Defining Diarrhea: Validating and improving caregiver-reported stool consistency as a measure of pediatric diarrhea in the Amhara region of Ethiopia

by

Kristen Aiemjoy

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedicated to Jeroen Ensink (1974 - 2015)
An exceptional mentor and scientist.
You are missed.

Acknowledgments

This thesis is a culmination of five years of doctoral work; it would not be possible without the support and encouragement of many individuals, communities and organizations. Thanks go first and foremost to my advisor, Dr. Jeremy Keenan, for enthusiastically mentoring a doctoral student interested in diarrheal disease despite being an ophthalmologist. Jeremy's passion for improving public health via rigorously conducted clinical research is inspiring and are his seemingly endless energy and enthusiasm. It has been a privilege to learn from his example.

Next, I gratefully acknowledge the participation of communities, households and individuals in these studies. I hope that the results can help improve the measurement of diarrhea with the ultimate goal of reducing diarrheal disease in the Amhara region of Ethiopia and beyond.

Thank you to the incredible study coordinators, Dr. Solomon Aragie, Dionna Fry and Sintayehu Gebresillasie, along with the nurses, laboratory technicians, drivers and community members that make up our in-country field teams and the crucial support from The Carter Center and the Amhara Regional Health Bureau.

A special thank you to Dr. Ben Arnold for your collaboration, mentorship and guidance over the past five years. Thank you to my internal committee members, Dr. George Ruthorford for your career guidance and Dr. John Neahaus for your interest and support in studying misclassification.

Thank you to Dr. Eric Delwart and Eda Altan with the Blood Systems Research Institute for running viral metagenomic sequencing on the stool samples and for the terrific collaboration and learning opportunity.

Many thanks to Dr. Maria Glymour for always having a moment to answer questions and the administrative support from the PhD program along with the Department of Epidemiology and Biostatistics.

Thank you to Dr. Jeff Martin, an inspirational teacher, for giving me the support and encouragement to teach.

I am grateful for the unwavering support and encouragement from my parents, Eileen and Tom Knud-

son, and my brilliant partner, Glenn Aiemjoy. And finally, I am so thankful for my kind-hearted and wise son, for his support and understanding of all the late nights and travel despite being three-years-old and for teaching me just about everything there is to know about stool consistency.

Defining Diarrhea: Validating and improving caregiver-reported stool consistency as a measure of pediatric diarrhea in the Amhara region of Ethiopia

Kristen Aiemjoy

Abstract

Background: Diarrhea is a leading cause of death among children under five globally. Most studies of pediatric diarrhea rely on caregiver-reported stool consistency and frequency to define disease.

Methods: We collected stool samples from 2398 children participating in two water-improvement intervention trials in the Amhara region of Ethiopia. In the smaller trial we examined the pediatric enteric virome across stool consistency to evaluate differences in species richness and community composition using metagenomic sequencing. We also measured caregiver-reported pediatric diarrhea as an outcome for this trial, and performed bias analysis to estimate the impact of misclassified diarrhea on the trial outcome. The consistency of each stool sample was graded by the child's caregiver and two trained laboratory technicians according to an illustrated stool consistency scale. We assessed the reliability of graded stool consistency among the technicians, then compared the caregiver's grade to the first technician's grade. Caregivers were asked to report if children had three or more loose or watery stools in a 24-hour period anytime in the past seven days.

Results: The sensitivity of caregiver-reported loose or watery stool was 15.5% [95% CI: 9.7, 24.2] and the specificity was 98.4% [95%CI 97.1, 99.1]. Species richness was highest in watery-consistency stool and decreased as stool consistency became firmer. The greatest differential abundance comparing loose or watery to formed stool was for norovirus GII (7.64, 95% CI 5.8, 9.5) followed by aichivirus A (5.93, 95% CI 4.0, 7.89) and adeno-associated virus 2 (5.81, 95%CI 3.9, 7.7).

Conclusions: Caregiver reported stool consistency using the terms 'loose or watery' does not accurately describe stool consistency as graded by trained laboratory technicians. Given the predominance of using caregiver-reported stool consistency to define diarrheal disease, the low sensitivity identified in this study suggests that the burden of diarrheal disease may be underestimated and intervention effects could be biased. We documented a difference in pediatric enteric virome according to mBSFS-C stool consistency category. Bias analysis did not reveal a corrected protective effect of the water-improvement intervention.

Contents

Figures

Tables

1

A population-based validation study of caregiver-reported stool consistency

1.1 Abstract

Diarrhea is a leading cause of death among children under five globally. Most studies of pediatric diarrhea rely on caregiver-reported stool consistency and frequency to define disease. Research on the validity of caregiver-reported diarrhea is sparse. We collected stool samples from 2398 children participating in two clinical trials in the Amhara region of Ethiopia. The consistency of each stool sample was graded by the child's caregiver and two trained laboratory technicians according to an illustrated stool consistency scale. We assessed the reliability of graded stool consistency among the technicians, then compared the caregiver's grade to the technician's grade. We also tested if the illustrated stool consistency scale could improve the validity of caregiver's report. The weighted kappa measuring agreement between the two laboratory technicians reached 0.90 after 500 stool samples were graded. The sensitivity of caregiver-reported loose or watery stool was 15.5% [95% CI: 9.7, 24.2] and the specificity was 98.4% [95%CI 97.1, 99.1]. With the illustrated scale the sensitivity was 68.5% [95%CI: 58.5, 77.1] and the specificity was 86.1% [95% CI: 79.3, 90.9]. The results indicate caregiver reported stool consistency using the terms 'loose or watery' does not accurately describe stool consistency as graded by trained laboratory technicians. Given the predominance of using caregiver-reported stool consistency to define diarrheal disease, the low sensitivity identified in this study suggests that the burden of diarrheal disease may be underestimated and intervention effects could be biased. The illustrated scale is a potential low-lost tool to improve the validity of caregiver-reported stool consistency.

1.2 Background

Diarrhea is a leading cause of childhood morbidity and mortality with an estimated 2 billion cases and 525,000 deaths annually[1]. In Ethiopia, diarrhea is the second-leading cause of death among children under

five, causing more deaths than HIV, TB and malaria combined[2,3]. Most trials and epidemiologic studies of childhood diarrhea use caregiver reports of stool consistency and frequency to characterize disease status[4]. The World Health Organization (WHO) definition of diarrhea is three or more loose or watery stools in a 24-hour period[5]. In three recent systematic reviews of water, sanitation and hygiene (WASH) intervention trials to prevent pediatric diarrheal disease, all 22 studies used caregiver-reported symptoms to classify diarrhea[6–10]. In a 2015 review of 55 studies of water quality interventions for reducing diarrheal disease, 36 used the WHO definition and 11 used another symptoms-based report[11].

Despite its widespread use, there is limited research on the validity of various definitions of diarrhea and the component items in those definitions (stool consistency and frequency)[4,12–15]. This knowledge gap is important given misclassified caregiver-reported stool consistency could introduce measurement error and bias. True protective effects of interventions on diarrheal disease may be undetectable when measurement error is present. Underreporting of moderate and severe diarrhea may underestimate the disease burden and resulting cost-effectiveness of interventions designed to mitigate that burden.

Visual and descriptive stool consistency scales may standardize and improve the accuracy of reported stool consistency. A widely used stool form scale, The Bristol Stool Form Scale (BSFS) was developed in the late 1980s to measure gut transit time[16,17] and later simplified to a five-level scale: the modified Bristol Stool Form Scale for children (mBSFS-C)[18]. The mBSFS-C could be used as a tool for eliciting self-reported or caregiver-reported diarrhea from epidemiological studies. However, to our knowledge, it has never been evaluated in a research setting in Africa.

Our objective was threefold: measure the inter-rater reliability of mBSFS-C among laboratory technicians, validate caregiver-reported stool consistency using laboratory technician-graded consistency as the reference standard and determine if the mBSFS-C can improve the validity of caregiver-reported stool consistency.

1.3 Methods

1.3.1 Study population

This study was conducted within two cluster-randomized trials studying water, sanitation and hygiene (WASH) interventions in rural Ethiopia: In one, 14 communities in the East Gojjam zone were randomized to receive either a hand dug well or no intervention[19–21] (labelled Trial I in this report; clinicaltrials.gov NCT02373657). In the other, 40 communities in the WagHimra zone were randomized to either a comprehensive WASH package or no intervention (labelled Trial II in this report; clinicaltrials.gov NCT02754583). The primary outcome of both trials was ocular chlamydia. The present study collected data at the final study visit of Trial I (April 2016) and the baseline visit of Trial II (January of 2016).

In each study, we conducted a door-to-door population census approximately three weeks before the study visit to enumerate households and children eligible to participate in data collection. Based on samples size calculations for the primary outcome, In Trial I, all censused children aged zero to five years were eligible to participate. In Trial II, a random sample of 40 children aged zero to five years (up to their sixth birthday) and 40 children aged six to nine years per community were eligible; if less than 40 children of a specific age group were censused then all children in that age group were sampled. In each case, the sampling strategy was based on power calculations for the primary outcome of the trial.

1.3.2 Measurements

Stool sample grading: Reference standard

The Modified Bristol Stool Form Scale for Children (mBSFS-C) is a five-level adaptation of the original seven-level BSFS[18,22]. It was five stool consistency categories with both cartoon depictions and descriptors: (1) hard lumps, (2) sausage-shaped but lumpy, (3) sausage-shaped and soft, (4) loose and (5)

		Original mBSFS-C	Amharic	Back-translated to English
Type 1		Separate hard lumps, like nuts (hard to pass)	በጣም የደረቀ ሰገራ (በጠጥ) ለመውጣት የሚስቸግር	Very dry stool (small and round like sheep feces), hard to pass
Type 2		Sausage shaped but lumpy	የደረቀ ሰገራ (የተያያዘ በጠጥ ይመስላል)	Dry stool (a single mass of small round feces, like sheep feces formed together)
Type 3		Like a sausage or snake, smooth and soft	ለስላሳ እና ደረቅ ያልሆነ የአባብ ቅርፅ ያለው	Soft, not dry and its shape is like snake
Type 4		Fluffy pieces with ragged edges, a mushy stool	በጣም ለስላሳ ቅርፅ የሌለው	Very soft and irregular shaped
Type 5		Watery, no solid pieces	ቀጭን ተቀማጥ	Watery stool

Figure 1.1: The modified Bristol Stool Form Scale for Children (mBSFS-C), translated into Amharic

watery (Figure 1.1). An Ethiopian clinician fluent in the Amharic and English languages translated the mBSFS-C into Amharic. The translations were adapted to cultural norms through consensus with three Amharic/English-speaking clinicians. It was back-translated into English to check comparability with the original scale. The descriptive words "nuts" and "sausage" were not used because these are not part of a standard diet in rural Ethiopia.

Laboratory technicians attended a 2-day classroom training on stool sample collection and consistency grading using the mBSFS-C and a 1-day field training with real stool samples before data collection started. We emphasized the importance of masking the consistency grades during both classroom and field training.

Stool sample collection

Data was collected during the regularly scheduled study visits for the parent trials, in a centralized location in each community. Caregivers were instructed to take their children to a semi-private outdoor place near the sample collection area and have their child defecate in a potty chair lined with a black plastic bag. For children unable to produce a stool within two hours, caregivers were asked to collect stool at home and bring it to a collection site the following day. All stool samples were graded in the field before they were set in a preservative and transported.

When the stool was returned to the field station, it was immediately inspected in the original collection container by two medical laboratory technicians and the consistency of the sample was independently graded according to the mBSFS-C. The technicians were masked to each other's grades. Masking was achieved by having the first technician silently enter their grade into custom-designed software on a smartphone; once entered this grade was immediately concealed in the application and impossible to change. The second technician then recorded their grade. Technicians were allowed to discuss their grades after both grades had been entered. The mBSFS-C diagram was available at the point of data collection in the data collection software and also as a laminated sheet.

Stool sample grading: Caregiver report of 'loose or watery' stool (index test 1)

After the stool sample was graded by both laboratory technicians, the caregiver was asked: "Did your child have a loose or watery stool?" The wording of the question was designed to mimic stool consistency element of the standard definition of diarrhea: three or more loose or watery stools in a 24-hour period.[5]

Stool sample grading: Caregiver mBSFS-C grade (Index test 2)

The caregiver was then shown the laminated color copy of the mBSFS-C with Amharic descriptions and asked to point to the consistency category most similar to their child's stool.

1.3.3 Statistical methods

Agreement of reference standard

We assessed agreement in the 5-level mBSFS-C stool consistency classification between the two graders using both an unweighted and quadratic-weighted kappa[23,24] and used a bootstrap with 1000 replicates to calculate bias-corrected 95% confidence intervals with resampling by community.

We used a K by K confusion matrix to visualize absolute agreement and partial agreement for the mBSFS-C, where the first technician's grades (in columns) are classified against the second technicians grades (in rows).

We also compared the unweighted and weighted kappa by the number of samples graded, in increments of 100, to evaluate a change in kappa over time. We specified a kappa of .9 or greater between graders to use the first laboratory technicians grade as the gold standard.

Validity of caregiver report of 'loose or watery' stool (index test 1)

We compared the caregiver report of 'loose or watery' stool consistency to the first laboratory technician grade (reference standard). We dichotomized the laboratory technician's mBSFS-C grade with types 4 and 5 qualifying as 'loose or watery' stool and types 1-3 as 'not loose or watery.' We calculated the sensitivity and specificity using separate logit models with the dichotomous result of the index test as the dependent variable conditional on the reference standard being positive (sensitivity) or negative (specificity). We used

Table 1.1: Definition of validity measures

Caregiver assessment		First technician's grade Reference standard		Total
Index test 1 'loose or watery' report	Index test 2 mBSFS-C grade	Loose/Watery mBSFS-C: Types 4&5	Not Loose/Watery mBSFS-C: Types 1-3	
Loose/Watery	mBSFS-C: Types 4&5	TP	FP	TP + FP
Not Loose/Watery	mBSFS-C: Type 1-3	FN	TN	FN + TN
Total		TP + FN	FP +TN	

mBSFS-C = Modified Bristol Stool Form Scale for Children
Index test 1= caregiver reported loose or watery stool: Did your child have a loose or watery stool?"
Index test 2 = caregiver grade after visual inspection of the stool sample using the mBSFS-C
Reference standard = First laboratory technician-graded stool consistency according to the mBSFS-C
True Positive (TP); False Positive (FP); False Negative (FN); True Negative (TN)
Sensitivity = TP/(TP+FN); Specificity = TN/(FP+TN)

a clustered sandwich estimator to adjust standard errors for clustering by community[25,26]. Definitions of the validity measures are given in Table 1.1.

We stratified the models by caregiver type (mother or father). To test for an interaction between caregiver grade and type we fit a logit model with the outcome of the index test as the dependent variable and the reference standard as an independent variable and an interaction term with caregiver type[27 28].

Validity and agreement of caregiver mBSFS-C grade (index test 2)

We dichotomized the caregivers mBSFS-C into types 4&5 (loose or watery) and types 1-3 (not loose or watery) and compared this against the similarly dichotomized grade of the first laboratory technician (the reference standard). We calculated the sensitivity and specificity using the same method as for index test 1, with separate logit models for sensitivity and specificity and a clustered sandwich estimator to account for clustering by community.

To test if the mBSFS-C improved the sensitivity and specificity of caregiver-reported stool consistency we used a logistic mixed-effects model with a random intercept for both community and child, to account for the paired comparison. We included an indicator variable for caregiver report with or without the

mBSFS-C[25,26], and used the Wald test for the coefficient of this indicator variable to evaluate statistical significance. The mixed-model yields estimates of median sensitivity and specificity conditional on the random effect, in this case, the child and the community[25].

We also evaluated agreement in 5-level mBSFS-C stool consistency grade between caregivers and the laboratory technician using both a weighted and quadratic-weighted kappa and used a bootstrap with 1000 replicates to calculate bias-corrected 95% confidence intervals with resampling by community.

Analyses were run in Stata 15 (StataCorp, College Station, TX). Figures were generated in R Studio using R Version 3.4.1 (Foundation of Open Source Statistics, Boston, MA, USA).

1.3.4 Ethics Statement

Ethical committees at the University of California (San Francisco, CA, USA); Emory University (Atlanta, GA, USA); The Food, Medicine and Health Care Administration and Control Authority of Ethiopia; and the Ethiopian Ministry of Science and Technology granted approval for this study. We obtained verbal informed consent in Amharic from all caregivers.

1.4 Results

1.4.1 Characteristics of graders

Trial I employed three medical laboratory technicians and one clinical nurse. Three worked at government clinics and one worked at a university clinic. The average number of years of experience was 5.5 years (SD 3.9, range 1-10). Trial II employed eight medical laboratory technologists as graders: seven from clinics operated by the ministry of health and one from a university clinic. On average, graders had 7.4 years of experience (standard deviation [SD] 3.7, range 2-14). One medical laboratory technician was employed in both studies (Table 1.2).

Table 1.2: Characteristics of graders in Trial I and Trial II

	Trial I n=4	Trail II n=8
Age in years, mean (SD)	27.5 (4.2)	28.9 (4.3)
Gender		
Female	1 (25%)	1 (13%)
Male	3 (75%)	7 (87 %)
Profession		
Clinical nurse	1 (25%)	0
Med. lab. Technician	3 (75%)	8 (100%)
Years of experience, mean (SD)	5.5 (3.9)	7.4 (3.7)
Employer		
University	1 (25%)	1 (13%)
Ministry of Health	3 (75%)	7 (87%)

Numbers are n (%) unless otherwise indicated

1.4.2 Characteristics of study population

A flow diagram of sampling and participation is shown in Figure 1.2. In Trial I, all 446 censused children were eligible to participate, 317 children presented for the study visit examination day and 271 provided stool samples. The mean age of children with stool samples was 2.7 years old, 48.3% (152/271) of children were female, and 63% (170/271) of caregivers were mothers. In Trial II, 2400 children age zero to nine were randomly sampled, 2362 children presented for the study visit examination day and 2127 children provided stool samples. 16 stool samples (.75%) were collected the day after the study visit for children unable to produce a stool on the day of the study visit. Of children with stool samples, the mean age was 5.1 years old, 51.2% (1233/2127) of children were female, and 69.6% (1678/2127) of caregivers were mothers. (Table 1.3)

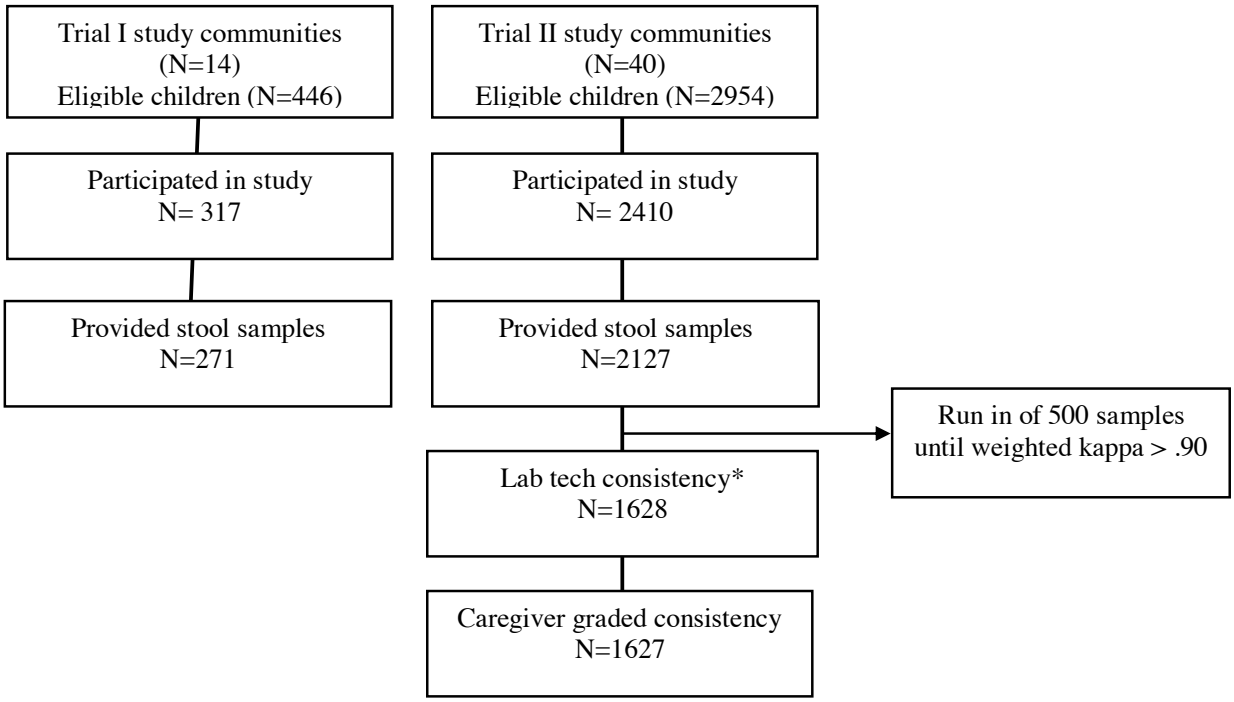


Figure 1.2: Selection and participation flow diagram

Table 1.3: Characteristics of study populations

	Trial I Study communities (N=14) Study population (N=271)	Trial II Study communities (N=40) Study population (N=2127)	Validation sample population* N=1627
Age in Years, Mean (SD)	2.8 (1.9)	5.1 (2.7)	5.3 (2.7)
Female	152 (48.3%)	1233 (51.2%)	834 (51.2%)
Caregiver Type			
Mother	170 (63%)	1678 (69.6%)	1344 (82.6%)
Father	26 (8.3%)	296 (12.3%)	206 (12.7%)
Aunt/Uncle	14 (4.4%)	46 (1.9%)	26 (1.6%)
Sibling	44 (14%)	18 (0.7%)	6 (0.4%)
Self	13 (4.1%)	78 (3.2%)	42 (2.6%)
Stool consistency type			
Type 1: Pellets	14 (4.4%)	381 (15.8%)	290 (17.8%)
Type 2: Lumpy	89 (28.3%)	677 (28.1%)	491 (30.2%)
Type 3: Smooth	61 (19.4%)	563 (23.4%)	412 (25.3%)
Type 4: Loose	81 (25.7%)	407 (16.9%)	376 (22.5%)
Type 5: Watery	29 (9.2%)	100 (4.1%)	68 (4.18%)

Values are n(%) unless otherwise indicated

*After 500 sample run in period

Table 1.4: K by K confusion matrix of caregiver-reported stool consistency using mBSFS-C versus lab technician graded stool consistency using mBSFS-C

	First lab technician					
Second lab technician	Type 1: pellets	Type 2: lumpy	Type 3: snake	Type 4: loose	Type 5: watery	Total
Type 1: pellets	261	8	2	1	1	273
Type 2: lumpy	21	460	20	3	0	504
Type 3: snake	8	20	376	16	2	422
Type 4: loose	0	1	12	341	4	358
Type 5: watery	0	2	2	6	61	71
Total	290	491	412	367	68	1628

After 500 sample run-in

1.4.3 Agreement of reference standard

In Trial I, the two laboratory technicians agreed on 169/271(62.6%) of the 5-level mBSFS-C grades, with an unweighted kappa of 0.50 (95% CI 0.42, 0.57) and a quadratic-weighted kappa of 0.70 (0.61, 0.77). In Trial II, 1870/2127 (87.9%) of grades were in agreement, with an unweighted kappa of 0.84 (95%CI 0.82, 0.86) and a quadratic-weighted kappa of 0.92 (95%CI: 0.90, 0.93). See Table 1.4 for the K by K confusion matrix.

Kappa increased with the number of samples graded (Figure 1.3). When restricted to the first 271 samples (the size of Trial I), the unweighted kappa for Trial II [0.56; 95%CI 0.49, 0.62] was comparable to Trial I [0.50 (95%CI: 0.49, 0.53) and the 95% confidence intervals overlapped. In Trial II, the unweighted kappa surpassed .90 after 600 samples were assessed and the weighted kappa surpassed .90 after 500 samples were assessed by 4 field teams, approximately 125 samples per team. In Trial I the weighted kappa did not reach .9.

To ensure robustness of our references standard (laboratory technician-graded stool consistency) for the validity research questions in this study, we opted to permit a run-in period of 500 samples until the weighted kappa exceeded .90. Thus, the remaining validation research questions evaluate samples 501 through 2127 in Trial II only (1627 samples in total). Characteristics of the study population in the

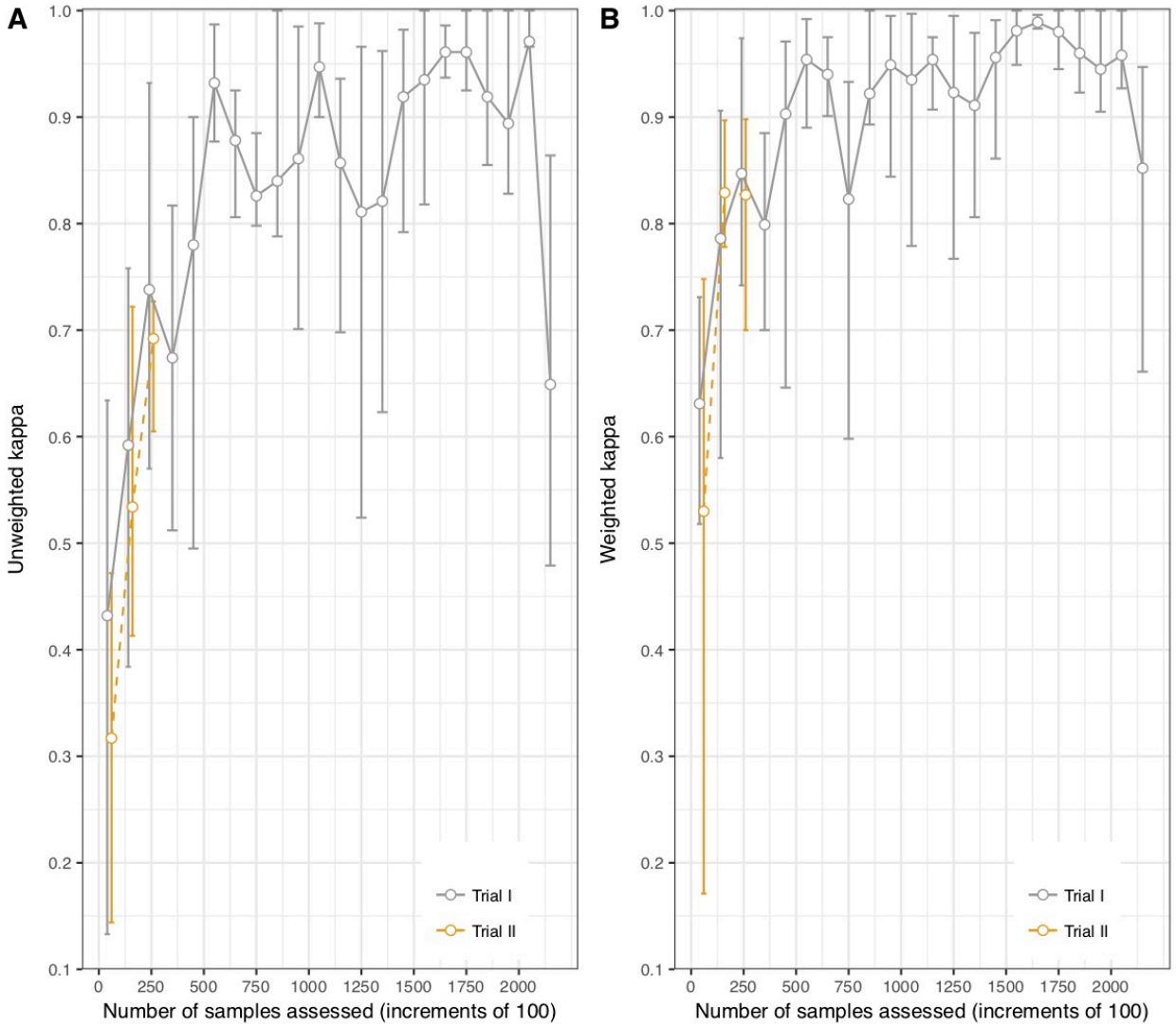


Figure 1.3: Both unweighted (A) and quadratic-weighted kappa (B) increase with the number of samples graded

validation sample (allowing for the 500-sample run-in) are displayed in Table 1.3.

1.4.4 Validity of caregiver report of 'loose or watery' stool (Index test 1)

Caregivers reported that 5.4% (87/1627) of samples were 'loose or watery' while laboratory technicians graded 26.7% (435/1628) of samples as mBSFS-C types 4 and 5 (i.e., the equivalent to 'loose or watery'). The overall sensitivity of caregiver-reported 'loose or watery' stool consistency was 15.6% (68/435; 95% CI: 9.7, 24.2) and the overall specificity was 98.4% (1173/1192; 95% CI: 97.1, 99.1). The sensitivity of

mother's reported stool consistency was higher than father's report 16.8% (66/392; 95% CI: 10.4, 26.1) versus 3.0% (1/33; 95% CI: 0.4, 20.4); The specificity was 98.8% (941/952; 95% CI: 97.7, 99.4) for mothers and 95.4%(165/173; 95% CI: 87.8, 98.3) for fathers. The p-value for the interaction term between caregiver type and grade was 0.004 (Table 1.5).

1.4.5 Validity and agreement of caregiver mBSFS-C grade (Index test 2)

When caregivers used the mBSFS-C to grade stool consistency, the overall sensitivity was 68.5% (298/435; 95%CI: 58.5, 77.1) and the specificity was 86.1% (1026/1192; 95% CI: 79.3, 90.9). The sensitivity was significantly higher when mothers used the mBSFS-C (index test 2) than when they did not (index test 1); P<0.00001 in a mixed model with random intercepts for child and community.

Sensitivity improved for both mothers (271/392; 69.1%, 95% CI: 58.1, 78.4) and fathers (20/33; 60.6%, 95% CI: 46.1, 73.4); Specificity was 86.8% (826/952; 95% CI: 79.6, 91.7) for mothers and 88.4% (153/173; 95% CI: 82.2, 92.7) for fathers (Table 7) See Figure 1.4 for an ROC space plot visualizing of the change in sensitivity with and without the mBSFS-C by caregiver type.

The unweighted kappa using all five levels of the mBSFS-C between caregiver's grade and the lab technician reference standard was 0.35 (95% CI: 0.32- 0.38); the quadratic weighted kappa was 0.49 (0.44 - 0.53). The K by K confusion matrix is presented in Table 1.6.

1.5 Discussion

We documented the validity of caregiver-reported 'loose or watery' stool consistency in two trials in the Amhara Region of Ethiopia using stool samples from 1627 children. Our findings indicate that caregiver-reported 'loose or watery' stool consistency has poor validity when compared to laboratory technician graded stool 'loose or watery' stool consistency. The low sensitivity is concerning given the terms 'loose

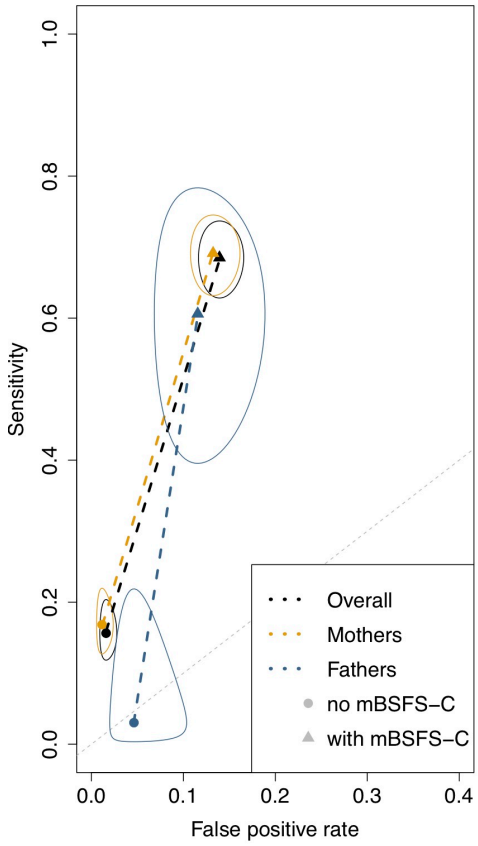


Figure 1.4: ROC space plot for caregiver-reported stool consistency with and without the mBSFS-C

Table 1.5: Sensitivity and specificity caregiver's report of 'loose or watery' stool consistency and caregiver's mBSFS-C grade

	TP	TN	FP	FN	Sensitivity	Specificity
Index test 1: Caregiver 'loose or watery' grade						
Overall (n=1627)	68	1173	19	367	15.6 (9.7, 24.2)*	98.4 (97.1, 99.1)**
Mothers (n=1347)	66	941	11	326	16.8 (10.4, 26.1)	98.8 (97.7, 99.4)
Fathers (n=203)	1	165	8	32	3.0 (0.4, 20.4)	95.4 (87.8, 98.3)
Index test 2: Caregiver mBSFS-C grade						
Overall (n=1627)	298	1026	166	137	68.5 (58.5, 77.1)*	86.1 (79.3, 90.9)**
Mothers (n=1347)	271	826	126	121	69.1 (58.1, 78.4)	86.8 (79.6, 91.7)
Fathers (n=203)	20	153	20	13	60.6 (46.1, 73.4)	88.4 (82.2, 92.7)

mBSFS-C = Modified Bristol Stool Form Scale for Children

After 500 sample run-in

Index test 1= caregiver reported loose or watery stool: Did your child have a loose or watery stool?"

Index test 2 = caregiver grade after visual inspection of the stool sample using the mBSFS-C

Reference standard = First laboratory technician-graded stool consistency according to the mBSFS-C

True Positive (TP); False Positive (FP); False Negative (FN); True Negative (TN)

Sensitivity = TP/(TP+FN); Specificity = TN/(FP+TN)

Sensitivity = TP/(TP+FN); Specificity = TN/(FP+TN)

*Z= 10.25, P<0.001; testing the difference in senstivity with and without the mBSFS-C

**Z= -8.83, P<0.001; testing the difference in specificity with and without the mBSFS-C

Table 1.6: K by K matrix of of caregiver-reported stool consistency using mBSFS-C versus lab technician graded stool consistency using mBSFS-C

	Lab technician graded stool consistency (referenece standard)					
Caregiver-reported stool consistency using mBSFS-C (Index text 2)	Type 1: pellets	Type 2: lumpy	Type 3: snake	Type 4: loose	Type 5: watery	Total (index test)
Type 1: pellets	136	68	10	57	2	273
Type 2: lumpy	98	237	99	36	5	475
Type 3: snake	35	76	267	36	1	415
Type 4: loose	10	51	26	117	16	220
Type 5: watery	11	58	10	121	44	244
Total (reference standad)	290	489	412	367	68	1627

or watery' are key components of the widely-used WHO definition of diarrhea: 'three or more loose or watery stools in a 24-hour period'. The degree of misclassification suggests epidemiologic studies, randomized control trials, and global burden of disease estimates that rely on caregiver-reported 'loose or watery' stool to define diarrheal disease in children may underestimate the prevalence of diarrhea and report potentially biased measures of association.

Symptom-based definitions of diarrhea are pervasive in epidemiology and clinical research, yet few studies have attempted to validate these measurements[4,12]. The WHO definition of diarrhea is based on a 1991 longitudinal study of 512 children in Bangladesh investigating four definitions of stool consistency and frequency against mothers' perception of diarrhea[29]. The definition 'three or more loose or watery stools' was chosen because it had the highest sensitivity (77.8%) compared to mother's perception of diarrhea. The definition was not compared to direct observation of the child, stool sample or a clinical diagnosis of diarrhea. A careful assessment of the WHO-definition is likely warranted given its pervasive use in epidemiology and clinical research.

We found that the mBSFS-C, an illustration of five stool consistency categories, had good agreement when used by local medical laboratory technicians to classify consistency of stool samples in a field-based research setting in Africa.

We specified a kappa of .9 or greater between laboratory technician graders in order to justify using the first technician's grade as the gold standard comparison. A kappa of .9 was reached after 500 samples were graded in the larger Trial I. However, in the smaller Trial II with only 271 samples in total, the highest weighted kappa reached was .83. Thus, all grades from the smaller trial were excluded from the validation study. The upward trajectory in kappa in the smaller Trial II mirrored that Trial I, signaling that a kappa of .9 may have been reached with more time and samples to grade.

The initial creation and assessment of the mBSFS-C in 2010 measured reliability between fourteen

physician-graders using stool photographs with an overall ICC of 0.85 (95%CI 0.78, 0.91)[18]. Intraclass correlation coefficients (ICCs) are dependent on the variation of stool consistency within a study population and thus are not directly comparable across study sites. Kappas have been reported for the original 7-level BSFS. A study published in 2016 was the first to assess reliability of the BSFS using real stool samples; when comparing patient and physician grades of stool consistency, 26% were in agreement with a weighted kappa of 0.67[30]. These results measure patient versus grader agreement rather than grader versus. grader agreement and are comparable with the weighted kappa comparing caregivers grade and the laboratory technician's grade using the mBSFS-C.

Laboratory technicians may not be the best gold standard grader for diarrheal disease. Caregivers, who are more familiar with their children's stool patterns may be better able to differentiate "diarrhea" from the normal stool pattern. Some iterations of the WHO definition diarrhea include three or more loose or watery stools more than is normal for the individual. Here we focus our validation study on the specific words "loose or watery". For this narrower definition, the trained laboratory technicians are an appropriate reference standard. For validating the WHO or other definitions of diarrhea, other gold standards may be more appropriate such as a clinician's diagnosis or presence of a specific enteric pathogen.

The mBSFS-C improved the validity of caregiver-reported 'loose or watery' stool consistency. The mBSFS-C is an option to improve the validity of caregiver-reported stool consistency when studies rely on caregiver-reported symptoms to define diarrheal disease. Reproducing the scale for use in epidemiologic studies and trials is easy to implement at low-cost. Care must be taken to ensure that the stool-consistency descriptor translations are understood and culturally appropriate. In addition, caregivers should be informed that the cartoon pictures depict stool consistency and not merely appearance.

Although the mBSFS-C shows promise in improving caregiver-reported stool consistency, researchers must carefully consider the underlying construct that is of interest. Loose or watery stools can have both

infectious and non-infectious causes. Moreover, enteric infections can be both symptomatic and asymptomatic. A pathogen-specific outcome is most likely more expensive than a symptoms-based outcome, but the specificity may be enough to outweigh the difference in costs. However, pathogen-specific outcomes may missclassify disease status because they include asymptomatic cases. Symptomatic outcomes may be contaminated with non-infectious symptomatic diarrhea cases, thus obscuring an effect of an intervention targeting infection pathways. The most appropriate outcome for measuring diarrheal disease depends on a study's objectives, budget and tolerance for misclassification.

We acknowledge several limitations of this study. First, we were unable to assess the accuracy of caregiver-reported stool frequency. Measurement error of reported stool frequency (number of bowel movements) is another threat to the validity of symptoms–based reporting of diarrhea and warrants further research. Second, we may have overestimated the sensitivity of caregiver-reported stool consistency by having caregivers report the consistency of a stool sample in front of them. When caregivers respond to traditional survey questions without actually observing their child's recent stools, we might expect more false negatives and thus the true sensitivity of caregiver-reported stool consistency may be even lower than what we observed in this study. Third, we did not differentiate the stool of breast-fed infants, which is typically looser than the stool of solid-fed infants. However, we do not expect the validity of the caregiver's report to be affected as the reference standard grade would still classify breastfed stool as 'loose'. Fourth, this study was population-based and thus represented a normal spectrum of stool consistencies. A critically ill or hospitalized patient population may be more appropriate to validate definitions or diarrhea against clinical diagnoses. However, this study population is still relevant to the many public health intervention trials, epidemiologic studies and surveys that use population-based samples and rely on caregiver report to define diarrheal disease. Finally, we did not assess intra-rater reliability of the mBSFS-C scale among laboratory technicians or caregivers. Stool samples were fixed in a liquid preservative immediately after

collection, and thus the consistency could not be graded again.

Despite these limitations, our study had several strengths. We collected stool samples from over two-thousand children from two distinct locations in Ethiopia. Each stool sample consistency was graded by a caregiver and two trained medical laboratory technicians according to an established stool consistency scale. We also evaluated the utility of a simple illustrative scale to improve the reporting of stool consistency. Our study was population-based, representing a normal spectrum of stool consistencies and thus demonstrating the utility of the mBSFS-C for classifying stool consistency for population-based research.

Our findings are disconcerting for researchers and public health professionals who use caregiver-reported stool consistency to quantify diarrheal disease. We found caregiver-reported 'loose and watery' stool, a key component of the WHO definition of diarrhea, does not accurately reflect 'loose or watery' stool as measured by the mBSFS-C. The degree of misclassification reported in this study would introduce substantial measurement error to studies quantifying diarrheal disease according to reported stool consistency. Specifically, the adjusted prevalence can be estimated as the sum of the measured prevalence and specificity, minus the reciprocal of the sensitivity plus specificity minus one[31]. It is not possible to directly calculate an adjusted prevalence of diarrhea for the present study, since we have diagnostic accuracy estimates only of stool consistency, and not frequency. However, as a thought exercise, if we assume that reported stool frequency is perfectly valid and that the error in reported frequency is independent of reported consistency, then the 13% seven-day prevalence of diarrhea observed in the present study would be equivalent to an adjusted prevalence of 83%. This example illustrates the potential impact of the diagnostic accuracy of a screening test on the final prevalence estimate.

Given the global public health importance of diarrheal disease and the predominance of using caregiver-reported symptoms to identify cases, the low sensitivity identified in this study suggests that the burden of diarrheal disease may be underestimated and intervention effects could be biased. Researchers should take

care when using caregiver reported loose or watery stool to define diarrheal disease. If caregiver-reported stool consistency is the only option to measure pediatric diarrhea, a pictorial scale like the mBSFS-C may improve the validity of caregiver-report. Replicating this study in other settings to determine if similarly low sensitivity obtains could have important implications for global estimates of diarrheal disease burden.

2

Viral species richness and composition in young children with loose or watery stool

2.1 Abstract

Stool consistency is an important diagnostic criterion in both research and clinical medicine and is often used to define diarrheal disease. We examine the pediatric enteric virome across stool consistencies to evaluate differences in richness and community composition using fecal samples collected from children participating in a clinical trial in the Amhara region of Ethiopia. The consistency of each sample was graded according to the modified Bristol Stool Form Scale for children (mBSFS-C) before a portion of stool was preserved for viral metagenomic analysis. Stool samples were grouped into 29 pools according to stool consistency type. Differential abundance was determined using negative-binomial modeling. Of 446 censused children who were eligible to participate, 317 presented for the study visit examination and 269 provided stool samples. The mean age of children with stool samples was 2.7 years old. Species richness was highest in watery-consistency stool and decreased as stool consistency became firmer (Spearman's r=-0.45, p=0.013). The greatest differential abundance comparing loose or watery to formed stool was for norovirus GII (7.64, 95% CI 5.8, 9.5) followed by aichivirus A (5.93, 95% CI 4.0, 7.89) and adeno-associated virus 2 (5.81, 95%CI 3.9, 7.7). In conclusion, we documented a difference in pediatric enteric viromes according to mBSFS-C stool consistency category, both in species richness and composition. Our results suggest that loose or watery stool, as measured by the mBSFS-C, may signal enteric viral infection in young children. Additional studies are warranted to confirm these findings.

2.2 Background

Stool consistency is an important diagnostic criterion in both research and clinical medicine[32]. Changes in stool consistency are used to measure many gastrointestinal disorders such as ulcerative colitis, irritable bowel syndrome and diarrhea[5,33–36]. Most epidemiologic studies of diarrheal disease internationally use stool consistency, specifically 'loose or watery stool' to classify diarrhea cases[4,5].

Visual and descriptive stool consistency scales may standardize and improve the accuracy of reported stool consistency. The most widely used stool form scale, The Bristol Stool Form Scale (BSFS), was developed in the late 1980s to measure gut transit time[16,17]. The BSFS classifies stool form into seven categories according to stool cohesion, surface cracking and consistency.

The BSFS was later simplified to a five-level scale and renamed the modified Bristol Stool Form Scale for children (mBSFS-C)[18], ranging from type 1 (hard pellets) to type 5 (watery stool).

While recent studies have described the bacterial microbiome of the colon and feces, there have been few parallel investigations of the enteric and stool virome[37–40]. Healthy gut and fecal bacterial microbiomes are characterized as having higher species richness. This relationship may not hold for the fecal virome, where higher species richness may signal more viral infections and disease.

Here we use metagenomics to examine the pediatric fecal virome across standardized stool consistency categories using stool samples from 269 children aged 0 to 5 years in rural Ethiopia. We evaluate potential associations between enteric virome composition, species richness and stool consistency.

2.3 Methods

2.3.1 Study Design

This study was conducted during the final visit of a clinical trial evaluating a water improvement intervention in the Amhara region of Ethiopia (clinicaltrials.gov NCT02373657)[19–21]. Methods for the parent trial are described in detail elsewhere[41,42]. A door-to-door population census was taken in all communities before the study visit. All children aged 0-5 years enumerated on the census were eligible to participate in the study and provide stool samples. The final study visit occurred in April 2016; April is the dry season in this region.

Stool sample collection and grading Caregivers were instructed to have their child defecate in a plastic child's potty chair lined with a black plastic bag. For children unable to produce a stool in the field, supplies were provided to the caregiver, with instructions to collect stool at home the following morning and bring it to a collection site the following day at a designated time.

When the stool was returned to the field station, it was independently inspected in the original collection container and graded according to the Modified Bristol Stool Form Scale for Children (mBSFS-C), which was available as a laminated sheet with both the cartoon images and Amharic translations (Figure 1.1). Methods describing the grading process and kappa evaluating agreement are described in detail elsewhere[41] .

After the stool sample consistency was graded, 0.5mL of stool was placed in a 1mL plastic tube, put on ice and transferred to a -20 Celsius freezer at the end of the day. At the completion of the sample collection, in early May 2016, all samples were transferred to Bahir Dar Regional Laboratory (Bahir Dar, Ethiopia) and kept at -20 Celsius until they were shipped to University of California, San Francisco in February 2017.

2.3.2 Laboratory methods

Stool specimens were combined into pools of 6 to 12, with sampling stratified by stool consistency grade. Pools were first clarified by 15,000g centrifugation for ten minutes, and supernatants filtered using a 0.45-*mu*m filter (Millipore). Nucleic acids in the filtrates were digested with a mixture of nuclease enzymes and viral nucleic acids were then extracted using a Maxwell 16 automated extractor (Promega)[43]. Random RT-PCR followed by Nextera™ XT Sample Preparation Kit (Illumina) were used to generate a library for Illumina MiSeq (2 x 250 bases) with dual barcoding as previously described[44,45].

2.3.3 Bioinformatic analyses

An in-house analysis pipeline was used to analyze sequence data. Raw data were first pre-processed by subtracting human and bacterial sequences, duplicate sequences, and low-quality reads. The reads were de novo assembled and contigs and singlet reads were aligned against a customized viral proteome database using BLASTx. Candidate viral hits were then compared to a non-virus non-redundant protein database to remove false positive viral hits.

2.3.4 Statistical methods

All statistical analyses were performed in R version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) using R Studio version 1.1.383. The number of viral reads along with the taxonomic assignments and sample characteristics were assembled using the phyloseq package. We define abundance as the raw number of reads for each species/genotype and relative abundance as the proportion of raw reads of each species/genotype in each pool. To calculate relative abundance, we divided the number or reads for each species/genotype in each pool by the total number of reads in that pool and multiplied

by 100. We evaluated species richness (Observed, Chao1) and alpha diversity measures (Simpson, Shannon, and Fisher) using the estimate_richness function of phyloseq. We used Spearman's rank order test to assess correlation between richness and mBSFS-C stool consistency category and determine statistical significance.

We determined differential abundance comparing loose and watery (mBSFS-C types 4&5) to formed stool (mBSFS-C types 1-3) at the species/genotype level using negative-binomial modeling in the DESeq2 package[46]. We used the Bonferroni method to adjust p-values for multiple comparisons. We explored both an unadjusted model and a model adjusted for age (median age per pool). Results are expressed as log2 fold change in loose or watery stool compared to formed stool.

2.3.5 Ethics Statement

Ethical committees at the University of California (San Francisco, CA, USA); Emory University (Atlanta, GA, USA); The Food, Medicine and Health Care Administration and Control Authority of Ethiopia; and the Ethiopian Ministry of Science and Technology granted approval for this study. We obtained verbal informed consent in Amharic from all caregivers.

2.4 Results

2.4.1 Characteristics of study population

Of 446 censured children who were eligible to participate, 317 children presented for the study visit examination and 269 provided stool samples. The mean age of children with stool samples was 2.7 years old, 53.2% (143/269) of children were female. A detailed description of the study population characteristics by stool consistency category is presented in (Table 2.1). The 269 samples were analyzed in 29 pools: 4 pools (29 samples) were watery/type 5, 8 pools (79 samples) were loose/type 4, 6 pools (59 samples) were

Table 2.1: Characteristics of the study population by mBSFS-C stool consistency category

	Watery	Loose	Smooth	Lumpy	Pellets
N children (pools)	29 (4)	79 (8)	59 (6)	88 (9)	14 (2)
Mean age (SD)	1 (1.14)	2.7 (1.6)	3.1 (2.1)	3.4 (1.81)	3.4 (1.65)
Female	15 (51.7%)	39 (49.4%)	33 (55.9%)	48 (54.5%)	8 (57.1%)
Blood in Stool*	4 (13.8%)	4 (5.1%)	2 (3.4%)	5 (5.7%)	0 (0%)
Fever**	9 (32.1%)	28 (36.8%)	20 (35.1%)	33 (38.8%)	6 (50%)

Numbers are N(%) unless otherwise indicated

*Caregiver reported observing blood in the stool any day in the past seven days

**Caregiver reported child had a fever any day in the past seven days

smooth/type 3, 9 pools (88 samples) were lumpy/type 2 and 2 pools (12 samples) were hard pellets/type1. The mean age for children with watery/type 5 stool was 1 year, younger than with loose/type 4 stool (2.7 years), smooth/type 3 (3.1 years), lumpy/type 2 (3.4 years) or pellet/type 1 (3.4 years). Caregivers reported observing blood in the stool in the past seven days for 4/29 (13.8%) children with watery stool samples, 4/79 (5.1%) of children with loose samples, 2/59 (3.4%) of children with smooth samples, 5/88 of (5.7%) children with lumpy samples and 0/14 (0%) of children with pellet samples.

2.4.2 Prevalence

The most prevalent viral reads belonged to anelloviruses (10 of 29 pools), picornaviruses in the species cosavirus A (9 of 29 pools), and salivirus A (8 of 29 pools). Several viruses/genotypes were more prevalent in watery or loose stool compared to formed stool (Figure 2.1). For example, Norovirus GII was detected in 5 of 12 (42 %) loose or watery pools and 0 of 17 formed pools.

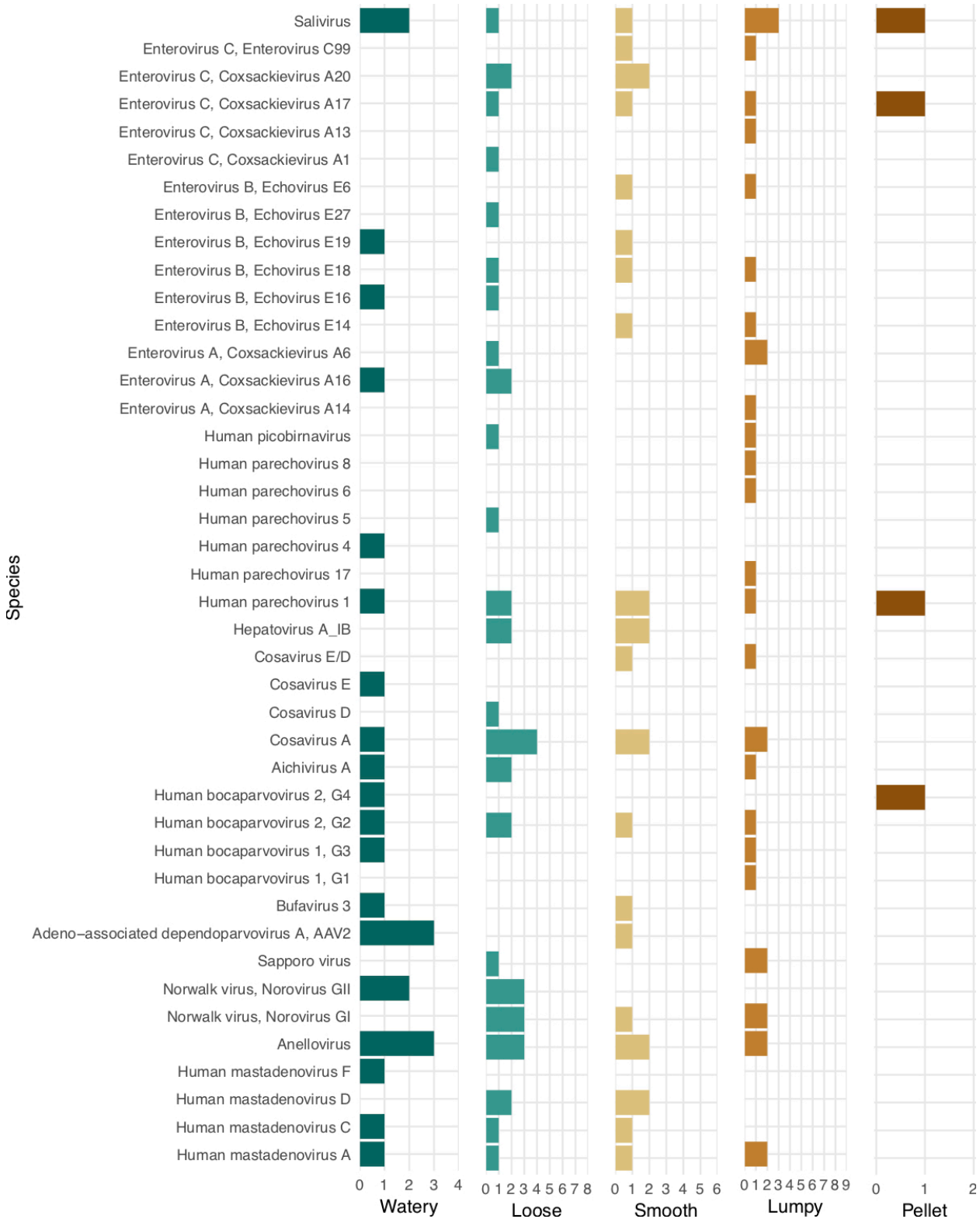


Figure 2.1: Prevalence of species/genotypes by mBSFS-C stool consistency category

Aichivirus A was detected in 3 of 12 (25 %) loose or watery pools and 1 of 17 (6%) formed pools. Human

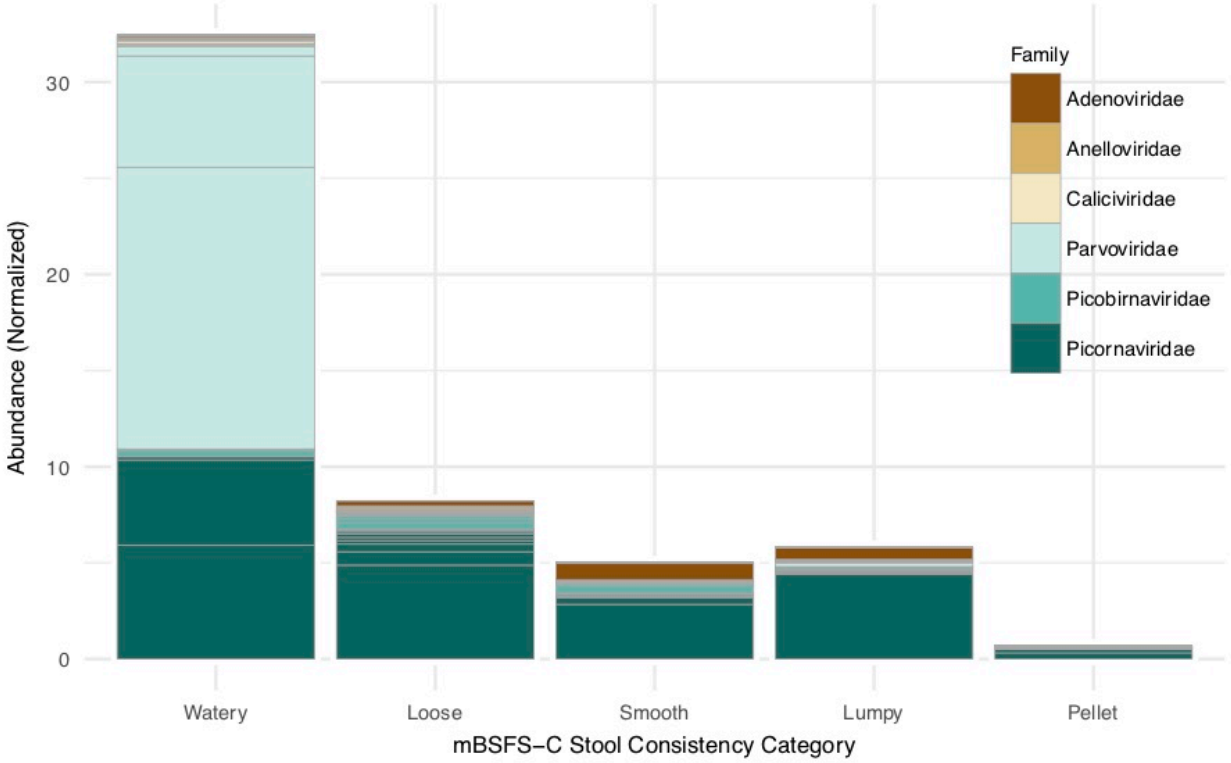


Figure 2.2: Viral abundance by mBSFS-C stool consistency category

mastadenovirus C was detected in 2 of 12 (17%) loose or watery pools and 1 of 17 (6%) formed pools. For some viruses, such as salivirus, Human parechovirus 1, and Saporro virus, there did not appear to be a difference in prevalence by stool consistency category. Human mastadenovirus D was only detected in 1 of 4 (25%) watery pools and no other pools. Rotavirus was not detected in any samples (Table 2.2).

2.4.3 Abundance

The most abundant reads belonged to the family Picornaviridae followed by Parvoviridae. Both Picornaviridae and Parvoviridae reads had the highest abundance in watery-consistency pools, followed by loose, smooth and lumpy consistency with the lowest abundance in pellet-consistency pools (Figure 2.2).

The most abundant species/genotype, both in terms of absolute and relative abundance, was adeno-

Table 2.2: Prevalence of species/genotype by mBSFS-C stool consistency category

N pools (children)	Watery N=4 (29)	Loose N=8(79)	Smooth N=6(59)	Lumpy N=9(88)	Pellets N=2(14)	Overall N=29(269)
Adeno-associated dependoparvovirus A, AAV2	3 (75%)	0 (0%)	1 (16.7%)	0 (0%)	0 (0%)	4 (13.8%)
Aichivirus A	1 (25%)	2 (25%)	0 (0%)	1 (11.1%)	0 (0%)	4 (13.8%)
Anellovirus	3 (75%)	3 (37.5%)	2 (33.3%)	2 (22.2%)	0 (0%)	10 (34.5%)
Bufavirus 3	1 (25%)	0 (0%)	1 (16.7%)	0 (0%)	0 (0%)	2 (6.9%)
Cosavirus A	1 (25%)	4 (50%)	2 (33.3%)	2 (22.2%)	0 (0%)	9 (31%)
Cosavirus D	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Cosavirus E	1 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Cosavirus E/D	0 (0%)	0 (0%)	1 (16.7%)	1 (11.1%)	0 (0%)	2 (6.9%)
Enterovirus A, Coxsackievirus A14	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Enterovirus A, Coxsackievirus A16	1 (25%)	2 (25%)	0 (0%)	0 (0%)	0 (0%)	3 (10.3%)
Enterovirus A, Coxsackievirus A6	0 (0%)	1 (12.5%)	0 (0%)	2 (22.2%)	0 (0%)	3 (10.3%)
Enterovirus B, Echovirus E14	0 (0%)	0 (0%)	1 (16.7%)	1 (11.1%)	0 (0%)	2 (6.9%)
Enterovirus B, Echovirus E16	1 (25%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	2 (6.9%)
Enterovirus B, Echovirus E18	0 (0%)	1 (12.5%)	1 (16.7%)	1 (11.1%)	0 (0%)	3 (10.3%)
Enterovirus B, Echovirus E19	1 (25%)	0 (0%)	1 (16.7%)	0 (0%)	0 (0%)	2 (6.9%)
Enterovirus B, Echovirus E27	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Enterovirus B, Echovirus E6	0 (0%)	0 (0%)	1 (16.7%)	1 (11.1%)	0 (0%)	2 (6.9%)
Enterovirus C, Coxsackievirus A1	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Enterovirus C, Coxsackievirus A13	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Enterovirus C, Coxsackievirus A17	0 (0%)	1 (12.5%)	1 (16.7%)	1 (11.1%)	1 (50%)	4 (13.8%)
Enterovirus C, Coxsackievirus A20	0 (0%)	2 (25%)	2 (33.3%)	0 (0%)	0 (0%)	4 (13.8%)
Enterovirus C, Enterovirus C99	0 (0%)	0 (0%)	1 (16.7%)	1 (11.1%)	0 (0%)	2 (6.9%)
Hepatovirus A, Hepatovirus A_IB	0 (0%)	2 (25%)	2 (33.3%)	0 (0%)	0 (0%)	4 (13.8%)
Human bocaparvovirus 1, Human bocavirus 1	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Human bocaparvovirus 1, Human bocavirus 3	1 (25%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	2 (6.9%)
Human bocaparvovirus 2, Human bocavirus 2	1 (25%)	2 (25%)	1 (16.7%)	1 (11.1%)	0 (0%)	5 (17.2%)
Human bocaparvovirus 2, Human bocavirus 4	1 (25%)	0 (0%)	0 (0%)	0 (0%)	1 (50%)	2 (6.9%)
Human mastadenovirus A	1 (25%)	1 (12.5%)	1 (16.7%)	2 (22.2%)	0 (0%)	5 (17.2%)
Human mastadenovirus C	1 (25%)	1 (12.5%)	1 (16.7%)	0 (0%)	0 (0%)	3 (10.3%)
Human mastadenovirus D	0 (0%)	2 (25%)	2 (33.3%)	0 (0%)	0 (0%)	4 (13.8%)
Human mastadenovirus F	1 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Human picobirnavirus	0 (0%)	1 (12.5%)	0 (0%)	1 (11.1%)	0 (0%)	2 (6.9%)
Norwalk virus, Norovirus GI	0 (0%)	3 (37.5%)	1 (16.7%)	2 (22.2%)	0 (0%)	6 (20.7%)
Norwalk virus, Norovirus GII	2 (50%)	3 (37.5%)	0 (0%)	0 (0%)	0 (0%)	5 (17.2%)
Parechovirus A, Human parechovirus 1	1 (25%)	2 (25%)	2 (33.3%)	1 (11.1%)	1 (50%)	7 (24.1%)
Parechovirus A, Human parechovirus 17	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Parechovirus A, Human parechovirus 4	1 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Parechovirus A, Human parechovirus 5	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	1 (3.4%)
Parechovirus A, Human parechovirus 6	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Parechovirus A, Human parechovirus 8	0 (0%)	0 (0%)	0 (0%)	1 (11.1%)	0 (0%)	1 (3.4%)
Salivirus	2 (50%)	1 (12.5%)	1 (16.7%)	3 (33.3%)	1 (50%)	8 (27.6%)
Sapporo virus	0 (0%)	1 (12.5%)	0 (0%)	2 (22.2%)	0 (0%)	3 (10.3%)

Number of positive pools (%)

AAV2 = Adeno-associated virus 2

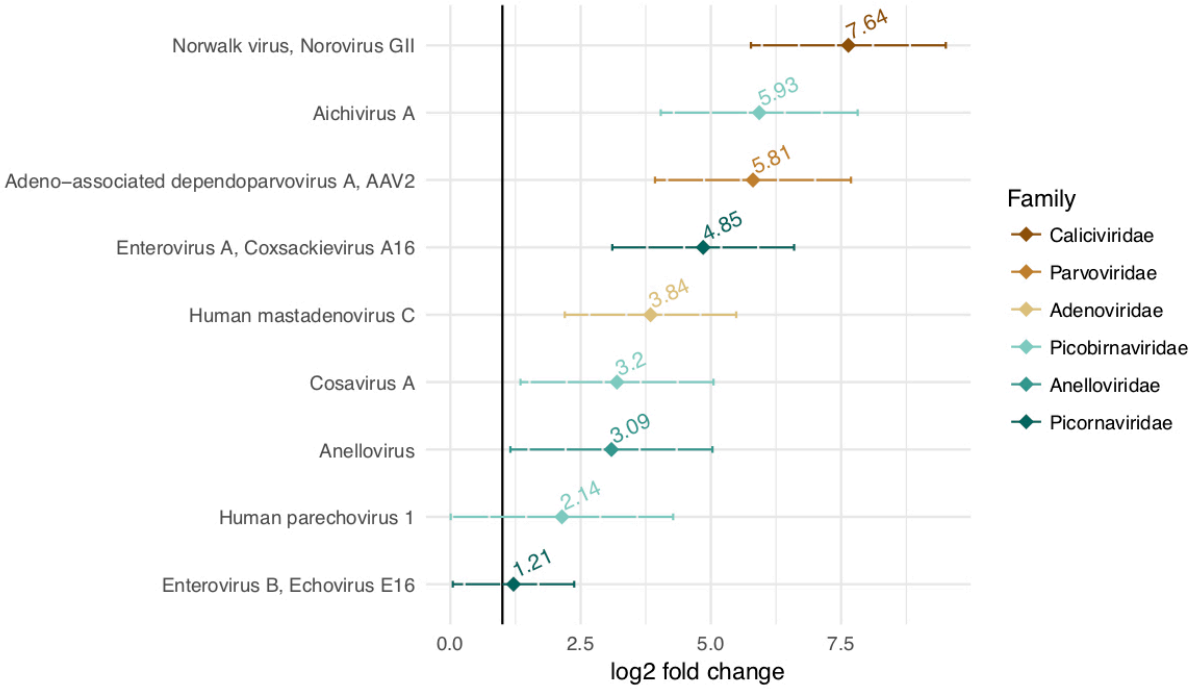


Figure 2.3: Differential abundance: Loose/watery stool compared to formed stool

associated dependoparvovirus A, adeno-associated virus 2 (AAV2) followed by saliviruses, human bocavirus 2, and echovirus E19. The greatest differential abundance comparing loose or watery stool to formed stool was for norovirus GII, followed by aichivirus A, AAV2, coxsackievirus A16, human mastadenovirus, cosavirus A and anellovirus (Figure 2.3).

For norovirus GII, the abundance was 2656 reads in loose in watery consistency pools and 0 reads in formed consistency with a differential abundance of 7.64 (95% CI 5.8 – 9.5, p=<0.001) comparing loose or watery stool to formed stool. For aichivirus A, the abundance was 5659 reads in loose or watery stool compared to 86 reads in formed stool, with a differential abundance of 5.93 (95% CI 4.0-7.8, p=<0.001). The abundance of AAV2 was 115,946 in loose or watery-consistency pools and 222 in formed-consistency pools with a differential abundance of 5.81 (95%CI 16.0 - 27.6, p=<0.001) (Table 2.3).

Table 2.3: Differential abundance and $\log_2$ fold change

Species	Abundance in loose stool (normalized value)	Abundance in formed stool (normalized value)	$\log_2$ fold change	p-value	adjusted p-value
Norwalk virus, Norovirus GII	2656 (0.26)	0 (0)	7.64 (5.8 , 9.5)	<0.001	<0.001
Aichivirus A	5659 (0.38)	86 (0.01)	5.93 (4.0 , 7.8)	<0.001	<0.001
Adeno-associated dependoparvovirus A, AAV2	115946 (14.73)	222 (0.03)	5.81 (3.9 , 7.7)	<0.001	<0.001
Enterovirus A, Coxsackievirus A16	2561 (0.22)	0 (0)	4.85 (3.1 , 6.6)	<0.001	<0.001
Human mastadenovirus C	3023 (0.27)	449 (0.03)	3.84 (2.2 , 5.5)	<0.001	<0.001
Cosavirus A	932 (0.09)	122 (0.02)	3.2 (1.3 , 5)	0.001	0.003
Anellovirus	2515 (0.28)	427 (0.06)	3.09 (1.2 , 5)	0.002	0.006
Parechovirus A, Human parechovirus 1	3545 (0.53)	1986 (0.24)	2.14 (0 , 4.3)	0.049	0.071
Enterovirus B, Echovirus E16	71447 (4.44)	0 (0)	1.21 (0 , 2.4)	0.040	0.064

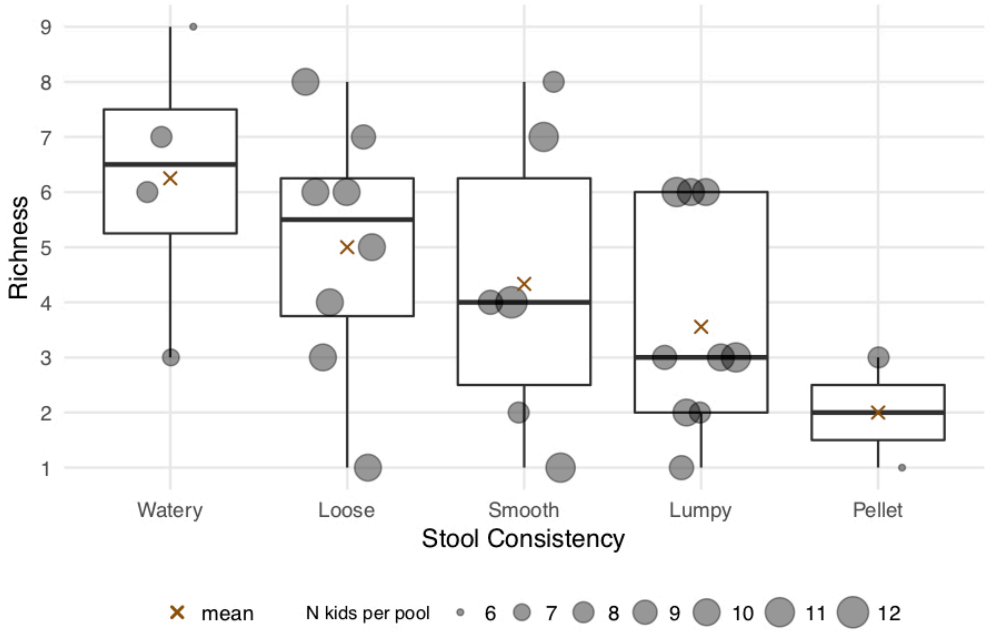


Figure 2.4: Species richness according to mBSFS-C stool consistency category

2.4.4 Richness and Alpha Diversity

Species richness was highest in watery-consistency stool and decreased consistently as stool consistency became firmer (Spearman's r=-0.45, p=0.013). The median number of distinct viruses was 6.5 (IQR 2.25) for watery stool, 5.5 (IQR 2.5) for loose stool, 4.0 (IQR 3.75) for smooth-consistency stool, 3.0 (IQR 4) for lumpy-consistency stool and 2.0 (IQR 1) for pellet-consistency stool (Figure 2.4).

There was no association between stool consistency and Simpson alpha diversity (Spearman's r=-0.05, p=0.80) or Shannon alpha diversity (Spearman's r=-0.08, p=0.67). With the Fisher alpha diversity met-

ric, loose and watery stool appeared to be more diverse than formed stool, but the difference was not statistically significant (Spearman's r=-22, p=0.238) (Figure 2.5).

2.5 Discussion

We document differences in the richness and composition of the pediatric enteric virome by stool consistency. There was a marked and statistically significant decrease in the number of distinct virus species as stool consistency became firmer, with the highest number of distinct species in watery (mBSFS-C type 5) stool and the lowest number of distinct species in firm pellet stool (mBSFS-C type 1). Loose and watery stool types were more likely to have norovirus GII, aichivirus A, adeno-associated dependoparvovirus A - AAV2, coxsackievirus A16, human mastadenovirus C, cosavirus A and annelovirus compared to formed stool.

Our findings regarding species richness contrast with many studies in the bacterial microbiome literature which report elevated species richness in healthy individuals and stool. For example, a study of the fecal bacterial microbiome according to self-reported BSFS stool consistency and found that women with looser stool had fewer distinct bacterial species compared to women with firmer stool[40]. Virus richness, unlike bacterial richness, may be a marker for disease rather than health. Stool consistency is often used in symptomatic definitions of diarrhea, in particular, the World Health Organization definition ("three or more loose or watery stools in a 24 hour period[5]"). Our findings of elevated species richness in watery and loose stool support the content validity of consistency-based definitions as they relate to infectious episodes of diarrhea.

Noroviruses are the leading cause of epidemic viral gastroenteritis globally[47]. We detected norovirus genotype II and genotype I in a fifth of pools, with genotype II showing the highest differential abundance in loose or watery stools compared to formed stools. Norovirus genotype II has been shown to be elevated

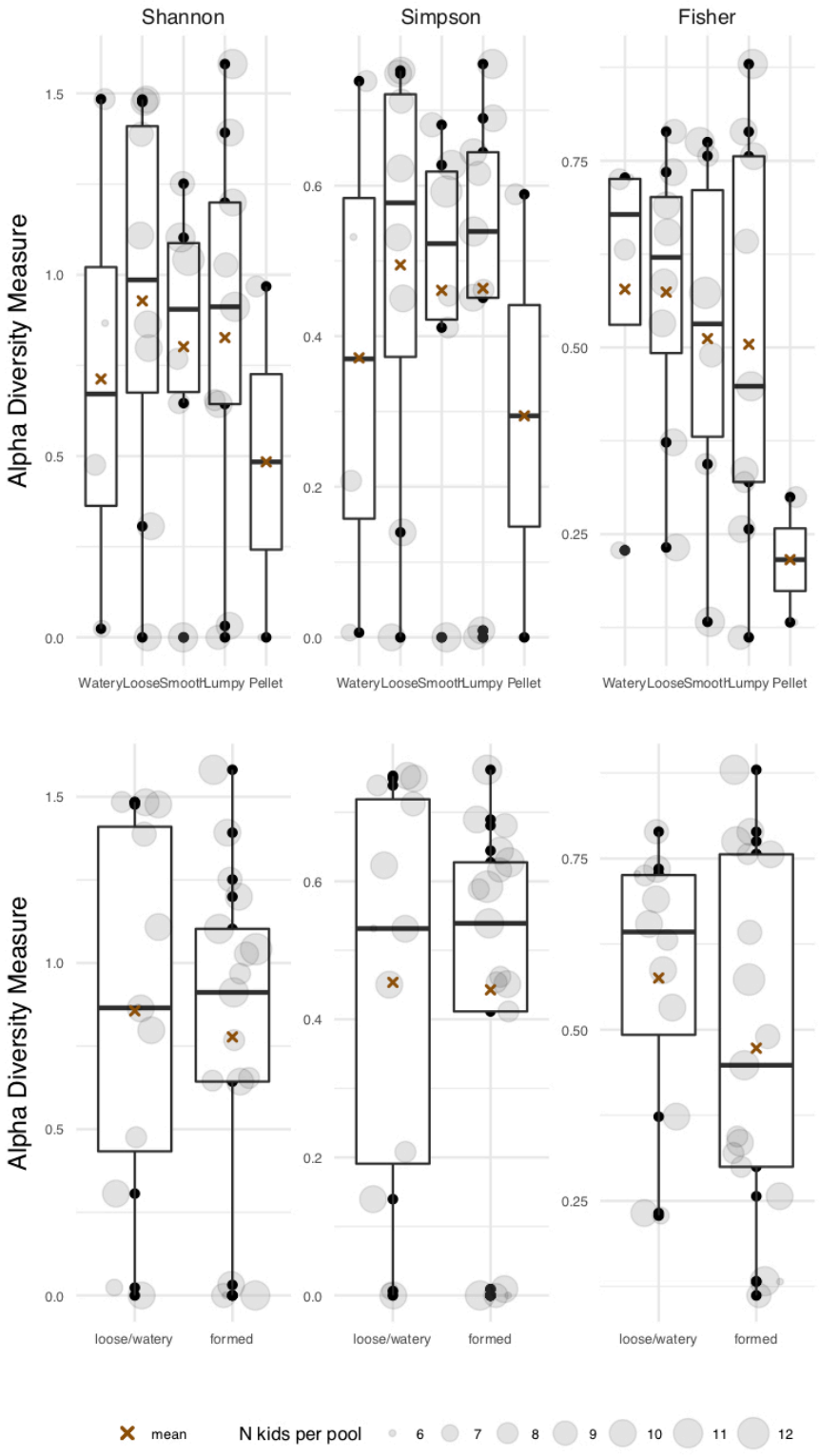


Figure 2.5: Alpha Diversity according to mBSFS-C stool consistency category

in children with diarrhea and or gastroenteritis in many regions, including Africa[48,49].

The second highest differential abundance we detected was for aichivirus A, which was detected 25% of loose or watery pools and 6% of formed pools. Aichivirus A has been detected in stool samples from children with diarrhea in France[50], Brazil[51], Tunisia[52] and several countries in Asia[53,54]. We likewise found an elevated abundance of adeno-associated dependoparvovirus A - AAV2 in loose or watery consistency pools compared to formed consistency pools. AAV2 needs a helper virus, commonly an adenovirus, to infect the gastrointestinal tract and it is not thought to be pathogenic or cause diarrheal disease[55]. However, adeno-associated viruses have been detected in children with diarrhea using metagenomic sequencing[56]. Indeed, we detected adenovirus in three of the four pools with detected AAV2 reads.

We also found an increased abundance of coxsackievirus A16 in loose or watery consistency pools compared to formed pools. Coxsackievirus A16 is a leading cause of hand foot and mouth disease. Hand-foot and mouth infections are common in young children and we suspect that age may confound the association between stool consistency and coxsackie A16 abundance. In the median age-adjusted model, the differential abundance of coxsackie A16 in loose or watery stools compared to formed stools decreased to 1.78, and the 95% confidence intervals included the null effect of 1.0.

Enteric human adenoviruses, primarily human mastadenovirus F, are considered the third leading cause of nonbacterial bacterial diarrhea globally[48,57]. We detected human mastadenovirus F in just one watery pool with too few reads to detect a statistically significant difference. We did detect a signal for higher differential abundance of human mastadenovirus C in loose or watery stool compared to formed stool although the p-value was not statistically significant. Studies in Albania, Korea and Asia have all detected human mastadenovirus C in children with diarrhea and gastroenteritis, but these studies lacked healthy controls to compare prevalence[58–60]. A study of human adenovirus in diarrhea children in Tanzania did not find a difference in the prevalence of human mastadenovirus F between children with and

without diarrhea but did report a human mastadenovirus C prevalence of 12.5% in diarrhea cases and 7.7% in controls (prevalence ratio of 1.6)[61]. Human mastadenovirus C may be an unrecognized cause of pediatric diarrhea in Africa and warrants additional research with a larger sample size.

Cosavirus A was first identified 2008 and has since been recognized as a cause of diarrhea in children[62–65]. In our sample, Cosavirus A was 3.2 times more abundant in loose or watery stool compared to formed stool. Additional un-pooled studies would help confirm if Cosavirus A may be a cause of loose or watery stools in this population. Annelovirus is a nearly ubiquitous virus in human blood and its presence in the pooled samples may indicate blood in the feces, particularly in loose and watery stool samples[66,67]. We did not detect any rotavirus, a leading cause of pediatric diarrhea. In 2013, Ethiopia initiated a country-wide vaccination campaign reaching a coverage of 85% by 2015[68]. Successful rotavirus vaccination may explain why no rotavirus was detected in loose, watery or formed stool.

There are several limitations of this study that are important to consider. First, our samples were pooled, reducing the effective sample size from 269 individual stool specimens to 29 pools. With a larger sample size (smaller pools) or individually run samples, we would have had more statistical precision for evaluating differences in the fecal virome by measured stool consistency. However, pooling has been shown to be an efficient strategy to accurately estimate prevalence when resources are limited[12,69,70] Second, the kappa measuring agreement between the laboratory technicians stool consistency grades was .72, introducing some possible misclassification of stool consistency into our analyses. This misclassification, likely non-differential, would on average bias any associations towards the null. Third, laboratory staff were not masked to the consistency of the pools when running the analyses. However, the samples were processed in an arbitrary order, not according to stool consistency. Fourth, we did not run any of the pools in duplicate to assess repeatability. Finally, fecal specimens were stored without media, although this should not affect identification of viruses as it would for bacteria.

Despite these limitations, our study had several strengths. Unlike many comparable studies using targeted PCR to detect specific pre-specified viruses, we used a metagenomic approach allowing systematic and unbiased characterization of the stool virome. We applied modern statistical tools developed for microbiome analysis, enabling a comprehensive view of the entire fecal viral community. Our study was also population-based, enrolling a representative sample of all children in the community, both healthy and sick.

In conclusion, we found differences in the pediatric fecal virome by stool sample consistency as measured by the modified Bristol Stool Form Scale for Children (mBSFS-C). Watery and loose stools had greater species richness compared to formed stools and were more abundant in norovirus GII, aichivirus A, AAV2 and human mastadenovirus C. Our results suggest that loose or watery stool, as measured by the mBSFS-C, may signal enteric viral infection in young children. Additional studies with larger samples sizes are warranted to confirm these findings.

3

Quantitative Bias Analysis for Misclassified Pediatric Diarrhea in a Cluster-Randomized Water Intervention Trial

3.1 Abstract

Diarrhea is a major contributor to global child morbidity and mortality. Most trials measure diarrhea using caregiver-reported symptoms, which are prone to substantial measurement error. Bias analysis can be used to estimate the direction, magnitude and uncertainty arising from misclassified outcomes. Our objective is to report the effect of a water intervention trial on caregiver-reported pediatric diarrhea, and to perform bias analysis to estimate the effect of the intervention accounting for misclassified diarrhea. Fourteen communities were selected for the trial, with half randomized to a water point intervention and the other half to control. Caregivers were asked to report if children had three or more loose or watery stools in a 24-hour period anytime in the past seven days. We used logistic regression, along with robust standard errors to accommodate clustering within communities, to estimate the effect of the intervention on caregiver-reported diarrhea and assessed bias using several methods: a two-by-two correction, probabilistic bias analysis with uniform, logit-normal and triangular probability functions, and a regression correction approach. From an internal validation study, the sensitivity of caregiver-reported stool consistency 53.3% (95 % CI: 40.6, 65.7) and the specificity was 94.2% (95 % CI: 87.6, 97.4). We used these parameters as estimates for misclassification of caregiver-reported diarrhea. The uncorrected odds ratio (OR) was 1.2 (95% CI: 0.62, 2.32); the two-by-two bias-corrected OR was 1.37 (95% CI: 0.57, 3.3); the probabilistic bias-corrected OR with the logit-normal distribution was 1.36 (95% CI: 0.81, 2.24) and regression correction OR was 1.36 (95% CI 0.77, 2.38). For both the observed and bias-corrected estimates children the intervention arm had a slightly higher odds of reported-diarrhea, however confidence intervals for all estimates included the null. The confidence intervals for all bias-corrected methods were wider than the uncorrected confidence intervals. Bias analysis did not reveal a corrected protective effect of the water-improvement intervention. Larger trials with more precise effect estimates may be more informative

targets for misclassification bias analysis.

3.2 Background

Diarrhea is a leading cause of global child morbidity and mortality with an estimated 2 billion cases and 525,000 deaths every year[1]. Ethiopia has the fourth largest burden of diarrhea mortality globally and the second highest in Africa[71]. In Ethiopia, diarrhea is the second-leading cause of death among children under five, causing more deaths than HIV, TB and malaria combined[2,3]. Most epidemiologic studies, surveillance programs and trials measure diarrhea using caregiver-reported symptoms often using the World Health Organization (WHO) definition of diarrhea "three or more loose or watery stools in a 24 hour period[5]." In three recent systematic reviews of water, sanitation and hygiene (WASH) intervention trials to prevent pediatric diarrheal disease, all 22 studies used caregiver-reported symptoms to classify diarrhea[6,7,9,72]. In a 2015 review of 55 studies of water quality interventions for reducing diarrheal disease, 36 used the WHO definition and 11 used another symptoms-based report[11]. Caregiver-reported symptoms are prone to substantial measurement error[41]. Often there are incentives for caregivers to misreport symptoms such as establishing rapport with the interviewer, the promise of treatment or avoiding embarrassment[6,14,73]. Additionally, questions about child's symptoms can be misunderstood or the caregiver may simply not know the answer. Moreover, in water, sanitation and hygiene infrastructure trials, it is often impossible to mask the intervention and control groups, presenting an opportunity for differential reporting of the outcome with respect to exposure[74]. Caregivers may overreport or underreport symptoms based on their intervention group. While water, sanitation, and hygiene improvements are widely thought to reduce and/or prevent pediatric diarrhea, evidence from rigorously conducted randomized trials is scarce. In fact, five recent randomized trials, in Mali, India, Kenya and Bangladesh found no effect of WASH interventions on pediatric diarrhea[75–79]. Misclassification of binary outcomes introduces

systematic bias that can mask true effects[80] and is one potential explanation for why trials do not discern a measurable effect of WASH interventions on pediatric diarrhea. When misclassification probabilities (ie sensitivity and specificity) are known, bias analysis can be used to estimate the direction, magnitude and uncertainty arising from misclassified outcomes[81]. Most bias analysis tools readily available and familiar to epidemiologists use summary two-by-two tables, which cannot account for clustering of standard errors or for covariates. In contrast, Neuhaus (1999, 2002) developed a general approach to incorporate outcome misclassification into any binary regression. The approach requires known estimates of sensitivity and specificity which can come from external validation studies. Despite these advantages, regression correction approaches have not been adapted into routine epidemiological workflows.

Our objective is to report the effect of a water intervention trial on caregiver-reported pediatric diarrhea, and to perform quantitative bias analysis to estimate the effect of the intervention accounting for misclassified diarrhea. We use three bias analysis approaches: 1) the traditional 2 by 2 correction, whereby a summary two-by-two table of exposure and outcome is corrected by re-allocating the numbers in each cell based on the sensitivity and specificity of the outcome[82] 2) probabilistic bias analysis, which uses probability distributions to describe the misclassification parameters then generates corrected distributions using Monte Carlo sampling[83], and 3) a regression correction approach whereby a binary regression is corrected for with specified sensitivity and specificity[84,85].

3.3 Methods

3.3.1 Trial design

We conducted a cluster-randomized trial of a water improvement intervention in Ethiopia from April 2014 through April 2016 (clinicaltrials.gov NCT02373657) The intervention was primarily targeted at

reducing trachoma, and the primary outcome was ocular chlamydia infection in children. Fourteen communities were selected for the trial, with half randomized to a water point intervention and the other half to control. The study design and methods for the trial are described in detail elsewhere. The intervention consisted of building a new hand dug well in each community. The baseline visit for the trial occurred in April 2014 and the final study visit occurred in April 2016. April is the dry season in this region.

3.3.2 Study sample and selection

The trial study took place in a rural agrarian region in the Goncha Siso Enese district (woreda) of Amhara, Ethiopia. Woredas in Ethiopia are divided into administrative units known as kebeles, and at the time of the study, kebeles were subdivided into government-defined units known as state teams. State teams, which consisted of approximately 275 people in our study area, are termed communities in this paper.

A door-to-door population census was taken in all communities approximately two years following initiation of the intervention. All children aged 0-5 years (i.e., up to but not including the sixth birthday) enumerated on the census were eligible to participate in the two-year monitoring visit, which was conducted approximately 4 weeks following the census.

3.3.3 Diarrhea Assessment

At the time of the monitoring visit, each child's caregiver was asked if the child had experienced three or more loose or watery stools either (A) today, (B) yesterday, or (C) during a 24-hour period over the past 1 week. If any of these were answered affirmatively, the number of stools and presence of blood in the stool was documented. The responses were used to generate three definitions of diarrhea: (1) three or more loose or watery stools, (2) blood in the stool and (3) three or more watery stools and fever) over three time points (today, yesterday and in the past seven days)[4]. The primary diarrhea outcome for the trial was

pre-specified as three or more loose or watery stools in a 24-hour period anytime in the past seven days. Trained field staff asked all the questions according to a standardized protocol.

3.3.4 Statistical methods

Trial outcome

We used binomial, logit-link generalized estimating equations (GEE) with robust standard errors to estimate the effect of the intervention on the nine caregiver-reported definitions of diarrhea. From the GEE logit coefficients, we calculated both the odds ratio and the prevalence ratio for the effect of the water intervention on reported diarrhea[86].

Misclassification parameters

We used sensitivity and specificity estimates for caregiver-reported stool consistency from an internal validation study to approximate the misclassification parameters for caregiver-reported diarrhea. Details for the validation study are described in detail elsewhere[41]. In brief, we collected stool samples from children and compared the caregiver's report of stool sample consistency to that of two trained laboratory technicians. the sensitivity of caregiver-reported stool consistency was 53.3% (95 % CI: 40.6, 65.7) and the specificity was 94.2% (95 % CI: 87.6, 97.4). The unweighted kappa for the two laboratory technician grades was 0.51 (95 % CI: 0.40, 0.61) and the quadratic-weighted kappa was 0.72 (95 % CI: 0.59, 0.85). We considered the possibility of differential misclassification; ie if validity of reported symptoms is different in intervention verses control communities. In the water intervention communities, the sensitivity was 49.0% (95 % CI: 37.1, 61.1) and the specificity was 96.0% (95 % CI: 87, 98.8) and in the control communities the sensitivity was 57.4% (95 % CI: 35.1, 77.1) and the specificity was 92.5% (95 % CI: 79.2, 97.6). However, the p-value evaluating a difference in sensitivity by study arm was 0.51 and for specificity it was 0.43. We

used 99% confidence intervals to guide the probability density distribution inputs.

Simple bias analysis

We calculated the bias-corrected odds ratio and prevalence ratio under two different scenarios: differential misclassification of the outcome (diarrhea) with respect to exposure (water intervention assignment) and non-differential misclassification. We followed the misclassification bias-correction methods described by Greenland (1996), Rothman (2008) and Lash (2011)[82,83,87]. In brief, each cell in the observed two-by-two summary table is corrected using the misclassification parameters (sensitivity and specificity) to estimate the expected truth given the misclassification probability[82]. For example, the cell that is outcome positive and exposure positive is multiplied by the sensitivity of the outcome measurement and added to the value of the cell that is disease negative and exposure positive multiplied by one minus specificity. We used bootstrapping with 5000 replicates to estimate the 95% confidence intervals for the corrected ORs and RRs.

Probabilistic bias analysis

We used probabilistic bias analysis to explore a range of misclassification probabilities with three probability distributions (uniform, logit-normal and triangular) under differential and non-differential scenarios[83]. The parameters defining these distributions are listed in Table 3.1. We then used Monte Carlo sampling from the misclassification probability distribution with 10,000 iterations to generate frequency distributions for the corrected measures of effect. The median value of the frequency distribution describes the corrected measure of association and the 2.5th percentile and 97.5th percentile of the distribution indicate the limits in which 95% of corrected measures of association lie. We then plotted the relationship between sensitivity/specificity and the corrected OR and PR.

Table 3.1: Probability distribution parameters

		Non-differential		Differential			
		Sensitivity	Specificity	Sensitivity		Specificity	
				Intervention	Control	Intervention	Control
Uniform	Min	40.0%	70.0%	40%*	40%*	70%**	70%**
	Max	99.9%	99.9%	99.9%*	99.9%*	99.9%**	99.9%**
Logit-normal	Location	53.3%	94.2%	49.0%	57.4%	96.0%	92.5%
	Scale	0.9	0.9	0.9	0.9	0.9	0.9
	Lower bound shift	32.5%	80.2%	29.7%	22.5%	73.8%	63.2%
	Upper bound shift	73.0%	98.5%	68.7%	86.2%	99.5%	98.9%
Triangular	Lower limit	32.5%	80.2%	29.7%	22.5%	73.8%	63.2%
	Upper limit	73.0%	98.5%	68.7%	86.2%	99.5%	98.9%
	Mode	53.3%	94.2%	49.0%	57.4%	96.0%	92.5%

*correlation for: sensitivity = .5; **correlation for specificity = .5

Triangular and Logit-Normal distributions based on 99% CIs from validation study

Regression correction

We also estimated the bias-corrected odds ratios and prevalence ratio using a regression correction method developed by Neuhaus (1999 & 2002)[84,85] for both a generalized linear model (GLM) and a GEE with robust standard errors to account for clustering by community. Details describing the link function specification can be found in the appendix. For both the GLM and GEE, we specified custom forward and inverse logit link functions with the parameter $\gamma_0 = 1 - \textit{Specificity}$, and $\gamma_1 = 1 - \textit{Sensitivity}$ (Equations 3.1 & 3.2). We adapted the original SAS GLM approach to run in R and developed the GEE approach in R.

$$\textit{Forward logit link} : g(\mu) = \log\left(\frac{\mu - \gamma_0}{1 - \gamma_1 - \mu}\right) \tag{3.1}$$

$$\textit{Inverse logit link} : g(\eta) = \frac{1 - \gamma_1 - \gamma_0}{1 + e^{-}} + \gamma_0 \tag{3.2}$$

Misclassification parameters (sensitivity and specificity) were estimated in in Stata 15 (StataCorp, College Station, TX). All other statistical analyses were performed in R version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) using R Studio version 1.1.383. We used the episenr package for the simple and probabilistic bias analyses[88].

3.3.5 Ethics statement

Ethical committees at the University of California (San Francisco, CA, USA); Emory University (Atlanta, GA, USA); The Food, Medicine and Health Care Administration and Control Authority of Ethiopia; and the Ethiopian Ministry of Science and Technology granted approval for this study. We obtained verbal informed consent in Amharic from all caregivers.

3.4 Results

3.4.1 Characteristics of the study population

Of 446 censured children who were eligible to participate, 317 children presented for the study visit examination and 315 caregivers responded to questions about diarrhea symptoms. The mean age of children with stool samples was 2.8 years old, 51.7% (163/269) of children were male. Population characteristics by study arm are described in Table 2.

Table 2: Study population characteristics

Table 3.2: Study population characteristics

	Treatment arm	
	Water point n=7 communities	Control n=7 communities
Total number of children	155	160
Children per community (median (min - max)	18 (16-38)	14 (10-39)
Age in yrs (mean (sd))	2.93 (2.02)	2.67 (1.71)
Male	68 (43.9)	84 (52.5)
Caregive role		
Mother	85 (57.0)	97 (63.0)
Father	16 (10.7)	11 (7.1)
Aunt	5 (3.4)	5 (3.2)
Uncle	4 (2.7)	7 (4.5)
Guardian	1 (0.7)	1 (0.6)
Other	32 (21.5)	26 (16.9)
Self	6 (4.0)	7 (4.5)

3.4.2 Trial Results

Using the prespecified diarrhea definition of an episode of three or more loose or watery stools in 24 hours at any point in the past seven days, 28.6% (42/155) of children in the water intervention arm and 24.8% (38/160) of children in the control arm had a reported an event of diarrhea (prevalence ratio (PR) = 1.17 (95% CI: 0.6, 2.26); OR = 1.2 (95% CI 0.62, 2.32); p = 0.571). The ICC was 0.0271. There was no indication for a protective effect of the water intervention on any of the nine diarrhea definitions (Table 3). The difference in age-adjusted diarrhea prevalence between intervention and control arms is depicted in Figure 3.1.

Table 3: Reported Diarrhea Today, Yesterday and in the Past 7 days

3.4.3 Simple bias analysis

For the primary diarrhea outcome, the corrected estimate for the PR was 1.20 (95% CI: 0.72, 1.97) and the OR was 1.37 (95% CI: 0.57, 3.3) under non-differential assumptions. Under differential assumptions, the

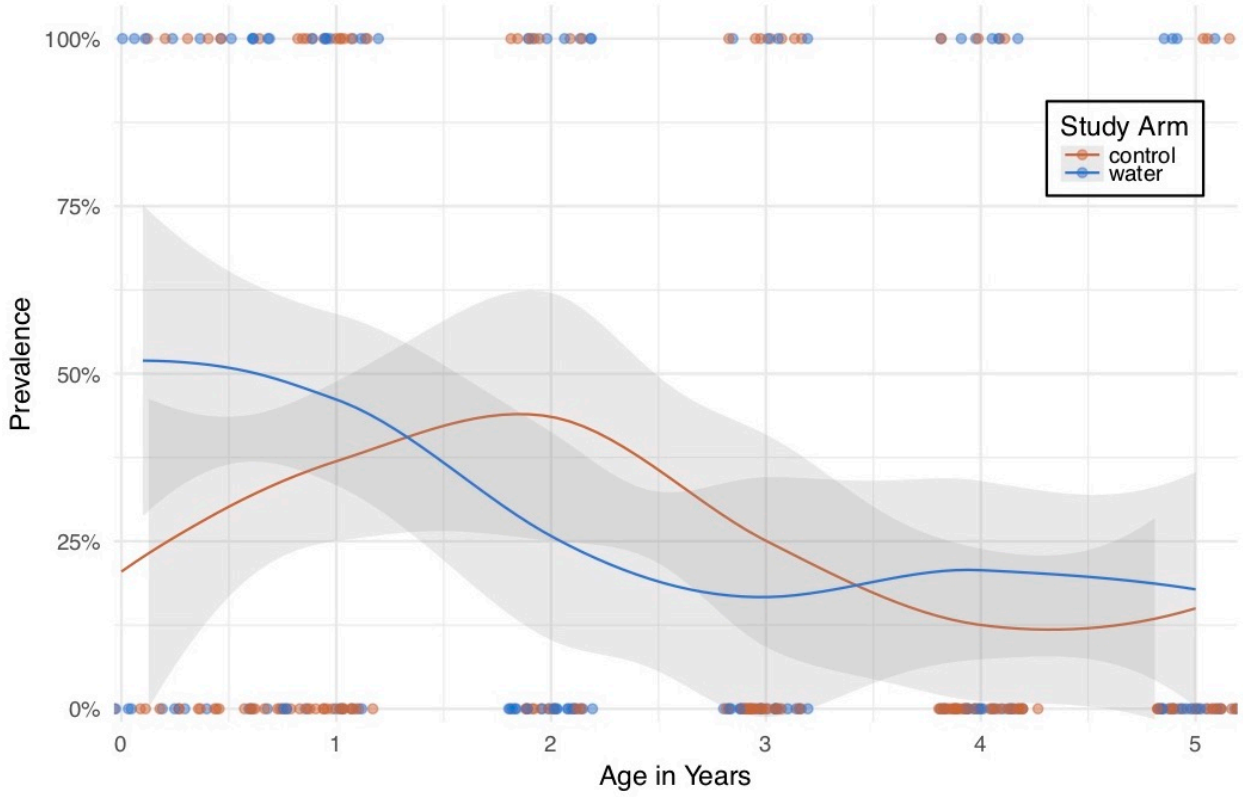


Figure 3.1: Age-adjust diarrhea prevalence in the past seven days, using WHO definition (three or more loose or watery stools in a 24-hour period)

Table 3.3: Probabilistic Bias Analysis Results

		PR (95% CI)	OR (95% CI)
Observed		1.17 (0.6, 2.26)	1.2 (95% CI: 0.62, 2.32)
Uniform	Non-differential	1.36 (0.87, 4.88)	1.50 (0.82, 5.44)
	Differential	1.19 (0.16, 9.98)	1.27 (0.13, 13.68)
Logit-normal	Non-differential	1.21 (0.83, 1.74)	1.36 (0.81, 2.24)
	Differential	1.73 (1.10, 2.79)	2.29 (1.23, 4.34)
Triangular	Non-differential	1.26 (0.85, 1.90)	1.41 (0.84, 2.44)
	Differential	1.64 (0.72, 13.12)	2.07 (0.65, 20.82)

PR = Prevalence Ratio; OR = Odds Ratio

corrected PR was 1.64 (95% CI: 0.96, 2.75) and the corrected OR was 2.49 (95% CI: 0.93, 6.36).

3.4.4 Probabilistic bias analysis

Results for the probabilistic bias analysis using uniform, logit-normal and triangular distributions are presented in Table 3.3. Confidence intervals for nearly all estimates included 1, except for logit-normal differential corrections for both the PR and PR. OR estimates and upper bounds were farther from the null compared to the PR. Upper bound estimates were more extreme for differential misclassification under a uniform and triangular distribution. In the non-differential misclassification analysis, corrected PR and OR estimates were highly dependent on the specificity of the outcome measurement (Figure 3.2). In the differential misclassification analysis, corrected PRs and ORs increased with increasing absolute differences in sensitivity and specificity (Figure 3.3). The absolute difference in specificity strongly influenced the corrected effect estimate. The majority of corrected estimates lay within the 95% CI for the observed estimate, which included the null value (1.0).

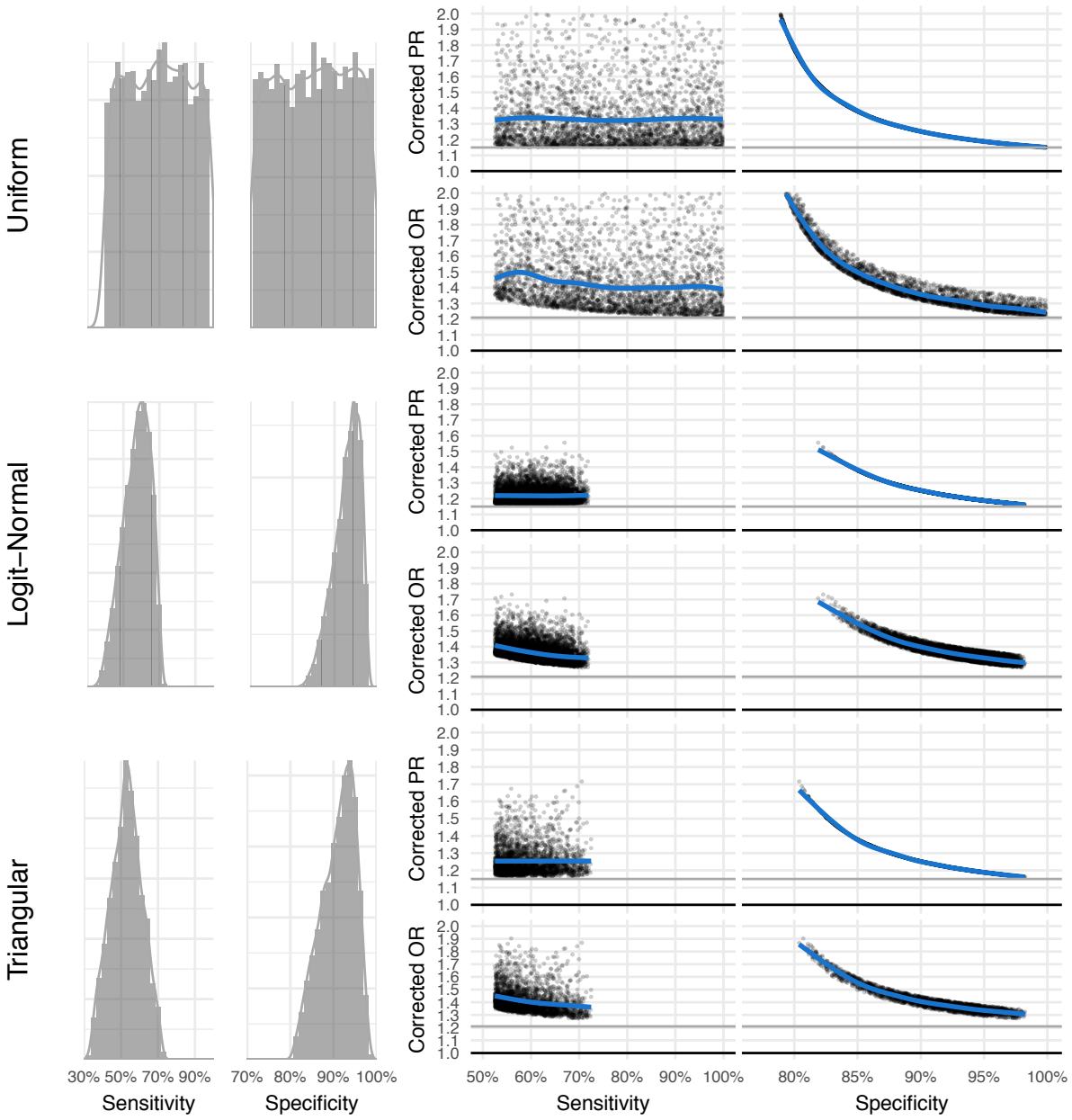


Figure 3.2: Non-Differential probabilistic bias analysis results

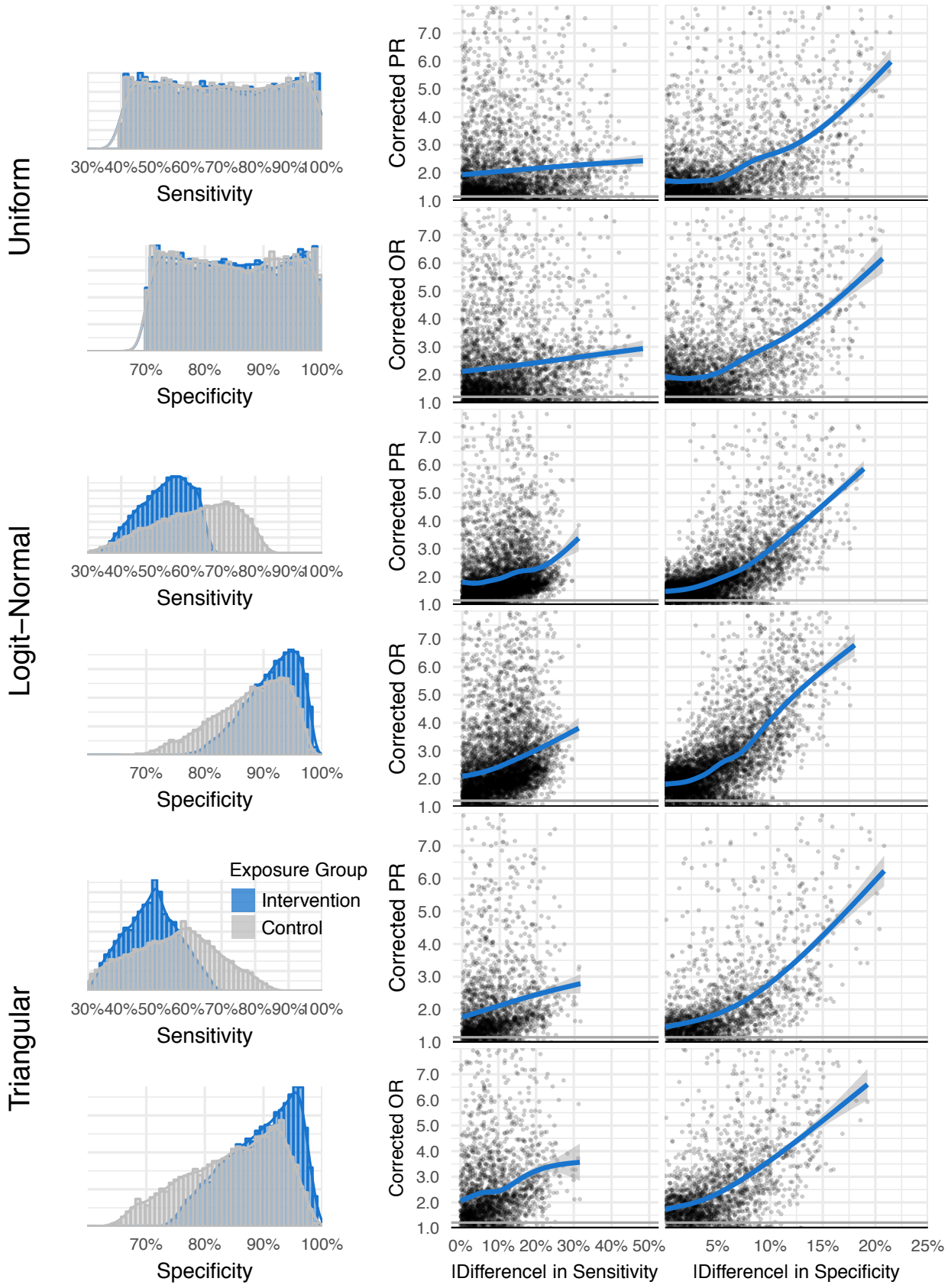


Figure 3.3: Differential probabilistic bias analysis results

3.4.5 Regression correction

For the GLM, the corrected PR was 1.27 (95% CI: 0.87, 1.53) and the corrected OR was 1.38 (95% CI 0.88, 2.13). For the GEE with robust standard errors, the corrected PR was 1.26 (95% CI: 0.71, 2.21) and the corrected OR was 1.36 (95% CI 0.77, 2.38).

3.5 Discussion

We found no evidence for a reduction in reported diarrhea symptoms after two-year water-improvement intervention in rural Ethiopia. When we corrected for error in reported diarrhea symptoms, the effect of the water intervention remained null. The uncorrected point estimate suggested children in the control arm had 1.17 times higher prevalence of diarrhea, however the 95% confidence intervals were wide and included both a protective effect (.66) with an upper limit of 2.26. The misclassification corrected point estimates were all farther from the null (1.3, 1.4, 1.5) and the confidence intervals were wider than the uncorrected estimate. The bias-corrected estimate and 95% confidence intervals from the GEE with the regression correction method is likely the most accurate because they account for clustering by community.

The lack of evidence supporting a reduction in diarrhea after the two-year water intervention is consistent with recent large cluster-randomized water, sanitation and hygiene trials. In a two-year randomized trial of a comprehensive water, sanitation and hygiene intervention in Kenya, Null and colleagues found that the post-intervention 7-day prevalence of diarrhea was 27.0% in the control arm and 27.7% in the intervention arm with a prevalence ratio of 1.02 (0.92-1.14)[77]. The sister trial in Bangladesh similarly found no evidence supporting a reduction in diarrhea with prevalence ratio of 0.89 (95%CI: 0.70-1.13)[76]. Both trials used caregiver-reported symptoms to measure diarrhea.

This study had several limitations that are worth considering. First, Of the 6 constructed wells, 3 were functional year-round, 2 were functional in the wet season only, and 1 ceased functioning after 3 months. While this productivity is sub-optimal, it is comparable to many real-word water interventions and thus the results can be interpreted as a reasonable implementation effect. Second, we have no baseline diarrhea measurements from the control communities. The relatively small number of clusters increase the possibility of finding these results simply by chance. However, the reported diarrhea prevalence in the intervention communities alone did not decrease over the two-year intervention period. Third, we used estimates of misclassification in caregiver-reported stool consistency taken from an internal validation study to estimate misclassification of caregiver-reported diarrhea. Diarrhea is a composite outcome incorporating both reported stool consistency and frequency. It is unlikely the misclassification of reported consistency would be the same as reported frequency.

Despite these limitations, our study had several strengths. We present experimental evidence for the effect of a water intervention on reported diarrhea in the Amhara region of Ethiopia. After an internal validation study demonstrated misclassification in the caregiver-reported stool consistency outcomes, we decided to report both the standard caregiver-reported diarrhea outcome and implemented quantitative bias analysis tools to estimate the direction and impact of misclassified diarrhea on point estimates and confidence intervals. We used three different bias correction approaches included a regression correction approach, which can adjust the standard errors for clustering by community. These methods can also be implemented in observational studies where adjustment for covariates is necessary. We adapted these techniques for use in the R programming language and have made our code freely available for others to implement.

The extremely low sensitivity of caregiver-reported stool consistency is a possible explanation why so many trials do not discern an impact of water, sanitation and hygiene improvements on childhood diar-

rhea[41]. However, even when accounting for misclassified symptoms we did not discern a protective point estimate. Symptom-based diarrhea outcomes are compromised in other ways. Most notably, diarrhea has both infectious and non-infectious etiologies which cannot be distinguished using symptom-based measures[48]. Moreover, many enteric pathogens produce both symptomatic and asymptomatic infections,[89] focusing on symptomatic diarrhea misses asymptomatic infections. Measuring enteric infections directly from stool rather than relying on symptoms alone may be an informative alternative outcome measure for clinical trials and epidemiologic studies of diarrheal disease.

references

[1] World Health Organization. Diarrheal disease factsheet. 2017:Fact sheet Number 330, Date Accessed: October 2017. URL http://www.who.int/mediacentre/factsheets/fs330/en/.

[2] Institute for Health Metrics and Evaluation. *Global Health Data Exchange*, Date Accessed: 07/2017. URL http://ghdx.healthdata.org/.

[3] World Health Organization. Ethiopia:who statistical profile. 2012, updated January 2015, Date Accessed: October 2017. URL http://www.who.int/gho/countries/eth.pdf?ua=1.

[4] Bradley C Johnston, Larissa Shamseer, Bruno R da Costa, Ross T Tsuyuki, and Sunita Vohra. Measurement issues in trials of pediatric acute diarrheal diseases: a systematic review. *Pediatrics*, 126(1):e222–e231, 2010. ISSN 0031-4005.

[5] World Health Organization. Diarrhea fact sheet, 2016. URL http://www.who.int/topics/diarrhoea/en/.

[6] Sandy Cairncross, Caroline Hunt, Sophie Boisson, Kristof Bostoen, Val Curtis, Isaac CH Fung, and Wolf-Peter Schmidt. Water, sanitation and hygiene for the prevention of diarrhea. *International Journal of Epidemiology*, 39(suppl 1):i193–i205, 2010. doi: 10.1093/ije/dyq035. URL http://ije.oxfordjournals.org/content/39/suppl_1/i193.abstract.

[7] T. F. Clasen, K. Bostoen, W. P. Schmidt, S. Boisson, I. C. Fung, M. W. Jenkins, B. Scott, S. Sugden, and S. Cairncross. Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane Database Syst Rev*, (6):Cd007180, 2010. ISSN 1361-6137. doi: 10.1002/14651858.CD007180.pub2.

[8] Thomas Clasen, Wolf-Peter Schmidt, Tamer Rabie, Ian Roberts, and Sandy Cairncross. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ*, 334(7597):782, 2007. doi: 10.1136/bmj.39118.489931.BE. URL http://www.bmj.com/bmj/334/7597/782.full.pdf.

[9] Annette Prüss-Ustün, Jamie Bartram, Thomas Clasen, John M Colford, Oliver Cumming, Valerie Curtis, Sophie Bonjour, Alan D Dangour, Jennifer De France, and Lorna Fewtrell. Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health*, 19(8):894–905, 2014. ISSN 1365-3156.

[10] Jennyfer Wolf, Annette Prüss-Ustün, Oliver Cumming, Jamie Bartram, Sophie Bonjour, Sandy Cairncross, Thomas Clasen, John M Colford, Valerie Curtis, and Jennifer France. 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 & International Health*, 19(8):928–942, 2014. ISSN 1365-3156.

[11] Thomas F Clasen, Kelly T Alexander, David Sinclair, Sophie Boisson, Rachel Peletz, Howard H Chang, Fiona Majorin, and Sandy Cairncross. Interventions to improve water quality for preventing diarrhoea. *The Cochrane Library*, 2015. ISSN 1465-1858.

[12] Ian Sinha, Leanne Jones, Rosalind L Smyth, and Paula R Williamson. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med*, 5(4):e96, 2008. ISSN 1549-1676.

[13] R. E. Black, K. H. Brown, S. Becker, and M. Yunus. Longitudinal studies of infectious diseases and physical growth of children in rural bangladesh. i. patterns of morbidity. *Am J Epidemiol*, 115 (3):305–14, 1982. ISSN 0002-9262 (Print) 0002-9262.

[14] D. Blum and R. G. Feachem. Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *Int J Epidemiol*, 12(3):357–65, 1983. ISSN 0300-5771 (Print) 0300-5771.

[15] Frideric T Koster, DL Palmer, Joyasul Chakraborty, Thad Jackson, and George C Curlin. Cellular immune competence and diarrheal morbidity in malnourished bangladeshi children: a prospective field study. *The American journal of clinical nutrition*, 46(1):115–120, 1987. ISSN 0002-9165.

[16] S. J. Lewis and K. W. Heaton. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*, 32(9):920–4, 1997. ISSN 0036-5521 (Print) 0036-5521. doi: 10.3109/00365529709011203.

[17] LJ ODONNELL, J Virjee, and KW Heaton. Pseudo-diarrhea in the irritable bowel syndrome-patients records of stool form reflect transit-time while stool frequency does not. In *Gut*, volume 29, pages A1455–A1455. BMJ PUBLISHING GROUP BRITISH MED ASSOC HOUSE, TAVISTOCK SQUARE, LONDON WC1H 9JR, ENGLAND, 1988.

[18] Bruno P. Chumpitazi, Mariella M. Lane, Danita I. Czyzewski, Erica M. Weidler, Paul R. Swank, and Robert J. Shulman. Creation and initial evaluation of a stool form scale for children. *The*

Journal of pediatrics, 157(4):594–597, 2010. ISSN 0022-3476 1097-6833. doi: 10.1016/j.jpeds.2010.04.040. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937014/.

[19] K. Aiemjoy, N. E. Stoller, S. Gebresillasie, A. Shiferaw, Z. Tadesse, T. Sewnet, B. Ayele, M. Chanyalew, K. Callahan, A. Stewart, P. M. Emerson, T. M. Lietman, J. D. Keenan, and C. E. Oldenburg. 'if an eye is washed properly, it means it would see clearly': A mixed methods study of face washing knowledge, attitudes, and behaviors in rural ethiopia. *PLoS Negl Trop Dis*, 10(10): e0005099, 2016. ISSN 1935-2727. doi: 10.1371/journal.pntd.0005099.

[20] Kristen Aiemjoy, Nicole E Stoller, Sintayehu Gebresillasie, Ayalew Shiferaw, Zerihun Tadesse, Tegene Sewent, Bezuayehu Ayele, Melsew Chanyalew, Solomon Aragie, and Kelly Callahan. Is using a latrine "a strange thing to do"? a mixed-methods study of sanitation preference and behaviors in rural ethiopia. *The American Journal of Tropical Medicine and Hygiene*, pages 16–0541, 2016. ISSN 0002-9637.

[21] Sintayehu ; Stoller Nicole; Shiferaw Ayalew ; Tadesse Zerihun; Aragie Solomon; Callahan Kelly; Chanyalew Melsew ; Keenan Jeremy Aiemjoy, Kristen; Gebresillasie. Epidemiology of soil-transmitted helminth and intestinal protozoa infections in pre-school aged children in the amhara region of ethiopia. *American Journal of Tropical Medicine & Hygiene*, In Press, 2017.

[22] Mariella M Lane, Danita I Czyzewski, Bruno P Chumpitazi, and Robert J Shulman. Reliability and validity of a modified bristol stool form scale for children. *The Journal of pediatrics*, 159(3): 437–441. e1, 2011. ISSN 0022-3476.

[23] Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968. ISSN 1939-1455.

[24] AP Morton and AJ Dobson. Assessing agreement. *The Medical journal of Australia*, 150(7): 384–387, 1989. ISSN 0025-729X.

[25] T. S. Genders, S. Spronk, T. Stijnen, E. W. Steyerberg, E. Lesaffre, and M. G. Hunink. Methods for calculating sensitivity and specificity of clustered data: a tutorial. *Radiology*, 265(3):910–6, 2012. ISSN 0033-8419. doi: 10.1148/radiol.12120509.

[26] William Rogers. Regression standard errors in clustered samples. *Stata technical bulletin*, 3(13), 1994.

[27] Steven S. Coughlin, Bruce Trock, Michael H. Criqui, Linda W. Pickle, Deirdre Browner, and Mariella C. Tefft. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *Journal of Clinical Epidemiology*, 45(1):1–7, 1992. ISSN 0895-4356. doi: https://doi.org/10.1016/0895-4356(92)90180-U. URL http://www.sciencedirect.com/science/article/pii/089543569290180U.

[28] A. C. Janssens, Y. Deng, G. J. Borsboom, M. J. Eijkemans, J. D. Habbema, and E. W. Steyerberg. A new logistic regression approach for the evaluation of diagnostic test results. *Med Decis Making*, 25(2):168–77, 2005. ISSN 0272-989X (Print) 0272-989x. doi: 10.1177/0272989x05275154.

[29] A. H. Baqui, R. E. Black, M. Yunus, A. R. Hoque, H. R. Chowdhury, and R. B. Sack. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. *Int J Epidemiol*, 20(4):1057–63, 1991. ISSN 0300-5771 (Print) 0300-5771.

[30] M. R. Blake, J. M. Raker, and K. Whelan. Validity and reliability of the bristol stool form scale in healthy adults and patients with diarrhoea-predominant irritable bowel syndrome. *Aliment Pharmacol Ther*, 44(7):693–703, 2016. ISSN 0269-2813. doi: 10.1111/apt.13746.

[31] Walter J. Rogan and Beth Gladen. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 107(1):71–76, 1978. ISSN 0002-9262. doi: 10.1093/oxfordjournals.aje.a112510. URL http://dx.doi.org/10.1093/oxfordjournals.aje.a112510.

[32] G. F. Longstreth, W. G. Thompson, W. D. Chey, L. A. Houghton, F. Mearin, and R. C. Spiller. Functional bowel disorders. *Gastroenterology*, 130(5):1480–91, 2006. ISSN 0016-5085 (Print) 0016-5085 (Linking). doi: 10.1053/j.gastro.2005.11.061. URL https://www.ncbi.nlm.nih.gov/pubmed/16678561.

[33] G. L. Austin, C. B. Dalton, Y. M. Hu, C. B. Morris, J. Hankins, S. R. Weinland, E. C. Westman, W. S. Yancy, and D. A. Drossman. A very low-carbohydrate diet improves symptoms and quality of life in diarrhea-predominant irritable bowel syndrome. *Clinical Gastroenterology and Hepatology*, 7(6):706–708, 2009. ISSN 1542-3565. doi: 10.1016/j.cgh.2009.02.023. URL <GotoISI>://WOS:000267189400018.

[34] K. W. Heaton and L. J. D. Odonnell. An office guide to whole-gut transit - time patients recollection of their stool form. *Journal of Clinical Gastroenterology*, 19(1):28–30, 1994. ISSN 0192-0790. doi: Doi10.1097/00004836-199407000-00008. URL <GotoISI>://WOS:A1994NW35600007.

[35] R. J. Shulman, M. N. Eakin, D. I. Czyzewski, M. Jarrett, and C. N. Ou. Increased gastrointestinal permeability and gut inflammation in children with functional abdominal pain and irritable bowel syndrome. *J Pediatr*, 153(5):646–50, 2008. ISSN 1097-6833 (Electronic) 0022-3476 (Linking). doi: 10.1016/j.jpeds.2008.04.062. URL https://www.ncbi.nlm.nih.gov/pubmed/18538790.

[36] Jill Tinmouth, George Tomlinson, Gabor Kandel, Sharon Walmsley, Hillary A Steinhart, and Richard Glazier. Evaluation of stool frequency and stool form as measures of hiv-related diarrhea. *HIV clinical trials*, 2015.

[37] Elizabeth K Costello, Christian L Lauber, Micah Hamady, Noah Fierer, Jeffrey I Gordon, and Rob Knight. Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697, 2009. ISSN 0036-8075.

[38] Dirk Gevers, Rob Knight, Joseph F Petrosino, Katherine Huang, Amy L McGuire, Bruce W Birren, Karen E Nelson, Owen White, Barbara A Methé, and Curtis Huttenhower. The human microbiome project: a community resource for the healthy human microbiome. *PLoS biology*, 10 (8):e1001377, 2012. ISSN 1545-7885.

[39] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, and Takuji Yamada. A human gut microbial gene catalogue established by metagenomic sequencing. *nature*, 464(7285): 59, 2010. ISSN 1476-4687.

[40] Doris Vandeputte, Gwen Falony, Sara Vieira-Silva, Raul Y Tito, Marie Joossens, and Jeroen Raes. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*, pages gutjnl–2015–309618, 2015. ISSN 0017-5749.

[41] Kristen Aiemjoy, Solomon Aragie, Sintayehu Gebresillasie, Dionna M Fry, Adane Dagnew, Dagnachew Hailu, Melsew Chanyalew, Zerihun Tadesse, Aisha Stewart, Kelly Callahan, et al. Defining diarrhea: A population-based validation study of caregiver-reported stool consistency in the amhara region of ethiopia. *American Journal of Tropical Medicine & Hygiene*, 98(4), April

2018. doi: DOI:https://doi.org/10.4269/ajtmh.17-0806. URL https://www.ajtmh.org/content/journals/10.4269/ajtmh.17-0806.

[42] T. Gebre, B. Ayele, M. Zerihun, A. Genet, N. E. Stoller, Z. Zhou, J. I. House, S. N. Yu, K. J. Ray, P. M. Emerson, J. D. Keenan, T. C. Porco, T. M. Lietman, and B. D. Gaynor. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in ethiopia: a cluster-randomised trial. *Lancet*, 379(9811):143–51, 2012. ISSN 1474-547X (Electronic) 0140-6736 (Linking). doi: 10.1016/S0140-6736(11)61515-8. URL https://www.ncbi.nlm.nih.gov/pubmed/22192488.

[43] T. G. Phan, A. C. da Costa, J. Del Valle Mendoza, F. Bucardo-Rivera, J. Nordgren, M. O'Ryan, X. Deng, and E. Delwart. The fecal virome of south and central american children with diarrhea includes small circular dna viral genomes of unknown origin. *Arch Virol*, 161(4):959–66, 2016. ISSN 1432-8798 (Electronic) 0304-8608 (Linking). doi: 10.1007/s00705-016-2756-4. URL http://www.ncbi.nlm.nih.gov/pubmed/26780893.

[44] L. Li, X. Deng, E. T. Mee, S. Collot-Teixeira, R. Anderson, S. Schepelmann, P. D. Minor, and E. Delwart. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods*, 213:139–46, 2015. ISSN 1879-0984 (Electronic) 0166-0934 (Linking). doi: 10.1016/j.jviromet.2014.12.002. URL http://www.ncbi.nlm.nih.gov/pubmed/25497414.

[45] T. G. Phan, D. Mori, X. Deng, S. Rajindrajith, U. Ranawaka, T. F. Fan Ng, F. Bucardo-Rivera, P. Orlandi, K. Ahmed, and E. Delwart. Small circular single stranded dna viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology*, 482:98–104,

2015. ISSN 1096-0341 (Electronic) 0042-6822 (Linking). doi: 10.1016/j.virol.2015.03.011. URL http://www.ncbi.nlm.nih.gov/pubmed/25839169.

[46] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):550, 2014. ISSN 1474-760X.

[47] J. Joukje Siebenga, Harry Vennema, Du-Ping Zheng, Jan Vinjé, Bonita E. Lee, Xiao-Li Pang, Eric C. M. Ho, Wilina Lim, Avinash Choudekar, Shobha Broor, Tamar Halperin, Nassar B. G. Rasool, Joanne Hewitt, Gail E. Greening, Miao Jin, Zhao-Jun Duan, Yalda Lucero, Miguel O'Ryan, Marina Hoehne, Eckart Schreier, Rodney M. Ratcliff, Peter A. White, Nobuhiro Iritani, Gábor Reuter, and Marion Koopmans. Norovirus illness is a global problem: Emergence and spread of norovirus gii.4 variants, 2001–2007. *The Journal of Infectious Diseases*, 200(5):802–812, 2009. ISSN 0022-1899. doi: 10.1086/605127. URL http://dx.doi.org/10.1086/605127.

[48] J. Liu, J. A. Platts-Mills, J. Juma, F. Kabir, J. Nkeze, C. Okoi, D. J. Operario, J. Uddin, S. Ahmed, P. L. Alonso, M. Antonio, S. M. Becker, W. C. Blackwelder, R. F. Breiman, A. G. Faruque, B. Fields, J. Gratz, R. Haque, A. Hossain, M. J. Hossain, S. Jarju, F. Qamar, N. T. Iqbal, B. Kwambana, I. Mandomando, T. L. McMurry, C. Ochieng, J. B. Ochieng, M. Ochieng, C. Onyango, S. Panchalingam, A. Kalam, F. Aziz, S. Qureshi, T. Ramamurthy, J. H. Roberts, D. Saha, S. O. Sow, S. E. Stroup, D. Sur, B. Tamboura, M. Taniuchi, S. M. Tennant, D. Toema, Y. K. Wu, A. Zaidi, J. P. Nataro, K. L. Kotloff, M. M. Levine, and E. R. Houpt. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the gems case-control study. *Lancet*, 388(10051):1291–1301, 2016. ISSN 0140-6736. doi: Doi10.1016/S0140-6736(16)31529-X. URL <GotoISI>://WOS:000383704000032.

[49] N. Ouedraogo, J. Kaplon, I. J. Bonkoungou, A. S. Traore, P. Pothier, N. Barro, and K. Ambert-Balay. Prevalence and genetic diversity of enteric viruses in children with diarrhea in ouagadougou, burkina faso. *PLoS One*, 11(4):e0153652, 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0153652.

[50] K. Ambert-Balay, M. Lorrot, F. Bon, H. Giraudon, J. Kaplon, M. Wolfer, P. Lebon, D. Gendrel, and P. Pothier. Prevalence and genetic diversity of aichi virus strains in stool samples from community and hospitalized patients. *Journal of Clinical Microbiology*, 46(4):1252–1258, 2008. doi: 10.1128/jcm.02140-07. URL http://jcm.asm.org/content/46/4/1252.abstract.

[51] D-Y Oh, PA Silva, B Hauroeder, S Diedrich, DDP Cardoso, and E Schreier. Molecular characterization of the first aichi viruses isolated in europe and in south america. *Archives of virology*, 151(6): 1199–1206, 2006. ISSN 0304-8608.

[52] Khira Sdiri-Loulizi, Mouna Hassine, Hakima Gharbi-Khelifi, Nabil Sakly, Slaheddine Chouchane, Mohamed Neji Guediche, Pierre Pothier, Mahjoub Aouni, and Katia Ambert-Balay. Detection and genomic characterization of aichi viruses in stool samples from children in monastir, tunisia. *Journal of Clinical Microbiology*, 47(7):2275–2278, 2009. doi: 10.1128/jcm.00913-09. URL http://jcm.asm.org/content/47/7/2275.abstract.

[53] Ngan Thi Kim Pham, Pattara Khamrin, Tuan Anh Nguyen, Dey Shuvra Kanti, Tung Gia Phan, Shoko Okitsu, and Hiroshi Ushijima. Isolation and molecular characterization of aichi viruses from fecal specimens collected in japan, bangladesh, thailand, and vietnam. *Journal of Clinical Microbiology*, 45(7):2287–2288, 2007. doi: 10.1128/jcm.00525-07. URL http://jcm.asm.org/content/45/7/2287.abstract.

[54] Teruo Yamashita, Kenji Sakae, Shinichi Kobayashi, Yuichi Ishihara, Takashi Miyake, Agboatwalla Mubina, and Shin Isomura. Isolation of cytopathic small round virus (aichi virus) from pakistani children and japanese travelers from southeast asia. *Microbiology and immunology*, 39(6):433–435, 1995. ISSN 1348-0421.

[55] CSN Schmaljohn and ST Fields. Virology. *Fields virology. 5th ed. Philadelphia: Lippincott Williams & Wilkins*, pages 1741–89, 2007.

[56] Stacy R Finkbeiner, Adam F Allred, Phillip I Tarr, Eileen J Klein, Carl D Kirkwood, and David Wang. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS pathogens*, 4(2):e1000011, 2008. ISSN 1553-7374.

[57] P. Chhabra, D. C. Payne, P. G. Szilagyi, K. M. Edwards, M. A. Staat, S. H. Shirley, M. Wikswo, W. A. Nix, X. Lu, U. D. Parashar, and J. Vinje. Etiology of viral gastroenteritis in children <5 years of age in the united states, 2008-2009. *J Infect Dis*, 208(5):790–800, 2013. ISSN 0022-1899. doi: 10.1093/infdis/jit254.

[58] G. La Rosa, S. Della Libera, S. Petricca, M. Iaconelli, D. Donia, P. Saccucci, F. Cenko, G. Xhelilaj, and M. Divizia. Genetic diversity of human adenovirus in children with acute gastroenteritis, albania, 2013–2015. *BioMed Research International*, 2015:142912, 2015. ISSN 2314-6133 2314-6141. doi: 10.1155/2015/142912. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4538309/.

[59] J. I. Lee, G. C. Lee, J. Y. Chung, T. H. Han, Y. K. Lee, M. S. Kim, and C. H. Lee. Detection and molecular characterization of adenoviruses in korean children hospitalized with acute gastroenteritis. *Microbiol Immunol*, 56(8):523–8, 2012. ISSN 0385-5600. doi: 10.1111/j.1348-0421.2012.00469.x.

[60] L. Li, T. G. Phan, T. A. Nguyen, K. S. Kim, J. K. Seo, H. Shimizu, E. Suzuki, S. Okitsu, and H. Ushijima. Molecular epidemiology of adenovirus infection among pediatric population with diarrhea in asia. *Microbiol Immunol*, 49(2):121–8, 2005. ISSN 0385-5600 (Print) 0385-5600.

[61] Sabrina John Moyo, Kurt Hanevik, Bjørn Blomberg, Oyvind Kommedal, Svein Arne Nordbø, Samuel Maselle, and Nina Langeland. Prevalence and molecular characterisation of human adenovirus in diarrhoeic children in tanzania; a case control study. *BMC Infectious Diseases*, 14:666, 2014. ISSN 1471-2334. doi: 10.1186/s12879-014-0666-1. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4266963/.

[62] Siwar Ayouni, Marie Estienney, Sabeur Hammami, Mohamed Neji Guediche, Pierre Pothier, Mahjoub Aouni, Gael Belliot, and Alexis De Rougemont. Cosavirus, salivirus and bufavirus in diarrheal tunisian infants. *PloS one*, 11(9):e0162255, 2016. ISSN 1932-6203.

[63] Lori R Holtz, Stacy R Finkbeiner, Carl D Kirkwood, and David Wang. Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. *Virology journal*, 5(1):159, 2008. ISSN 1743-422X.

[64] Amit Kapoor, Joseph Victoria, Peter Simmonds, Elizabeth Slikas, Thaweesak Chieochansin, Asif Naeem, Shahzad Shaukat, Salmaan Sharif, Muhammad Masroor Alam, and Mehar Angez. A highly prevalent and genetically diversified picornaviridae genus in south asian children. *Proceedings of the National Academy of Sciences*, 105(51):20482–20487, 2008. ISSN 0027-8424.

[65] Beatrix Kapusinszky, Tung G Phan, Amit Kapoor, and Eric Delwart. Genetic diversity of the genus cosavirus in the family picornaviridae: a new species, recombination, and 26 new genotypes. *PLoS One*, 7(5):e36685, 2012. ISSN 1932-6203.

[66] Beatrix Kapusinszky, Philip Minor, and Eric Delwart. Nearly constant shedding of diverse enteric viruses by two healthy infants. *Journal of clinical microbiology*, 50(11):3427–3434, 2012. ISSN 0095-1137.

[67] G. Sun, Q. Zang, Y. Gu, G. Niu, C. Ding, and P. Zhang. Viral metagenomics analysis of picobirnavirus-positive feces from children with sporadic diarrhea in china. *Arch Virol*, 161(4): 971–5, 2016. ISSN 0304-8608. doi: 10.1007/s00705-015-2726-2.

[68] G. Weldegebriel, J. M. Mwenda, J. Chakauya, F. Daniel, B. Masresha, U. D. Parashar, and J. E. Tate. Impact of rotavirus vaccine on rotavirus diarrhoea in countries of east and southern africa. *Vaccine*, 2017. ISSN 1873-2518 (Electronic) 0264-410X (Linking). doi: 10.1016/j.vaccine.2017.10.050. URL http://www.ncbi.nlm.nih.gov/pubmed/29102168.

[69] J. Diamant, R. Benis, J. Schachter, J. Moncada, F. Pang, H. C. Jha, R. C. Bhatta, T. Porco, and T. Lietman. Pooling of chlamydia laboratory tests to determine the prevalence of ocular chlamydia trachomatis infection. *Ophthalmic Epidemiol*, 8(2-3):109–17, 2001. ISSN 0928-6586 (Print) 0928-6586 (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/11471080.

[70] J. M. Sacks, S. R. Bolin, and S. V. Crowder. Prevalence estimation from pooled samples. *Am J Vet Res*, 50(2):205–6, 1989. ISSN 0002-9645 (Print) 0002-9645 (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/2785780.

[71] Cynthia Boschi-Pinto, ClaudioF Lanata, and RobertE Black. *The Global Burden of Childhood Diarrhea*, book section 13, pages 225–243. Springer US, 2009. ISBN 978-0-387-89244-3. doi: 10.1007/b106524_13. URL http://dx.doi.org/10.1007/b106524_13.

[72] J. Wolf, A. Pruss-Ustun, O. Cumming, J. Bartram, S. Bonjour, S. Cairncross, T. Clasen, Jr. Colford, J. M., V. Curtis, J. De France, L. Fewtrell, M. C. Freeman, B. Gordon, P. R. Hunter, A. Jeandron, R. B. Johnston, D. Mausezahl, C. Mathers, M. Neira, and J. P. Higgins. 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 Health*, 19(8):928–42, 2014. ISSN 1365-3156 (Electronic) 1360-2276 (Linking). doi: 10.1111/tmi.12331. URL https://www.ncbi.nlm.nih.gov/pubmed/24811732.

[73] W. P. Schmidt, B. F. Arnold, S. Boisson, B. Genser, S. P. Luby, M. L. Barreto, T. Clasen, and S. Cairncross. Epidemiological methods in diarrhoea studies–an update. *Int J Epidemiol*, 40 (6):1678–92, 2011. ISSN 1464-3685 (Electronic) 0300-5771 (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/22268237.

[74] T. Clasen and S. Boisson. Assessing the health impact of water quality interventions in low-income settings: Concerns associated with blinded trials and the need for objective outcomes. *Environ Health Perspect*, 124(7):886–9, 2016. ISSN 1552-9924 (Electronic) 0091-6765 (Linking). doi: 10.1289/ehp.1510532. URL https://www.ncbi.nlm.nih.gov/pubmed/26685286.

[75] Thomas Clasen, Sophie Boisson, Parimita Routray, Belen Torondel, Melissa Bell, Oliver Cumming, Jeroen Ensink, Matthew Freeman, Marion Jenkins, and Mitsunori Odagiri. 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(11):e645–e653, 2014. ISSN 2214-109X.

[76] S. P. Luby, M. Rahman, B. F. Arnold, L. Unicomb, S. Ashraf, P. J. Winch, C. P. Stewart, F. Be-

gum, F. Hussain, J. Benjamin-Chung, E. Leontsini, A. M. Naser, S. M. Parvez, A. E. Hubbard, A. Lin, F. A. Nizame, K. Jannat, A. Ercumen, P. K. Ram, K. K. Das, J. Abedin, T. F. Clasen, K. G. Dewey, L. C. Fernald, C. Null, T. Ahmed, and Jr. Colford, J. M. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural bangladesh: a cluster randomised controlled trial. *Lancet Glob Health*, 6(3):e302–e315, 2018. ISSN 2214-109X (Electronic) 2214-109X (Linking). doi: 10.1016/S2214-109X(17)30490-4. URL https://www.ncbi.nlm.nih.gov/pubmed/29396217.

[77] C. Null, C. P. Stewart, A. J. Pickering, H. N. Dentz, B. F. Arnold, C. D. Arnold, J. Benjamin-Chung, T. Clasen, K. G. Dewey, L. C. H. Fernald, A. E. Hubbard, P. Kariger, A. Lin, S. P. Luby, A. Mertens, S. M. Njenga, G. Nyambane, P. K. Ram, and Jr. Colford, J. M. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural kenya: a cluster-randomised controlled trial. *Lancet Glob Health*, 6(3):e316–e329, 2018. ISSN 2214-109X (Electronic) 2214-109X (Linking). doi: 10.1016/S2214-109X(18)30005-6. URL https://www.ncbi.nlm.nih.gov/pubmed/29396219.

[78] Sumeet R Patil, Benjamin F Arnold, Alicia L Salvatore, Bertha Briceno, Sandipan Ganguly, John M Colford Jr, and Paul J Gertler. The effect of india's total sanitation campaign on defecation behaviors and child health in rural madhya pradesh: a cluster randomized controlled trial. *PLoS medicine*, 11(8):e1001709, 2014. ISSN 1549-1676.

[79] Amy J Pickering, Habiba Djebbari, Carolina Lopez, Massa Coulibaly, and Maria Laura Alzua. 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(11):e701–e711, 2015. ISSN 2214-109X.

[80] Bruce A Barron. The effects of misclassification on the estimation of relative risk. *Biometrics*, pages 414–418, 1977. ISSN 0006-341X.

[81] Timothy L Lash, Matthew P Fox, Richard F MacLehose, George Maldonado, Lawrence C McCandless, and Sander Greenland. Good practices for quantitative bias analysis. *International journal of epidemiology*, 43(6):1969–1985, 2014. ISSN 1464-3685.

[82] Kenneth J Rothman, Sander Greenland, and Timothy L Lash. Modern epidemiology. 2008.

[83] Timothy L Lash, Matthew P Fox, and Aliza K Fink. *Applying quantitative bias analysis to epidemiologic data*. Springer Science & Business Media, 2011. ISBN 0387879595.

[84] J. M. Neuhaus. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855, 1999. ISSN 0006-3444. doi: DOI10.1093/biomet/86.4.843. URL <GotoISI>://WOS:000084833000008.

[85] J. M. Neuhaus. Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58(3):675–83, 2002. ISSN 0006-341X (Print) 0006-341X (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/12230004.

[86] P. Cummings. Estimating adjusted risk ratios for matched and unmatched data: An update. *Stata Journal*, 11(2):290–298, 2011. ISSN 1536-867x. URL <GotoISI>://WOS:000294275400008.

[87] Sander Greenland. Basic methods for sensitivity analysis of biases. *International journal of epidemiology*, 25(6):1107–1116, 1996. ISSN 1464-3685.

[88] 2017. URL https://cran.r-project.org/web/packages/episensr/episensr.pdf.

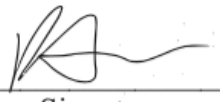
[89] K. L. Kotloff, J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu, S. O. Sow, D. Sur, R. F. Breiman, A. S. Faruque, A. K. Zaidi, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B. Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F. Quadri, R. A. Adegbola, M. Antonio, M. J. Hossain, A. Akinsola, I. Mandomando, T. Nhampossa, S. Acacio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K. Muhsen, H. Sommerfelt, R. M. Robins-Browne, and M. M. Levine. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, gems): a prospective, case-control study. *Lancet*, 382(9888):209–22, 2013. ISSN 1474-547X (Electronic) 0140-6736 (Linking). doi: 10.1016/S0140-6736(13)60844-2. URL https://www.ncbi.nlm.nih.gov/pubmed/23680352.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

10-4-2018

Date