RESEARCH ARTICLE

The seasonality of diarrheal pathogens: A retrospective study of seven sites over three years

Dennis L. Chao[1]*, **Anna Roose**[2], **Min Roh**[1], **Karen L. Kotloff**[2], **Joshua L. Proctor**[1]

1 Institute for Disease Modeling, Bellevue, Washington, United States of America, **2** Center for Vaccine Development and Global Health, University of Maryland, Baltimore, Maryland, United States of America

* dennisc@idmod.org

Abstract

Background

Pediatric diarrhea can be caused by a wide variety of pathogens, from bacteria to viruses to protozoa. Pathogen prevalence is often described as seasonal, peaking annually and associated with specific weather conditions. Although many studies have described the seasonality of diarrheal disease, these studies have occurred predominantly in temperate regions. In tropical and resource-constrained settings, where nearly all diarrhea-associated mortality occurs, the seasonality of many diarrheal pathogens has not been well characterized. As a retrospective study, we analyze the seasonal prevalence of diarrheal pathogens among children with moderate-to-severe diarrhea (MSD) over three years from the seven sites of the Global Enteric Multicenter Study (GEMS), a case–control study. Using data from this expansive study on diarrheal disease, we characterize the seasonality of different pathogens, their association with site-specific weather patterns, and consistency across study sites.

Methodology/Principal findings

Using traditional methodologies from signal processing, we found that certain pathogens peaked at the same time every year, but not at all sites. We also found associations between pathogen prevalence and weather or "seasons," which are defined by applying modern machine-learning methodologies to site-specific weather data. In general, rotavirus was most prevalent during the drier "winter" months and out of phase with bacterial pathogens, which peaked during hotter and rainier times of year corresponding to "monsoon," "rainy," or "summer" seasons.

Conclusions/Significance

Identifying the seasonally-dependent prevalence for diarrheal pathogens helps characterize the local epidemiology and inform the clinical diagnosis of symptomatic children. Our multi-site, multi-continent study indicates a complex epidemiology of pathogens that does not reveal an easy generalization that is consistent across all sites. Instead, our study indicates the necessity of local data to characterizing the epidemiology of diarrheal disease.

Recognition of the local associations between weather conditions and pathogen prevalence suggests transmission pathways and could inform control strategies in these settings.

Author summary

Diarrheal disease is one of the leading causes of death among young children in the developing world. It is difficult to determine which of a wide variety of pathogens are most responsible for disease, since this differs by location and time of year. Here, we study the seasonal prevalence of several pathogens among children with moderate-to-severe diarrhea across study sites in Africa and South Asia. We found that several pathogens, including rotavirus, had regular annual peaks. Some pathogens were associated with weather conditions, such as heat or rain, or with general seasons of the year, such as summer or winter. We believe that describing the seasonal epidemiology of these pathogens could enable better diagnoses of symptomatic children based on the time of the year. Additionally, weather can be a major driver of diarrheal pathogen transmission, and identifying the conditions associated with each pathogen could help us infer pathogen transmission pathways, predict large outbreaks, and develop intervention strategies.

Introduction

Pediatric diarrheal disease is caused by a wide variety of pathogens [1–3]. Various studies have found that some pathogens are seasonal, peaking at different times of the year [4–6]. Frequently, the seasonal periodicity of diarrheal disease is attributed to weather, which could drive incidence by diverse mechanisms. For example, weather conditions can favor the survival and replication of pathogens on fomites [7], the transmission between human hosts through flooding and contamination of drinking water [8], and the prevalence of vectors that transmit disease between hosts [9, 10]. Weather has been shown to be statistically correlated with infectious diseases incidence including diarrhea from pathogens such as *Vibrio cholerae* and rotavirus [11, 12].

However, most studies of disease seasonality have been conducted in temperate climates, and substantially less is known about the seasonality of diseases in tropical countries [13, 14], where diarrheal disease is one of the leading causes of morbidity and mortality among children [15]. The wide variety of climates and variations in populations in the tropics make it challenging to uncover general patterns in the epidemiology of diarrheal disease. Compounding these challenges, most studies are limited to sites within a single country focused on a specific disease. Characterizing the seasonal epidemiology of these pathogens could enable clinicians to better diagnose children based on the time of the year. Additionally, identifying the weather conditions associated with each pathogen could help us infer pathogen transmission pathways, predict large outbreaks, and develop intervention strategies.

In this article, we perform a secondary analysis of the Global Enteric Multicenter Study (GEMS), a large, multi-country study of moderate-to-severe diarrhea (MSD) among children younger than five years of age [2], to investigate the underlying patterns of pathogen-specific seasonality for diarrheal disease in resource-limited settings. Utilizing a variety of mathematical techniques, we analyze the GEMS data to measure the strength of pathogen seasonality at

each site, to link local weather conditions to pathogen prevalence, and to reveal site-specific seasons associated with pathogens. The data from GEMS affords a rare opportunity to compare the seasonality of many pathogens across sites in different countries using data from a single coordinated study.

We use traditional signal processing analysis methods to detect periodicity in time series data, such as the Lomb–Scargle periodogram [16, 17] and Fast Fourier transforms, to estimate the annual, biannual, and other seasonal signals for each pathogen at each site; these have been used previously for analysis of time series data of incidence [18]. These methods enable the quantification of uncertainty of each of these estimates to mitigate the impact of intrinsic stochasticity due to study design. We also investigate the association between local weather conditions with pathogen prevalence data for each site. We broaden these association analyses by identifying site-dependent seasons through the use of machine-learning methodologies performed on local weather data. These data-driven seasons offer several advantages over using weather data: sites can be qualitatively compared by season instead of requiring a specific weather value threshold to hold across sites; seasons are more robust to linking disease prevalence versus day-to-day weather fluctuations; and seasons more naturally capture the combination of weather values that may drive certain pathogens. With each of these techniques and data from GEMS, we are able to broadly characterize the seasonality of multiple enteric pathogens across multiple study sites and identify associations to environmental factors.

Materials and methods

GEMS data

The Global Enteric Multicenter Study (GEMS) was conducted for 36 consecutive months from December 2007 to March 2011 at sites in seven communities in sub-Saharan Africa (Basse, The Gambia; Bamako, Mali; Nyaza Province, Kenya; Manhiça, Mozambique) and south Asia (Karachi [Bin Qasim Town], Pakistan; Kolkata, India; Mirzapur, Bangladesh) [2]. Children 0–59 months of age with moderate-to-severe diarrhea (MSD) who lived inside of the site's enumerated catchment area were eligible to enroll. For a child to be included, the diarrhea episode needed to be new (onset after at least seven days without diarrhea), acute (onset within seven days), and had to meet at least one clinical criteria for MSD. Clinical criteria for MSD included the following: the child presents clinical signs of dehydration, i.e., sunken eyes or loss of skin turgor, assessed by clinician and confirmed by mother, prescription or use of intravenous hydration; dysentery identified by blood in stool; or admission to the hospital for diarrhea or dysentery [2]. To limit the number of enrollments and ensure balanced enrollment by age, only the first 8–9 children in each age strata (0–11 months, 12–23 months, 24–59 months) were recruited each fortnight (14 days) at each site. Stool samples from enrolled children were tested for pathogens. In total, GEMS enrolled 9,439 out of 14,753 eligible children with MSD during the study period (Table 1). Matched controls, who did not have diarrhea, were not included in our analyses.

We limited our analyses to pathogens associated with the plurality of attributable diarrheal illnesses in GEMS [2]: rotavirus, Cryptosporidium, *Shigella* spp., typical EPEC (tEPEC), and enterotoxigenic *Escherichia coli* encoding heat-stable enterotoxin with or without genes encoding heat-labile enterotoxin (ST-ETEC). Two additional major pathogens identified in a subsequent analysis were included [19]: adenovirus 40/41 and *Campylobacter* spp. We also included three pathogens that were identified as prevalent in multiple sites during at least one of the seasons: *V. cholerae*, norovirus GI, and norovirus GII.

Table 1. Estimated prevalence of pathogens among children with moderate-to-severe diarrhea. We estimated the number and percentage of eligible children positive for each pathogen based on the proportion of tested cases who were positive.

	The Gambia	Mali	Kenya	Mozambique	Pakistan	India	Bangladesh
all MSD	2059	3611	2074	1405	1630	2290	1684
Campylobacter	73 (3.5%)	62 (1.7%)	293 (14.1%)	39 (2.8%)	392 (24.0%)	319 (13.9%)	287 (17.0%)
Shigella	214 (10.4%)	56 (1.6%)	144 (6.9%)	68 (4.8%)	161 (9.9%)	136 (5.9%)	714 (42.4%)
V cholerae	0 (0%)	0 (0%)	9 (0.4%)	28 (2.0%)	138 (8.5%)	105 (4.6%)	63 (3.7%)
ST-ETEC	255 (12.4%)	171 (4.7%)	198 (9.5%)	124 (8.8%)	143 (8.8%)	142 (6.2%)	41 (2.4%)
tEPEC	185 (9.0%)	402 (11.1%)	170 (8.2%)	105 (7.5%)	125 (7.7%)	65 (2.8%)	135 (8.0%)
Cryptosporidium	212 (10.3%)	452 (12.5%)	232 (11.2%)	225 (16.0%)	193 (11.8%)	381 (16.6%)	117 (6.9%)
Rotavirus	447 (21.7%)	615 (17.0%)	319 (15.4%)	373 (26.5%)	286 (17.5%)	596 (26.0%)	269 (16.0%)
Adenovirus	45 (2.2%)	64 (1.8%)	45 (2.2%)	33 (2.3%)	33 (2.0%)	98 (4.3%)	43 (2.6%)
Norovirus GI	77 (3.7%)	48 (1.3%)	59 (2.8%)	19 (1.4%)	95 (5.8%)	23 (1.0%)	76 (4.5%)
Norovirus GII	225 (10.9%)	52 (1.4%)	93 (4.5%)	47 (3.3%)	131 (8.0%)	166 (7.2%)	46 (2.7%)

https://doi.org/10.1371/journal.pntd.0007211.t001

Estimating pathogen prevalence among all eligible visits for two-week periods

To estimate the number of children who would have tested positive for a pathogen had all eligible children been enrolled, we assume that the proportion of children testing positive for a pathogen is equal among enrollees and those eligible within the same age strata (0–11, 12–23, and 24–59 months old) and fortnight of clinic visit. Fortnights are defined as consecutive 14-day periods starting with the first enrollment at each site, so there were about 80 fortnights of data per study site. For example, Fig 1(A) illustrates the difference between rotavirus confirmed enrolled children (purple) and the total estimated rotavirus prevalence among eligible children (black lines) in Bangladesh. Thus, the overall estimated pathogen prevalence includes the laboratory-positive children among the enrolled plus the estimated proportion among the eligible, reflecting the true age distribution of MSD cases. Fig 1(A) does not show a large difference between confirmed rotavirus cases and estimated rotavirus cases due to the small differences between the number of eligible and enrolled children. S1 Fig for the Mali study site illustrates a wider difference between the number of eligible and enrolled children, resulting in a wider discrepancy between confirmed and estimated rotavirus cases.

Weather data for GEMS study sites

Daily historical weather data from weather stations were acquired from the Global Surface Summary of the Day (GSOD) [20]. These data include daily measurements of temperature, precipitation, and dew point. For some GEMS study sites, the closest weather station does not have complete data for the entire study interval (2007–2011). In these cases, we choose the closest weather station with nearly complete data; for example, the weather station for The Gambia site is in Tambacounda, Senegal which is approximately 77 kilometers away and is the largest distance between site and station used in this study (S1 Table). Both the raw data downloaded from [20] and our filtered data files for the analysis can be found at [21].

Relative humidity was computed using the following empirical relationship, originally defined by Bosen in 1958 [23]:

$$RH = 100 \times ((112 - (0.1 \times T) + DP)/(112 + (0.9 \times T)))^8, \qquad (1)$$

where T is temperature in Celsius, DP is dewpoint in Celsius, and RH is relative humidity in

A. Child enrollment by fortnight in the Bangladesh study site

B. Rotavirus incidence exhibits strong yearly seasonality

C. Environmental covariates cluster into site-specific seasons

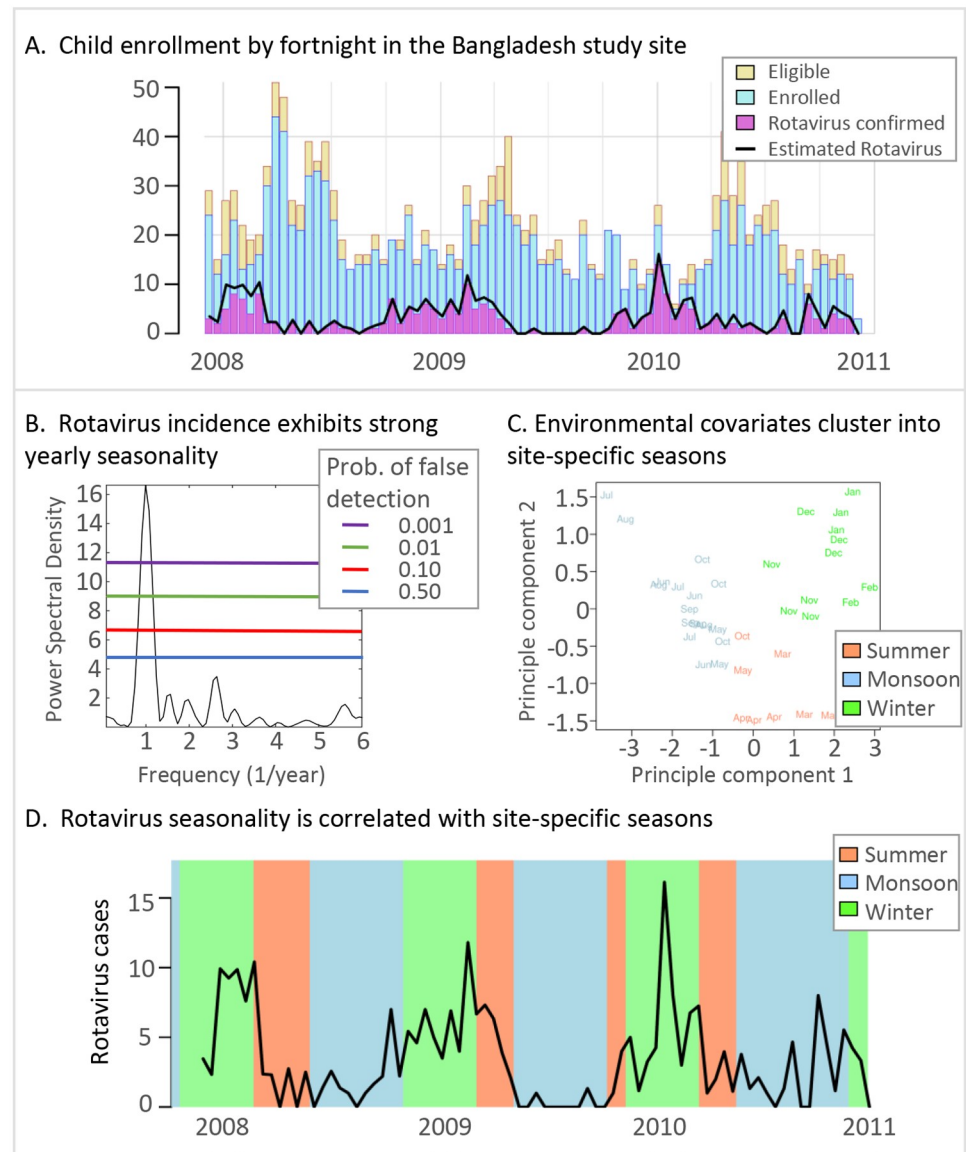### D. Rotavirus seasonality is correlated with site-specific seasons

Fig 1. Analysis overview. A. The number of eligible and enrolled children for the Bangladesh study site each fortnight (gold and blue bars, respectively). The number of rotavirus-confirmed samples from enrolled children are in purple. The estimated rotavirus prevalence among all eligible children which includes confirmed positive and estimated positive from the eligible children is illustrated by the black line. B. The power spectral density of the estimated rotavirus positive prevalence time series. Each colored line indicates the probability of a random time series signal with the same length and sampling being misinterpreted as a true signal. C. Seasons were identified using PCA to reduce the dimensionality of the weather variables and k-means clustering. The color of each month corresponds to the cluster and season. For Bangladesh, we call each season by the colloquial names "Summer," "Monsoon," and "Winter". D. The estimated number of cases positive for a pathogen each fortnight (black line). Each background color corresponds to the data-driven season identified in C.

https://doi.org/10.1371/journal.pntd.0007211.g001

percentage. We estimated specific humidity:

$$P \approx 101.3 \times e^{-\text{EL}/8200}, \tag{2}$$

$$E_s = 0.622 \times e^{(17.502 \times T)/(T+240.97)}, \tag{3}$$

$$E \quad = \quad (0.01 \times RH) \times E_s \,, \tag{4}$$

$$W \quad = \quad (0.622 \times E)/(P - E) \,, \tag{5}$$

$$\mathrm{SH} \quad = \quad 1000 \times W(1 + W) \,, \tag{6}$$

where EL is elevation in meters, P is air pressure in kPa, E_s is saturation vapor pressure in kPA, E is the vapor pressure in kPa, W is the mixing ratio, and SH is the specific humidity in g/kg [personal communication from James Tamerius]. S2 Fig shows the temperature, rain, and specific humidity for each GEMS site during the study period. Days with missing rainfall data were assumed to have no rainfall, and days with missing temperature or humidity data were not included when computing average temperature or humidity for the fortnight or month containing them.

Detecting periodicity and relative phase in disease incidence time series

We estimate the seasonality of each pathogen by using standard techniques for analyzing time series data; we leverage power spectral density methods, such as the Lomb-Scargle periodogram and Fast Fourier transforms, to identify the periodicity and within-season peak for each incidence time series. These methods have been used for analysis of disease incidence previously [18] and are generalizations of more focused analyses such as harmonic regression. Here, we utilize the Lomb-Scargle periodogram [16–18] to identify the periodicity, or spectral power, for each resolvable frequency in the time series which includes annual and biannual periods and accounts for the constraint by the temporal length of GEMS and our two-week sampling interval. The Lomb-Scargle periodogram is especially useful for analyzing constrained time series; the technique quantifies the uncertainty of detecting a specific periodicity [24] even with unevenly sampled data [17]. Fig 1(B) shows the standard Lomb-Scargle periodogram which is scaled by two times the variance of the pathogen time series. Each horizontal line indicates a threshold corresponding to a probability of a random time series measurement of being mischaracterized as a true periodic signal. We quantify the uncertainty of the periodicity of a time series as a probability of false detection denoted as $p_{fd} <$ threshold. Fig 2 broadens the scope of Fig 1(B) by illustrating all sites, multiple pathogens, and statistical significance of the annual periodicity. We used the MATLAB function *plomb* from version 2014A to estimate the spectral density; see [21] for all computational scripts.

For each combination of pathogen and site, we used the fast Fourier transform (FFT) to extract the relative phase of when each pathogen annually peaks. The FFT and periodogram are mathematically connected. Here, we use the FFT to identify the relative timing of the peak of disease incidence across diseases. We performed this transformation with respect to the first day in the two-week period. Computing the peak incidence from the FFT of the disease incidence time series was implemented in MATLAB version 2014A; the computational scripts can be found at [21].

Identifying associations between environmental factors and pathogens

We estimated the strength of association between weather covariates (i.e., cumulative precipitation, average temperature, average relative humidity, and average specific humidity) over each fortnight and the estimated number of children positive by pathogen and by country. Specifically, we identify the highest and lowest quartiles for the number of children with MSD estimated to be positive by fortnight. Quartiles were chosen due to low sample numbers for certain pathogens; for example, some pathogens were detected in only 1/4 of the fortnights. In
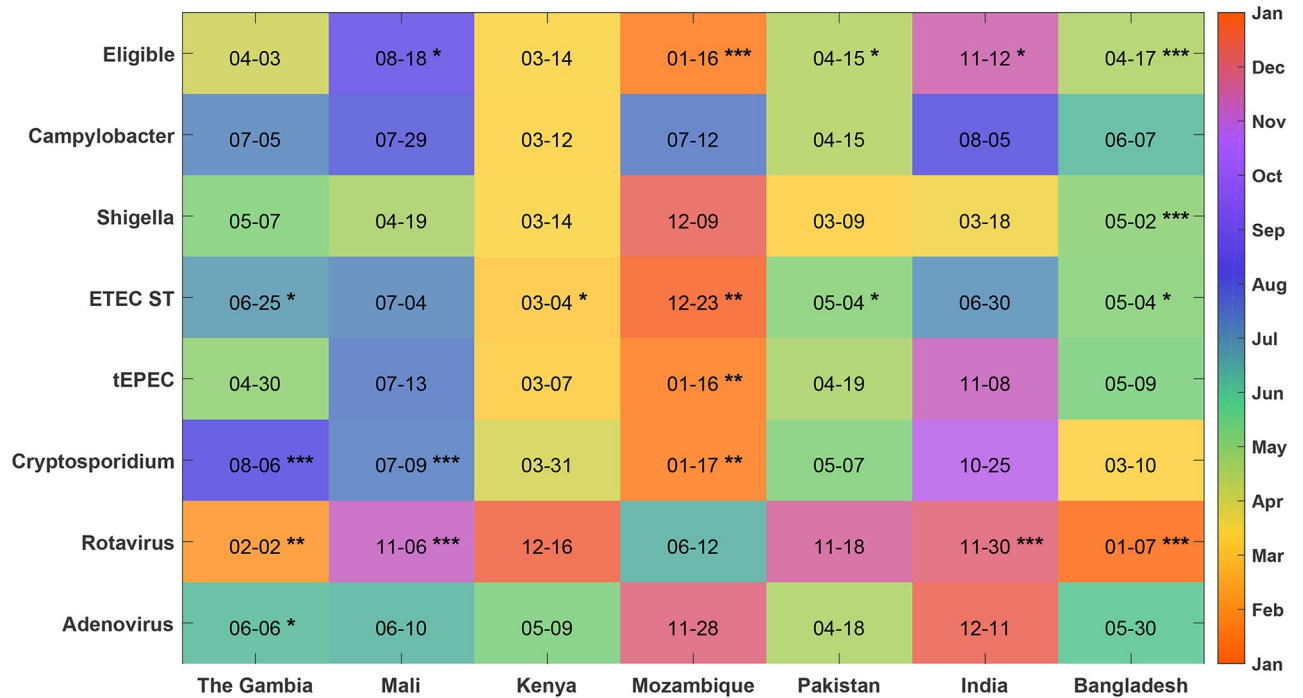
Fig 2. Annual peak dates of pathogens. We used the phase information from the FFT of the number of eligible children estimated to be positive each fortnight to determine the date of the annual peak of each pathogen in each country (indicated as month–day in each square and color-coded by month of year). The significance (false-detection rate p_{fd}) of the annual cycle was obtained from the Lomb–Scargle periodogram of the most significant period between 11 and 13 months, indicated by **** for <0.1%, *** for <1%, ** for <5%, and * for <10%.

https://doi.org/10.1371/journal.pntd.0007211.g002

these cases, the median could not be used to distinguish high and low prevalence, and the weather during fortnights with no positive cases was compared to the weather during fortnights with cases. The associations between pathogen prevalence and environmental values are computed using a Wilcoxon rank-sum test. The computations were performed in the R scientific computing environment [22] and the code can be found in [21].

Identifying data-driven seasons directly from weather data

We identify seasons directly from the monthly weather data derived from GSOD, instead of using pre-defined seasons from the literature. These seasons are *data-driven* in the sense that our algorithm clusters months that have similar weather covariates. We transform daily environmental values to monthly values by summing rainfall and taking the average values for temperature, relative humidity, and specific humidity. However, the year-to-year variability has a substantial impact on the beginning and end of seasons; see S2 Fig for an illustration of the season variability when using actual weather data to identify seasons.

To define seasons using weather data, we used principal component analysis (PCA) to reduce the number of weather covariates from four to three. The dimensionality reduction with principal components analysis eliminates redundant variables [25]. In particular, since the four weather covariates studied were derived from only three measured weather-related variables, the PCA procedure reduced the dimensionality appropriately. We chose not to reduce dimensionality further, since the first two principal components only explained 88 to 96% of the variance in the weather data. We implemented k-means clustering to group weather–months by similarity; three clusters best described the groupings of weather–months. Fig 1(C) shows each of the month's weather data projected on the leading two principal

components colored by cluster. We call these clusters data-driven seasons. This procedure identified reasonable quantitative definitions of seasons that are generally consistent with those informally identified in previous publications (S3 Table). Some seasons can occur more than once per year, e.g., there can be two rainy seasons in one calendar year. Fig 1(D) illustrates the timing of seasons in Bangladesh. We used the R scientific computing environment [22] and provide the analysis code at the following repository [21].

Determining significant differences in pathogen-positive cases by season

To determine if the estimated number of children positive for pathogens each month had significant differences among our data-driven seasons, we used the Kruskal-Wallis test. We then used Dunn's test to identify the pairs of seasons with significantly different numbers of children positive for pathogens. The tests were applied using the R scientific computing environment [22].

Ethics approval

The GEMS study protocol was approved by ethics committees at the University of Maryland, Baltimore and at each field site. Parents or caregivers of participants provided written informed consent, and a witnessed consent was obtained for illiterate parents or caretakers.

Results

Annual and biannual cycles are detected for several diarrheal pathogens and study sites

We studied the periodicity of the estimated number of eligible children positive for eight pathogens: *Campylobacter*, *Shigella*, ST-ETEC, tEPEC, *Cryptosporidium*, rotavirus, and adenovirus 40/41. *V. cholerae*, norovirus GI, and GII did not have statistically significant annual seasonality at any of the sites. The number of children with MSD in Mali, Mozambique, Pakistan, India, and Bangladesh had significant annual periodicity ($p_{fd} <$10%) (Fig 2). Each of the seven sites had 1–5 pathogens with significant annual periodicity ($p_{fd} <$10%). Rotavirus has significant annual periodicity in The Gambia, Mali, India, and Bangladesh (all $p_{fd} <$1%). *Cryptosporidium* has significant annual periodicity ($p_{fd} <$1%) in The Gambia, Mali, and Mozambique. *Shigella* ($p_{fd} <$0.1%) and EAEC ($p_{fd} <$1%) had significant annual periodicity only in Bangladesh. Fig 3 plots the number of study-eligible children (children with MSD) visiting the clinics each fortnight and the estimated number of them positive for rotavirus, *Cryptosporidium*, and *Shigella*. ST-ETEC and tEPEC had significant annual periodicity at $p_{fd} <$1% only in Mozambique. We also tested for 6-month periodicity (5.5–6.5 month period) using the same procedure as above. Of the seven sites, only Kenya had pathogens with a significant 6-month periodicity: rotavirus and norovirus GII ($p_{fd} <$10%).

Broad statistical association between fortnightly pathogen prevalence and weather

We found that fortnights with a high number of MSD clinic visits tended to have significantly higher rainfall and humidity in Mali and Mozambique than fortnights with a low number of MSD visits (Wilcoxon rank-sum test $p <$0.05). In contrast, fortnights with a high number of MSD visits were less rainy in India and less humid in The Gambia. Fig 4 illustrates the statistically significant associations between weather and pathogen prevalence by site. There were substantially more statistically significant associations between weather covariates and pathogen prevalence than detection of annual periodic cycles; see Figs 2 and 4 for a visual
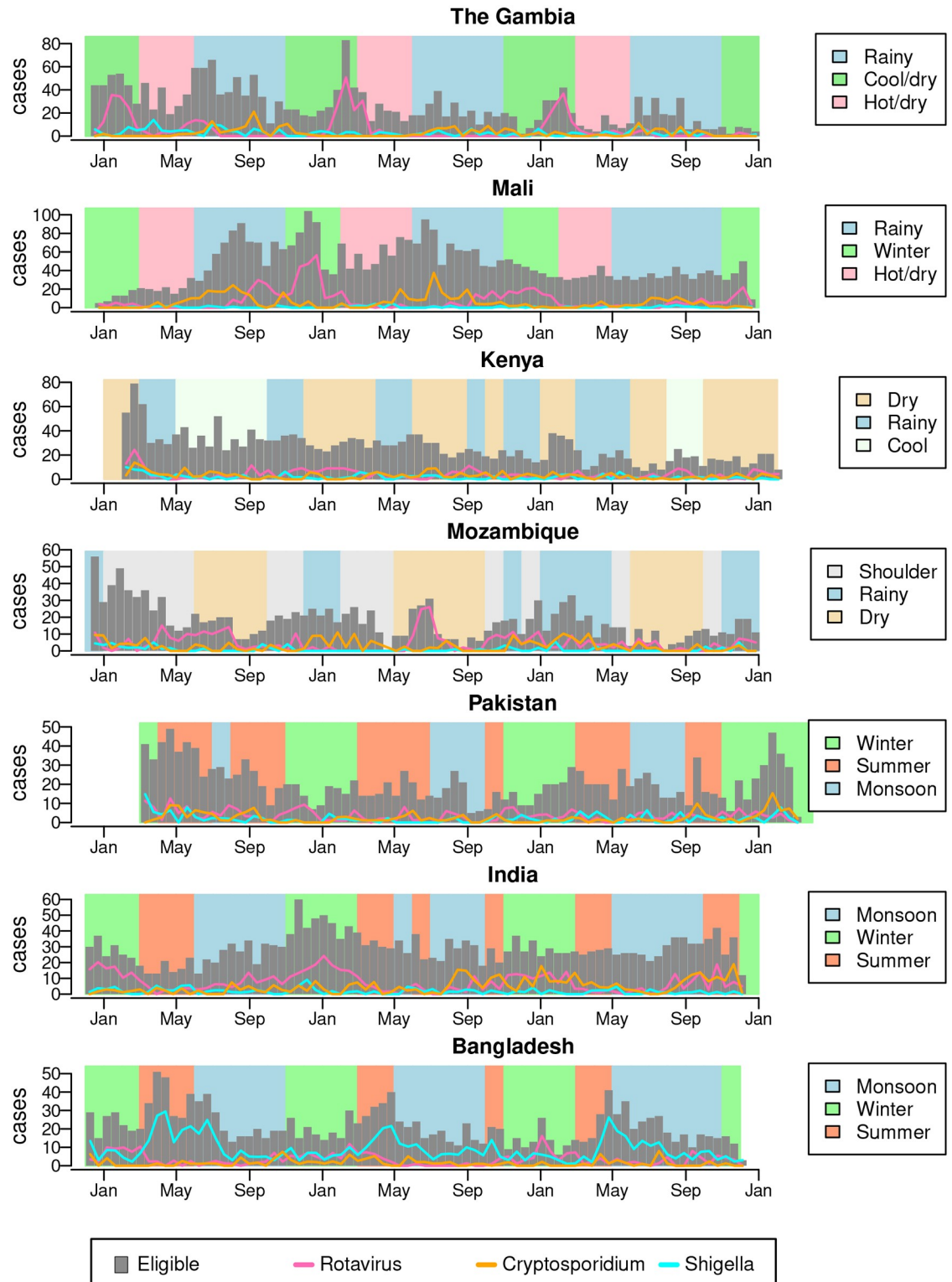
Fig 3. Estimated fortnightly cases positive for rotavirus, _Shigella_, and _Cryptosporidium_. Gray bars plot the number of children with MSD visiting the study clinics each fortnight. The lines show the estimated number of these children positive for rotavirus, _Shigella_, and _Cryptosporidium_. The background shading indicates the timing of data-driven seasons at each site.

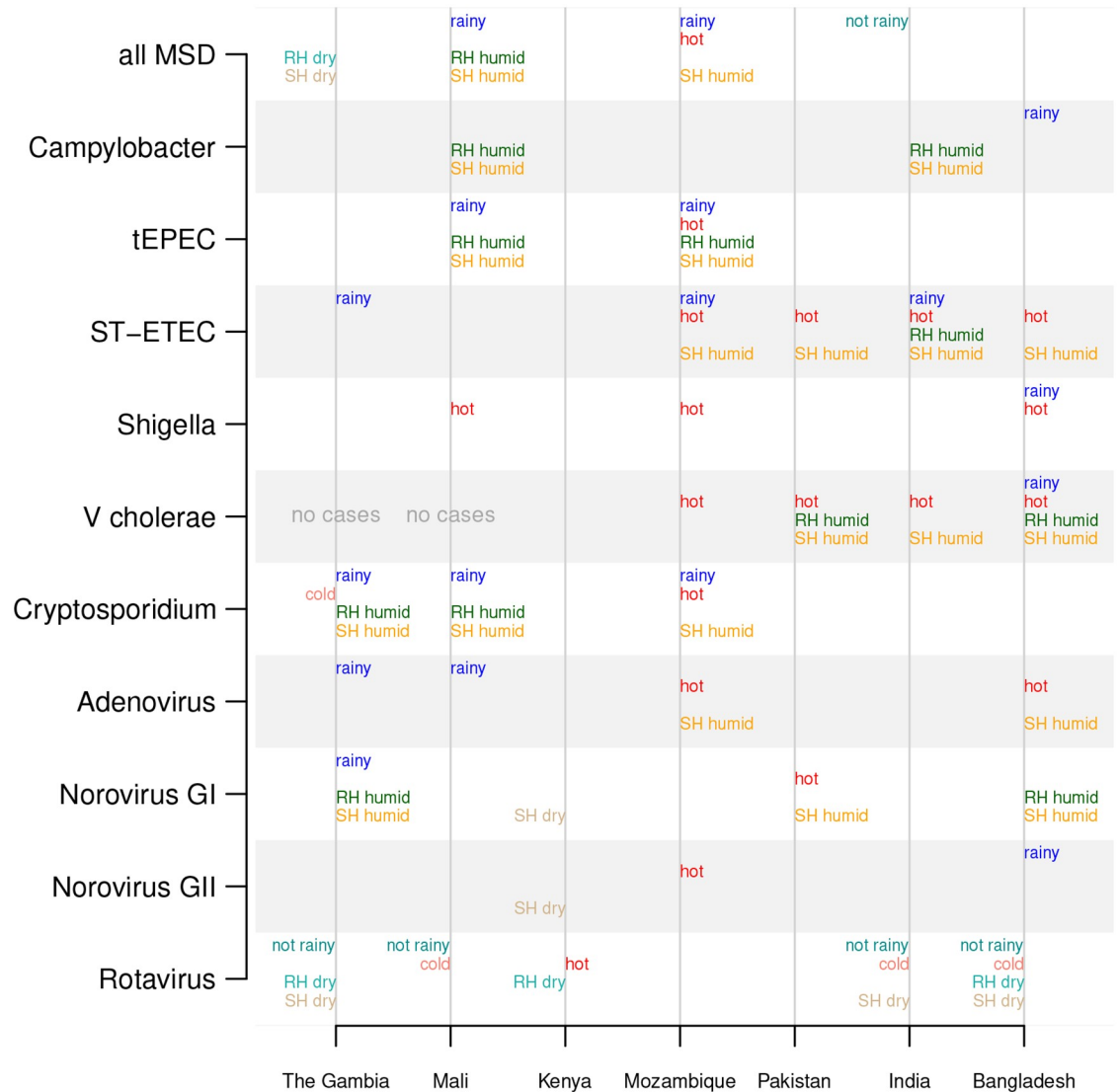https://doi.org/10.1371/journal.pntd.0007211.g003

Fig 4. Significant differences in weather of fortnights with high vs low pathogen prevalence. The weather covariates with significant differences (p<0.05 using the Wilcoxon rank–sum test) between fortnights with high vs low prevalence of pathogens are printed in the chart. Weather covariates that are higher during high-prevalence weeks are printed to the right of the vertical lines ("rainy" indicates higher rain in pathogen-associated fortnights, "hot" means higher temperatures, "RH humid" means higher relative humidity, and "SH humid" means higher specific humidity), and those that are lower during high-prevalence weeks are printed on the left (i.e., "not rainy", "cold", "RH dry", and "SH dry").

comparison across site and pathogen. We also detected a significant association between increased numbers of estimated rotavirus positive cases and lower humidity, rain, and temperature. Hot weather was also associated with high numbers of *Shigella*-positive cases (Fig 4). ST-ETEC, tEPEC, and *V. cholerae* were associated with hot and humid weather. Rainy and humid fortnights had significantly more *Cryptosporidium*-positive cases than dry fortnights.

Site-specific seasons can be identified from weather data

We identified three data-driven seasons for each GEMS site using rainfall, temperature, and humidity data. Fig 1(C) illustrates the clustering of monthly weather data, collected from the

closest weather station in Bangladesh during 2008 to 2011, projected on to the first two principal components. The color of each month indicates the cluster (*season*) to which the month belongs. S2 Table summarizes average weather conditions for each season at each site. Fig (1D) illustrates how the clustering is reflected temporally during the study time period. This methodology allows for seasons to vary in duration and initiation time and is based on site-specific weather data from the duration of study enrollment (S2 Fig).

We attached labels to these data-driven seasons to aid in interpretability. The labeling procedure consists of two components: the characteristics arising from clustering the weather data and region-specific characteristics. For example, periods encompassing the times of year with the most rainfall were labeled "Monsoon" in regions that experience monsoons and "Rainy" elsewhere. The hottest periods of the year were labeled "Summer" and the coldest periods "Winter" when those terms are used in the region; otherwise, the labels "Hot" and "Cool" were assigned. We used descriptive names in some cases, such as "Hot/dry." The seasons were generally consistent with those described in prior publications (S3 Table). Because the seasons were identified separately for each site, the actual weather for the same season label across sites could be quite different. Note that season labels are internally consistent within a site (e.g., "Summer" is always hotter than "Winter" at a site), but not across sites ("Summer" at one site might be cooler than "Summer" at another) (S2 Table).

Pathogen prevalence is strongly associated with data-driven seasons

We found rotavirus tended to have significantly higher prevalence in winter or cool/dry seasons than during rainy seasons (Fig 5). For example, the estimated rotavirus prevalence in The Gambia was significantly lower during the rainy compared to the two other seasons, but not statistically different between the cool/dry and hot/dry season. We did not find a country–pathogen combination with statistically different prevalences among all three seasons. We found *Shigella* was most prevalent during the summer in Bangladesh and the hot/dry season in Mali (Fig 5). ST-ETEC was most prevalent during the rainy season in The Gambia and Mozambique and the summer and monsoon seasons in Bangladesh and India. *Cryptosporidium* had highest prevalence in the rainy season in The Gambia, Mali, and Mozambique. Bacterial pathogens and *Cryptosporidium* had lower prevalence in winter or cool/dry seasons than in rainy or summer seasons.

The most prevalent pathogens differ by season and child age

Because different pathogens were associated with different seasons, the relative prevalence (and ranking) of pathogens among children with MSD differed by season. Among children 0 to 23 months old, rotavirus was the pathogen most frequently detected among eligible children in the winter, dry, or cool/dry seasons at all sites (22–57% of cases, Fig 6, left panel). In the rainy or monsoon months, rotavirus was less dominant, and *Shigella*, *Campylobacter*, and *Cryptosporidium* were the most often detected. Among those 24 to 59 months old, rotavirus was the most frequently detected pathogen at three of the seven sites in winter or cool/dry seasons (Fig 6, right panel). For these older children, *V. cholerae* is one of the top three pathogens at some sites during the monsoon or summer seasons.

Annual cycles indicate a birth month is a risk factor for diarrheal illness

If MSD or pathogen associated with MSD has a highly significant annual period, the age at which a child is most likely to get diarrhea depends on the birth month of a child. For example, if diarrheal disease peaks in the first quarter of the year, a child born in the second quarter might not be at high risk until six to nine months of age, while a child born before the first
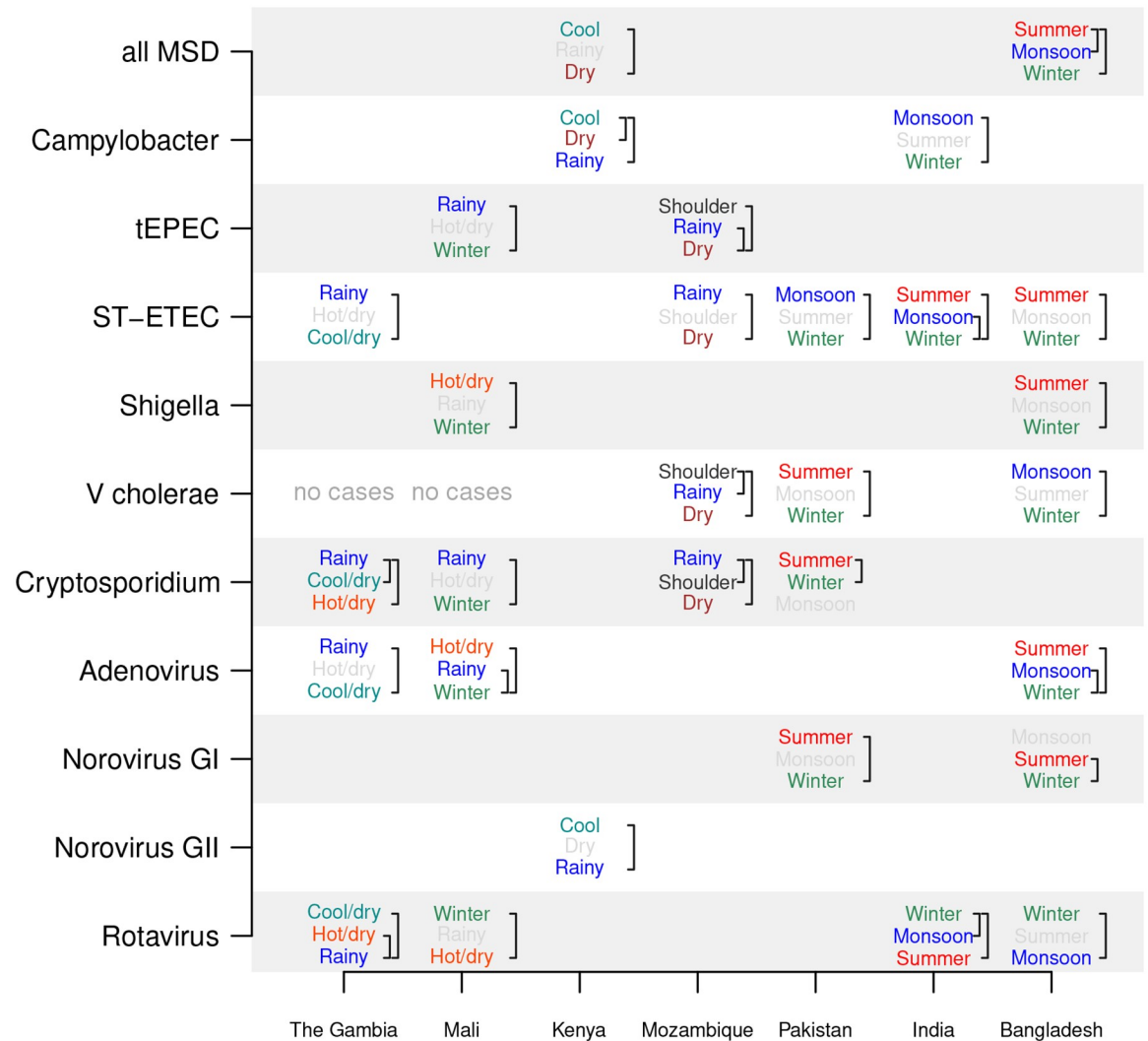
Fig 5. Relative prevalence of pathogens across seasons. The Dunn test is used to compare the monthly cases between pairs of seasons. If there is a significant difference between seasons, their names are printed, with the season with the most cases per month on top and fewest on the bottom. Brackets indicate pairs of seasons with significant differences, and season names are color-coded (summer, winter, monsoon, hot/dry, cool/dry, etc) when there is a statistically significant difference between seasons (p<0.05 using a Bonferroni correction) and light gray otherwise. No country–pathogen combination had significant prevalence differences among all three seasons.

quarter could be at high risk of exposure soon after birth. This effect can be seen in The Gambia, as shown in the heat maps of birth month vs age of MSD visit shown in S3 Fig. Strong seasonality is visible as diagonal "stripes" in these heatmaps. Similar effects can be seen among rotavirus-positive MSD visits when rotavirus has a significant annual periodicity (S4 Fig).

Discussion

In a systematic manner, we have characterized the seasonality of diarrheal pathogen prevalence at seven study sites in the Global Enteric Multicenter Study (GEMS). The GEMS study provided an unprecedented perspective on the population-based burden of moderate-to-severe diarrhea among children in resource-constrained settings spanning multiple countries and continents [2]. In this work, we leverage study data collected over a three year period to
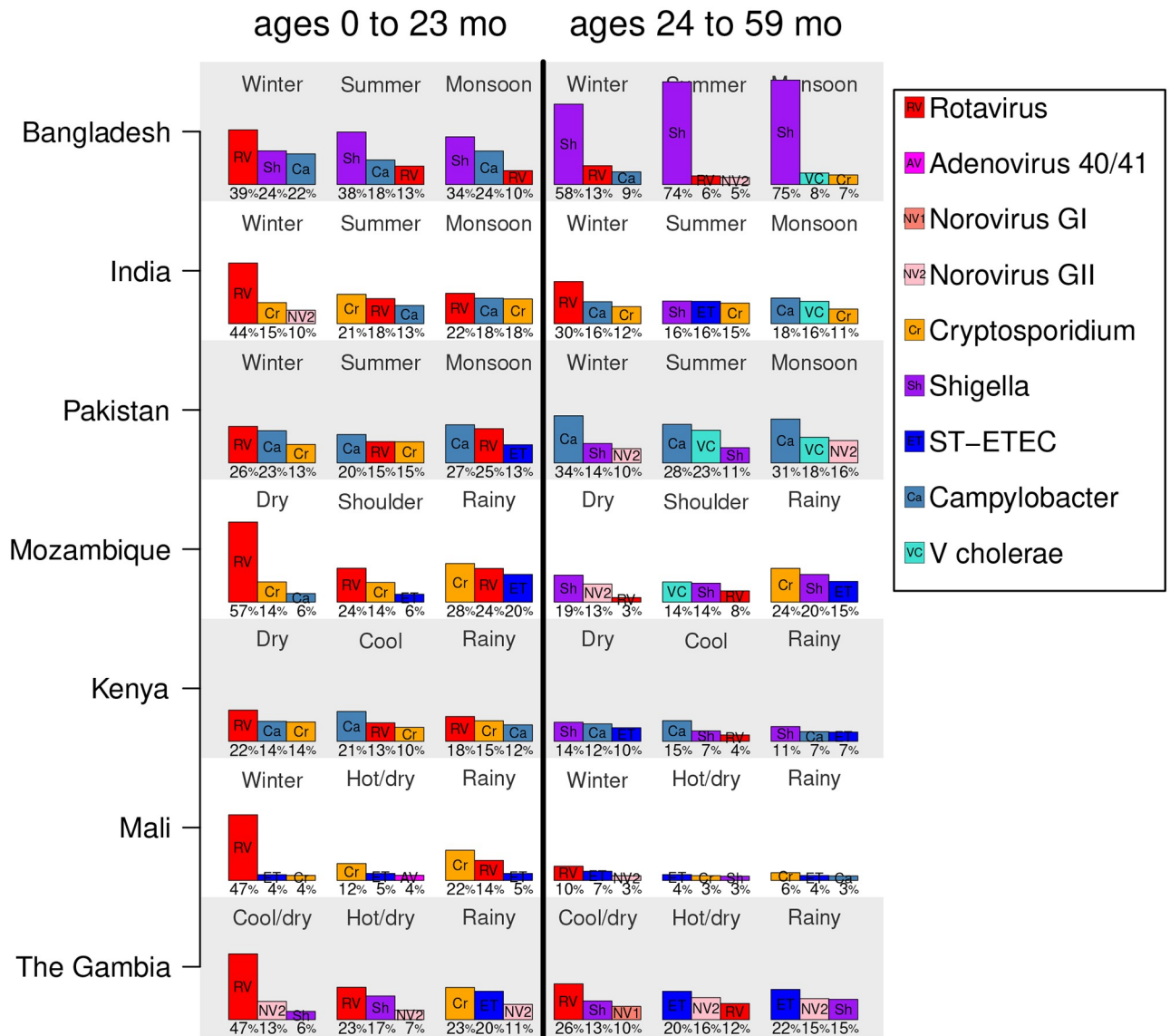
Fig 6. The top three pathogens by season. For each country, the top three most frequently detected pathogens (out of the nine considered) among eligible children ages 0 to 23 months (left half of the plot) and 24 to 59 months old (right half) are shown. Bar heights are proportional to the percent of cases positive for a pathogen (which is also printed below each bar), and the pathogens are indicated by abbreviations and color as summarized in the legend. Pathogens considered are: rotavirus, adenovirus 40/41, norovirus GI, norovirus GII, *Cryptosporidium*, *Shigella*, ST-ETEC, *Campylobacter*, and *V. cholerae*. Note that the seasons are not plotted in a particular order and may be non-sequential.

https://doi.org/10.1371/journal.pntd.0007211.g006

better understand the seasonality of the most important diarrhea-associated pathogens as well as the relationship of prevalence to site-specific environmental weather variables. We utilize standard mathematical methodologies to test for seasonality as well as more modern machine-learning algorithms to identify data-driven seasons from detailed environmental data to correlate with the periodicity of pathogen incidence. This work provides a unique perspective on characterizing the seasonal epidemiology of diarrheal pathogens due to the diversity of study sites in resource constrained settings, consistent study protocol, and large sample size.

Rotavirus was the only pathogen with highly significant annual prevalence peaks at most of the study sites. Previous studies have found that rotavirus diarrhea is seasonal and tends to

occur in the winter, but these study sites have been primarily in temperate regions [4–6]. Recent studies also suggest rotavirus seasonality might be weaker in the tropics, with some evidence that it peaks during the cooler and drier seasons [4–6, 26]. Introduction of rotavirus vaccine may attenuate the annual peaks in diarrheal disease associated with rotavirus [27]. We found less consistent periodicity of other pathogens across GEMS sites; however, there is less evidence in the scientific literature on the global seasonality of diarrheal pathogens other than rotavirus. We hypothesized that weather drives the transmission of many diarrheal pathogens at GEMS sites, and the lack of significant annual periodicity may be the result of a milder and more variable weather drivers found in these tropical regions.

We found associations between diarrheal pathogen prevalence and weather, despite the lack of significant annual periodicity of pathogen prevalence time series. Several bacterial pathogens were more prevalent during hot and rainy weather, which could favor the growth of bacteria in the environment or the contamination of water sources [28]. We found ST-ETEC was generally associated with warmer weather (in Mozambique, Pakistan, India, and Bangladesh), consistent with the previously observed association between ETEC and higher temperatures but not rainfall [29, 30]. A significant association between *Cryptosporidium* and rainy weather was identified in The Gambia, Mali, and Mozambique, which is consistent with the finding that precipitation may drive cryptosporidiosis in the tropics [31]. In contrast, rotavirus was found to be more prevalent during the drier winter months, out of phase with *Cryptosporidium*; this is broadly consistent with reviews of rotavirus seasonality studies in the tropics [5, 26, 32]. However, that some studies have detected secondary peaks of rotavirus activity during monsoon seasons in tropical settings [26, 33–35]. Our results for Bangladesh generally agree with the analysis by Das et.al. [36], which reported that rotavirus peaked in the winter, cholera in the monsoon, and ETEC in the summer at three sites in Bangladesh. However, it is difficult to identify the combination of components of weather that drives each pathogen, since weather covariates can be highly correlated (e.g., heat and humidity).

We found that the weather conditions at each site seemed to fall into a few distinct classes, such as "warm, humid, and rainy", "hot and dry", and "cool and rainy", and these classes broadly agree with informally identified site-specific seasons, such as "summer" and "winter". We used modern machine-learning techniques to formalize this observation and cluster the environmental data by site to identify data-driven seasons. To our knowledge, associating pathogen prevalence with data-driven seasons is an innovation for revealing site-specific weather patterns that are potential drivers of pathogen prevalence. We found that pathogen prevalence was often significantly different across seasons, and that pathogens might lack strict annual periodicity because of the year-to-year variability in the timing and length of seasons. We believe that these seasons serve several purposes when studying pathogen association with weather conditions: 1) Seasons last for a few months while weather conditions can change daily so associations with seasons are more robust to disease reporting lags and frequency; 2) Some pathogens may be driven by a combination of weather conditions (e.g., heat and humidity), which are captured by seasons but not by individual weather covariates; and 3) Seasons give us a common terminology to use across different sites.

The site-specific seasons defined here enable public health officials and clinicians to parse out population pathogen prevalence changes across seasons. For examples, Fig 6 shows how the top three pathogens change across seasons, sites, and age of child. Among children with MSD 0 to 23 months of age, rotavirus was the pathogen most frequently detected, particularly during winter or cool/dry seasons. *Cryptosporidium* was the top pathogen among 0 to 23 month-olds during the rainy season in some countries. Among children 24 to 59 months old, *Campylobacter*, *Shigella*, and ST-ETEC were the most frequently detected pathogens. Detailed site-specific descriptions could help with differential diagnosis and treatment choices for

patients with diarrhea if laboratory services are unavailable. The GEMS 1A study found a similar spectrum of pathogens among children with less severe diarrhea at the GEMS sites [37]. The timing of seasons can shift from year to year, which could complicate their use in clinical decision-making. However, we believe that by assigning seasons intuitive names, giving the approximate dates of the seasons, and describing the general weather conditions associated with each season, a local clinician could identify the appropriate season when making diagnoses. These weather-defined seasons were not optimized for studying disease epidemiology. A simpler classification of seasons, such as rainy vs dry season, could be easier to use in practice for some pathogens.

For sites with strong annual rotavirus seasonality, the birth month of a child is associated with age-dependent risk of rotavirus diarrhea. A similar result was reported in a study of rotavirus in England and Wales, a high-income country, where children born in the summer had a higher risk of rotavirus diarrhea in the first year of life compared to those born in the winter [38]. Broadly, this data and analysis can used to assess risk of a diarrheal disease by birth month; from a public health perspective, these results could be included in supply chain and operational planning for clinics and hospitals.

The smaller seasonal changes in temperature and humidity in some tropical settings compared to temperate ones could make it difficult to study weather as a driver of diarrheal disease. We found the association between pathogens and weather is less pronounced at sites with less seasonal variation in weather, such as the study sites in Kenya and Mozambique. Because temperate and tropical climates have different ranges of weather conditions, the relationship between pathogens and weather covariates could differ [39]. Shorter seasons may make detecting the association between weather and pathogens difficult, since the pathogen would have less time to respond (e.g., amplify, transmit) to changes in weather conditions. The Kenya site has two rainy and two dry seasons per year, thus weather covariates had bi-annual rather than annual periodicity, and rotavirus and norovirus GII prevalence had significant bi-annual periodicity, but previous studies noted that the seasonality of rotavirus disease is subtle in Kenya [40].

Weather might also mediate complex transmission pathways [12]; for example, *Shigella* had been observed to peak in April–June at the GEMS study site in Bangladesh and was associated with seasonal peaks in housefly density in February and March [41]. Therefore, peaks in disease prevalence could be driven by the weather during a preceding season. This in part motivated our decision to define seasons at a monthly resolution instead of a finer weekly or bi-weekly scale. Weather's effects on pathogen transmission may also interact with population density, so that adjacent urban and rural areas can experience differing pathogen seasonality [34], which could make it difficult to generalize the relationships between weather and pathogen prevalences. The effects of weather on infection can be affected by other socioeconomic factors. For example, Bhavnani et al found that heavy rainfall could increase the risk of diarrhea in households with unimproved water sources but that dry conditions could increase the risk of diarrhea in households with unimproved sanitation [42]. We only made empirical observations about the relationship between weather and children presenting with MSD without considering causality. We also need to consider the factors that influence care-seeking behavior and how representative the cases we measured are of the general population and of the most vulnerable sub-populations [43].

It is mathematically challenging to determine if a pathogen is causally associated with seasonal environmental factors. This study used three years of observations, so variation in weather between years could mitigate the confounding of climate and calendar-based covariates [11, 12]. We determined associations between weather and pathogen prevalence providing statistical support for our conclusions; moreover, we describe the probability of

mis-identifying a seasonal signal by site and pathogen, based on the study duration and surveillance sampling. We did not, however, attempt to identify *causally* linked environmental factors. There is the potential that additional years of data or spatial weather covariates within sites could further improve our ability to link weather to disease prevalence, particularly if years with extreme weather are included [44]. More sensitive pathogen detection assays could also improve our ability to analyze seasonality. A recent reanalysis of a subset of the samples from GEMS revealed higher prevalence of some pathogens among cases, particularly *Shigella*, adenovirus 40/41, ST-ETEC, and *Campylobacter* [19]. Even with more sensitive assays and longer time series, determining the etiology of diarrheal disease is difficult, since multi-pathogen infections are common and disease could be caused by (or even mitigated by) one of the pathogens. We primarily focused on pathogens that have strong associations with diarrhea (e.g., low minimum infectious dose) to mitigate this challenge. One result of this work is to provide caution to the global health community, especially given the current trend of estimating burden at a fine-scale spatial resolution with an underlying statistical model that relies too heavily on data from a nonrepresentative region.

Notwithstanding many of these limitations, our study was unique in studying the seasonality of multiple pathogens across multiple countries at the same time using the same study design. Although the primary purpose of GEMS was to identify the most prevalent and virulent pathogens associated with MSD at the study sites, the association of certain pathogens with weather covariates was strong enough to study. We believe that identifying the environmental conditions that facilitate transmission of pathogens could help us understand the mechanisms by which they spread in human populations and choose the most effective interventions to reduce transmission [42, 45–47]. We believe that this study will better inform the global health community around pathogen prevalence in different resource constrained settings. Moreover, the identification of age-dependent risk of pathogen and population prevalences by season could lead to better clinical diagnoses and allocation of public resources.

Supporting information

S1 Checklist. STROBE checklist of items for case–control studies.
(DOC)

S1 Table. Locations of weather stations.
(TEX)

S2 Table. Weather conditions during data-driven seasons. Months indicate when these seasons have occurred during the study, which can vary slightly by year. "% time" is the % of the study months assigned to each season. Mean rainfall, temperature, relative humidity, and specific humidity are displayed for each season, and the interquartile range of monthly conditions is printed in parentheses below the means.
(TEX)

S3 Table. Season labels and timing from the literature.
(TEX)

S1 Fig. Estimating prevalence of rotavirus among children with MSD. Children with MSD were eligible for the study (gold bars). A sample of them were enrolled in GEMS each fortnight (blue bars) and stools samples were tested for rotavirus and other pathogens. The estimated number of eligible children positive for a rotavirus (black line), is extrapolated from the positivity among enrolled children (purple bars) during the same fortnight.
(PDF)

S2 Fig. Seasons and monthly temperature, rainfall, and specific humidity for each site from 2007–2011. Seasons were defined by a k-means clustering algorithm on PCA-transformed monthly weather covariates. Relative humidity not shown.
(PDF)

S3 Fig. Age and birth month distribution of eligible children. The numbers in the plot are the number of eligible children born in month x and visited the clinic at age y. Yellow areas are low numbers of eligible children, while red numbers are high. Note the diagonal elements in countries where there is a distinct diarrheal disease season.
(PDF)

S4 Fig. Age and birth month distribution of enrolled cases who were rotavirus-positive. The numbers in the plot are the number of enrolled cases born in month x and were enrolled at age y who were rotavirus-positive. Note the diagonal elements in countries where there is a distinct rotavirus season. Enrollment rates could differ by age group, so one should not compare numbers across age groups (0–11m, 12–23m, 24m and older).
(PDF)

Author Contributions

Conceptualization: Dennis L. Chao, Anna Roose, Karen L. Kotloff, Joshua L. Proctor.

Data curation: Dennis L. Chao, Anna Roose, Karen L. Kotloff.

Formal analysis: Dennis L. Chao, Anna Roose, Min Roh, Joshua L. Proctor.

Investigation: Dennis L. Chao, Anna Roose, Karen L. Kotloff, Joshua L. Proctor.

Methodology: Dennis L. Chao, Anna Roose, Min Roh, Karen L. Kotloff, Joshua L. Proctor.

Software: Dennis L. Chao, Min Roh, Joshua L. Proctor.

Visualization: Dennis L. Chao, Min Roh, Joshua L. Proctor.

Writing – original draft: Dennis L. Chao, Anna Roose, Min Roh, Karen L. Kotloff, Joshua L. Proctor.

Writing – review & editing: Dennis L. Chao, Anna Roose, Min Roh, Karen L. Kotloff, Joshua L. Proctor.

References

1. Leclerc H, Schwartzbrod L, Dei-Cas E. Microbial agents associated with waterborne diseases. Critical Reviews in Microbiology. 2002; 28(4):371–409. https://doi.org/10.1080/1040-840291046768 PMID: 12546197

2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. 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. The Lancet. 2013; 382(9888):209–222. https://doi.org/10.1016/S0140-6736(13)60844-2

3. Platts-Mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). Lancet Glob Health. 2015; 3(9):e564–75. https://doi.org/10.1016/S2214-109X(15)00151-5 PMID: 26202075

4. Cook S, Glass R, LeBaron C, Ho MS. Global seasonality of rotavirus infections. Bulletin of the World Health Organization. 1990; 68(2):171. PMID: 1694734

5. Levy K, Hubbard AE, Eisenberg JN. Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. International Journal of Epidemiology. 2009; 38(6):1487–1496. https://doi.org/10.1093/ije/dyn260 PMID: 19056806

6. Patel MM, Pitzer VE, Alonso WJ, Vera D, Lopman B, Tate J, et al. Global seasonality of rotavirus disease. Pediatr Infect Dis J. 2013; 32(4):e134–47. https://doi.org/10.1097/INF.0b013e31827d3b68 PMID: 23190782

7. Ansari SA, Springthorpe VS, Sattar SA. Survival and vehicular spread of human rotaviruses: possible relation to seasonality of outbreaks. Reviews of Infectious Diseases. 1991; 13(3):448–461. https://doi.org/10.1093/clinids/13.3.448 PMID: 1866549

8. Watkins RR. Gastrointestinal infections in the setting of natural disasters. Current Infectious Disease Reports. 2012; 14(1):47–52. https://doi.org/10.1007/s11908-011-0225-5 PMID: 22116641

9. Levine OS, Levine MM. Houseflies (*Musca domestica*) as mechanical vectors of shigellosis. Reviews of Infectious Diseases. 1991; 13(4):688–696. https://doi.org/10.1093/clinids/13.4.688 PMID: 1925289

10. Graczyk TK, Knight R, Tamang L. Mechanical transmission of human protozoan parasites by insects. Clinical Microbiology Reviews. 2005; 18(1):128–132. https://doi.org/10.1128/CMR.18.1.128-132.2005 PMID: 15653822

11. Metcalf CJE, Walter KS, Wesolowski A, Buckee CO, Shevliakova E, Tatem AJ, et al. Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead. Proc R Soc B. 2017; 284(1860):20170901. https://doi.org/10.1098/rspb.2017.0901 PMID: 28814655

12. Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. PLoS Pathog. 2018; 14(11):e1007327. https://doi.org/10.1371/journal.ppat.1007327 PMID: 30408114

13. Lal A, Hales S, French N, Baker MG. Seasonality in human zoonotic enteric diseases: a systematic review. PLoS one. 2012; 7(4):e31883. https://doi.org/10.1371/journal.pone.0031883 PMID: 22485127

14. Iacono GL, Armstrong B, Fleming LE, Elson R, Kovats S, Vardoulakis S, et al. Challenges in developing methods for quantifying the effects of weather and climate on water-associated diseases: A systematic review. PLoS Neglected Tropical Diseases. 2017; 11(6):e0005659. https://doi.org/10.1371/journal.pntd.0005659 PMID: 28604791

15. Troeger C, Colombara DV, Rao PC, Khalil IA, Brown A, Brewer TG, et al. Global disability-adjusted life-year estimates of long-term health burden and undernutrition attributable to diarrhoeal diseases in children younger than 5 years. The Lancet Global Health. 2018; 6(3):e255–e269. https://doi.org/10.1016/S2214-109X(18)30045-7 PMID: 29433665

16. Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science. 1976; 39(2):447–462. https://doi.org/10.1007/BF00648343

17. Scargle JD. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal. 1982; 263:835–853. https://doi.org/10.1086/160554

18. Melamed RD, Khiabanian H, Rabadan R. Data-driven discovery of seasonally linked diseases from an Electronic Health Records system. BMC Bioinformatics. 2014; 15(6):S3. https://doi.org/10.1186/1471-2105-15-S6-S3 PMID: 25078762

19. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case–control study. The Lancet. 2016; 388(10051):1291–1301. https://doi.org/10.1016/S0140-6736(16)31529-X

20. Lott N. Global surface summary of day. National Climatic Data Center, Asheville, NC (https://datanoaagov/dataset/global-surface-summary-of-the-day-gsod. accessed 8-22-2016.

21. Chao D, Roh M, Proctor JL. Seasonality of MSD from GEMS sites; 2019. https://github.com/InstituteforDiseaseModeling/SeasonalityStudyOfGEMSdata.

22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria. 2019. https://www.R-project.org.

23. Bosen JF. An approximation formula to compute relative humidity from dry bulb and dew point temperatures. Monthly Weather Review. 1958; 86(12):486–486. https://doi.org/10.1175/1520-0493(1958)086%3C0486:AAFTCR%3E2.0.CO;2

24. Horne JH, Baliunas SL. A prescription for period analysis of unevenly sampled time series. The Astrophysical Journal. 1986; 302:757–763. https://doi.org/10.1086/164037

25. Hotelling H. Analysis of a complex of statistical variables with principal components. Journal of Educational Psychology. 1933; 24(6):417–441. https://doi.org/10.1037/h0071325

26. Colston JM, Ahmed AMS, Soofi SB, Svensen E, Haque R, Shrestha J, et al. Seasonality and within-subject clustering of rotavirus infections in an eight-site birth cohort study. Epidemiol Infect. 2018; 146(6):688–697. https://doi.org/10.1017/S0950268818000304 PMID: 29534766

27. Masukawa MdLT, de Souza EM, Gimenes E, Uchimura NS, Moriwaki AM, Uchimura TT. Time series investigation of changes in seasonality of acute diarrhea hospitalizations before and after rotavirus vaccine in Southern Brazil. Cad Saúde Pública. 2016; 32(10):e00080515. https://doi.org/10.1590/0102-311X00080515 PMID: 27783754

28. Carlton EJ, Woster AP, DeWitt P, Goldstein RS, Levy K. A systematic review and meta-analysis of ambient temperature and diarrhoeal diseases. Int J Epidemiol. 2016; 45(1):117–30. https://doi.org/10.1093/ije/dyv296 PMID: 26567313

29. Gonzales-Siles L, Sjoling A. The different ecological niches of enterotoxigenic Escherichia coli. Environ Microbiol. 2016; 18(3):741–51. https://doi.org/10.1111/1462-2920.13106 PMID: 26522129

30. Philipsborn R, Ahmed SM, Brosi BJ, Levy K. Climatic drivers of diarrheagenic Escherichia coli incidence: a systematic review and meta-analysis. The Journal of Infectious Diseases. 2016; 214(1):6–15. https://doi.org/10.1093/infdis/jiw081 PMID: 26931446

31. Jagai JS, Castronovo DA, Monchak J, Naumova EN. Seasonality of cryptosporidiosis: A meta-analysis approach. Environ Res. 2009; 109(4):465–78. https://doi.org/10.1016/j.envres.2009.02.008 PMID: 19328462

32. Cunliffe N, Kilgore P, Bresee J, Steele A, Luo N, Hart C, et al. Epidemiology of rotavirus diarrhoea in Africa: a review to assess the need for rotavirus immunization. Bulletin of the World Health Organization. 1998; 76(5):525. PMID: 9868844

33. Sarkar R, Kang G, Naumova EN. Rotavirus seasonality and age effects in a birth cohort study of southern India. PLoS One. 2013; 8(8):e71616. https://doi.org/10.1371/journal.pone.0071616 PMID: 23977089

34. Martinez PP, King AA, Yunus M, Faruque A, Pascual M. Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world. Proceedings of the National Academy of Sciences. 2016; 113(15):4092–4097. https://doi.org/10.1073/pnas.1518977113

35. Mohan VR, Karthikeyan R, Babji S, McGrath M, Shrestha S, Shrestha J, et al. Rotavirus infection and disease in a multisite birth cohort: results from the MAL-ED study. The Journal of Infectious Diseases. 2017; 216(3):305–316. https://doi.org/10.1093/infdis/jix199 PMID: 28472348

36. Das S, Begum D, Ahmed S, Ferdous F, Farzana F, Chisti M, et al. Geographical diversity in seasonality of major diarrhoeal pathogens in Bangladesh observed between 2010 and 2012. Epidemiology & Infection. 2014; 142(12):2530–2541. https://doi.org/10.1017/S095026881400017X

37. Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, et al. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case–control study as a follow-on to the Global Enteric Multicenter Study (GEMS). Lancet Glob Health. 2019; 7(5):e568–e584. https://doi.org/10.1016/S2214-109X(19)30076-2 PMID: 31000128

38. Atchison CJ, Tam CC, Lopman BA. Season of birth and risk of rotavirus diarrhoea in children aged <5 years. Epidemiol Infect. 2009; 137(7):957–60. https://doi.org/10.1017/S0950268808001933 PMID: 19144248

39. Atchison C, Tam C, Hajat S, van Pelt W, Cowden J, Lopman B. Temperature-dependent transmission of rotavirus in Great Britain and The Netherlands. Proceedings of the Royal Society of London B: Biological Sciences. 2010; 277:933–942. https://doi.org/10.1098/rspb.2009.1755

40. Omore R, Tate JE, O'Reilly CE, Ayers T, Williamson J, Moke F, et al. Epidemiology, seasonality and factors associated with rotavirus infection among children with moderate-to-severe diarrhea in rural western Kenya, 2008–2012: the Global Enteric Multicenter Study (GEMS). PloS one. 2016; 11(8):e0160060. https://doi.org/10.1371/journal.pone.0160060 PMID: 27494517

41. Farag TH, Faruque AS, Wu Y, Das SK, Hossain A, Ahmed S, et al. Housefly population density correlates with shigellosis among children in Mirzapur, Bangladesh: a time series analysis. PLoS Neglected Tropical Diseases. 2013; 7(6):e2280. https://doi.org/10.1371/journal.pntd.0002280 PMID: 23818998

42. Bhavnani D, Goldstick JE, Cevallos W, Trueba G, Eisenberg JN. Impact of rainfall on diarrheal disease risk associated with unimproved water and sanitation. The American Journal of Tropical Medicine and Hygiene. 2014; 90(4):705–711. https://doi.org/10.4269/ajtmh.13-0371 PMID: 24567318

43. Nasrin D, Wu Y, Blackwelder WC, Farag TH, Saha D, Sow SO, et al. Health care seeking for childhood diarrhea in developing countries: Evidence from seven sites in Africa and Asia. Am J Trop Med Hyg. 2013; 89(1 Suppl):3–12. https://doi.org/10.4269/ajtmh.12-0749 PMID: 23629939

44. Checkley W, Epstein LD, Gilman RH, Figueroa D, Cama RI, Patz JA, et al. Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. Lancet. 2000; 355 (9202):442–50. https://doi.org/10.1016/s0140-6736(00)82010-3 PMID: 10841124

45. Clasen T, Roberts I, Rabie T, Schmidt W, Cairncross S. Interventions to improve water quality for preventing diarrhoea. Cochrane Database of Systematic Reviews. 2006; 3(3):CD004794.

46. Berendes D, Leon J, Kirby A, Clennon J, Raj S, Yakubu H, et al. Household sanitation is associated with lower risk of bacterial and protozoal enteric infections, but not viral infections and diarrhoea, in a

cohort study in a low-income urban neighbourhood in Vellore, India. Tropical Medicine & International Health. 2017; 22(9):1119–1129. https://doi.org/10.1111/tmi.12915

47. Wang Y, Moe CL, Null C, Raj SJ, Baker KK, Robb KA, et al. Multipathway quantitative assessment of exposure to fecal contamination for young children in low-income urban environments in Accra, Ghana: the SaniPath analytical approach. The American Journal of Tropical Medicine and Hygiene. 2017; 97(4):1009–1019. https://doi.org/10.4269/ajtmh.16-0408 PMID: 29031283