Health effects of sanitation facilities: A Bayesian semi-parametric analysis of compositional data

Samuel I. Watson

University of Warwick, Coventry, United Kingdom.

E-mail: s.watson.1@warwick.ac.uk

Jo Sartori

University of Warwick, Coventry, United Kingdom.

Olalekan Uthman

University of Warwick, Coventry, United Kingdom.

Richard J. Lilford

University of Warwick, Coventry, United Kingdom.

Summary. Diarrhoeal disease is one of the leading causes of illness and death in young children, a problem exacerbated by a lack of access to safe sanitation facilities. But the effects of different types of sanitation facilities, and the relationship between their levels of coverage in an area and disease risk remain uncertain. We specify a hierarchical model that allows us to estimate the disease risk response surface across the multi-dimensional space of the composition of sanitation facilities in an area. This permits a non-linear relationship between coverage of improved sanitation, differential effects by sanitation type, and unobserved differences between countries. The model is used to estimate the change in risk associated with different types of sanitation facility, from which it is estimated that only increases in the coverage of sewerage and private improved sanitation facilities will achieve clinically meaningful reductions in the risk of adverse outcomes.

1. Introduction

The increasing pace of urbanisation over the last few decades has led to the substantial growth of slums and informal settlements. It is estimated over one billion people will soon

live in such areas world wide, with almost all of these in low and middle income countries (LMICs) (Ezeh et al., 2016). Africa is predicted to become 56% urban by 2050 (UN-HABITAT, 2014). As a result there has been an increase in the prevalence of health and social issues associated with living in slum conditions (Ezeh et al., 2016). A lack of clean water and effective sanitation has meant childhood diarrhoea is now one of the leading causes of death in the under fives, and the single greatest cause in urban areas (Naghavi et al., 2015). Goal 6 of the United Nations's Sustainable Development Goals (SDGs) is to "ensure availability and sustainable management of water and sanitation for all." (UN, 2018). However, in many urban areas worldwide there has been a notable lack of progress in achieving this goal. Of the 2.6 billion people living without 'improved' sanitation worldwide, around 800 million are in urban areas, a figure set to rise substantially (WHO/UNICEF, 2014).

The 'Gold standard' of sanitation facilities is often considered the private flush toilet connected to a sewer system that transports waste to treatment plants or other means of disposal. However, these systems have a relatively high capital cost to install, and require a constant supply of water and maintenance to remain effective. These barriers can be prohibitive in lower income countries. Alternative solutions often involve improvements to the basic latrine by, for example: adding ventilation to reduce the presence of flies; improving the slab to facilitate use, enable flushing, and cover the pit; or by improving the pit to allow for composting or leaching. And while latrines with improvements linked to sewers, septic pits, and tankers, are collectively referred to as 'improved' sanitation (WHO/UNICEF, 2014), they represent a broad range of interventions each likely have a different level of effectiveness in reducing adverse endpoints. The interventions are even ranked in a 'ladder' by international agencies (WHO/UNICEF, 2014).

Owing to the transmissible nature of the diseases and resulting disease dynamics, the type of sanitation facility an individual uses is perhaps less important than the type of facility used by the community around them (Geruso and Spears, 2014). Transmissible diseases often exhibit non-linear relationships with the level of coverage of preventative measures. Despite this, the majority of previous empirical work in the area has examined the effect of dichotomous 'treatments' either at the household level - whether the

household has access to 'improved' sanitation - or at the cluster level - whether the level of 'improved' sanitation coverage is at a certain level, which is often case for cluster randomised controlled trials (cRCT).

Two comprehensive systematic reviews have been published on the topic, Clasen et al. (2010) surveyed only trials, including 13 cRCTs, while the more recent Wolf et al. (2018) also included observational studies (this was an update of their earlier review Wolf et al. (2014)), and compared 22 results from 19 studies. Cluster trials in this area typically investigate an intervention of providing a high level of coverage of improved latrines (including ventilated latrines, slab latrines, latrines connected to sewers, etc.) compared to the status quo 'standard' sanitation, which could be anywhere between 0 and 100% improved latrines (e.g. Pickering et al. (2015); Luby et al. (2018); Clasen et al. (2014)), depending on the context. Statistical inference is then on the basis of either unadjusted or adjusted comparisons, of treatment and control clusters or households, of risk ratios for diarrhoea typically among young children.

Methods of analysis used for observational data are similarly limited. Most studies are retrospective analyses of household survey data, particularly single-country Demographic and Health Surveys (DHS). Fan and Mahal (2011); Kumar and Vollmer (2013); Begum et al. (2011a); and Capuno et al. (2015) used DHS (or similar household level survey) data and propensity score matching methods to compare households with 'improved sanitation' (typically improved latrines) to households with unimproved sanitation. Komarulzaman et al. (2017) similarly used an adjusted logistic regression approach and Godfrey et al. (2014) use a difference-in-differences method to estimate the effect of improved sanitation on diarrhoea risk at the household level. Aziz et al. (1990); Garrett et al. (2008); Moraes et al. (2003); and Messou et al. (1997) conducted non-experimental intervention studies and, with the exception of Moraes et al. (2003), who use longitudinal data and reasonably comprehensive adjustment, all evaluate the effect of their interventions with crude before-after analyses of a diarrhoea risk ratio. As far as we are aware, most if not all studies are conducted at household or similar level, and compare individual sanitation access with no assessment of the effect of overall community coverage, except Kumar and Vollmer (2013), who use socio-economic status as a proxy for

community coverage. Similarly, different types of sanitation are not compared directly in multi-intervention studies and the privacy or level of sharing is not assessed. In a now retracted study, due to data coding errors, Hunter and Prüss-Ustün (2016) did look at the effect of coverage of improved sanitation at the country level using a semi-parametric fixed-effects modelling approach.

There are two aspects of sanitation that remain understudied. The effect of the *type* of facility in the community, despite differences in effectiveness by type being assumed by agencies such as the WHO, and their *coverage* in the community. At the community level there is an overall composition of sanitation facilities: the proportions of people using each of a number of different types of facility add up to one and thus form a simplex. Analysis of compositional data, either as outcome or independent variable in a model, has a long history in other fields, particularly geosciences. Pearson (1896) was one of the first to point out the difficulty in analysing data of this type due to the 'spurious correlation' created due to the necessary change in all elements of compositional data when any one element changes. Appropriate methods for compositional data analysis have therefore been developed over time. Examples in public health are relatively rare, but are finding increasing application, for example Trinh et al. (2018) explore the role of socio-economic factors on the macronutrient composition of diet in Vietnam. In this article, we propose to try to address these issues to explore the effectiveness of the composition of sanitation facilities in urban communities in Sub-Saharan Africa.

2. Data description

The sample used in this study comprises all urban clusters in Sub-Saharan Africa from the DHS for which household, individual, and location data could be matched and which surveyed both sanitation type and whether the sanitation was shared. The DHS program conducts or supports nationally representative surveys across the world. The surveys cover a broad range of demographic, health, and socioeconomic data. The DHS comprises surveys conducted with households located in clusters, selected through a probability sampling approach using the most recent national census as a sampling frame. The specific DHS surveys included are listed in Table A in the Supplementary

Information. In total, data from 7,209 clusters from 29 countries between 1992 and 2015 were extracted. With two exceptions, surveys prior to 2000 did not ask about sanitation sharing.

2.1. Outcome variables

Three outcome variables are used that relate to adverse health outcomes likely to be affected by poor sanitation. In the DHS data on these variables are only captured for children under the age of five. The outcome variables are: at least one episode of diarrhoea in the last two weeks, at least one episode of bloody diarrhoea (dysentery) in the last two weeks, and childhood stunting, which is defined as being below the fifth percentile of height for age. For each cluster and year we derive the number of children with each outcome and total number of eligible children.

2.2. Sanitation variables

From the DHS we extracted the type of sanitation facility used by each household and whether or not it was shared with other households. Typically, empirical studies of sanitation facilities compare 'improved' to 'unimproved' facilities. However, we are concerned here also with the type of sanitation facility and how it is used since there may be different levels of effectiveness associated with each type or use. Aggregated to the cluster level, the types of facilities used by the household are a composition, the analysis of which we discuss in Section 3.1. We consider two ways of categorising the sanitation facilities. Firstly, we define three types of sanitation facility: unimproved, improved latrines, and sewers (the 'unimproved-latrine-sewer' composition). 'Sewer' facilities include any response where the household indicates the use of a flush toilet or piped sewerage system. 'Improved latrines' include latrines installed with a slab and ventilated improved latrines. 'Unimproved' generally includes all other facilities, in particular outdoor defecation and pit latrines. Secondly, we break the composition into: unimproved, shared improved facilities, and private improved facilities (the 'unimproved-shared-private' composition). Sanitation facilities are often not considered fully safe or improved if they are shared between many households (WHO/UNICEF, 2018). 'Improved' includes sewers

and improved latrines as defined above.

2.3. Other variables

Explanatory variables were extracted from the DHS on the basis of previous studies that have examined potential correlates with childhood diarrhoea and stunting (Carlton et al., 2016; Bhavnani et al., 2014; Genser et al., 2006; Walker et al., 2007). For each household we extracted the number of children in the household, the average age of the children in the household, a binary variable indicating whether the mother had completed at least primary education, and a binary variable indicating whether the household was in the top two wealth quintiles. We aggregated these data to the cluster level by taking arithmetic means.

On the basis of the location of the cluster we also determined the average rainfall and temperature for the month in which the survey took place. NASA provide model and observation based estimates on very fine spatial (3 arc-second, ∼5km) and temporal (3-hourly or monthly) grids from the Global Land Data Assimilation System (GLDAS). We extracted data on the period 1990 to 2015 to determine average rainfall and temperature. These climate variables are intended to control for some of the seasonal and geospatial variation in weather and climate that has previously been shown to be correlated with risk of diarrhoea in children (Bhavnani et al., 2014; Carlton et al., 2016). To preserve anonymity the DHS randomly displaces the location of urban clusters by up to 2km. Given the resolution of the spatial grid on which the climate data are provided, we do not envisage this to lead to any erroneous inferences. However, some clusters are randomly displaced into bodies of water for which climate data are not provided - for these clusters we assign average rainfall and temperature from the nearest land-based cluster location.

Access to 'improved' water sources may also have an effect on the outcomes we are considering and be correlated with the availability of different types of sanitation. For example, for sewage systems to be effective there is a minimum amount of water required to flow through the system, thus necessitating piped water. However, issues about what constitutes 'improved' water are at least as complicated as for sanitation. While piped

water is generally considered the highest standard, there are many instances of broken or damaged pipes leading to unsafe water. And there is a multitude of different water facilities, including tubewells, boreholes, standpipes, and pipes. Water is not the focus of this study, and so we define a single variable indicating the proportion of household with access to what is typically described as 'improved' water.

3. Statistical Methods

The analysis is carried out in a hierarchical Bayesian framework at the cluster level. The aim of this study is to estimate the effect of different levels of coverage of different types of sanitation, however this is complicated by a number of issues. At an aggregate level the mix of sanitation facilities in an community forms a composition, which cannot be treated as a simple multivariate object despite being *intrinsically* multivariate since the effect of any one part of the composition cannot be interpreted in the absence of the other components. The functional form of the response surface across the simplex is also unknown and may well be linear with respect to any of the components.

3.1. Compositional data

An observation $\mathbf{x}$ of a D-part compositional variable $\mathbf{X}$ consists of a D-dimensional vector belonging to the simplex,

$$\mathcal{S}^D = \left\{ (x_1, ..., x_d, ..., x_D) : x_d > 0, d = 1, ..., D; \sum_{d=1}^{D} x_d = c \right\}$$

where c is positive constant. At issue with the analysis of compositional data is that the components only carry relative information so that no one component can be interpreted in the absence of any other. This hinders interpretation of results from multivariate regression approaches. However, given that the components provide relative information, Aitchison (1982) proposed log-ratio transformations of compositions, which mapped the simplex to a real space. For example, a centered log-ratio transformation is often used in a regression framework as each coefficient can be related to an original component. However, as Bruno et al. (2016) argue the centered log-ratio is inappropriate in non-parametric settings as it generates a singular design matrix that requires a sum-to-zero

constraint for model estimation. So following Bruno et al. (2016) we adopt an isometric log-ratio (ilr) transformation, which defines an isometry between $\mathcal{S}^D$ and $\mathbb{R}^{D-1}$. The ilr transformed variable is $\mathbf{w} = ilr(\mathbf{x}) = [w_1, w_2, ..., w_{D-1}]$ where

$$w_i = \frac{1}{\sqrt{i(i+1)}} log\left(\frac{\prod_{j=1}^{i} x_j}{(x_i+1)i}\right).$$

Since ilr preserves distances between the components of the composition, it can be used to estimate the relationship between the composition and the response variable of interest. As the transformation requires all components of the composition to be in $(0, 1)$, i.e. strictly greater than zero and less than one, we set zero-components to 0.65 times the smallest non-zero value for the component, which we consider to be a lower detectable limit, as suggested by Aitchison (1982). The same amount was subtracted from components equal to one. In this analysis we consider a three dimensional simplex and the ilr transformation therefore defines a two dimensional variable.

3.2. Model Specification

The analysis is conducted at the cluster level to account for the aggregate effects of sanitation compositions on infectious disease. For each cluster $j = 1, ..., J$ at time $t = 1, ..., T$ in country $k = 1, ..., K$ we have the number of cases of the outcome of interest (diarrhoea, dysentery, or stunting) $y_{jkt} \in \mathbb{N}$ and the number of children under five $n_{jkt} \in \mathbb{N}$. We specify a binomial model with logisitic link function:

$$\begin{aligned} y_{jkt} &\sim Binomial(n_{jkt}, p_{jkt}) \\ p_{jkt} &= \Lambda\left[z'_{1jkt}\gamma + \alpha_j + \tau_t + g(\mathbf{x}_{jkt})\right] \end{aligned} \tag{1}$$

where z_{1jkt} is a $p \times 1$ vector of mean-centered standardised explanatory covariates (see Section 2.3) including an intercept term, γ is a $p \times 1$ vector parameters, $\alpha_j \sim N(0, \sigma_\alpha^2)$ is a country-specific random effect, τ_t are year 'fixed' effects. Time dummies were specified into five year periods (1991-5, 1996-2000, 2001-5, 2006-10, and 2010-16) given the high correlation between individual years and countries, and the low and zero outcome counts for some years for rare outcomes, such as dysentery.

There are a number of choices for the specification of the function $g(.)$:

(a) *Dichotomous regressor*: A dichotomous 'treatment' variable defined on the basis of the level of 'improved' sanitation, $g(\mathbf{x}_{jkt}) = \Delta T_{jkt}$, where T_{jkt} is equal to one if there is greater than, say, 50% coverage of any improved sanitation and zero otherwise, and Δ is a parameter to be estimated.

(b) *Linear* ilr *transformed components*: *ilr* components specified as additive, linear effects: $g(\mathbf{x}_{jkt}) = \Delta_1 w_{1,jkt} + \Delta_2 w_{2,jkt}$, where Δ_1, Δ_2 are parameters to be estimated.

(c) *Univariate smooth functions*: Additive non-parametric smooth functions of each of the *ilr*-transformed variables: $g(\mathbf{x}_{jkt}) = h_1(w_{1,jkt}) + h_2(w_{2,jkt})$. The univariate functions are estimated using a low rank thin plate spline, which are a low rank approximation to a penalised spline with penalty defined by a solution to the thin plate smoothing problem (Wood, 2003).

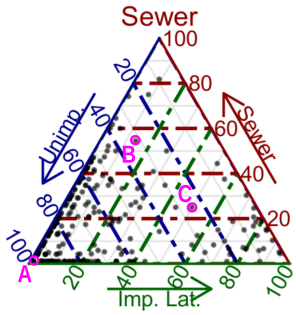
(d) *Bivariate smooth function*: a bivariate smooth surface: $g(\mathbf{x}_{jkt}) = h(\mathbf{w}_{jkt})$. The bivariate function is determined by the tensor product of the univariate thin plate smooth functions (Wood, 2003).

The choice among possible specifications for $g(.)$ is a trade-off. The dichotomous regressor approach, which reflects the method of many previous studies, may introduce bias and show poor predictive performance. However, while the bivariate smooth specification imposes the least structure, it may add significant computational complexity to the model without improving model fit or predictive performance. To select among the specifications we conduct a short simulation study and a series of model checks in Section 4. For comparison we also estimate 'unadjusted' models (i.e. $\gamma = 0$).

3.3. Prior distributions

Weakly informative priors are specified for the model parameters. Following Gelman et al. (2008) continuous covariates are rescaled to have zero mean and 0.5 standard deviation. Unlike 'uninformative priors' which specify an equal probability mass on extreme values of parameters and values near zero, 'weakly informative' priors are intended to provide a degree of regularization and stabilize computation while providing little information about the location of the parameter. Coefficients for covariates are given standard Cauchy priors. This implies a 95% prior credible interval for each parameter of

Fig. 1. Ternary diagram showing composition values for clusters in Benin.



[-12.7, 12.7] on the logit scale, equivalent to odds ratios in $[6 \times 10^{-6}, 1.6 \times 10^{5}]$, which is considered highly plausible. The low rank thin plate spline functions are specified as per (Wood, 2003); the parameters for each basis function are assigned hierarchical, normally distributed coefficients with unknown standard deviation (Wood, 2017). The priors for the standard deviation of these coefficients and the other hierarchical parameters are assigned half-t_4 priors (Gelman, 2006), which imply a prior 95% CrI of [0,3.5]. Less informative priors were considered, including half-Cauchy(0,25) priors, however, these distributions have relatively large tails and it was considered that the scale parameter was not likely to be large enough to justify such a prior.

Estimation was conducted using Stan implemented through the `rstan` package in R (Carpenter et al., 2017). Stan is a probabilistic programming language that provides Bayesian inference using Hamiltonian Monte Carlo (Carpenter et al., 2017; Betancourt, 2018). We used four chains for a minimum of 2,000 iterations each. Convergence of the chains was assessed graphically through trace plots and using the R-hat statistic. Code is available in the Supplementary Information.

3.4. Plotting results

Data on a three-dimensional simplex can be plotted on a 'ternary diagram.' Figure 1 plots the composition values for all clusters in Benin, as an illustration. Three points, A, B, and C, are marked. For each point the value of the three components of the composition can be found by tracing along the relevant colour-coded lines. Point A

indicates a point with 100% unimproved sanitation, Point B indicates a point with 30% unimproved, 55% sewers, and 15% improved latrines, and Point C indicates a point with 25% unimproved, 25% sewers, and 50% improved latrines. Ternary diagrams were plotted using the R package `ggtern`.

4. Simulation and Model Checking

The data generating mechanism for the simulated data sets is based on that posited in the main study. In particular, we simulate data uniformly across the three-dimensional simplex $\mathbf{x} = \{(x_1, x_2, x_3) : x_d > 0, d = 1, 2, 3; \sum_{d=1}^{3} x_d = 1\}$. We also simulate two normally distributed covariates $z = [z_1, z_2]$ from $N(0, 0.5^2)$ and set the hierarchical effect standard deviation to $\tau = 0.01$. Parameters $\beta_1 = [\beta_{1,1}, \beta_{1,2}]$ were generated randomly from $N(0, 0.5)$, and an intercept β_0 was chosen to give an approximate mean outcome probability of 20%. Data were then simulated for cluster $j = 1, ..., J$ clustered in country $k = 1, ...K$:

$$\begin{aligned} \alpha_j|\tau &\sim N(0, \tau) \\ p_{jk}|x_{jk}, z_{jk}, \alpha_k, \beta &= \Lambda(\beta_0 + \beta_1 \times z_{jk} + f(x_{jk}) + \alpha_k) \\ y_{jk}|p_{jk}, n_{jk} &\sim Bin(n_{jk}, p_{jk}) \end{aligned} \tag{2}$$

the number of individuals in each cluster was assumed to be uniformly distributed between 20 and 50. We considered two different data generating functions f:

(a) Planar surface: $f(x) = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3$.
(b) Parabolic surface: $f(x) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2$.

For each simulation we generated data of sample sizes 500, 1,000, and 5,000, each clustered equally into 50 countries. The parameters λ_1, λ_2, λ_3 were randomly generated for each function from a $N(0, 0.75^2)$ distribution, which would give an average most extreme difference of around 20 percentage points on the simplex. For each simulated data set we estimated the model described above with each of the specifications of $g(.)$ described in Section 3.2. Thus, for four model specifications, three sample sizes, and two different data generating functions, for which we simulated 500 data sets each, there were 12,000 models estimated. Given the time required to estimate each model using

MCMC (in the order of hours), mean field variational inference estimation was used in the simulations (Blei et al., 2017).

4.1. Posterior predictive model checks

To assess the models we considered their predictive performance for values across the composition simplex. Given observed (or simulated) data Y and set of model parameters Θ the posterior predictive distribution of new data given an assumed model M is

$$p_{\hat{Y}}(\hat{Y}|Y,M) = \int_{\Theta} p_{\hat{Y}}(\hat{Y}|\Theta,Y,M)p_{\Theta}(\Theta|Y,M)d\Theta \tag{3}$$

For each model and replication r with simulated data set $Y^{(r)}$ we estimated the posterior predictive mean for each point on a regularly-spaced lattice, $\mathbf{x}_l$, $l = 1, ..., L$, across the simplex at intervals of 0.02:

$$\hat{p}_{\hat{Y}|lr} = p_{\hat{Y}}(\hat{Y}|Y^{(r)}, M, \mathbf{x} = \mathbf{x}_l, z_{1,jk} = 0) \tag{4}$$

The median absolute deviation (MAD) for each replication r is then:

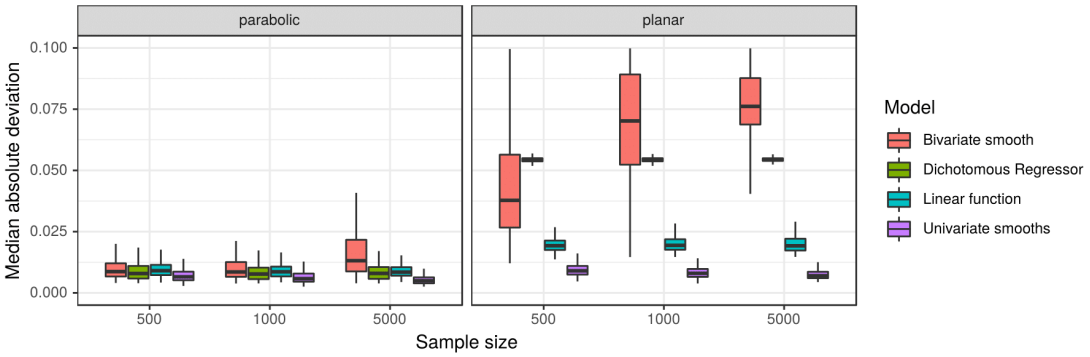
$$MAD_r = median_{l=1,...,L}(|\hat{p}_{\hat{Y}|lr} - f(x_l)|)$$

We also conduct a series of posterior predictive model checks, using posterior predictive p-values. For a test statistic of the data $T(Y)$ the posterior predictive p-value is defined as:

$$p_c(Y) = p_{\hat{Y}}(T(\hat{Y}) \geq T(Y)|Y,M) \tag{5}$$

We estimate the posterior predictive p-value for each simulated data set. The distribution of the p-value should be concentrated around 0.5 if the models are a reasonable fit to the data (Gelman, 2013). Values close to zero or one indicate the actual data fall at the extremes of data the model would produce, indicating poor model fit. We consider the mean and standard deviation of the posterior predictive distribution of the probability of experiencing the outcome. We conduct a graphical check of the p-value distributions for the simulated data.

Fig. 2. Median absolute deviation of posterior predictive mean across the simplex



4.2. Simulation results

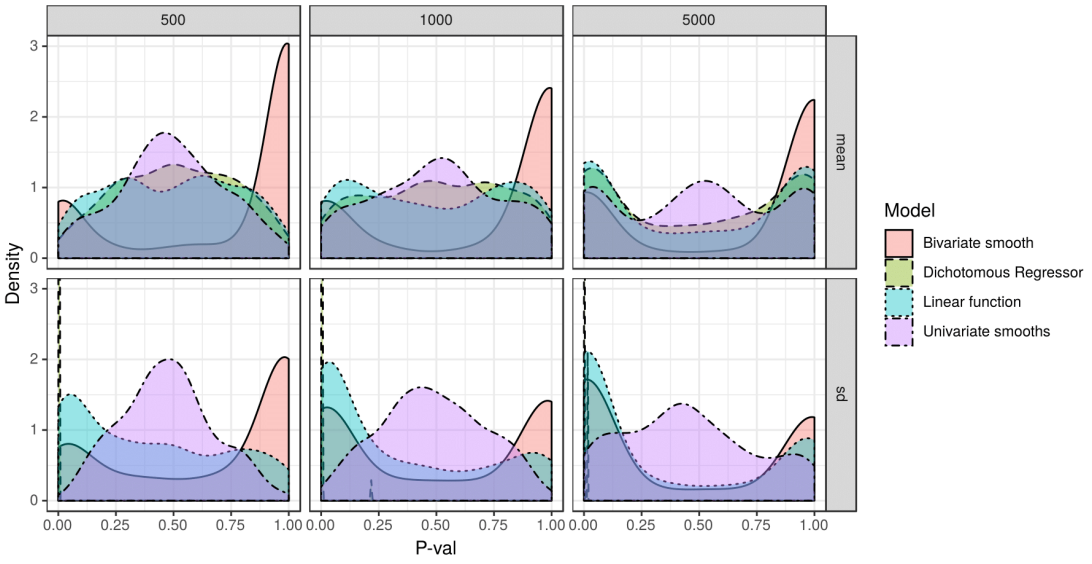
As Figure 2 shows, the univariate smooths model demonstrates the best performance with the simulated data with a mean MAD at a sample size of 5,000 in the parabolic case ($\lambda = [-0.12, -0.50, -0.42]$) of 0.008 and 0.013 in the planar case ($\lambda = [-0.35, 0.32, -1.89]$). All other specifications had higher MADs. Unexpectedly the bivariate smooth specification is the worst performing of all models, with performance getting worse with larger sample size. The MADs are larger in the planar case given the larger magnitude of the randomly generated λ parameters. The p-value distributions also reflect this finding with the univariate smooth specification p-value densities consistently demonstrating a concentration around 0.5 whereas the other specifications' posterior predictive distributions tending to over- or underestimate the standard deviation (Figure 3) and mean (Figure 7, Supplementary Information).

4.3. Model selection and checking

On the basis of the above simulation results the model specification with univariate smooth terms of the individual *ilr* components was selected as the primary specification. To examine whether the primary model specification provides a reliable fit to the actual data we conduct a graphical posterior predictive model check of the density of outcome probabilities with the actual density in the data and examine posterior predictive p-values for the mean and standard deviation of the outcome probabilities.

The models showed good fit for dysentery and stunting outcomes, but relatively

Fig. 3. Density of posterior predictive p-values for standard deviation with planar function data generating process



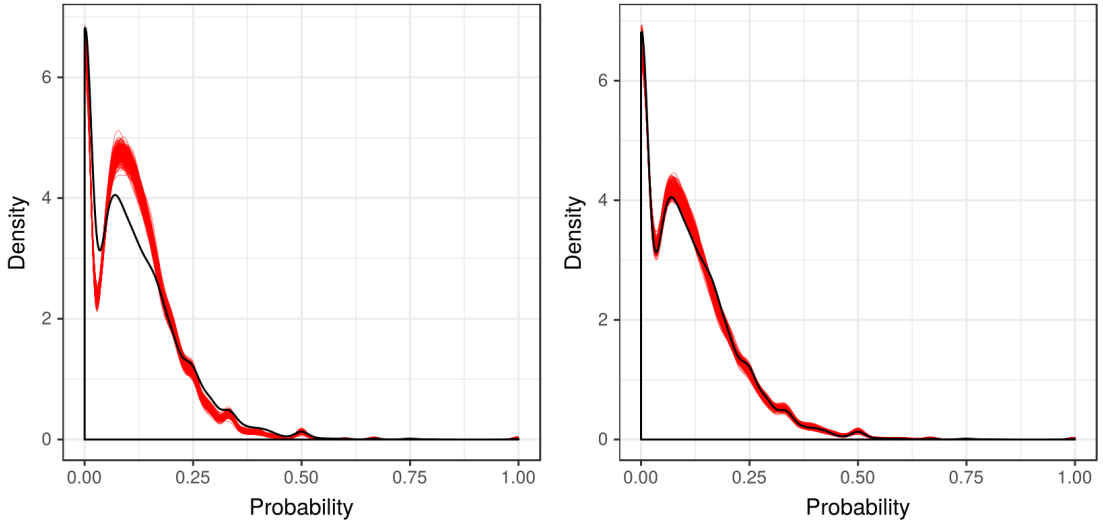


Fig. 4. Density of diarrhoea probability from draws from the posterior predictive distribution (red lines) and actual data (black line) from main model specification (left) and with added overdispersion parameter (right).

poor fit for diarrhoea. As Figure 4 (left) shows for the diarrhoea outcome, the originally specified model appears to underestimate the variation in outcomes. This is supported by the p-values, which do not suggest a poor fit for the mean (p=0.70), but underestimation of the standard deviation (p<0.01). Similar potential overdispersion was observed for the other outcomes as well. To account for this overdispersion we re-estimate the model with the following specification, which includes an observation-level random effect:

$$\begin{aligned} p_{jkt} &= \Lambda\left[z'_{1jkt}\gamma + \alpha_j + \tau_t + g_1(w_{1,jkt}) + g_2(w_{2,jkt}) + e_{jkt}\right] \\ e_{jkt} &\sim N(0, \sigma_e^2) \end{aligned} \tag{6}$$

Figure 4 (right) shows that this appears to provide a much better fit to the data, also reflected by the p-values for the mean (p=0.68) and standard deviation (p=0.85). However, examining the Widely Applicable Information Criterion (WAIC)–an estimate of pointwise out-of-sample predictive performance with smaller values indicating better performance–reveals the overdispersed model to have worse out-of-sample performance. The WAIC values for the original and overdispersed models are 27,426 and 53,043, respectively, for the diarrhoea outcome, with similar differences estimated for the other outcomes. The WAIC is equal to (-2 times) the expected log pointwise density (ELPD) for new observations less a penalty term for the effective number of parameters. While the ELPD was greater in the overdispersed model (-26,521 versus -13,713), the overdispersed model had a far higher penalty term (15,606 versus 393). This suggests the overdispersed model is overfitting, which accounts for its in-sample performance. A beta-binomial model was also investigated but no reasonable convergence of the HMC chains could be achieved. As a result we adopt the non-overdispersed model for all outcomes.

5. Results

5.1. Descriptive statistics

Overall 7,209 clusters from 29 countries were included in the sample, urban clusters in the DHS comprise an average of 24 households. Table 1 provides descriptive statistics of the key outcome and explanatory variables by general composition of sanitation facility. The majority of clusters, 51.4%, in the sample have a majority of households using

unimproved sanitation facilities. The general composition of sanitation facilities shows little relationship with household characteristics, except that those clusters with a majority of households using a sewage system of some type or private improved facilities are generally better educated and wealthier than other clusters (Table 5.1). For example, 84% of mothers have secondary education or higher in majority sewage system clusters compared to 51% in majority unimproved clusters and 53% in majority improved latrine clusters.

Mean proportions of children with diarrhoea and dysentery are lower in clusters with majority sewage systems (9.8% and 0.6%, respectively) than in majority improved latrine (10.8% and 1.0%) and unimproved sanitation (10.9% and 1.1%) clusters. Dysentery point prevalence in clusters with majority sewers (0.6%) or private improved sanitation (0.6%) is less than half that in clusters with unimproved sanitation (1.4%). The crude mean differences in the outcomes between clusters with almost exclusive coverage of each type of sanitation are reported in Table 2.

5.2. Main results

For the unimproved-latrine-sewer composition, Figure 5 shows the estimated response functions for the adjusted and unadjusted models. Figure 6 shows the same for the unimproved-shared-private composition. Table 3 reports estimated parameters from the model with the unimproved-latrine-sewer composition treatment variables. Results from the models with the unimproved-shared-private composition variables are qualitatively highly similar as would be expected (Table A, Supplementary Information).

For both the umimproved-latrine-sewer and unimproved-shared-private compositions, evidence of reduced risk is only apparent for high (>60%) levels of coverage of sewer systems and private sanitation. The unadjusted models show more variation across different compositions of sanitation than the adjusted models, suggesting much of the difference is accounted for by differences in observed covariates between survey clusters. This is also reflected by the estimated parameter estimates in Table 3.

Table 4 reports the estimated 'treatment effects' of altering the coverage of sanitation. High levels of coverage of 'improved latrines' compared to 'unimproved' has little

Table 1. Summary statistics of the study sample of urban clusters by majority type of sanitation facility. Values are mean (95% CrI) unless stated otherwise.

Variable	**Unimproved** > 50%	**Improved latrine** > 50%	**Sewer** > 50%	**Private improved** > 50%	**Shared improved** > 50%
N (%)	3,708 (51.4)	1,457 (20.2)	1,410 (19.6)	1,105 (15.3)	1,515 (21.0)
Diarrhoea (%)	10.9 (0.0, 33.3)	10.8 (0.0, 38.1)	9.8 (0.0, 39.8)	9.1 (0.0, 33.3)	11.0 (0.0, 40.0)
Dysentery (%)	1.1 (0.0, 10.0)	1.0 (0.0, 10.0)	0.6 (0.0, 7.7)	0.6 (0.0, 7.7)	0.9 (0.0, 10.0)
Stunting (%)	28.9 (0.0, 64.3)	29.2 (0.0, 68.7)	29.9 (0.0, 71.4)	28.4 (0.0, 66.7)	31.1 (0.0, 75.0)
Children per household	1.3 (1.0, 1.8)	1.3 (1.0, 1.8)	1.3 (1.0, 1.7)	1.3 (1.0, 1.8)	1.3 (1.0, 1.7)
Mean child age (months)	19.2 (2.8, 34.2)	20.5 (4.7, 36.4)	20.2 (1.6, 36.3)	21.2 (2.4, 36.3)	20.4 (4.0, 37.5)
Secondary education (%)	51.2 (0.0, 100.0)	53.3 (0.0, 100.0)	83.6 (20.0, 100.0)	76.1 (10.0, 100.0)	66.0 (0.0, 100.0)
Top wealth quintiles (%)	64.7 (0.0, 100.0)	78.7 (0.0, 100.0)	93.7 (13.5, 100.0)	87.8 (0.0, 100.0)	85.3 (0.0, 100.0)
Rainfall ($kgm^{-2}s^{-1}$)	0.9 (0.1, 2.2)	1.0 (0.0, 2.4)	0.9 (0.0, 2.7)	0.8 (0.0, 2.4)	1.0 (0.0, 2.6)
Temperature (C)	23.6 (15.0, 28.9)	24.3 (16.5, 28.3)	23.6 (18.0, 26.7)	22.6 (17.9, 28.0)	24.0 (16.5, 27.7)
Improved water (%)	81.3 (8.3, 100.0)	85.9 (21.2, 100.0)	92.5 (44.1, 100.0)	90.2 (33.9, 100.0)	88.6 (29.6, 100.0)
1991 - 1995	137 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
1996 - 2000	95 (1.3)	0 (0.0)	15 (0.2)	12 (0.1)	1 (0.0)
2001 - 2005	352 (4.9)	444 (6.2)	99 (1.4)	152 (2.1)	359 (5.0)
2006 - 2010	803 (11.1)	497 (6.9)	305 (4.2)	270 (3.7)	440 (6.1)
2010 - 2016	2,321 (32.2)	516 (7.2)	991 (13.7)	671 (9.3)	715 (9.9)

Table 2. Crude mean differences. A: unimproved sanitation $> 95\%$, B: improved latrines $> 95\%$, C: sewers $> 95\%$, D: shared improved $> 95\%$, E: private improved $> 95\%$. Effects are reported as percentage point changes.

Sanitation	**Diarrhoea**	**Dysentery**	**Stunting**
B-A	0.9	-0.5	-2.3
C-A	-0.9	-0.6	2.8
C-B	-1.9	0.0	5.0
D-A	0.9	-0.5	4.2
E-A	-3.3	-0.8	-7.3
E-D	-4.2	-0.3	-3.1

Table 3. Parameter estimates and model diagnostics

Parameter	**Diarrhoea**	**Dysentery**	**Stunting**
(Intercept)	-0.67 (-1.02, -0.33)	-8.79 (-10.47, -7.26)	-0.64 (-0.92, -0.37)
N. children	-0.65 (-0.74, -0.55)	-0.61 (-0.89, -0.32)	-0.65 (-0.71, -0.58)
Mean age	-0.01 (-0.01, -0.01)	-0.02 (-0.03, -0.01)	0.05 (0.05, 0.05)
Mother education	-0.42 (-0.51, -0.32)	-0.62 (-0.90, -0.35)	-0.09 (-0.16, -0.02)
Top two wealth	-0.06 (-0.14, 0.01)	0.07 (-0.16, 0.29)	-0.06 (-0.11, 0.01)
Rainfall	-0.08 (-0.13, -0.03)	0.03 (-0.09, 0.15)	-0.03 (-0.06, 0.01)
Temperature	0.01 (0.00, 0.02)	0.02 (-0.01, 0.05)	-0.01 (-0.02, 0.00)
Imp. water	0.04 (-0.04, 0.13)	-0.05 (-0.29, 0.19)	-0.02 (-0.08, 0.03)
1991 - 1995	Ref.	Ref.	Ref.
1996 - 2000	-0.68 (-0.91, -0.45)	-0.82 (-5.07, 2.73)	0.50 (0.34, 0.65)
2001 - 2005	-0.24 (-0.41, -0.08)	3.71 (2.51, 5.11)	0.16 (0.04, 0.27)
2006 - 2010	-0.38 (-0.56, -0.18)	5.28 (4.03, 6.78)	-0.04 (-0.17, 0.09)
2010 - 2016	-0.36 (-0.55, -0.18)	5.23 (4.00, 6.70)	-0.05 (-0.17, 0.05)
σ_α	0.12 (0.07, 0.21)	0.38 (0.21, 0.69)	0.14 (0.08, 0.23)
WAIC	27,426	7,904	30,984
p_c, mean	0.71	0.59	0.14
p_c, SD	$<$0.01	0.26	0.24

Fig. 5. Estimated response surface for risk of stunting for an unimproved, improved latrine, and sewer composition for adjusted (top row) and unadjusted (bottom row) models

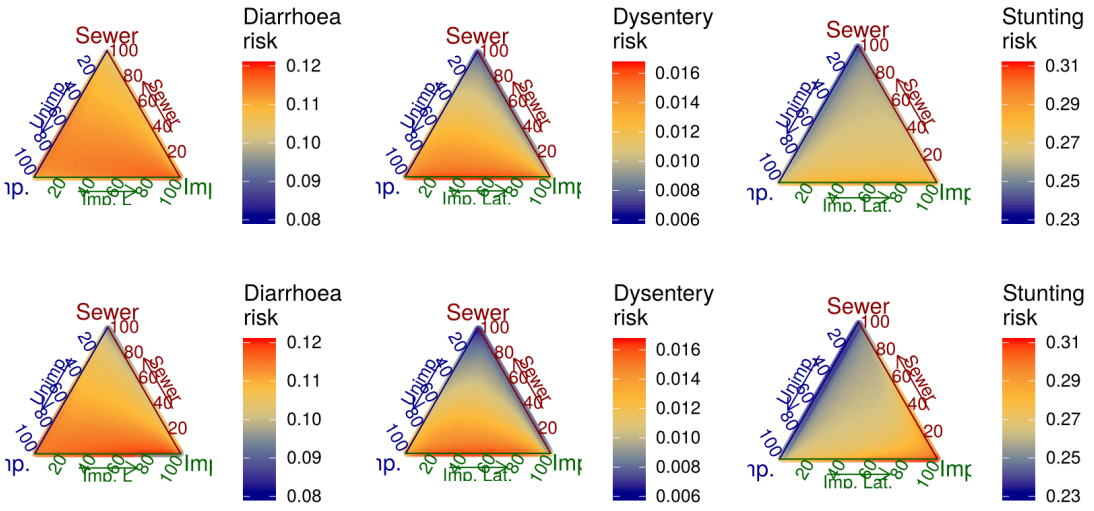


Fig. 6. Estimated response surface for risk of stunting for an unimproved, shared, and private improved sanitation composition for adjusted (top row) and unadjusted (bottom row) models

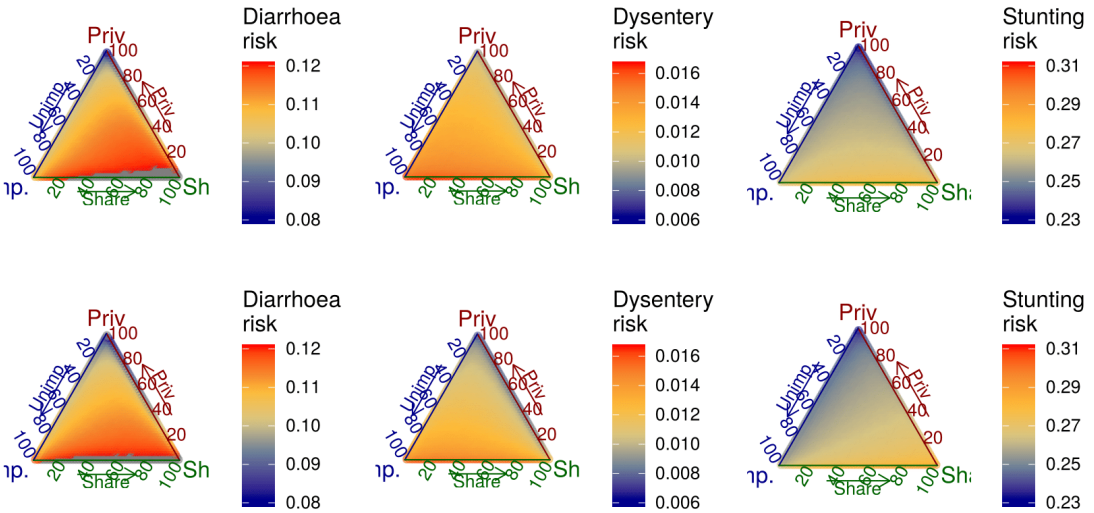


Table 4. Treatment effect estimates, posterior mean (95% CrI). A: unimproved sanitation $> 95\%$, B: improved latrines $> 95\%$, C: sewers $> 95\%$, D: shared improved $> 95\%$, E: private improved $> 95\%$. Effects are reported as percentage point changes.

Sanitation	Diarrhoea	Dysentery	Stunting
A → B	0.5 (-2.7, 4.2)	-0.2 (-2.2, 1.1)	0.8 (-3.6, 5.6)
A → C	-0.6 (-4.1, 2.5)	-0.7 (-3.8, 0.3)	-2.2 (-7.4, 1.7)
B → C	-1.1 (-5.0, 1.9)	-0.5 (-3.0, 0.5)	-3.0 (-8.6, 1.2)
A → D	1.8 (-1.4, 6.4)	-0.2 (-2.2, 1.1)	0.8 (-3.6, 5.7)
A → E	-1.9 (-6.0, 1.0)	-0.5 (-3.3, 0.5)	-2.4 (-7.7, 1.6)
D → E	-3.6 (-9.4, 0.0)	-0.3 (-2.4, 0.8)	-3.2 (-9.1, 1.1)

association with a reduction in risk of any outcome, whereas high levels of coverage of sewerage systems is associated with reductions in all outcomes. High coverage of sewage systems is associated with a change of -0.7 percentage points (pp) (95% CrI: -3.8, 0.3) for dysentery, compared to unimproved sanitation against a mean value of 1.0% (0.0%, 9.7%) across all clusters; a -2.2 pp (-7.4, 1.7) for stunting, which has a mean value of 29.1% (0.0%, 66.7%); and, -0.6 pp (-4.1, 2.5) for diarrhoea when compared to improved latrines against a sample mean of 10.7% (0.0%, 35.7%). Similar effect sizes are estimated for clusters with high levels of private santitation compared to unimproved or shared sanitation.

5.3. Temporal Confounding

While there are dummies in the model for five year periods, there may still be a degree of temporal confounding given the rapidly changing nature of sanitation policy over the last two decades, and scale of other water, sanitation, and hygeine interventions. To explore whether this affects our inferences, we re-estimate the models using data from only the period 2010 to 2015 (inclusive). These data include 4,828 clusters from 26 countries. The mean proportions of children reporting diarrhoea, dysentery, and stunting in this sample were 10.9%, 1.2%, and 27.2%, respectively.

The posterior predicted mean function across the simplex was qualitatively similar to those using the complete data set. Treatment effects from these models are reported

Table 5. Treatment effect estimates with 2010-15 data only, posterior mean (95% CrI). A: unimproved sanitation $> 95\%$, B: improved latrines $> 95\%$, C: sewers $> 95\%$, D: shared improved $> 95\%$, E: private improved $> 95\%$. Effects are reported as percentage point changes.

Sanitation	Diarrhoea	Dysentery	Stunting
A $\rightarrow$ B	0.0 (-3.9, 4.3)	-0.1 (-1.7, 1.4)	0.8 (-4.0, 7.1)
A $\rightarrow$ C	-2.1 (-7.9, 2.9)	-0.5 (-3.6, 1.2)	-2.9 (-10.9, 3.8)
B $\rightarrow$ C	-2.1 (-8.8, 2.9)	-0.5 (-3.5, 1.2)	-3.8 (-13.2, 3.7)
A $\rightarrow$ D	-0.1 (-4.0, 3.5)	-0.2 (-2.2, 1.3)	-0.9 (-7.0, 4.0)
A $\rightarrow$ E	-2.2 (-8.3, 2.9)	-0.6 (-3.9, 1.2)	-2.9 (-10.9, 3.8)
D $\rightarrow$ E	-2.1 (-8.3, 2.9)	-0.4 (-2.8, 1.2)	-2.1 (-9.6, 3.8)

in Table 5, which show a similar pattern to the complete data set, albeit with greater estimated improvement associated with high sewerage coverage.

5.4. Dichotomous regressor specification

We compare the results above to the more common specification of a dichotomous regressor for 'improved' sanitation. The treatment variable was specified as equal to one if the cluster had greater than 50% improved sanitation and zero otherwise. These results are reported in Table 6. They suggest that increasing the level of improved sanitation to over 50% is associated with a reduction in the risk of dysentery and stunting, but not diarrhoea. Conceptually, this specification estimates a weighted average effect across the different levels of sanitation. The estimated treatment effects suggest little benefit of improved sanitation. However, on the basis of the WAIC these models appear to have moderately better predicted out-of-sample performance.

6. Discussion

In this article we have examined the effects of different types of sanitation on the risk of three key childhood outcomes thought to be sensitive to the effects of improved sanitation. Generally, we have shown that high levels of coverage of private, improved sanitation facilities are associated with the lowest risk of all three outcomes net of the

Table 6. Treatment effect estimates from dichotomous regressor specification and model diagnostics. A: unimproved sanitation $> 50\%$, B: improved sanitation $\geq 50\%$.

Sanitation	**Diarrhoea**	**Dysentery**	**Stunting**
A $\rightarrow$ B	0.0 (-3.5, 3.2)	-0.3 (-3.0, 0.7)	-0.6 (-5.1, 3.7)
σ_α	0.10 (0.06, 0.17)	0.35 (0.19, 0.60)	0.12 (0.07, 0.20)
WAIC	23,959	6,952	28,964
p_c, mean	0.85	0.65	0.15
p_c, SD	$<$0.01	0.28	0.11

effects of socio-economic characteristics, weather, or country and cluster level unobserved heterogeneity. In relative terms, private improved sanitation was associated with an approximate 40% drop in diarrhoea incidence, 30% in dysentery, and 10% in stunting, compared to unimproved sanitation. However, there was a fair degree of uncertainty about effect sizes. Sewers and improved latrines were lower risk than unimproved facilities, but the differences in risk were smaller in magnitude that those between unimproved and private, improved facilities. This provides some evidence to support the scales of improvement used by organisations such as the WHO/UNICEF Joint Monitoring Program (WHO/UNICEF, 2018).

Previous studies of sanitation interventions generally point to sanitation improvements as being protective against diarrhoeal disease (Wolf et al., 2018; Fewtrell et al., 2005; Clasen et al., 2010). Observed pooled effect sizes can be compared to those presented here. For example, Wolf et al. (2018) reports a pooled relative risk (95% CI) for diarrhoea of unimproved sanitation versus sewer connections of 0.60 (0.39, 0.92), whereas this study suggests an equivalent relative risk of approximately 0.8. When unimproved is compared to private improved though, the relative risk is approximately 0.6 for diarrhoea risk. Some of the heterogeneity between studies included in that review may be then explained by the sharing of sanitation facilities or differing socio-economic status between included households. Certainly, there have been a small number of previous studies to have used DHS or similar household survey data to address the question of sanitation effectiveness (e.g. Begum et al. (2011b); Capuno et al. (2015); Fan and Ma-

hal (2011); Kumar and Vollmer (2013)). A strength of this study is that it compares different types of sanitation intervention, different levels of coverage of those facilities, allows for a non-linear relationship with health outcomes, and pools data from multiple countries. On the basis of these results, one may suspect that small effects observed in some studies could be a result of either insufficient coverage of improved sanitation, or perhaps more importantly that the sanitation was shared rather than private, and vice versa.

The etiology of childhood diarrhoeal disease is complex and multi-faceted. While sanitation interventions play a key role in its prevention, type and coverage of sanitation does not appear to explain much of the variation in outcomes. The estimated response surfaces across the sanitation simplex presented in this article reveal variation by type and coverage of intervention masked in simpler models like the 'dichotomous regressor' specification. Nevertheless, the dichotomous specification had better predicted out-of-sample performance according to the WAIC, which penalises models for additional parameters. The models for diarrhoea showed poor overall fit to the data, under-predicting the variance in outcomes.

A wealth of previous evidence has shown a relationship between socio-economic status and disease risk. Better educated or wealthier areas have longer life expectancies and suffer less ill health. We have endeavoured to control for these differences, which are likely to be correlated with access to type of sanitation facility. In particular, we took account of maternal education, wealth, allowed for unobserved heterogeneity between countries, and examined heterogeneous effects by education and wealth levels. Nevertheless, our results may well not represent the causal effect. The causes and correlates of infectious disease are numerous and complex, and the statistical adjustment is not perfect. However, these analyses, as we have argued, represent an improvement on previous observational work, conform to our *a priori* expectations of effectiveness, and are similar to previous studies where they are comparable.

There may be further weaknesses with the analysis resulting from the nature of the data. Diarrhoea and dysentery are both self-reported outcomes and are subject to recall bias (Alam et al., 1989; Boyer, 1995). We have used the DHS standard two week recall

period, but this has been shown to not be as reliable as shorter recall periods (Alam et al., 1989). Diarrhoea is also arguably a subjective outcome and may be 'reactive' to upgrades in water, sanitation, or hygiene improvements in the sense that mothers may not want to reveal childhood sickness. We also use stunting as an outcome, which is defined on the basis of anthropometric measurements as is therefore not subject to the aforementioned problems. However, stunting manifests as a result of influences over the child's life course, such as malnutrition or other disease, and not just the impact of diarrhoea or other factors amenable to better sanitation. A 'better' clinical outcome is certainly warranted to investigate the effects of better sanitation, for example analyses of pathogenic microbes, but these data are often not available, especially in observational contexts.

The results of this study tentatively agree with the current sanitation 'ladder' used by monitoring agencies. Our results are of a similar order of magnitude to other studies of sanitation interventions and provide some potentially useful context to heterogeneity between results from different studies. Beyond a simple improved/unimproved dichotomy, the use of the facilities and how they're shared is important, as well as the level of coverage in clusters of households. Extensions could be to extend to a multinomial-Dirichtlet model to allow for correlation across outcomes, and to embed the model in a larger structural equation framework to allow for endogeneity in sanitation compositions. We suggest this as an area for future research.

There remains a lot of unexplained variation in childhood adverse outcomes, and even with private improved sanitation, levels of the outcomes were predicted to still be high by any conventional standard. This points to a need to better understand disease transmission mechanisms in order to better design public health interventions, particularly when, in resource poor settings, providing private improved sanitation universally is currently infeasible.

References

Aitchison, J. (1982) The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**, 139–177.

Alam, N., Henry, F.J. and Rahaman, M.M. (1989) Reporting errors in one-week diarrhoea recall surveys: experience from a prospective study in rural Bangladesh. *International journal of epidemiology*, **18**, 697–700.

Ali, M. et al (2004) Use of satellite imagery in constructing a household GIS database for health studies in Karachi, Pakistan. *International Journal of Health Geographics*, **3**, 20.

Aziz, K.M.A. et al (1990) Reduction in diarrheal diseases in children in rural Bangladesh by environmental and behavioural-modifications. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **84**, 433–438.

Begum, S., Ahmed, M. and Sen, B. (2011a) Do Water and Sanitation Interventions Reduce Childhood Diarrhoea? New Evidence from Bangladesh. *Bangladesh Development Studies*, **34**, 1–30.

— (2011b) Do Water and Sanitation Interventions Reduce Childhood Diarrhoea? New Evidence from Bangladesh. *Bangladesh Development Studies*, **34**, 1–30.

Betancourt, M. (2018) A Conceptual Introduction to Hamiltonian Monte Carlo.

Bhavnani, D. et al (2014) Impact of rainfall on diarrheal disease risk associated with unimproved water and sanitation. *American Journal of Tropical Medicine and Hygiene*, **90**, 705–711.

Blei, D.M., Kucukelbir, A. and McAuliffe, J.D. (2017) Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, **112**, 859–877.

Bouzid, M., Cumming, O. and Hunter, P.R. (2018) What is the impact of water sanitation and hygiene in healthcare facilities on care seeking behaviour and patient satisfaction? A systematic review of the evidence from low-income and middle-income countries. *BMJ Global Health*, **3**, e000648.

Boyer, G.S. (1995) Discrepancies Between Patient Recall and the Medical Record. *Archives of Internal Medicine*, **155**, 1868.

Bruno, F., Greco, F. and Ventrucci, M. (2016) Non-parametric regression on compositional covariates using Bayesian P-splines. *Statistical Methods & Applications*, **25**, 75–88.

Capuno, J.J., Tan, C.A.R. and Fabella, V.M. (2015) Do piped water and flush toilets prevent child diarrhea in rural Philippines? *Asia-Pacific Journal of Public Health*, **27**, NP2122–NP2132.

Carlton, E.J. et al (2016) A systematic review and meta-analysis of ambient temperature and diarrhoeal diseases. *International Journal of Epidemiology*, **45**, 117–130.

Carpenter, B. et al (2017) Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, **76**.

Clasen, T. et al (2014) Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: A cluster-randomised trial. *The Lancet Global Health*, **2**, e645–e653.

Clasen, T.F. et al (2010) Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane Database of Systematic Reviews*.

Ezeh, A. et al (2016) The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet*.

Fan, V.Y.M. and Mahal, A. (2011) What prevents child diarrhoea? The impacts of water supply, toilets, and hand-washing in rural India. *Journal of Development Effectiveness*, **3**, 340–370.

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**, 42–52.

Fink, G., Günther, I. and Hill, K. (2011) The effect of water and sanitation on child health: evidence from the demographic and health surveys 19862007. *International Journal of Epidemiology*, **40**, 1196–1204.

Garrett, V. et al (2008) Diarrhoea prevention in a high-risk rural Kenyan population through point-of-use chlorination, safe water storage, sanitation, and rainwater harvesting. *Epidemiology and Infection*, **136**, 1463–1471.

Gelman, A. (2006) Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515–533.

— (2013) Understanding posterior p-values. *Electronic Journal of Statistics*.

Gelman, A. et al (2008) A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2**, 1360–1383.

Genser, B. et al (2006) Risk factors for childhood diarrhea incidence: dynamic analysis of a longitudinal study. *Epidemiology (Cambridge, Mass.)*, **17**, 658–667.

Geruso, M. and Spears, D. (2014) Sanitation and health externalities: Resolving the Muslim mortality paradox.

Godfrey, S. et al (2014) Impact study of the One Million Initiative rural water and sanitation programme in Mozambique. *Waterlines*, **33**, 35–44.

Hunter, P.R. and 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**, e0164571.

Komarulzaman, A., Smits, J. and de Jong, E. (2017) Clean water, sanitation and diarrhoea in Indonesia: Effects of household and community factors. *Global Public Health*, **12**, 1141–1155.

Kumar, S. and Vollmer, S. (2013) Does access to improved sanitation reduce childhood diarrhea in rural India? *Health Economics*, **22**, 410–427.

Luby, S.P. et al (2018) Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: A cluster randomised controlled trial. *The Lancet Global Health*, **6**, e302—-e315.

Messou, E. et al (1997) [Effect of hygiene measures, water sanitation and oral rehydration therapy on diarrhea in children less than five years old in the south of Ivory Coast]. *Bulletin de la Societe de pathologie exotique (1990)*, **90**, 44–47.

Moraes, L.R.S. et al (2003) Impact of drainage and sewerage on diarrhoea in poor urban areas in Salvador, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **97**, 153–158.

Naghavi, M. et al (2015) Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, **385**, 117–171.

Pearson, K. (1896) Mathematical Contributions to the Theory of Evolution.–On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs. *Proceedings of the Royal Society of London (1854-1905)*, **60**, 489–498.

Pickering, A.J. et al (2015) Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial. *The Lancet Global Health*, **3**, e701———e711.

Trinh, H.T. et al (2018) Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: New insights using compositional data analysis. *Statistical Methods in Medical Research*, 096228021877022.

UN (2018) Sustainable Development Goals.

UN-HABITAT (2014) The state of African cities: Re-imagining sustainable urban transitions.

Walker, S.P. et al (2007) Child development: risk factors for adverse outcomes in developing countries. *The Lancet*, **369**, 145–157.

WHO/UNICEF (2014) Progress on sanitation and drinking-water: 2014 update. WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation. *Tech. rep.*

— (2018) Sanitation. https://washdata.org/monitoring/sanitation.

Wolf, J. et al (2018) Impact of drinking water, sanitation and handwashing with soap on childhood diarrhoeal disease: updated meta-analysis and meta-regression. *Tropical Medicine and International Health*, **23**, 508–525.

— (2014) Systematic review: Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: Systematic review and meta-regression. *Tropical Medicine and International Health*, **19**, 928–942.

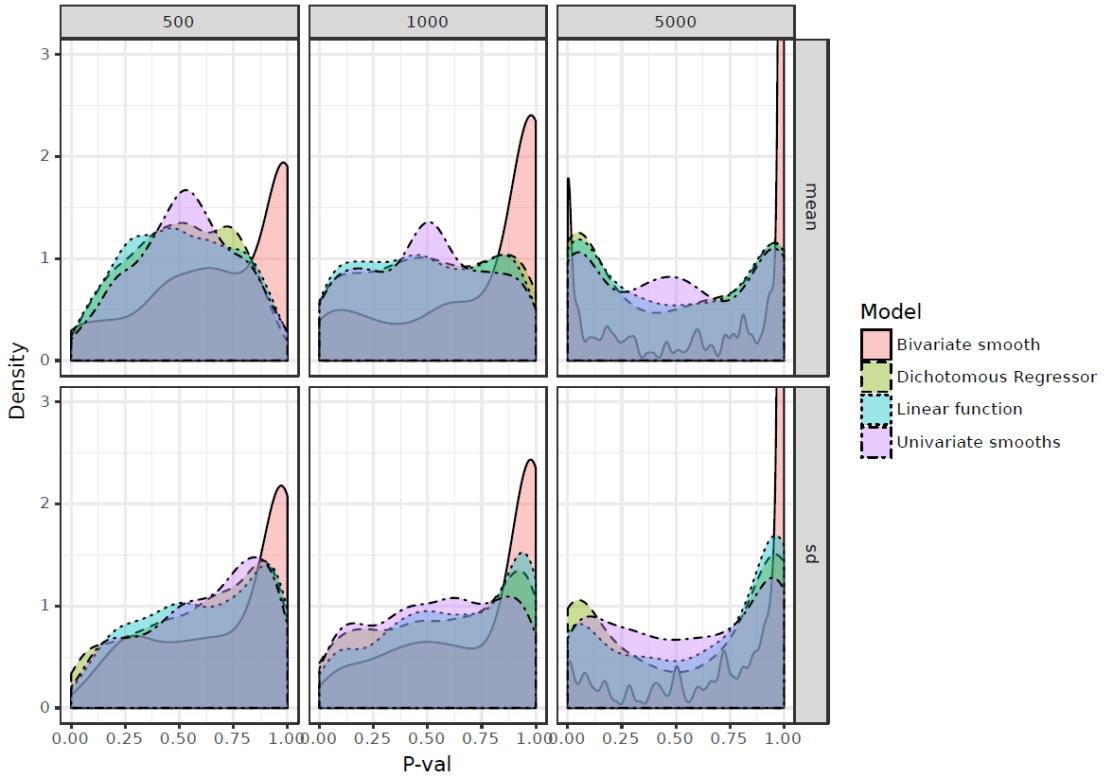
Wood, S.N. (2003) Thin plate regression splines.

— (2017) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2nd edn.

A. Additional results

This supplementary text provides some additional results and plots. The next section discusses heterogeneous effects.

Fig. 7. Density of posterior predictive p-values for mean



B. Heterogeneous Effects

As an extension to the models analysed so far, we now consider heterogeneous effects through interaction. Parental education, in particular that of the mother, has been shown in a number of previous studies to be associated with access to sanitation, childhood health outcomes, and the effectiveness of, and interaction with, sanitation facilities (e.g. Ali et al. (2004); Bouzid et al. (2018); Fink et al. (2011); Kumar and Vollmer (2013)). However, reflecting our earlier discussion, these effects in individual-level or household-level models may well be confounded with cluster-level coverage. As a result

Table 7. Included DHS surveys

Country	Survey year(s)	N
Burkina Faso	2010	176
Benin	2001 & 2011-12	428
Burundi	2010-11	75
Chad	2014-15	163
Comoros	2012	108
Cote d'Ivoire	2011-12	160
Cameroon	2011	295
Dem. Rep. Congo	2013-14	161
Ethiopia	1992 & 2003	323
Gabon	2012	183
Ghana	2003 & 2008 & 2014	567
Guinea	2012	107
Kenya	2003 & 2008-9 & 2014	398
Liberia	2006-7 & 2013	175
Lesotho	2009-10 & 2014	210
Madagascar	2008-9	149
Mali	2001 & 2006 & 2012-13	392
Malawi	2000 & 2004-5 & 2010 & 2015-16	506
Mozambique	2011	256
Nigeria	2013	371
Namibia	2013	256
Rwanda	2005 & 2007-8	124
Sierra Leone	2008 & 2013	303
Senegal	2010-11	147
Togo	2013-14	128
Tanzania	2009-10	79
Uganda	2006 & 2011	143
Zambia	2007 & 2013-14	421
Zimbabwe	2005-6 & 2010-11 & 2015	461
TOTAL		7,209

Fig. 8. Estimated response surface for risk of diarrhoea for an unimproved, improved latrine, and sewer composition

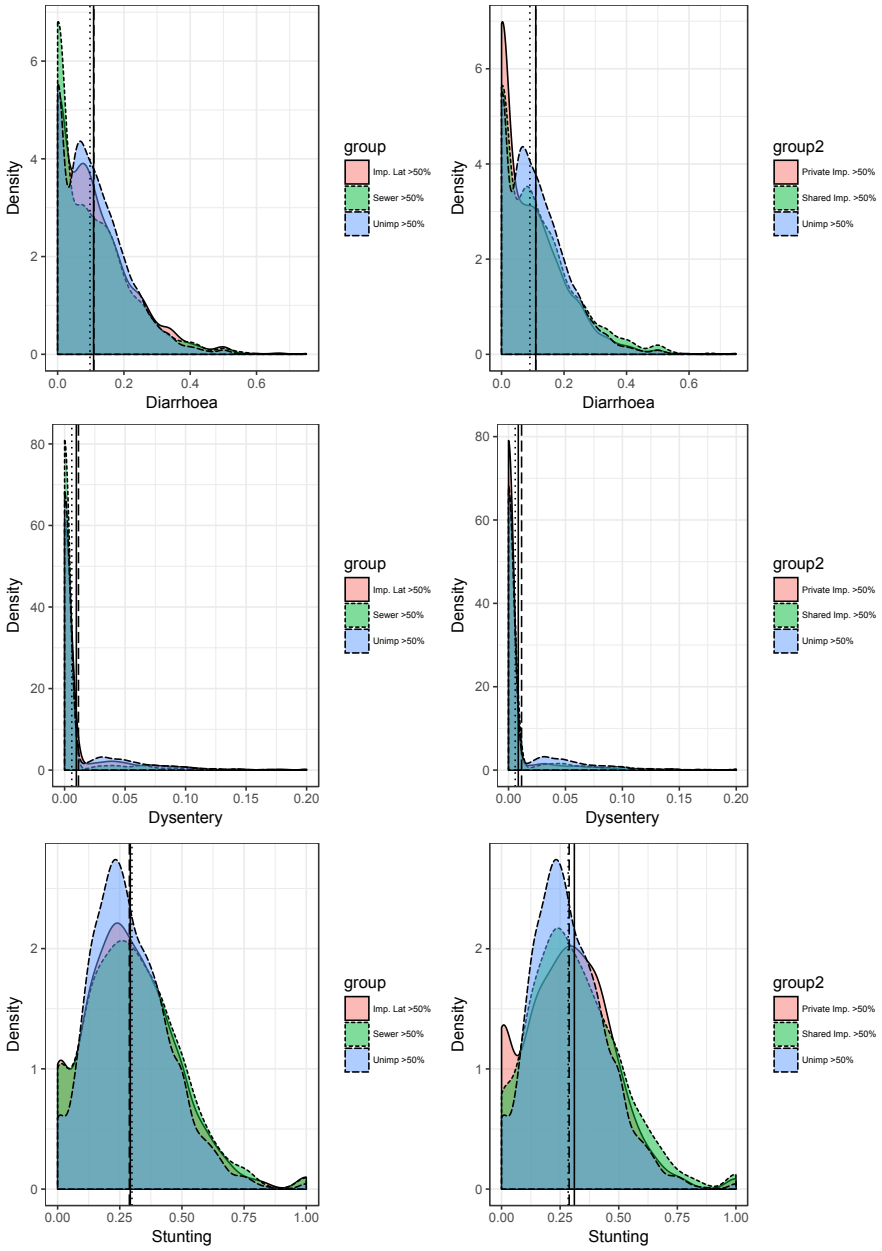


Figure (8) shows the distribution of each outcome for clusters by general type of sanitation facility.

Table 8. Parameter estimates and model diagnostics for unimproved-shared-private composition model

Parameter	**Diarrhoea**	**Dysentery**	**Stunting**
(Intercept)	-0.69 (-1.04, -0.34)	-8.73 (-10.53, -7.17)	-0.65 (-0.93, -0.38)
N. children	-0.64 (-0.73, -0.55)	-0.60 (-0.89, -0.31)	-0.65 (-0.71, -0.58)
Mean age	-0.01 (-0.01, -0.01)	-0.02 (-0.03, -0.01)	0.05 (0.05, 0.05)
Mother education	-0.40 (-0.47, -0.31)	-0.69 (-0.96, -0.41)	-0.13 (-0.20, -0.06)
Top two wealth	-0.07 (-0.14, -0.05)	-0.03 (-0.25, 0.19)	-0.08 (-0.13, -0.03)
Rainfall	-0.10 (-0.14, 0.05)	0.03 (-0.09, 0.15)	-0.04 (-0.07, 0.01)
Temperature	0.01 (0.00, 0.02)	0.02 (-0.01, 0.05)	-0.01 (-0.02, 0.00)
Imp. water	0.04 (-0.04, 0.12)	-0.02 (-0.27, 0.22)	-0.02 (-0.07, 0.04)
1991 - 1995	Ref.	Ref.	Ref.
1996 - 2000	-0.72 (-0.94, -0.49)	-0.85 (-5.23, 2.74)	0.47 (0.32, 0.62)
2001 - 2005	-0.27 (-0.43, 0.10)	3.82 (2.60, 5.31)	0.20 (0.09, 0.31)
2006 - 2010	-0.41 (-0.59, -0.23)	5.39 (4.11, 6.93)	0.01 (-0.12, 0.13)
2010 - 2016	-0.38 (-0.56, -0.20)	5.31 (4.03, 6.82)	-0.02 (-0.19, 0.10)
σ_α	0.12 (0.07, 0.21)	0.39 (0.21, 0.69)	0.13 (0.07, 0.22)
WAIC	27,386	232	31,007
p_c, mean	0.68	0.57	0.16
p_c, SD	<0.01	0.64	0.72

we extend the model to allow the smooth terms to vary by maternal education. The smooth function $g(w)$ can be represented as a linear combination of P basis splines: $\sum_{p=1}^{P} \delta_p b_p(w)$. Interaction with a variable z is specified as $zg(w) = \sum_{p=1}^{P} \delta_p z b_p(w)$ (which can be implemented in R using the `mgcv` package as `s(w,bs="tp",by=z)`).

Figure 9 (Supplementary Information) shows the posterior predicted mean across the simplex for the unimproved-latrine-sewer composition at two levels of maternal education. There is a large difference in average risk of each outcome at the two levels of education. For dysentery and stunting a similar pattern of predicted risk is observed across the simplex, with the lowest risk associated with higher sewer coverage, although there is more variation for the high education group. This pattern doesn't appear to be explained by low coverage, given clusters with both high and low sewerage coverage observed in low education clusters (Figure 10, Supplementary Information). For diarrhoea, little to no variation is observed across the simplex at either level of education.

Fig. 9. Estimated response surface for risk of stunting for an unimproved, improved latrine, and sewer composition

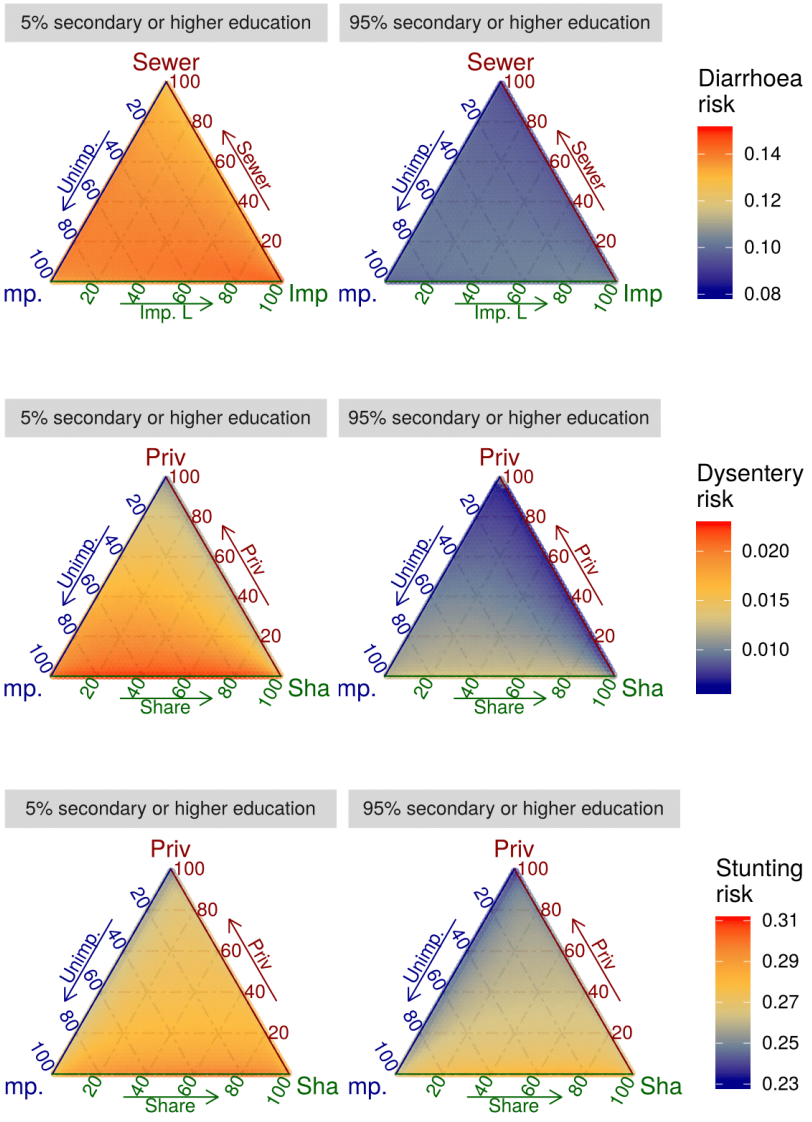


Fig. 10. Hexagonal binning plot showing frequency of sewerage coverage and maternal education