2.3	0.024	0.957
	Only safe water	1.3	0.013	0.540
	Both absent	2.5	0.026	1.051
Rural	Both present	1.3	0.013	0.523
	Only sanitation	2.3	0.023	0.953
	Only safe water	2.1	0.021	0.868
	Both absent	2.8	0.028	1.154
Caste				
Scheduled Caste	Both present	3.2	0.033	1.348
	Only sanitation	2.1	0.022	0.880
	Only safe water	2.1	0.021	0.855
	Both absent	2.8	0.029	1.162
Scheduled Tribe	Both present	0.7	0.007	0.278
	Only sanitation	1.5	0.015	0.602
	Only safe water	1.7	0.017	0.703
	Both absent	2.2	0.023	0.931
Other Backward Class	Both present	1.6	0.017	0.678
	Only sanitation	2.3	0.024	0.957
	Only safe water	1.9	0.019	0.783
	Both absent	2.8	0.028	1.154
Other	Both present	1.4	0.014	0.586
	Only sanitation	2.5	0.026	1.059
	Only safe water	2.0	0.020	0.821
	Both absent	3.0	0.031	1.240
Wealth quintile				
Poorest	Both present	0.9	0.009	0.357
	Only sanitation	2.2	0.023	0.914
	Only safe water	1.8	0.019	0.762
	Both absent	2.9	0.030	1.227
Poorer	Both present	0.7	0.008	0.303
	Only sanitation	1.8	0.018	0.745
	Only safe water	2.6	0.026	1.072
	Both absent	2.7	0.027	1.115

(*Continued*)

Table 3. (*Continued*)

Characteristic	Sanitation and safe water	Prevalence of diarrhoea (any) in households with at least one child	Odds of Diarrhoea	Adjusted/Prevalence Odds ratios
Middle	Both present	2.0	0.020	0.830
	Only sanitation	2.4	0.025	1.004
	Only safe water	1.3	0.013	0.519
	Both absent	2.2	0.022	0.910
Richer	Both present	1.9	0.019	0.766
	Only sanitation	2.6	0.026	1.068
	Only safe water	1.8	0.018	0.745
	Both absent	2.6	0.027	1.098
Richest	Both present	1.6	0.017	0.674
	Only sanitation	2.2	0.022	0.893
	Only safe water	1.7	0.018	0.716
	Both absent	4.9	0.051	2.077

Figure 2. Prevalence odds ratios of diarrhoea and clustering of diarrhoea in households with at least one child by socioeconomic characteristics and presence and absence of sanitation and safe water, India, 2015–16.

Table 4. Multinomial logistic regression of the effect of safe water and sanitation on the clustering of diarrhoea in households with two or more children, India, 2015–16

	Model 1		Model 2	
Safe water and sanitation	One child with diarrhoea RRR (95% CI)	Two or more children with diarrhoea RRR (95% CI)	One child with diarrhoea RRR (95% CI)	Two or more children with diarrhoea RRR (95% CI)
Both present (Ref.)				
Only sanitation	1.331*** (1.229–1.442)	1.509*** (1.24–1.837)	1.269*** (1.169–1.378)	1.438*** (1.176–1.759)
Only safe water	1.097 (0.997–1.206)	1.158 (0.914–1.467)	1.119** (1.009–1.24)	1.259 (0.977–1.622)
Both absent	1.371*** (1.276–1.472)	1.672*** (1.402–1.993)	1.348*** (1.236–1.47)	1.74*** (1.408–2.149)

The reference category for the outcome variable was 'No diarrhoea in the household'. Model 1 examined the effect of water and sanitation on the occurrence of diarrhoea in a household and Model 2 examined the same effect adjusted for caste, place of residence and wealth quintiles. ***$p<0.01$; **$p<0.05$.

The benefits of access to sanitation and safe drinking water are nearly universal across rural and urban areas, across almost all caste groups and across all economic status groups (Fig. 2). This provides a definite advantage against clustering of diarrhoea in households compared with the prevalence of diarrhoea within households in rural areas. The presence of sanitation and safe drinking water also serves to depress the prevalence of diarrhoea clustering in Scheduled Tribes and other groups when compared with the prevalence of diarrhoea. This reduction in diarrhoea clustering was stark when both sanitation and safe drinking water were available among the richest households.

The results of the multinomial logistic regression presented in Table 4 using the unadjusted Model 1 indicate that having access to only improved sanitation increased the risk of a single child having diarrhoea in households having two or more children by 33%, and when both improved sanitation and safe water were absent the risk increased by 37%. This was in comparison to having both improved sanitation and safe water in the household. Having access to only safe water increased the risk by 10% but this was not statistically significant. Similar results were observed when two or more children in the household experienced diarrhoea episodes. However, the risk increased by almost 51% when households had access to only improved sanitation, and when both were absent the risk of clustering increased by almost 67% when compared with households that had both. Model 2, which adjusted for socioeconomic characteristics of the household, indicated an adjustment in the risk of clustering of diarrhoea to 44% from 51% (unadjusted risk ratios) when only sanitation was present, 26% from 16% (unadjusted risk ratios) when only safe water was present and 74% from 67% (unadjusted risk ratios) when both were absent. It should be noted that some of the cells for computing the adjusted risk ratios were the result of a very small number of cases and cannot be as robust. This is a serious limitation for the multivariate analysis undertaken. In so far as these adjusted risk ratios did not contradict the emerging patterns from the unadjusted model, the underlying associations indicated held.

Discussion

This study demonstrates the potential for diarrhoea to cluster in households with more than one child in India. While its prevalence in the present sample extended to just 2.4% of households with two or more children, it represents a significant departure from an assumption of uniform prevalence of diarrhoea across households. There is clearly a potential for diarrhoea to cluster in

households with multiple children. It may be possible to reduce the prevalence of diarrhoea in households by targeting those households with more than one child in the under-five age group, and by providing them with safe water and improved sanitation.

The presence of both sanitation and safe water has a strong effect in reducing both diarrhoea prevalence and its clustering within households, particularly in rural areas, amongst Scheduled Tribe and Other caste groups and among the poorest and richest households. Safe water alone has a greater impact in reducing the prevalence of diarrhoea and household clustering of diarrhoea in the absence of improved sanitation when compared with the presence of improved sanitation. The effect of the presence of safe drinking water in the household alone is nearly as good as having both sanitation and safe drinking water.

The relevance of universalization of access to safe drinking water to diarrhoea prevention among under-five children has been established in previous studies using other data and methods (Nandi *et al.*, 2017; Nilima *et al.*, 2018; Ramanathan & Vijayan, 2019). The provision of piped water to rural households planned under the Jal Jeevan Mission of the Government of India (Ministry of Jal Shakthi, 2019) could contribute to the reduction in diarrhoea prevalence and its clustering in rural areas of India. This is not to say that access to improved sanitation is not important. In areas that are yet to receive sanitation improvement schemes there is a need to advocate for the use of safe water for consumption and to continue to emphasize the importance of this, even in areas where there is improved sanitation. When combined with the Swachh Bharat Mission for universalizing sanitation access (Ministry of Drinking Water and Sanitation, 2017), the impact on the household prevalence of diarrhoea among under-five children in India would be very significant because the present analysis indicates a joint potential to reduce diarrhoea clustering within households by about 75%, even after controlling for the socioeconomic conditions of the household.

References

Arulampalam W and Bhalotra S (2006) Sibling death clustering in India: state dependence versus unobserved heterogeneity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(4), 829–848.

Cairncross S, Hunt C, Boisson S, Bostoen K, Curtis V, Fung IC and Schmidt WP (2010) Water, sanitation and hygiene for the prevention of diarrhoea. *International Journal of Epidemiology* **39** (Supplement 1), i193–205.

Das Gupta M (1990) Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India. *Population Studies* **44**(3), 489–505.

Edvinsson S and Janssens A (2012) Clustering of deaths in families: infant and child mortality in historical perspective. *Biodemography and Social Biology* **58**(2), 75–86.

Gupta MD (1997) Socio-economic status and clustering of child deaths in rural Punjab. *Population Studies* **51**(2), 191–202.

IIPS and ICF (2017) *India National Family Health Survey NFHS-4 2015–16.* IAKR74DT.DTA, IAHR74DT.DTA. International Institute for Population Sciences and ICF. URL: http://dhsprogram.com (accessed 28th August 2018).

Katz J, Carey VJ, Zeger SL and Sommer A (1993) Estimation of design effects and diarrhea clustering within households and villages. *American Journal of Epidemiology* **138**(11), 994–1006.

Lee J (1994) Odds ratio or relative risk for cross-sectional data? *International Journal of Epidemiology* **23**(1), 201–203.

Ministry of Drinking Water and Sanitation (2017) *Guidelines for Swachh Bharat Mission (Gramin).* URL: http://swachhbharatmission.gov.in/sbmcms/writereaddata/images/pdf/Guidelines/Complete-set-guidelines.pdf.

Ministry of Jal Shakthi (2019) *Jal Jeevan Mission (JJM). Department of Drinking Water and Sanitation.* URL: https://jalshakti-ddws.gov.in/sites/default/files/JJM_note.pdf (accessed 7th December 2019).

Nandi A, Megiddo I, Ashok A, Verma A and Laxminarayan R (2017) Reduced burden of childhood diarrheal diseases through increased access to water and sanitation in India: a modeling analysis. *Social Science & Medicine* **180**, 181–192.

Nilima, Kamath A, Shetty K, Unnikrishnan B, Kaushik S and Rai SN (2018) Prevalence, patterns, and predictors of diarrhea: a spatial-temporal comprehensive evaluation in India. *BMC Public Health* **18**(1), 1288.

R Core Team (2017) *R Development Core Team.* URL: https://www.r-project.org/.

Ramanathan M and Vijayan B (2019) Covariates of diarrhoea among under-five children in India: are they level dependent? *PLoS One* **14**(8), e0221200.

Ronsmans C (1995) Patterns of clustering of child mortality in a rural area of Senegal. *Population Studies* **49**(3), 443–461.

Saha UR and van Soest A (2011) Infant death clustering in families: magnitude, causes, and the influence of better health services, Bangladesh 1982–2005. *Population Studies* **65**(3), 273–287.

Shabani J, Lutambi AM, Mwakalinga V and Masanja H (2010) Clustering of under-five mortality in Rufiji Health and Demographic Surveillance System in rural Tanzania. *Global Health Action* **3**(1), 5264.

StataCorp (2018) *Stata Statistical Software: Release 15.* Stata Corporation.

Thompson ML, Myers JE and Kriebel D (1998) Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? *Occupational and Environmental Medicine* **55**(4), 272–277.

UNICEF (2018) *UNICEF Data: Monitoring the Situation of Children and Women.* URL: https://data.unicef.org/topic/child-health/diarrhoeal-disease/ (accessed 2nd November 2018).

Wickham H (2009) *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag, New York. URL: http://ggplot2.org.

WHO and UNICEF (2018) *JMP Methodology : 2017 Update & SDG Baselines.* WHO, Geneva. URL: https://washdata.org/sites/default/files/documents/reports/2018-04/JMP-2017-update-methodology.pdf (accessed 2nd November 2018).

Zocchetti C, Consonni D and Bertazzi PA (1997) Relationship between prevalence rate ratios and odds ratios in cross-sectional studies. *International Journal of Epidemiology* **26**(1), 220–223.