Enteropathogen seroepidemiology among children in low-resource settings

Benjamin F. Arnold [1 *], Diana L. Martin [2], Jane Juma [3], Harran Mkocha [4], John B. Ochieng [3], Gretchen M. Cooley [2], Richard Omore [3], E. Brook Goodhew [2], Jamae F. Morris [5], Veronica Costantini [6], Jan Vinjé [6], Patrick J. Lammie [2,7], Jeffrey W. Priest [8]

1. Division of Epidemiology and Biostatistics, University of California, Berkeley, Berkeley, CA, USA
2. Division of Parasitic Diseases and Malaria, United States Centers for Disease Control and Prevention, Atlanta, GA, USA
3. Kenya Medical Research Institute, Centre for Global Health Research, Kisumu, Kenya
4. Kongwa Trachoma Project, Kongwa, Tanzania
5. Department of African-American Studies, Georgia State University, Atlanta, GA, USA
6. Division of Viral Diseases, United States Centers for Disease Control and Prevention, Atlanta, GA, USA
7. Neglected Tropical Diseases Support Center, Task Force for Global Health, Decatur, GA, USA
8. Division of Foodborne, Waterborne, and Environmental Diseases, United States Centers for Disease Control and Prevention, Atlanta, GA, USA

* Corresponding author:
Benjamin F. Arnold, PhD
Division of Epidemiology and Biostatistics
School of Public Health
University of California, Berkeley
2121 Berkeley Way
Berkeley, CA 94602
+1 (510) 219 - 2787
benarnold@berkeley.edu

Abstract

Little is known about enteropathogen seroepidemiology among children in low-resource settings. We measured serological IgG response to eight enteropathogens (*Giardia duodenalis, Cryptosporidium parvum, Entamoeba histolytica, Salmonella enterica*, enterotoxigenic *Escherichia coli*, *Vibrio cholerae, Campylobacter jejuni*, norovirus) using multiplex bead assays in cohorts from Haiti, Kenya, and Tanzania. By age 2 years, most children had evidence of exposure by IgG response to the pathogens studied. We discovered a shift in IgG distributions for many pathogens as children age, caused by boosting and waning from repeated exposures, which complicates interpretation of seroprevalence among older children. Longitudinal profiles revealed important variation in enteropathogen IgG response above seropositivity cutoffs, underscoring the importance of longitudinal designs to estimate seroincidence rates as a measure of force of infection. In longitudinal cohorts there was a linear relationship between seroprevalence and prospective seroincidence rates, suggesting the two measures provide similar information about variation in pathogen transmission.

Introduction

A broad set of viral, bacterial, and parasitic enteropathogens are leading causes of the global infectious disease burden, with the highest burden among young children living in lower income countries [1]. Infections that result in acute diarrhea and related child deaths drive disease burden estimates attributed to enteropathogens, but recent advances in molecular assays have shown that asymptomatic infections are extremely common [2,3]. Understanding the full picture of enteropathogen exposure and infection in populations is important to assess transmission, disease burden, naturally acquired immunoprotection, as well as to design public health prevention measures, and measure intervention effects.

One approach to developing a higher-resolution picture of exposure to enteropathogens is to measure antibody responses in saliva or serum [4–7]. Since many enteropathogen infections are asymptomatic, antibody-based measures provide a more complete assessment of population exposure compared with clinical surveillance. For example, antibody-based estimates of incident exposure to *Salmonella enterica* and *Campylobacter jejuni* were 2-6 orders of magnitude higher than standard case-based surveillance in European populations [8–11], and a recent study of *Salmonella enterica* serotype Typhi in Fiji found similarly high discordance between antibody-based cumulative incidence and case-based surveillance [12]. Beyond surveillance, studies have started to use enteropathogen-specific antibody responses to measure the effects of household- and school-based water, sanitation, and handwashing interventions [13–15] and to examine heterogeneity in exposure across environmental conditions [16–19].

Many enteropathogens elicit a transiently high antibody response after infection that wanes over time. In lower transmission settings where antibody responses could be monitored longitudinally after distinct infections, *Salmonella enterica*, *Campylobacter jejuni*, *Cryptosporidium parvum,* and *Giardia duodenalis* immunoglobulin G (IgG) levels in blood have been shown to wane over a period of months since infection; IgM and IgA levels decline even more quickly [10,20–24]. Using information about a population's mean antibody decay profile from the time of infection has inspired methods to estimate seroincidence rates from cross-sectional studies in low transmission settings [10,25]. As examples in this study will illustrate, in higher transmission settings frequent exposure and antibody boosting prevails, which could favor longitudinal designs to estimate seroincidence.

To our knowledge, there has been no detailed characterization of enteropathogen seroepidemiology among children in low-resource settings where transmission is intense beginning early in life [3]. Drawing on cohorts of young children in Haiti, Tanzania, and Kenya that measured serological antibody response to eight enteropathogens using multiplex bead

assays [6,7,26,27], our objectives were to identify general seroepidemiologic patterns across pathogens and populations and to evaluate methods that estimate epidemiologic parameters of interest, such as the force of infection. Our results provide new insights into the seroepidemiology of enteropathogens among children living in low-resource settings, provide guidance for future study design and analysis, and highlight key remaining knowledge gaps.

Results

Study populations

The analysis included measurements from cohorts in Haiti, Kenya, and Tanzania. Blood specimens were tested for IgG levels to eight enteropathogens using a multiplex bead assay on the Luminex platform (Table 1). The Haitian cohort included repeated measurements among children enrolled in a study of lymphatic filariasis transmission in Leogane from 1990–1999 [28,29]. Leogane is a coastal agricultural community west of Port au Prince. At the time of the study its population was approximately 15,000, most homes had no electricity and none had running water. In total, the Haiti study tested 771 finger prick blood specimens collected from 142 children ages birth to 11 years old, with each measurement typically separated by one year (median measurements per child: 5; range: 2 to 9). In Kenya, a 2013 prospective trial of locally-produced, in-home ceramic water filters enrolled 240 children in a serological substudy [26]. Study participants were identified through the Asembo Health and Demographic Surveillance System, which is located in a rural part of Siaya County, western Kenya along the shore of Lake Victoria. Only 29% of the population had piped drinking water (public taps), water source contamination with *E. coli* prevailed (93% of samples tested), and the average age children began consuming water was 4 months [26]. Children aged 4 to 10 months provided blood specimens at enrollment, and again 6 to 7 months later (n=199 children measured longitudinally). In Tanzania, 96 independent clusters across 8 trachoma-endemic villages in the Kongwa region were enrolled in a randomized trial to study the effects of annual azithromycin distribution on *Chlamydia trachomatis* infection [27]. The population is very rural, and water is scarce in the region: at enrollment, 69% of participants reported their primary drinking water source, typically an unprotected spring, was >30 minutes walk one-way. From 2012 to 2015, the Tanzania study tested dried blood spots from between 902 and 1,577 children ages 1–9 years old in annual cross-sectional surveys (total measurements: 4,989). Although children could have been measured repeatedly over the four-year study in Tanzania, they were not tracked longitudinally. There was no evidence that the Kenya and Tanzania interventions reduced enteropathogen antibody response (Supplementary Information File 1), so this analysis pooled measurements over arms in each population.

Age-dependent shifts in population antibody distributions

We estimated seropositivity cutoffs using receiver operator characteristic (ROC) curve analyses for *Giardia, Cryptosporidium,* and *Entamoeba histolytica* including a panel of external, known positive and negative specimens, as previously reported [6,26]. We used two within-sample approaches to estimate seropositivity cutoffs for pathogens without ROC-based cutoffs: we used Gaussian mixture models [30] fit to measurements among children ages 0–1 year old (to ensure a sufficient number of unexposed children), and we used presumed seronegative distributions among children who experienced large increases in antibody levels (details in Methods). Classification agreement was high between the different estimation approaches (agreement >95% for most comparisons; Supplementary Information File 2). Mixture models failed to estimate realistic cutoff values if there was an insufficient number of unexposed children, which was the case for enterotoxigenic *Escherichia coli* heat labile toxin β subunit (ETEC LT β subunit) and *Vibrio cholerae* toxin β subunit in all cohorts, and for nearly all antigens in Tanzania where the study did not enroll children <1 year old (Table 1).

Among children <2 years old, antibody levels clearly distinguished subpopulations of seronegative and seropositive, but there were not distinct seronegative and seropositive subpopulations by age 3 years for most pathogens measured in Haiti (Figure 1) and Tanzania (Figure 1 - supplement 1). By age 3 years, the majority of children were seropositive to *Cryptosporidium*, ETEC LT β subunit, and norovirus GI.4 and GII.4; in all cases antibody distributions were shifted above seropositivity thresholds. In contrast, there was a qualitative change in the antibody response distributions to *Giardia*, *E. histolytica*, *Salmonella* and *Campylobacter* with increasing age, shifting from a bimodal distribution of seronegative and seropositive groups among children ≤1 year old to a unimodal distribution by age 3 years and older (Figure 1, Figure 1 - supplement 1). A direct comparison of age-dependent shifts in antibody distributions to *Giardia* VSP-3 antigen and *C. trachomatis* Pgp3 antigen in Tanzania illustrates stark differences in enteropathogen-generated immune responses versus pathogens like *C. trachomatis* that elicit a response that consistently differentiates exposed and unexposed subpopulations as children age (Figure 1 - supplement 2). Distributions of IgG levels in the younger Kenyan cohort (ages 4–17 months) showed distinct groups of seropositive and seronegative measurements for most antigens (Figure 1 - supplement 3). IgG responses to ETEC LT β subunit and cholera toxin β subunit were near the maximum of the dynamic range of the assay for nearly all children measured in the three cohorts (Figure 1, Figure 1 - supplement 1, Figure 1 - supplement 3), with evidence of waning IgG levels as children aged, presumably from adaptive immunity (Figure 1, Figure 1 - supplement 1).

Joint variation in antibody response

We examined joint variation in IgG responses to determine whether multiple recombinant antigens for the same pathogen provide additional information for individual-level immune response and to determine whether there was evidence for cross-reactivity between closely related antigens. Responses to *Giardia* recombinant VSP antigens were strongly correlated in Haiti (Spearman rank: ρ=0.99), Kenya (ρ=0.84) and Tanzania (ρ=0.97), which would be expected due to cross-reactivity in the antigens (Figure 2). *Cryptosporidium* (Cp17, Cp23) and *Campylobacter* (p18, p39) antigens were strongly correlated but high within-individual variability suggests that measuring multiple recombinant antigens for these pathogens, derived from different proteins, yields more information about exposure than measuring one alone (Figure 2). High correlation between *Salmonella* LPS Groups B and D, between norovirus GI.4 and GII.4, and between ETEC and *V. cholerae* likely reflected antibody cross-reactivity. Correlation could also result from multiple previous infections with different *Salmonella* serogroups or different norovirus genogroups. We excluded cholera β toxin antibody responses from remaining analyses because of the difficulty of interpreting its epidemiologic measures in light of high levels of cross-reactivity with ETEC heat labile toxin β subunit, since heat labile toxin-producing ETEC is very common among children in low-resource settings [3], and there was no documented transmission of cholera in the study populations during measurement periods. A broader comparison across all antigens in each cohort revealed no other antigen combinations with high correlation (Supplementary Information File 3).

Birth to three years of age: a key window of antibody acquisition and primary seroconversion

IgG levels increased rapidly with age for most pathogens in all three cohorts (Figure 3). There was some evidence of maternally-derived IgG among children under 6 months old with a drop in mean IgG levels by age, but this pattern was only evident for norovirus GI.4 in Haiti (Figure 3) and *Cryptosporidium* Cp17 and Cp23 in Kenya (Figure 3 - supplement 2). Age-dependent mean IgG responses to many pathogens declined from a young age, presumably from exposure early in life and acquired immunity. Mean IgG levels declined after age 1 year for *Giardia* in Haiti (Figure 3), *Giardia* and *Campylobacter* in Tanzania (Figure 3 - supplement 1).

We classified children as seropositive to *Giardia*, *Cryptosporidium*, *Campylobacter*, or *Salmonella* if antibody levels against either of the antigens from each pathogen were above estimated seropositivity cutoff values. For most pathogens, seroprevalence rose quickly and plateaued by ages 1 to 3 years in Haiti (Figure 4), Kenya (Figure 4 - supplement 1), and Tanzania (Figure 4 - supplement 2). In Haiti, seroprevalence ranged from 66% (*E. histolytica*) to 100%

(ETEC LT β subunit) by age 3 years (Figure 4), and in Tanzania, the majority of 1 year olds were already seropositive for *Giardia* (77%) and *Cryptosporidium* (85%) (Figure 4 - supplement 2).

Longitudinal antibody profiles

In Haiti and Kenya we examined longitudinal IgG profiles among children and compared them with seropositivity cutoffs. Vaccine immunogenicity and pathogen challenge studies among healthy adults often use a 4-fold increase in antibody levels (difference of +0.6 on the $\log_{10}$ scale) as a criterion for seroconversion [31–33]. In Kenya, 4-fold increases in IgG largely coincided with a change in status from seronegative to seropositive, presumably because increases in IgG followed primary exposure in the young cohort ages 4–17 months (Figure 5). Many Kenyan children exhibited >4-fold increases and decreases in IgG response to *Campylobacter* p18 and p39 antigens above the seropositivity cutoff, a result of earlier primary infection and/or additional infection and boosting during the study period (Figure 5). In the Haitian cohort, which was followed beyond the window of primary exposure, children commonly had >4-fold increases and decreases in IgG while remaining above seropositivity cutoffs—a pattern observed across pathogens but particularly clear for *Cryptosporidium* (Figure 5 - supplement 1).

Force of infection estimated by seroconversion rates

The seroconversion rate, an instantaneous rate of seroconversion among those who are susceptible, is one estimate of a pathogen's force of infection and a fundamental epidemiologic measure of transmission [34]. Longitudinal designs in Haiti and Kenya enabled us to use individual child antibody profiles to estimate average incidence rates of seroconversion and seroreversion during the study. We defined incident seroconversions and seroreversions as a change in IgG across a pathogen's seropositivity cutoff. In a secondary analysis, we defined incident boosting as a ≥4-fold increase in IgG to a final level above a seropositivity cutoff, and incident waning as ≥4-fold decrease in IgG from an initial level above a seropositivity cutoff. The secondary definition captured large changes in IgG above seropositivity cutoffs, which aligned with repeated boosting and waning observed in the Haitian cohort (Figure 5 - supplement 1).

Seroconversion rates in Kenya ranged from 0.1 seroconversions per child-year for *E. histolytica* to >3 seroconversions per child-year for ETEC and *Campylobacter* (Figure 6). Highest transmission pathogens (ETEC, *Campylobacter*) were dominated by seroconversion, whereas seroreversion rates were higher than seroconversion rates for *E. histolytica* and *Salmonella* (Figure 6). The proportion of children who changed serostatus over the 6 month follow-up period

reflected the same pattern as seroconversion rates, and exceeded 90% for ETEC and *Campylobacter* (Figure 6).

In Haiti, seroconversion rates ranged from 0.34 seroconversions per child-year (*E. histolytica*) to 1.13 seroconversions per child-year (ETEC) (Table 2). Seroincidence rates estimated from 4-fold changes in IgG led to more events identified and higher rates compared with those estimated from seroconversion alone (Table 2). For example, *Cryptosporidium* incident cases increased from 70 to 190 (a 2.7 fold increase) and the average rate increased from 0.64 (95% CI: 0.54, 0.77) to 0.80 (95% CI: 0.70, 0.90) per child-year (a 1.3 fold increase) when using a 4-fold IgG change criteria because of substantial IgG boosting and waning above the seropositivity cutoff (Figure 5 - supplement 1).

There was an approximately linear relationship between seroprevalence and prospectively estimated seroconversion rates in Kenya and Haiti (Figure 7). Overall levels and slope of the relationship differed considerably between cohorts, presumably because Kenya measurements were within a window of primary exposure for most children (4–17 months) whereas Haiti measurements extended from birth to 11 years and captured lower incidence periods with overall higher seroprevalence as children aged.

Since seroprevalence and seroconversion rates appeared to embed similar information about relative strength of enteropathogen transmission, we attempted to estimate seroconversion rates directly from age-structured seroprevalence in Kenya. We estimated seroconversion rates from seroprevalence curves using methods developed for cross-sectional ("current status") data, a common approach in serosurveillance of vaccine preventable diseases [34], malaria [35], and dengue [36] (Supplementary Information File 4). Seroconversion rates estimated from cross-sectional, semiparametric spline models were similar to rates estimated prospectively for all pathogens, but had substantially wider confidence intervals owing to loss of information by ignoring the longitudinal data structure (Figure 6 - supplement 1). Parametric approaches including an exponential proportional hazards model [37] and a reversible catalytic model [35], which both assumed a constant seroconversion rate over the age range, had narrower confidence intervals than the semiparametric spline model but tended to underestimate seroconversion rates compared with longitudinal estimates (Figure 6 - supplement 1).

Discussion

Across cohorts in Haiti, Kenya, and Tanzania, we identified consistent patterns in IgG responses that provide new insights into enteropathogen seroepidemiology among children in low-resource settings. We observed the most population-level heterogeneity in IgG levels and seroconversion between birth and 3 years, reflecting high transmission and early life primary exposure. In older

children, IgG levels either remained uniformly high, or began to wane through adaptive immunity. For particularly high transmission pathogens (e.g., ETEC, *Campylobacter*), most variation in IgG levels was observed among children <1 year old. A novel discovery was that for *Giardia, Cryptosporidium, E. histolytica,* and *Campylobacter* population antibody level distributions shifted with age: distributions among children <1 year were generally bimodal and reflected seronegative and seropositive subpopulations, but by age 3 years were unimodal and not centered clearly above or below seropositivity cutoffs. The shift of IgG distributions from bimodal to unimodal, resulting from a combination of antibody boosting, waning and adaptive immunity, complicates the interpretation of seroprevalence at older ages: among older children a seronegative response could either mean the children were never exposed or they were previously exposed but antibody levels waned below seropositivity cutoffs. Our findings suggest that among children in low-resource settings, seroprevalence and other derived epidemiologic parameters like seroconversion rates are most clearly interpretable when estimated within a window of primary infection, and the window likely closes by age 2–3 years. High levels of between-individual variability in antibody response identified through longitudinal profiles in Haiti and Kenya, with significant boosting and waning of antibody levels above seropositivity cutoffs, reinforce the value of longitudinal designs to derive antibody distributions among unexposed and to estimate seroincidence rates.

Measuring heterogeneity in exposure with mean IgG levels avoids the need to define seropositivity cutoffs and the difficulty of shifting interpretations of seroprevalence as children age, which are both germane considerations for enteropathogens. For comparisons within-study and within-pathogen, such as endpoints in randomized trials, comparisons based on quantitative IgG response will have high internal validity and require almost no assumptions [7]. Yet, the absence of global reference standards to translate arbitrary units of an assay into antibody titers can complicate comparisons across assays or studies; moreover, quantitative responses provide only indirect information about actual transmission in a population, which prevents direct comparisons between pathogens. Classifying responses as seronegative and seropositive to estimate seroprevalence and seroconversion rates resolves these limitations and extends inference to more epidemiologically interpretable parameters, such as force of infection.

The step of defining seropositivity cutoffs presents challenges for enteropathogens, but our findings provide guidance for future studies. Our results demonstrate that within-sample approaches to estimate seropositivity cutoffs are feasible for enteropathogens, but there are important considerations that must be addressed through design and analysis. Two-component mixture models fit the data and provided reasonable cutoff estimates only when restricted to an age range that included two clearly delineated subpopulations of seronegative and seropositive responses. For most pathogens studied, this required measurements among children <1 year old, an age range during which IgG responses still followed a bimodal distribution. The Tanzania

study enrolled children 1 to 9 years old, and it was impossible to characterize the distribution of unexposed IgG levels with a mixture model for any enteropathogens except *Giardia* and *E. histolytica* (Figure 1 - supplement 1). Even in the Haiti and Kenya cohorts, which measured children <1 year, the only reliable approach to estimate seropositivity cutoffs for the highest transmission pathogens like ETEC and *Campylobacter* was to estimate a distribution among presumed unexposed by identifying measurements among children who subsequently experienced a large increase in IgG (Figure 1, Figure 1 - supplement 3), a strategy only possible in a longitudinal design. High levels of agreement (>95%) across multiple cutoff approaches, consistent with findings from trachoma serology [38], supports a pragmatic approach that could include multiple strategies within the same study, depending on data availability and on the empirical distribution of IgG response. Measuring a sufficient number of young children before primary infection, preferably with longitudinal measurements, will help ensure that within-sample seropositivity cutoff estimation is possible.

The instantaneous seroconversion rate is considered a more direct measure of transmission than seroprevalence and is widely used as one measure of a pathogen's force of infection [34]. Pathogens with the fastest rising IgG and seroprevalence curves (e.g., ETEC, *Campylobacter*; Figure 3, Figure 4) also had the highest seroconversion rates measured prospectively over the study period. Moreover, average seroprevalence, equal to the area under the age-seroprevalence curve [7], was linearly related to seroconversion rates in both Haiti and Kenya (Figure 7). These results suggest that seroprevalence will likely be rank-preserving for comparing relative pathogen transmission if measured over an age range with increasing seroprevalence. Yet, differences in the slope of the relationship between seroprevalence and seroconversion rates across countries suggests that the mapping between epidemiologic measures is context dependent, and in this study was likely influenced by differences in age and rates of seroreversion across countries. Finally, our comparison of seroconversion rate estimates in the Kenya cohort derived from age-structured seroprevalence that ignored the longitudinal data structure showed that although the estimates were broadly comparable, cross-sectional methods were generally more variable, biased, or both compared with longitudinal estimates (Figure 6 - supplement 1). If enteropathogen seroincidence rates (force of infection) are a key parameter of interest, our results support the use of longitudinal designs that are more efficient and can estimate seroreversion rates, significant for many enteropathogens (Figure 6, Table 2); cross-sectional measurements technically contain no information about seroreversion (Supplementary Information File 4), which poses difficulties for enteropathogens.

Since many enteric infections are subclinical [2,3], differences between pathogens in seroprevalence or seroconversion rates would not necessarily translate to differences in disease burden. For example, about 60% of ETEC strains globally produce heat labile toxin (LT), but of those only 33% also produce heat stable toxin [39], which are the strains thought to contribute

more to severe, acute diarrheal illness [3]. Higher quantities of enteropathogen genetic material in stool has been shown to distinguish clinical disease from asymptomatic infection [40,41], but it is unlikely that serum IgG levels could be used in a similar way because of their potential role in protective immunity.

High correlation between *Salmonella* LPS Groups B and D, between ETEC LT β subunit and cholera toxin β subunit, and between norovirus GI and norovirus GII at the individual level (Figure 2) was consistent with cross-reactivity, and suggests that antibody-based measures of exposure will be less specific than molecular assays for these enteropathogens [40]. *Salmonella* LPS Groups B and D have antigenic overlap in their oligosaccharides [42], and ETEC LT β subunit and cholera toxin β subunit are known to be immunologically cross-reactive [43]. There was no known cholera transmission in the study populations so we assumed the elevated responses to the cholera toxin β subunit reflected exposure to LT-producing ETEC, but without confirmed measures of infection that remains an assumption. Norovirus GI and GII virus-like particles are antigenically different, but cross-reactivity between norovirus genogroups is possible [44]. In addition, a recent study from Uganda found high seroprevalence for both GI and GII norovirus suggesting repeated infections by viruses of the same or different genogroup which potentially may boost cross-reactive antibodies [45].

This study had limitations. First, when measurement timing is independent from known infections, measurements will typically fall at intermediate points in the IgG boosting and waning process, and will thus be more variable than if they were carefully timed to precede and follow infection. Second, the Haiti and Tanzania cohorts were not originally designed to assess enteropathogens, and antibody measurements were not paired with measures of clinical symptoms of diarrhea or with measures of patent infection in stool. In Kenya, stool testing for pathogens was limited and low prevalence prevented detailed comparisons between serological and molecular results: for example, only 14 children were positive by PCR to *Cryptosporidium* [26]. Third, without definitive measures of infection we could only infer differences in transmission through antibody-based measures of exposure. Analyses of *Plasmodium falciparum* [46,47] and dengue virus [48] illustrate how paired, longitudinal measurements of patent infection and antibody response enable richer characterizations of antibody dynamics and more comprehensive assessments of infectious disease transmission. High-resolution, longitudinal assessment of paired enteropathogen infection and antibody measurements among children could provide valuable, additional insights into the pathogen-specific antibody kinetics following primary- and secondary infections. Such studies could also assess whether approaches to estimate enteropathogen incidence from cross-sectional samples, which rely on estimates of antibody decay with time since infection [10,25], could be used among children in low-resource settings.

Conclusions

Among children in Haiti, Kenya, and Tanzania, antibody-based measures of enteropathogen exposure reflected high transmission with primary exposure to most pathogens occurring by age 1–2 years. For many enteropathogens, we observed a consistent shift in population-level IgG distributions from bimodal to unimodal between birth and age 3 years — a result of antibody boosting and waning from repeated exposure. This shift in distribution complicates the interpretation of seroprevalence among children beyond the window of primary exposure, and seroincidence rates must account for IgG boosting and waning above seropositivity cutoffs. Antibodies are a promising approach to measure population-level enteropathogen exposure, and seroepidemiologic measures of heterogeneity and transmission are central considerations for their use in trials or in serologic surveillance. Our findings show that for most enteropathogens, the ideal window to measure heterogeneity in antibody response closes by age two years in low-resource settings, and studies that plan to estimate seroincidence rates (force of infection) should favor longitudinal designs with multiple measurements in this early age window.

Materials and Methods

Ethics statement

In Haiti, the human subjects protocol was reviewed and approved by the Ethical Committee of St. Croix Hospital (Leogane, Haiti) and the institutional review board at the US Centers for Disease Control and Prevention (CDC). After listening to an overview of the study, individuals were asked for verbal consent to participate. Verbal consent was deemed appropriate by both review boards because of low literacy rates in the study population. With each longitudinal visit, the study team re-consented participants before specimen collection. Mothers provided consent for children under 7, and children 7 years and older provided additional verbal assent. In Kenya, the human subjects protocol was reviewed and approved by institutional review boards at the Kenya Medical Research Institute (KEMRI) and at the US CDC. Primary caretakers provided written informed consent for their infant child's participation in the trial and blood specimen collection and testing [26]. The original trial was registered at clinicaltrials.org (NCT01695304). In Tanzania, the human subjects protocol was reviewed and approved by the Institute for Medical Research Ethical Review Committee in Dar es Salaam, Tanzania and the institutional review board at the US CDC. Parents of enrolled children provided consent, and children 7 years and older also provided verbal assent before specimen collection.

Multiplex bead assays

Sera from the Haiti cohort study were diluted 1:400 and analyzed by multiplex bead assay as described in detail elsewhere [6,7,29]. Lipopolysaccharides (LPS) from Group D *Salmonella enterica* serotype Enteritidis, LPS from Group B *S. enterica* serotype Typhimurium, and recombinant heat labile toxin β subunit protein from enterotoxigenic *Escherichia coli* (ETEC LT β subunit) were purchased from Sigma Chemical (St. Louis, MO). Recombinant *Campylobacter* p18 and p39 antigens [49] were expressed and purified as previously described [14]. Virus-like particles from norovirus GI.4 and GII.4 New Orleans were purified from a recombinant baculovirus expression system [7,50]. These antigens were coupled to SeroMap (Luminex Corp, Austin, TX) beads in buffer containing 0.85% NaCl and 10 mM Na_2HPO_4 at pH 7.2 (PBS) using 120 micrograms for 1.25 x 10^7 beads using the methods described by Moss and colleagues [51]. Coupling conditions and externally defined cutoff values for the *Giardia*, *Cryptosporidium*, and *E. histolytica* antigens as well as for the *Schistosoma japonicum* glutathione-*S*-transferase (GST) negative control protein have been previously reported [6].

For the Kenya cohort study, an optimized bead coupling technique using less total protein was performed in buffer containing 0.85% NaCl and 25 mM 2-(N-morpholino)-ethanesulfonic acid at

pH 5.0. The β subunit protein from cholera toxin was purchased from Sigma Chemical. The GST negative control protein (15 μg), *Cryptosporidium* Cp17 (6.8 μg) and Cp23 (12.5 μg) proteins and the *Campylobacter* p39 protein (25 μg) were coupled to 1.25 x 10^7 beads using the indicated protein amounts [14]. The *Giardia*, *E. histolytica*, ETEC, cholera, and *Campylobacter* p18 proteins were coupled using 30 μg of protein per 1.25 x 10^7 beads. *Salmonella* LPS B and LPS D were coupled to the same number of beads using 60 μg and 120 μg, respectively. Blood spot elutions from the Kenya study were diluted to a final serum concentration of 1:400 (assuming 50% hematocrit) and analyzed by multiplex bead assay as described by Morris and colleagues [26].

For the Tanzania cohort study, the same conditions described in the Kenya cohort study were used to couple antigens from *Giardia*, *Cryptosporidium*, *E. histolytica*, ETEC β toxin subunit, cholera β toxin subunit, GST, and *Salmonella* LPS group B and LPS group D. *Campylobacter* p39 and p18 were both coupled at 25 μg per 1.25 x 10^7 beads in buffer containing 0.85% NaCl and 25 mM 2-(N-morpholino)-ethanesulfonic acid at pH 5.0. Dried blood spots were eluted in the casein-based buffer described previously [24] and samples were diluted to either 1) 1:400 serum dilution with 50 μl run per well for year 1 or 2) 1:320 serum dilution with 40 μl run per well for years 2-4. All samples were run in duplicate. The incubation steps, washes, and data collection methods used in the multiplex bead assay were performed as described previously [24]. All samples were run in duplicate, and the average median fluorescence intensity minus background (MFI-bg) value was recorded. The Tanzania study used different bead lots in year 1 and years 2-4; we confirmed that the use of different bead lots had no influence on the results (Supplementary Information File 1).

Antibody distributions and determination of seropositivity

In all analyses, we transformed IgG levels to the $\log_{10}$ scale because the distributions were highly skewed. Means on the log-transformed data represent geometric means. We summarized the distribution of $\log_{10}$ IgG response using kernel density smoothers. In the Tanzania and Haiti cohorts, where children were measured across a broad age range, we stratified IgG distributions by each year of age <3 years to examine age-dependent changes in the population distributions. To assess potential cross-reactivity between antigens, we estimated pairwise correlations between individual-level measurements in each cohort using a Spearman rank correlation [52] and visualized the relationship for each pairwise combination with locally weighted regression fits [53].

We compared three approaches to estimate seropositivity cutoffs. *Approach 1:* External known positive and negative specimens were used to determine seropositivity cutoffs for *Giardia* VSP-3 and VSP-5 antigens, *Cryptosporidium* Cp17 and Cp23 antigens, and *E. histolytica* LecA antigen.

Cutoffs were determined using ROC analysis as previously described [6,26] for all antigens except for LecA, VSP-3, and VSP-5 in Haiti; in these cases, the mean plus 3 standard deviations of 65 specimens from citizens of the USA with no history of foreign travel were used to estimate cutoffs [6]. *Approach 2:* We fit a 2-component, finite Gaussian mixture model [30] to the antibody distributions among children 0-1 years old, and estimated seropositivity cutoffs using the lower component's mean plus three standard deviations. The rationale for restricting the mixture model estimation in Haiti and Tanzania to children 0-1 years old was based on initial inspection of the age-stratified IgG distributions that revealed a shift from bimodal to unimodal distributions by age 3 (Figure 1), and to ensure that there was a sufficiently large fraction of unexposed children in the sample to more clearly estimate a distribution among seronegative children. *Approach 3:* In the longitudinal Haiti and Kenya cohorts we identified children <1 year old who presumably seroconverted, defined as an increase in MFI-bg values of +2 or more on the $\log_{10}$ scale. A sensitivity analysis showed that an increase of 2 on the $\log_{10}$ scale was conservative to identify seroconversion for most antibodies considered in this study; an increase of between 0.3 to 2.16 MFI-bg lead to optimal agreement with ROC-based and mixture model-based classifications in Kenya, and an increase of 0.92 to 2.41 led to optimal agreement across antigens and references in Haiti (Supplementary Information File 5). We then used the distribution of measurements before seroconversion to define the distribution of IgG values among the presumed unexposed. We used the mean $\log_{10}$ MFI-bg plus three standard deviations of the presumed unexposed distribution as a seropositivity cutoff. We summarized the proportion of observations that were in agreement between the three classification approaches, and estimated Cohen's Kappa [54]. Additional details and estimates of seropositivity cutoff agreement are reported in Supplementary Information File 2.

In analyses of seroprevalence and seroconversion, we classified measurements as seropositive using ROC-based cutoffs if available, and mixture model-based cutoffs otherwise. There were three exceptions. By age 1 year, a majority of children across the cohorts had IgG levels near the maximum of the assay's dynamic range for enterotoxigenic *Escherichia coli* heat labile toxin β subunit (ETEC LT β toxin) and *Vibrio cholerae* toxin β subunit. The absence of a sufficient number of unexposed children to ETEC LT β toxin, *V. cholerae*, and in some cases *Campylobacter* led mixture models either to not converge or to estimate unrealistically high seropositivity cutoffs beyond the range of quantifiable levels. For these pathogens, we used seropositivity cutoffs estimated from presumed unexposed measurements in the longitudinal Haiti and Kenya cohorts (approach 3, above). High levels of agreement between classifications (Supplementary Information File 2) meant results were insensitive to choice of approach in these cohorts.

Age-dependent antibody levels and seroprevalence curves

We estimated population mean IgG levels and seroprevalence by age using semiparametric cubic splines in a generalized additive model, specifying binomial errors for seroprevalence, and random effects for children or clusters in the case of repeated observations [55,56]. We also estimated the relationships by age using a stacked ensemble approach called "super learner" that included a broader and more flexible library of machine learning algorithms [7,57,58], and found similar fits to cubic splines. We estimated approximate, simultaneous 95% credible intervals around the curves using a parametric bootstrap from posterior estimates of the model parameter covariance matrix [59]. Supplementary Information File 6 includes additional details.

Seroincidence and force of infection

In the longitudinal cohorts in Kenya and Haiti, we estimated prospective seroconversion rates by dividing the number of children who seroconverted by the person-time at risk between measurements. We defined incident seroconversions and seroreversions as a change in IgG across a pathogen's seropositivity cutoff. In a secondary analysis aimed at capturing events that might occur above a pathogen's seropositivity cutoff, we defined incident boosting episodes as a ≥4-fold increase in IgG to a final level above a seropositivity cutoff, and incident waning episodes as ≥4-fold decrease in IgG from an initial level above a seropositivity cutoff. In the secondary definition, individuals were considered at risk for incident boosting episode if they were seronegative, if they experienced a ≥4-fold increase in IgG in their first measurement period, or if they experienced a ≥4-fold decrease in IgG in a preceding period (Haiti). We assumed that incident changes in serostatus were interval-censored at the midpoint between measurements. We estimated 95% confidence intervals for incidence rates with 2.5 and 97.5 percentiles of a nonparametric bootstrap [60] distribution that resampled children with replacement to account for repeated observations.

In Kenya, we estimated the proportion of children who seroconverted over a 6-7 month period by dividing the number of children who seroconverted by the number at risk at the beginning of the period. This analysis excluded 6 of 199 children measured prospectively whose measurements were separated by < 6 months (n=5) or by 8 months (n=1). We estimated 95% confidence intervals for incidence proportions using an exact binomial distribution.

We considered three additional estimators of the seroconversion rate in the Kenyan cohort using methods for "current status" (cross-sectional) data to see if methods based on age-structured seroprevalence curves alone were consistent with prospectively estimated rates. Current status approaches included: semiparametric cubic splines that estimate age-specific rates [34], which

were averaged over the empirical age distribution, an exponential proportional hazards model that assumed a constant rate over ages [37], and a reversible catalytic model with constant seroconversion and seroreversion rate parameters [35]. The first two methods assume no seroreversion. Supplementary Information File 4 includes additional details.

Data availability and replication files

Analyses were not pre-specified since our objectives were descriptive and methodologic. Analyses were conducted in R version 3.5.0. We have reported all analyses conducted. All data and computational notebooks used to complete the analyses are available through GitHub and the Open Science Framework (osf.io/r4av7).

Acknowledgements

The authors are grateful to Drs. Ciara E. O'Reilly and Jennifer L. Murphy for oversight and field assistance with dried blood spot collection in the Kenya study, and to Katy Hamlin for assistance with the analysis of specimens in the Haiti study.

Competing Interests

The authors declare they have no competing interests. Use of trade names is for identification only and does not imply endorsement by the Public Health Service or by the U.S. Department of Health and Human Services. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding

This work was supported by National Institute of Allergy and Infectious Disease (grant K01-AI119180) and by the Bill & Melinda Gates Foundation (grant OPP1022543). Studies in Haiti were supported by Centers for Disease Control, the National Institutes of Health and the United Nations Development Programme/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases (grants #920528 and #940441). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary Information Files

Supplementary Information File 1.
No effect of intervention or bead lot on enteropathogen antibody response in Kenya and Tanzania (osf.io/vdp9a).

Supplementary Information File 2.
Classification agreement between different seropositivity cutoff approaches (osf.io/7x6sw).

Supplementary Information File 3.
Joint distributions of antibody response (osf.io/wchzq).

Supplementary Information File 4.
Estimation of seroconversion rates assuming cross-sectional ("current status") data in Kenya (osf.io/eb4xt).

Supplementary Information File 5.
Sensitivity analysis of change in IgG used to identify presumed unexposed measurements in Haiti and Kenya (osf.io/gq9px).

Supplementary Information File 6.
Estimation of age-dependent means and seroprevalence using multiple approaches (osf.io/azsbf).

References

1. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390: 1260–1344.

2. 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. Lancet. 2016;388: 1291–1301.

3. Platts-Mills JA, Liu J, Rogawski ET, Kabir F, Lertsethtakarn P, Siguas M, et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. Lancet Glob Health. 2018; doi:10.1016/S2214-109X(18)30349-8

4. Griffin SM, Chen IM, Fout GS, Wade TJ, Egorov AI. Development of a multiplex microsphere immunoassay for the quantitation of salivary antibody responses to selected waterborne pathogens. J Immunol Methods. 2011;364: 83–93.

5. Exum NG, Pisanic N, Granger DA, Schwab KJ, Detrick B, Kosek M, et al. Use of Pathogen-Specific Antibody Biomarkers to Estimate Waterborne Infections in Population-Based Settings. Current environmental health reports. Springer; 2016;3: 322–334.

6. Moss DM, Priest JW, Hamlin K, Derado G, Herbein J, Petri WA Jr, et al. Longitudinal evaluation of enteric protozoa in Haitian children by stool exam and multiplex serologic assay. Am J Trop Med Hyg. 2014;90: 653–660.

7. Arnold BF, van der Laan MJ, Hubbard AE, Steel C, Kubofcik J, Hamlin KL, et al. Measuring changes in transmission of neglected tropical diseases, malaria, and enteric pathogens from quantitative antibody levels. PLoS Negl Trop Dis. 2017;11: e0005616.

8. Simonsen J, Strid MA, Mølbak K, Krogfelt KA, Linneberg A, Teunis P. Sero-epidemiology as a tool to study the incidence of Salmonella infections in humans. Epidemiol Infect. 2008;136: 895–902.

9. Falkenhorst G, Simonsen J, Ceper TH, van Pelt W, de Valk H, Sadkowska-Todys M, et al. Serological cross-sectional studies on salmonella incidence in eight European countries: no correlation with incidence of reported cases. BMC Public Health. 2012;12: 523.

10. Teunis PFM, van Eijkeren JCH, Ang CW, van Duynhoven YTHP, Simonsen JB, Strid MA, et al. Biomarker dynamics: estimating infection rates from serological data. Stat Med. 2012;31: 2240–2248.

11. Teunis PFM, Falkenhorst G, Ang CW, Strid MA, De Valk H, Sadkowska-Todys M, et al. Campylobacter seroconversion rates in selected countries in the European Union. Epidemiol Infect. 2013;141: 2051–2057.

12. Watson CH, Baker S, Lau CL, Rawalai K, Taufa M, Coriakula J, et al. A cross-sectional

seroepidemiological survey of typhoid fever in Fiji. PLoS Negl Trop Dis. 2017;11: e0005786.

13. Crump JA, Mendoza CE, Priest JW, Glass RI, Monroe SS, Dauphin LA, et al. Comparing serologic response against enteric pathogens with reported diarrhea to assess the impact of improved household drinking water quality. Am J Trop Med Hyg. 2007;77: 136–141.

14. Zambrano LD, Priest JW, Ivan E, Rusine J, Nagel C, Kirby M, et al. Use of Serologic Responses against Enteropathogens to Assess the Impact of a Point-of-Use Water Filter: A Randomized Controlled Trial in Western Province, Rwanda. Am J Trop Med Hyg. 2017;97: 876–887.

15. Chard AN, Trinies V, Moss DM, Chang HH, Doumbia S, Lammie PJ, et al. The impact of school water, sanitation, and hygiene improvements on infectious disease using serum antibody detection. PLoS Negl Trop Dis. 2018;12: e0006418.

16. Vargas P, Krolewiecki AJ, Echazú A, Juarez M, Cajal P, Gil JF, et al. Serologic Monitoring of Public Health Interventions against Strongyloides stercoralis. Am J Trop Med Hyg. 2017;97: 166–172.

17. Mosites E, Miernyk K, Priest JW, Bruden D, Hurlburt D, Parkinson A, et al. Giardia and Cryptosporidium antibody prevalence and correlates of exposure among Alaska residents, 2007-2008. Epidemiol Infect. 2018; 1–7.

18. Wade TJ, Augustine SAJ, Griffin SM, Sams EA, Oshima KH, Egorov AI, et al. Asymptomatic norovirus infection associated with swimming at a tropical beach: A prospective cohort study. PLoS One. 2018;13: e0195056.

19. Egorov AI, Griffin SM, Ward HD, Reilly K, Fout GS, Wade TJ. Application of a salivary immunoassay in a prospective community study of waterborne infections. Water Res. 2018;142: 289–300.

20. Strid MA, Engberg J, Larsen LB, Begtrup K, Mølbak K, Krogfelt KA. Antibody responses to Campylobacter infections determined by an enzyme-linked immunosorbent assay: 2-year follow-up study of 210 patients. Clin Diagn Lab Immunol. 2001;8: 314–319.

21. Priest JW, Li A, Khan M, Arrowood MJ, Lammie PJ, Ong CS, et al. Enzyme immunoassay detection of antigen-specific immunoglobulin g antibodies in longitudinal serum samples from patients with cryptosporidiosis. Clin Diagn Lab Immunol. 2001;8: 415–423.

22. Priest JW, Moss DM, Visvesvara GS, Jones CC, Li A, Isaac-Renton JL. Multiplex assay detection of immunoglobulin G antibodies that recognize Giardia intestinalis and Cryptosporidium parvum antigens. Clin Vaccine Immunol. 2010;17: 1695–1707.

23. Falkenhorst G, Ceper TH, Strid MA, Mølbak K, Krogfelt KA. Serological follow-up after non-typhoid salmonella infection in humans using a mixed lipopolysaccharide ELISA. Int J Med Microbiol. 2013;303: 533–538.

24. Hjøllo T, Bratland E, Steinsland H, Radunovic M, Langeland N, Hanevik K. Longitudinal cohort study of serum antibody responses towards Giardia lamblia variant-specific surface proteins in a non-endemic area. Exp Parasitol. 2018;191: 66–72.

25. Simonsen J, Mølbak K, Falkenhorst G, Krogfelt KA, Linneberg A, Teunis PFM. Estimation of incidences of infectious diseases based on antibody measurements. Stat Med. 2009;28: 1882–1895.

26. Morris JF, Murphy J, Fagerli K, Schneeberger C, Jaron P, Moke F, et al. A Randomized Controlled Trial to Assess the Impact of Ceramic Water Filters on Prevention of Diarrhea and Cryptosporidiosis in Infants and Young Children-Western Kenya, 2013. Am J Trop Med Hyg. 2018;98: 1260–1268.

27. Martin D, Wiegand R, Goodhew B, Lammie P, Mkocha H, Kasubi M. Impact of Ivermectin Mass Drug Administration for Lymphatic Filariasis on Scabies in Eight Villages in Kongwa District, Tanzania. Am J Trop Med Hyg. 2018;99: 937–939.

28. Lammie PJ, Reiss MD, Dimock KA, Streit TG, Roberts JM, Eberhard ML. Longitudinal analysis of the development of filarial infection and antifilarial immunity in a cohort of Haitian children. Am J Trop Med Hyg. 1998;59: 217–221.

29. Hamlin KL, Moss DM, Priest JW, Roberts J, Kubofcik J, Gass K, et al. Longitudinal monitoring of the development of antifilarial antibodies and acquisition of Wuchereria bancrofti in a highly endemic area of Haiti. PLoS Negl Trop Dis. 2012;6: e1941.

30. Benaglia T, Chauveau D, Hunter D, Young D. mixtools: An R package for analyzing finite mixture models. J Stat Softw. 2009;32: 1–29.

31. Bernstein DI, Atmar RL, Lyon GM, Treanor JJ, Chen WH, Jiang X, et al. Norovirus vaccine against experimental human GII.4 virus illness: a challenge study in healthy adults. J Infect Dis. 2015;211: 870–878.

32. Jin C, Gibani MM, Moore M, Juel HB, Jones E, Meiring J, et al. Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. Lancet. 2017;390: 2472–2480.

33. Chakraborty S, Harro C, DeNearing B, Brubaker J, Connor S, Maier N, et al. Impact of lower challenge doses of enterotoxigenic Escherichia coli on clinical outcome, intestinal colonization and immune responses in adult volunteers. PLoS Negl Trop Dis. 2018;12: e0006442.

34. Hens N, Shkedy Z, Aerts M, Damme CFPV, Beutels P. Modeling Infectious Disease Parameters Based on Serological and Social Contact Data. Springer; 2012.

35. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends Parasitol. 2007;23: 575–582.

36. Katzelnick LC, Ben-Shachar R, Mercado JC, Rodriguez-Barraquer I, Elizondo D, Arguello S, et al. Dynamics and determinants of the force of infection of dengue virus from 1994 to 2015 in Managua, Nicaragua. Proc Natl Acad Sci U S A. 2018;115: 10762–10767.

37. Jewell NP, Laan MV. Generalizations of current status data with applications. Lifetime Data Anal. 1995;1: 101–109.

38. Migchelsen SJ, Martin DL, Southisombath K, Turyaguma P, Heggen A, Rubangakene PP, et al. Defining Seropositivity Thresholds for Use in Trachoma Elimination Studies. PLoS Negl Trop Dis.

2017;11: e0005230.

39. Isidean SD, Riddle MS, Savarino SJ, Porter CK. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. Vaccine. 2011;29: 6167–6178.

40. Liu J, Kabir F, Manneh J, Lertsethtakarn P, Begum S, Gratz J, et al. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. Lancet Infect Dis. 2014;14: 716–724.

41. Platts-Mills JA, Gratz J, Mduma E, Svensen E, Amour C, Liu J, et al. Association Between Stool Enteropathogen Quantity and Disease in Tanzanian Children Using TaqMan Array Cards: A Nested Case-Control Study. Am J Trop Med Hyg. 2014;90: 133–138.

42. Grimont PAD, Weill F-X, Others. Antigenic formulae of the Salmonella serovars. WHO collaborating centre for reference and research on Salmonella. Institut Pasteur Paris, France; 2007;9: 1–166.

43. Smith NW, Sack RB. Immunologic cross-reactions of enterotoxins from Escherichia coli and Vibrio cholerae. J Infect Dis. 1973;127: 164–170.

44. Tamminen K, Huhti L, Koho T, Lappalainen S, Hytönen VP, Vesikari T, et al. A comparison of immunogenicity of norovirus GII-4 virus-like particles and P-particles. Immunology. 2012;135: 89–99.

45. Thorne L, Nalwoga A, Mentzer AJ, de Rougemont A, Hosmillo M, Webb E, et al. The First Norovirus Longitudinal Seroepidemiological Study From Sub-Saharan Africa Reveals High Seroprevalence of Diverse Genotypes Associated With Host Susceptibility Factors. J Infect Dis. 2018;218: 716–725.

46. White MT, Griffin JT, Akpogheneta O, Conway DJ, Koram KA, Riley EM, et al. Dynamics of the antibody response to Plasmodium falciparum infection in African children. J Infect Dis. 2014;210: 1115–1122.

47. Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. Proc Natl Acad Sci U S A. 2015;112: E4438–47.

48. Salje H, Cummings DAT, Rodriguez-Barraquer I, Katzelnick LC, Lessler J, Klungthong C, et al. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. Nature. 2018;557: 719–723.

49. Schmidt-Ott R, Brass F, Scholz C, Werner C, Gross U. Improved serodiagnosis of Campylobacter jejuni infections using recombinant antigens. J Med Microbiol. 2005;54: 761–767.

50. Jiang X, Wang M, Graham DY, Estes MK. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J Virol. 1992;66: 6527–6532.

51. Moss DM, Priest JW, Boyd A, Weinkopff T, Kucerova Z, Beach MJ, et al. Multiplex bead assay for serum samples from children in Haiti enrolled in a drug study for the treatment of lymphatic filariasis. Am J Trop Med Hyg. 2011;85: 229–237.

52. Zar JH. Spearman rank correlation. Encyclopedia of Biostatistics. Wiley Online Library; 2005; Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a15150

53. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc. amstat.tandfonline.com; 1988;83: 596–610.

54. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. SAGE Publications Inc; 1960;20: 37–46.

55. Wood S. Generalized Additive Models: An Introduction with R. 2nd ed. Taylor & Francis; 2017.

56. Wood S. mgcv:Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation [Internet]. University of Bath; 2012. Available: http://opus.bath.ac.uk/32507/

57. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007;6: Article25.

58. Polley E, Erin CL, Kennedy C, Lendle S, van der Laan M. SuperLearner: Super Learner Prediction [Internet]. 2018. Available: https://github.com/ecpolley/SuperLearner

59. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. Cambridge University Press; 2003.

60. Wasserman L. All of statistics: a concise course in statistical inference. Springer; 2004.

Table 1. Number of children and samples tested, and estimated seropositivity cutoffs by country and IgG antigen included in the seroepidemiologic analyses.

			Seropositivity cutoff, log10 IgG (MFI-bg) *		
	N children	N samples	External Reference	Mixture Model	Presumed Unexposed
Leogane, Haiti					
Giardia VSP-3	142	771	2.42	1.64	2.11
Giardia VSP-5	142	771	2.31	1.46	1.88
Cryptosporidium Cp17	142	771	2.26	2.00	2.58
Cryptosporidium Cp23	142	771	2.70	2.75	2.57
E. histolytica LecA	142	771	2.48	2.30	1.93
Salmonella LPS group B	142	771		1.60	1.37
Salmonella LPS group D	142	771		1.48	2.48
ETEC LT β subunit	142	771			2.86
Norovirus GI.4	142	771		2.51	2.09
Norovirus GII.4.NO	142	771		2.04	2.24
Asembo, Kenya					
Giardia VSP-3	240	439	2.81	1.60	1.69
Giardia VSP-5	240	439	2.65	1.81	1.68
Cryptosporidium Cp17	240	439	2.63	2.58	2.37
Cryptosporidium Cp23	240	439	3.14	3.27	2.36
E. histolytica LecA	240	439		1.89	
Salmonella LPS group B	240	439		1.38	
Salmonella LPS group D	240	439		1.42	
ETEC LT β subunit	240	439			2.80
Cholera toxin β subunit	240	439			2.96
Campylobacter p18	240	439			2.11
Campylobacter p39	240	439		2.62	2.57
Kongwa, Tanzania					
Giardia VSP-3	4989	4989	2.23	2.04	
Giardia VSP-5	4989	4989	2.15	2.27	
Cryptosporidium Cp17	4989	4989	2.26		
Cryptosporidium Cp23	4989	4989	2.58		
E. histolytica LecA	4989	4989	1.97	2.50	
Salmonella LPS group B†	902	902			
Salmonella LPS group D†	902	902			
ETEC LT β subunit	4989	4989			
Cholera toxin β subunit‡	4087	4087			
Campylobacter p18†	902	902			
Campylobacter p39†	902	902			

* Seropositivity cutoffs determined using external reference samples (typically ROC curves except for Giardia and E. hystolitica in Haiti), finite Gaussian mixture models, or distribution among the presumed unexposed (see Methods for details). External reference cutoffs reported from years (2013-2015) in Tanzania, estimated among 4,087 samples. Cutoff values are missing if they could not be estimated in each method; cutoff values based on the presumed unexposed required longitudinal measurements within individual children and therefore could not be estimated for any antigen in the repeated cross-sectional design in Tanzania.

† Measured only in year 1 of the study (2012)

‡ Measured only in years 2-4 of the study (2013-2015)

Table 2. Seroincidence rates per child-year among children ages 0-11 years in Leogane, Haiti, 1990-1999. Rates estimated using two definitions of incident changes in serostatus, one based on seroconversion and a second based on 4-fold change in IgG levels. Ratio of incident cases and seroincidence rates divides the second definition by the first.

	Seropositivity cutoff *			4 Fold Change in IgG levels †				
Pathogen	**Child-years**	**Incident cases**	**Rate (95% CI)**	**Child-years**	**Incident cases**	**Rate (95% CI)**	**Ratio of cases**	**Ratio of rates**
Seroconversion / boosting								
Giardia VSP-3 or VSP-5	269.6	108	0.40 (0.34, 0.48)	275.9	118	0.43 (0.34, 0.54)	1.1	1.1
Cryptosporidium Cp17 or Cp23	109.3	70	0.64 (0.54, 0.77)	236.6	190	0.80 (0.70, 0.92)	2.7	1.3
E. histolytica LecA	283.7	97	0.34 (0.28, 0.42)	299.9	120	0.40 (0.32, 0.50)	1.2	1.2
Salmonella LPS groups B or D	132.1	75	0.57 (0.47, 0.68)	224.4	140	0.62 (0.51, 0.75)	1.9	1.1
ETEC LT β subunit	9.7	11	1.13 (0.75, 1.82)	32.1	33	1.03 (0.72, 1.53)	3.0	0.9
Norovirus GI.4	213.0	80	0.38 (0.30, 0.47)	257.0	118	0.46 (0.36, 0.57)	1.5	1.2
Norovirus GII.4.NO	105.8	67	0.63 (0.51, 0.80)	148.5	109	0.73 (0.58, 0.94)	1.6	1.2
Seroreversion / waning								
Giardia VSP-3 or VSP-5	441.6	91	0.21 (0.17, 0.25)	291.9	126	0.43 (0.35, 0.54)	1.4	2.1
Cryptosporidium Cp17 or Cp23	586.1	29	0.05 (0.03, 0.07)	276.9	166	0.60 (0.51, 0.70)	5.7	12.1
E. histolytica LecA	395.2	43	0.11 (0.08, 0.15)	307.8	70	0.23 (0.17, 0.29)	1.6	2.1
Salmonella LPS groups B or D	544.3	25	0.05 (0.03, 0.07)	349.1	86	0.25 (0.19, 0.30)	3.4	5.4
ETEC LT β subunit	702.1	2	0.00 (0.00, 0.01)	649.7	22	0.03 (0.02, 0.05)	11.0	11.9
Norovirus GI.4	464.9	28	0.06 (0.03, 0.09)	359.9	58	0.16 (0.11, 0.22)	2.1	2.7
Norovirus GII.4.NO	574.3	19	0.03 (0.02, 0.05)	476.9	40	0.08 (0.06, 0.11)	2.1	2.5

* Incident changes in serostatus defined by crossing seropositivity cutoffs.

† Incident changes in serostatus defined by a 4-fold increase or decrease in IgG levels (MFI-bg), with incident boosting episodes restricted to changes that ended above the seropositivity cutoff and incident waning episodes restricted to changes that started from above the seropositivity cutoff.

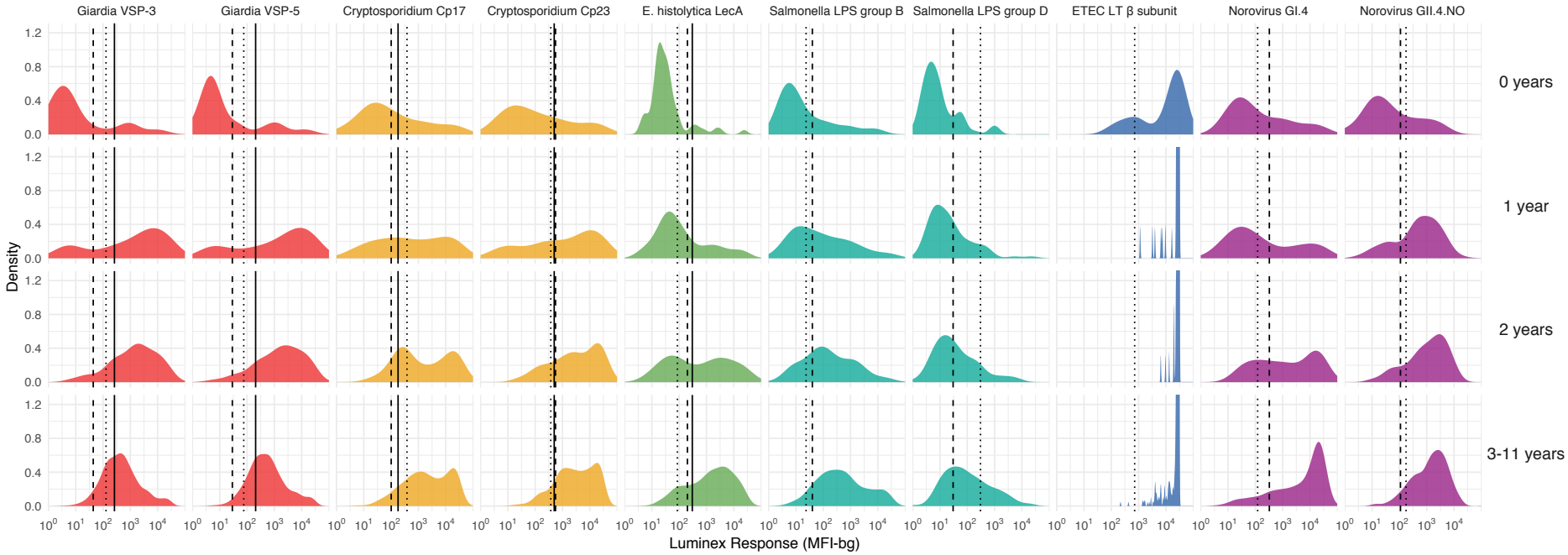

Figure 1. Age-stratified, IgG distributions among a longitudinal cohort of 142 children ages birth to 11 years in Leogane, Haiti, 1990–1999. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform in 771 specimens. Vertical lines mark seropositivity cutoffs are based on ROC analyses (solid), finite Gaussian mixture models (heavy dash), or distribution among presumed unexposed (light dash). Mixture models failed to converge for ETEC heat labile toxin β subunit. Created with notebook (`https://osf.io/dk54y`) and data (`https://osf.io/3nv98`). Supplement 1 shows similar distributions from the Tanzania study. Supplement 2 contrasts *Giardia* VSP-3 distributions with *C. trachomatis* pgp3 distributions. Supplement 3 shows distributions from the Kenyan cohort.

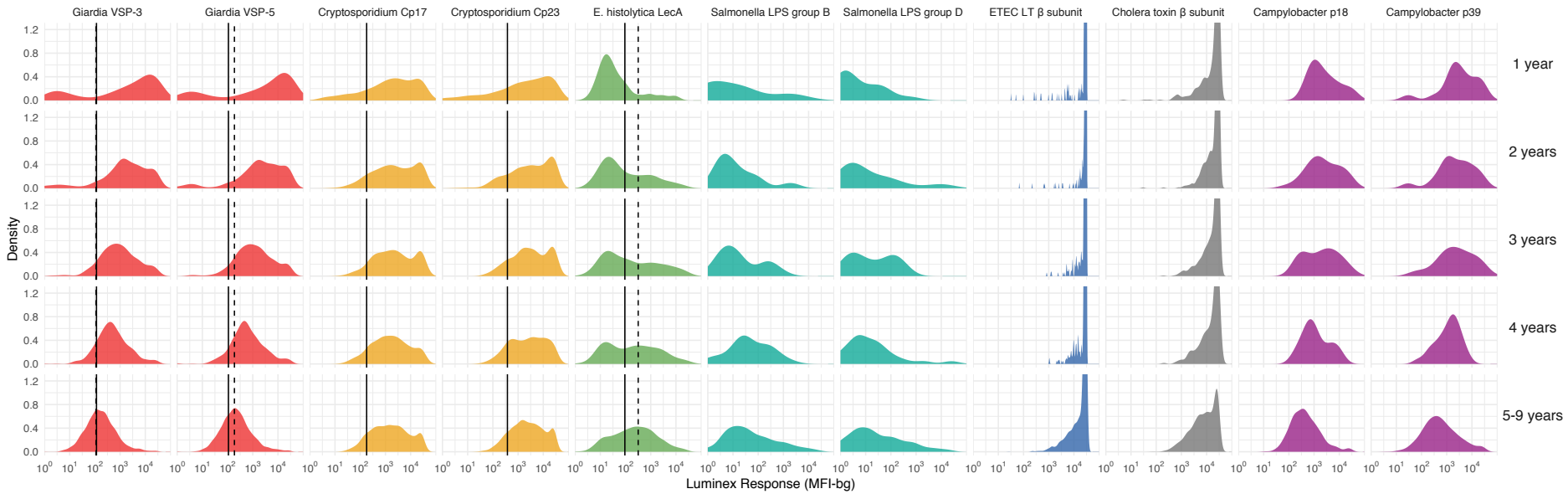

Figure 1 - supplement 1. Age-stratified, IgG distributions among 4,989 children ages 1 to 9 years old in Kongwa, Tanzania, 2012–2015. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. Vertical lines mark seropositivity cutoffs based on ROC analyses (solid) and finite Gaussian mixture models (heavy dash). Mixture models failed to converge for all antibody distributions except for *Giardia sp.* and *E. histolytica*. Created with notebook (`https://osf.io/dt9zu`) and data (`https://osf.io/kv4d3`).

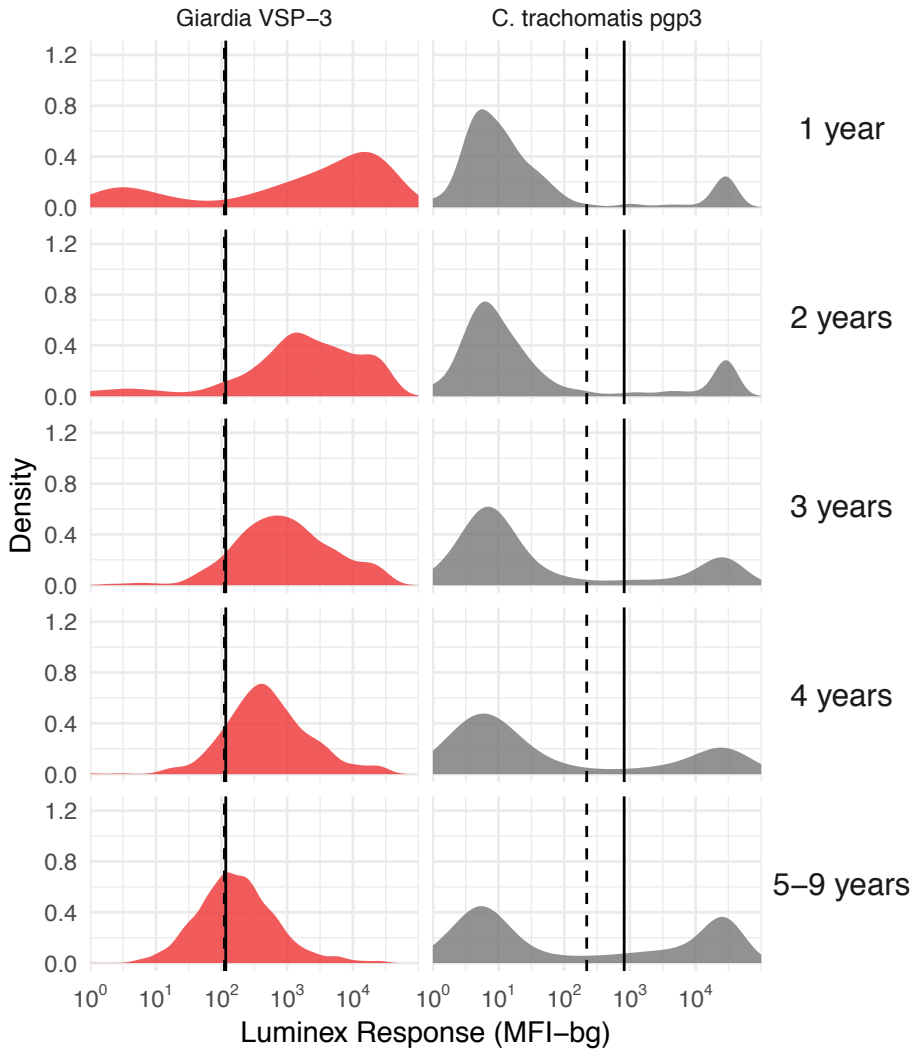

Figure 1 - supplement 2. Contrasting age-dependent changes in distributions of IgG levels to *Giardia sp.* VSP-3 and *C. trachomatis* pgp3 antigens among 5,057 children ages 1 to 9 years old in Kongwa, Tanzania, 2012–2015. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. Vertical lines mark seropositivity cutoffs based on ROC analyses (solid) and finite Gaussian mixture models (dash). Created with notebook (`https://osf.io/dt9zu`) and data (`https://osf.io/kv4d3`).

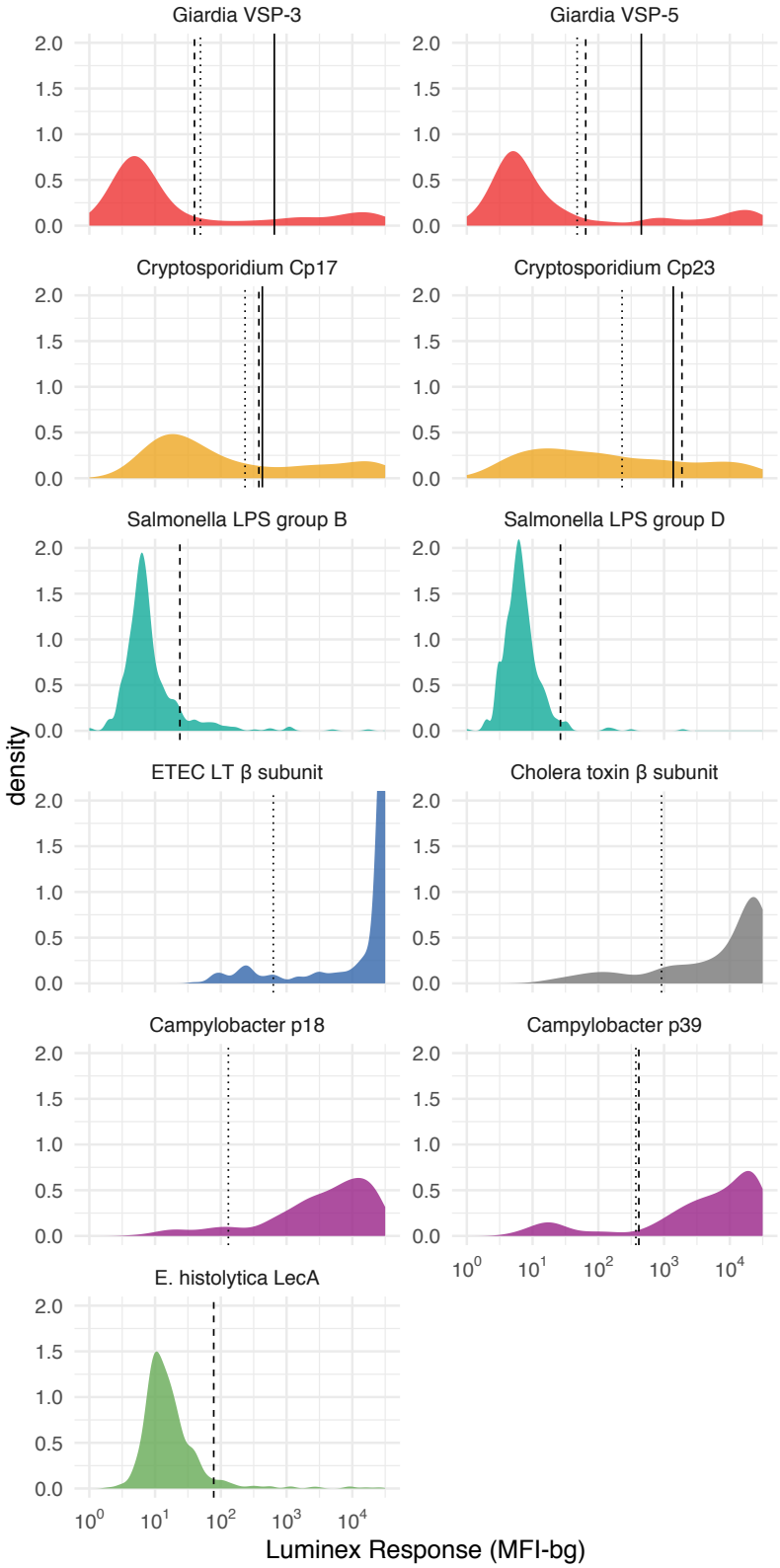

FIgure 1 - supplement 3. IgG distributions among children ages 4 to 17 months old in Asembo, Kenya, 2013. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. N=439 samples from 240 children. Vertical lines mark seropositivity cutoffs based on ROC analyses (solid), finite Gaussian mixture models (heavy dash), or distribution among presumed unexposed (light dash). Mixture models failed to converge for ETEC heat labile toxin β subunit and *Campylobacter spp.* p18. Created with notebook (`https://osf.io/456jp`) and data (`https://osf.io/2q7zg`).

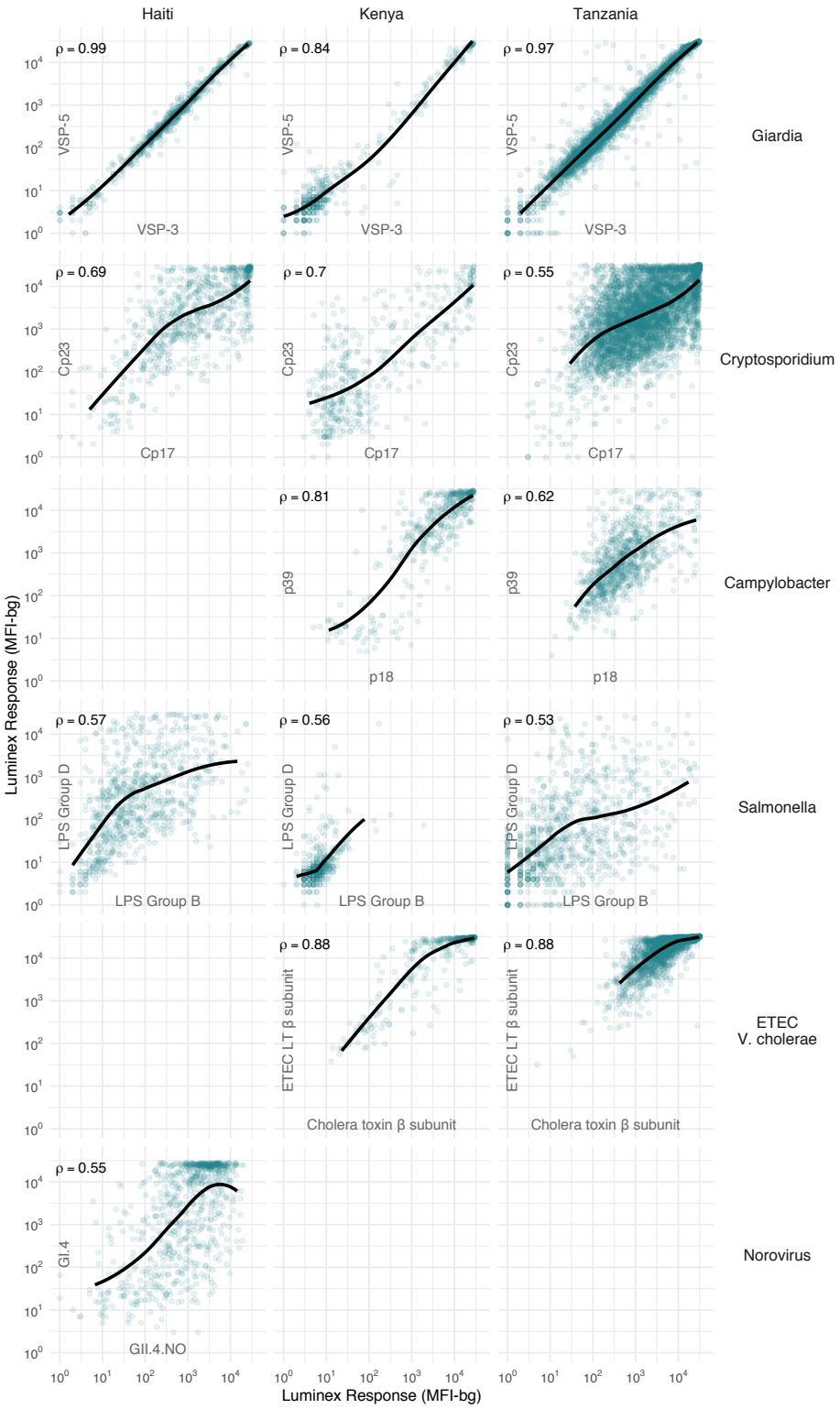

Figure 2. Joint distributions of select enteric pathogen antibody responses among children in three cohorts from Haiti, Kenya, and Tanzania. Each panel includes Spearman rank correlations (ρ) and locally weighted regression smoothers with default parameters, trimmed to 95% of the data to avoid edge effects. Antibody response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. Empty panels indicate that the antibodies were not measured in that cohort. Supplementary Information File 3 includes all pairwise comparisons. Created with notebook (`https://osf.io/hv9ce`) and data (`https://osf.io/3nv98`, `https://osf.io/2q7zg`, `https://osf.io/kv4d3`).

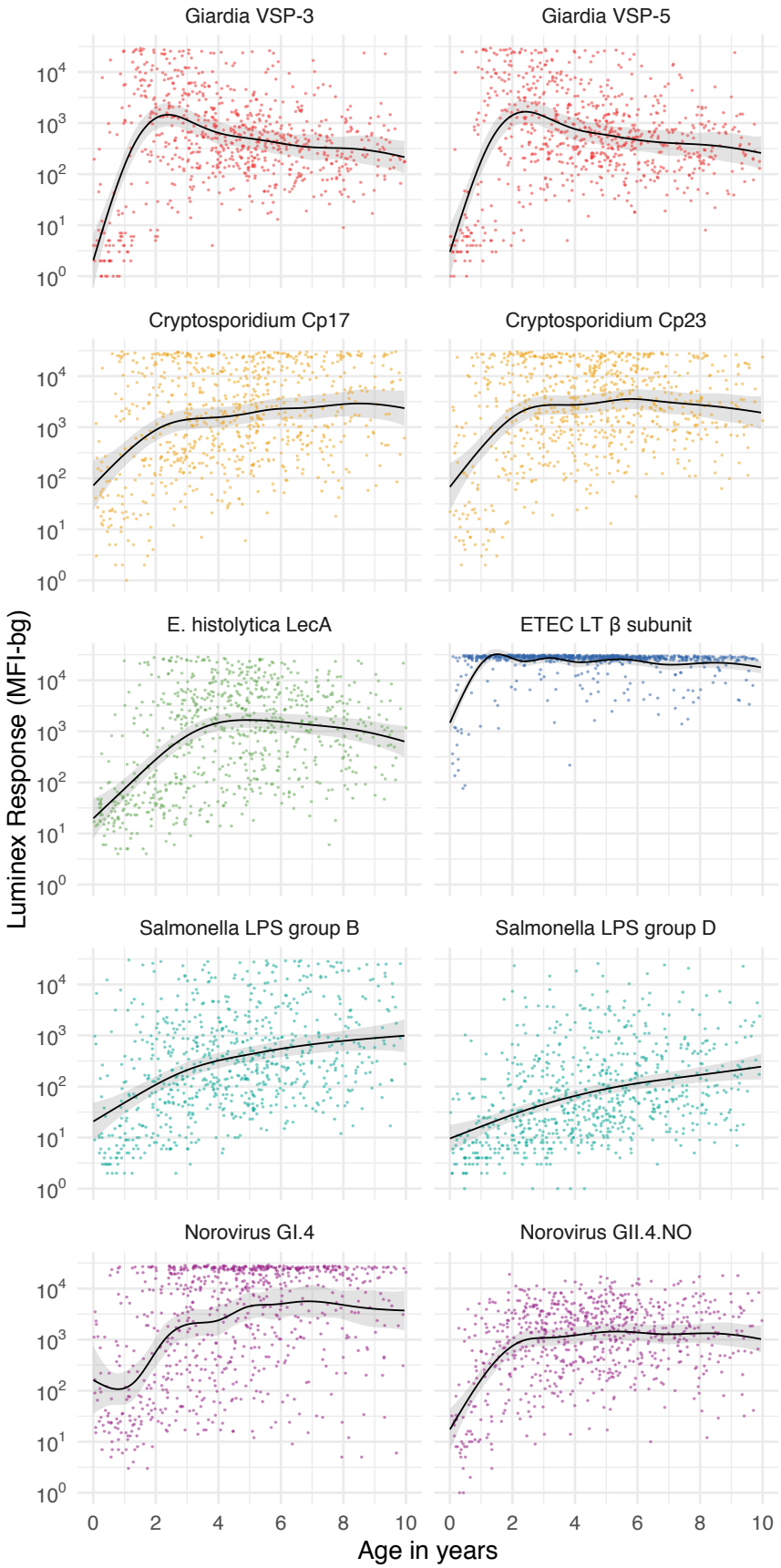
Giardia VSP-3
Giardia VSP-5
Cryptosporidium Cp17
Cryptosporidium Cp23
E. histolytica LecA
ETEC LT β subunit
Salmonella LPS group B
Salmonella LPS group D
Norovirus GI.4
Norovirus GII.4.NO
Luminex Response (MFI-bg)
Age in years
10^0
10^1
10^2
10^3
10^4
0
2
4
6
8
10

Figure 3. Age dependent mean curves fit with cubic splines among children ages birth to 10 years in Leogane, Haiti 1990-1999. Shaded bands are approximate, simultaneous 95% credible intervals. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform (N=771 measurements from 142 children). Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/3nv98`). Data for some antigens measured among children <5 years previously published [7]. Supplement 1 includes similar curves from Tanzania; Supplement 2 includes similar curves from Kenya.

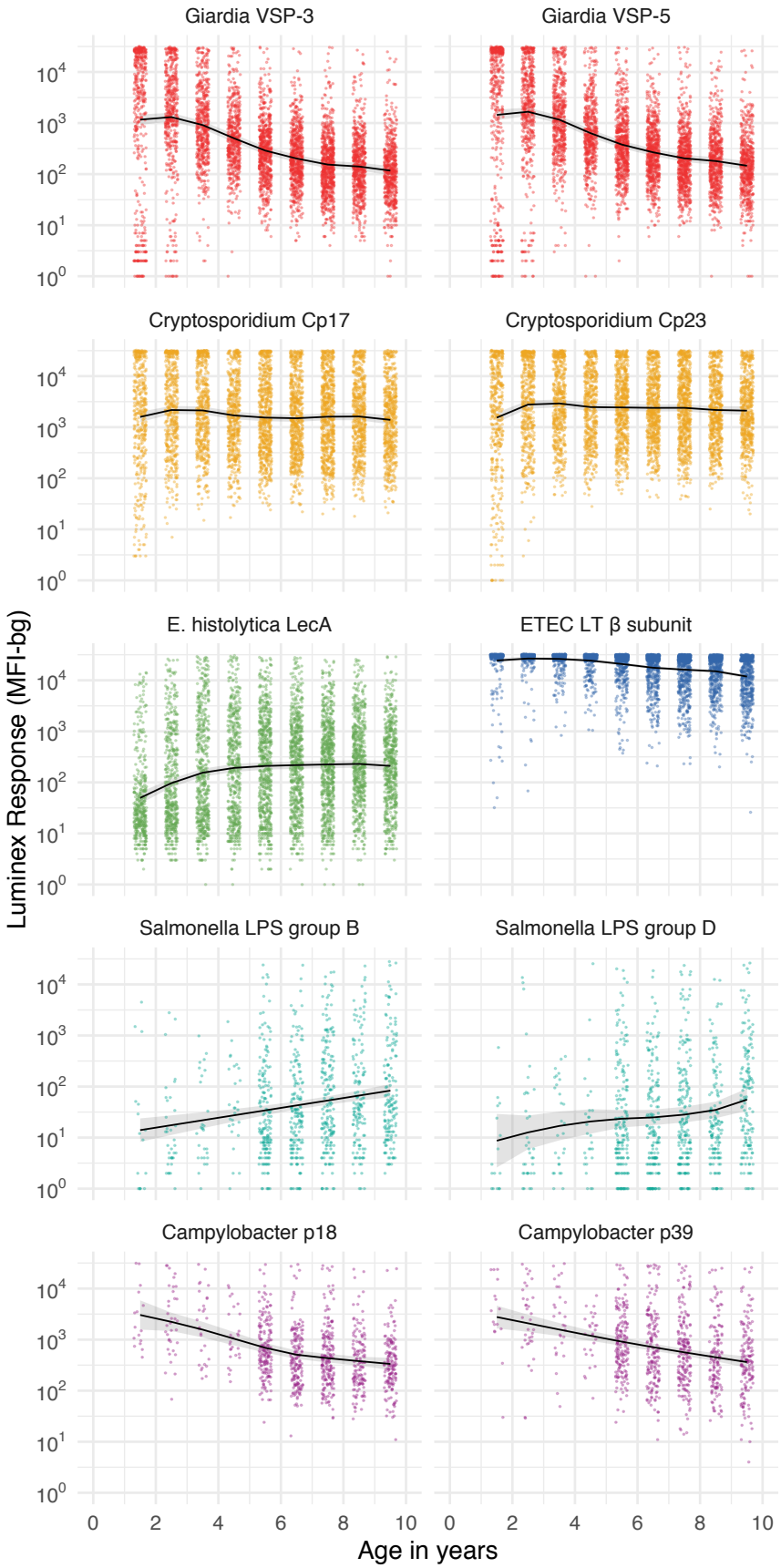
Giardia VSP-3
Giardia VSP-5
Cryptosporidium Cp17
Cryptosporidium Cp23
E. histolytica LecA
ETEC LT β subunit
Salmonella LPS group B
Salmonella LPS group D
Campylobacter p18
Campylobacter p39
Luminex Response (MFI-bg)
10^4
10^3
10^2
10^1
10^0
0
2
4
6
8
10
Age in years

Figure 3 - supplement 1. Age dependent mean curves fit with cubic splines among children ages 1 to 9 years in Kongwa, Tanzania, 2012-2015. Shaded bands are approximate, simultaneous 95% credible intervals. Ages were measured in years completed (rounded) so points in the figure are jittered. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform in 4,989 specimens. *Salmonella spp.* and *Campylobacter spp.* antigens were only included in the multiplex in 2012 (N=902 specimens). Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/kv4d3`).

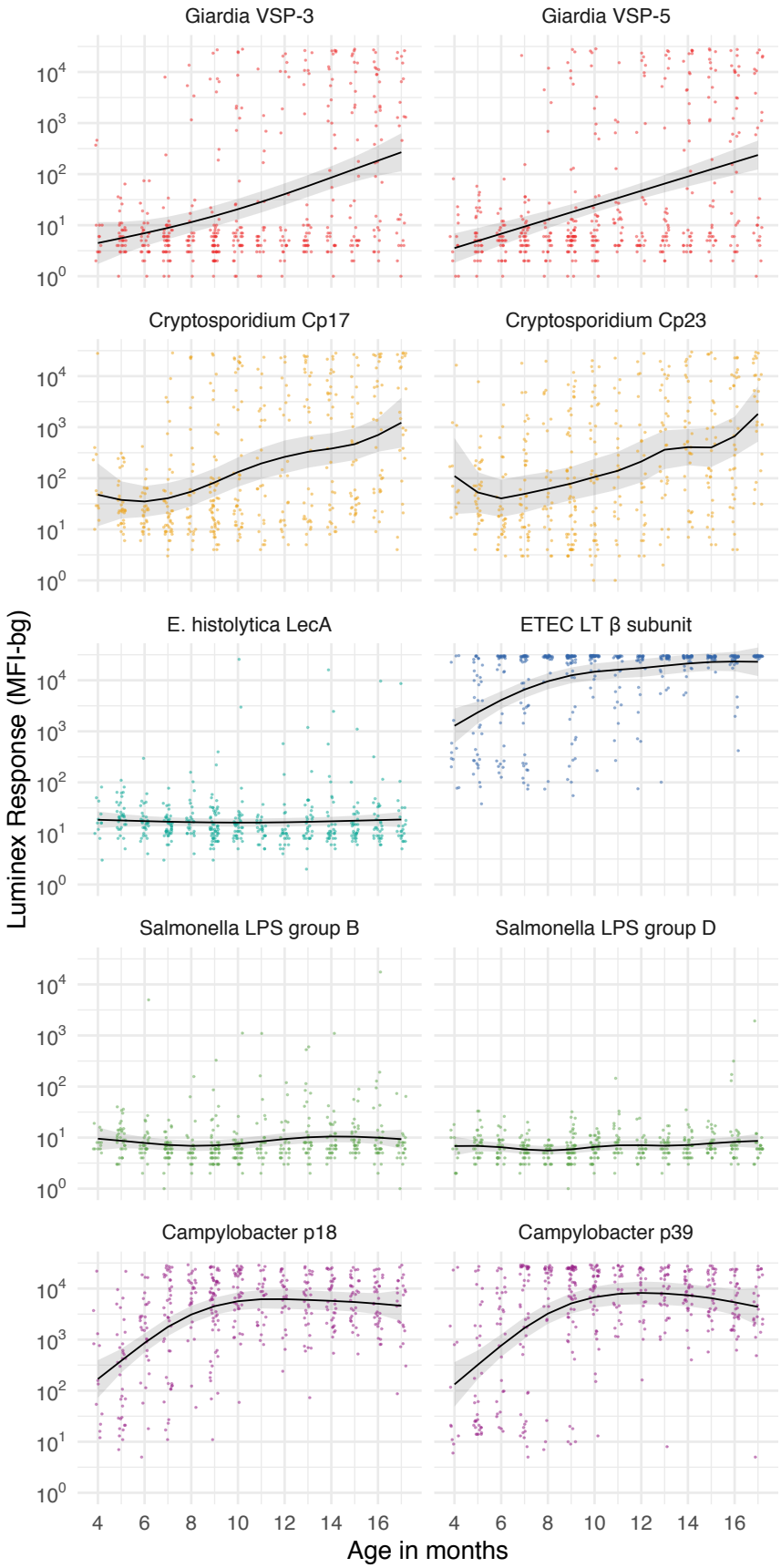
Giardia VSP-3
Giardia VSP-5
Cryptosporidium Cp17
Cryptosporidium Cp23
E. histolytica LecA
ETEC LT β subunit
Salmonella LPS group B
Salmonella LPS group D
Campylobacter p18
Campylobacter p39
Luminex Response (MFI-bg)
10^0
10^1
10^2
10^3
10^4
4
6
8
10
12
14
16
Age in months

Figure 3 - supplement 2. Age dependent mean curves fit with cubic splines among children ages 4 to 17 months in Asembo, Kenya, 2013. Shaded bands are approximate, simultaneous 95% credible intervals. Ages were measured in months completed (rounded) so points in the figure are jittered. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform (N=439 measurements from 240 children). Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/2q7zg`).

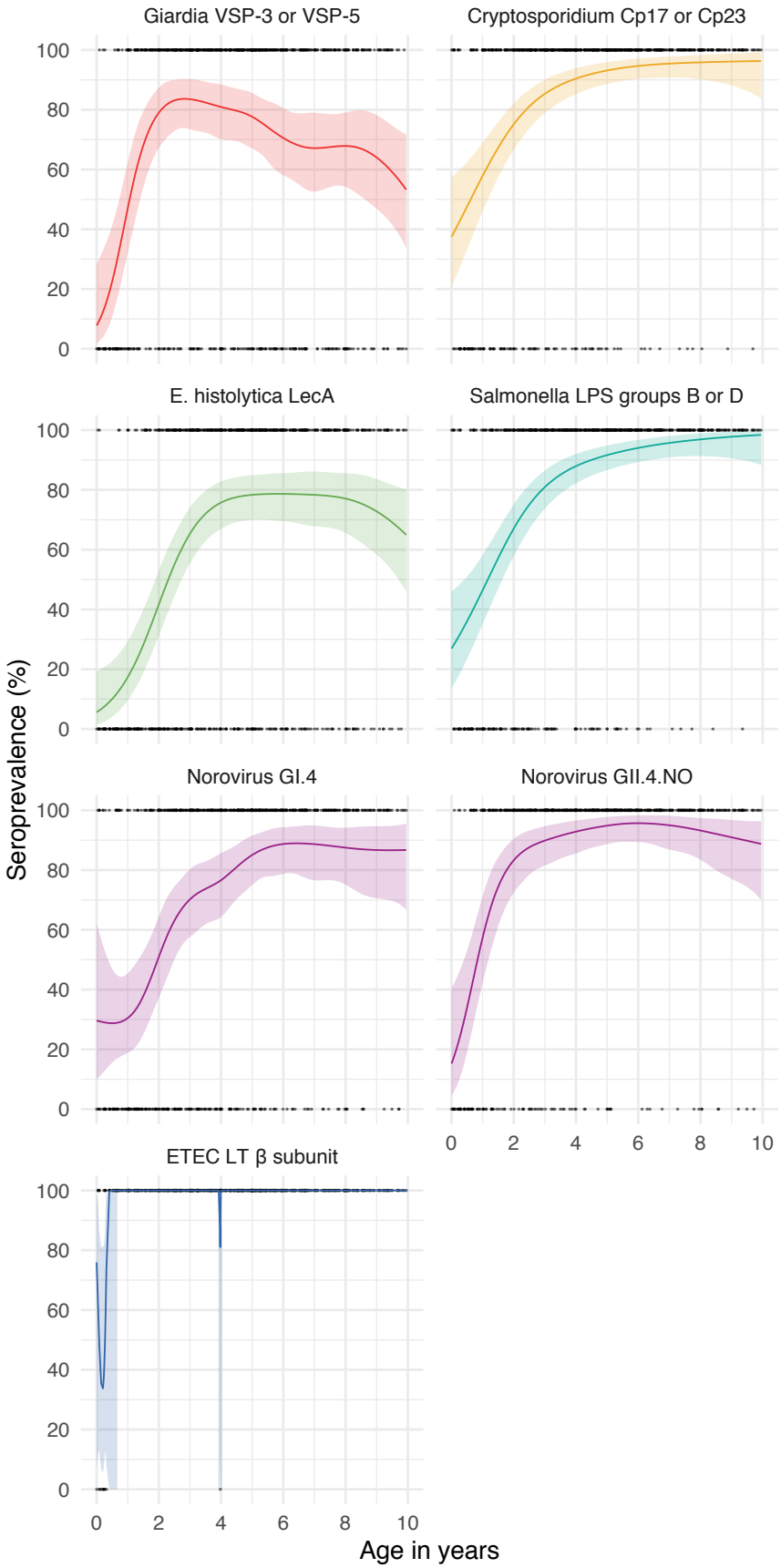
Giardia VSP-3 or VSP-5
Cryptosporidium Cp17 or Cp23
E. histolytica LecA
Salmonella LPS groups B or D
Norovirus GI.4
Norovirus GII.4.NO
ETEC LT β subunit
Seroprevalence (%)
Age in years
0
20
40
60
80
100
0
2
4
6
8
10

Figure 4. Age dependent seroprevalence curves fit with cubic splines among children ages birth to 10 years in Leogane, Haiti, 1990-1999. Shaded bands are approximate, simultaneous 95% credible intervals. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform (N=771 measurements from 142 children). Points at 0 and 100 mark individual outcome measurements. Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/3nv98`). Supplement 1 includes similar curves from Kenya; Supplement 2 includes similar curves from Tanzania.

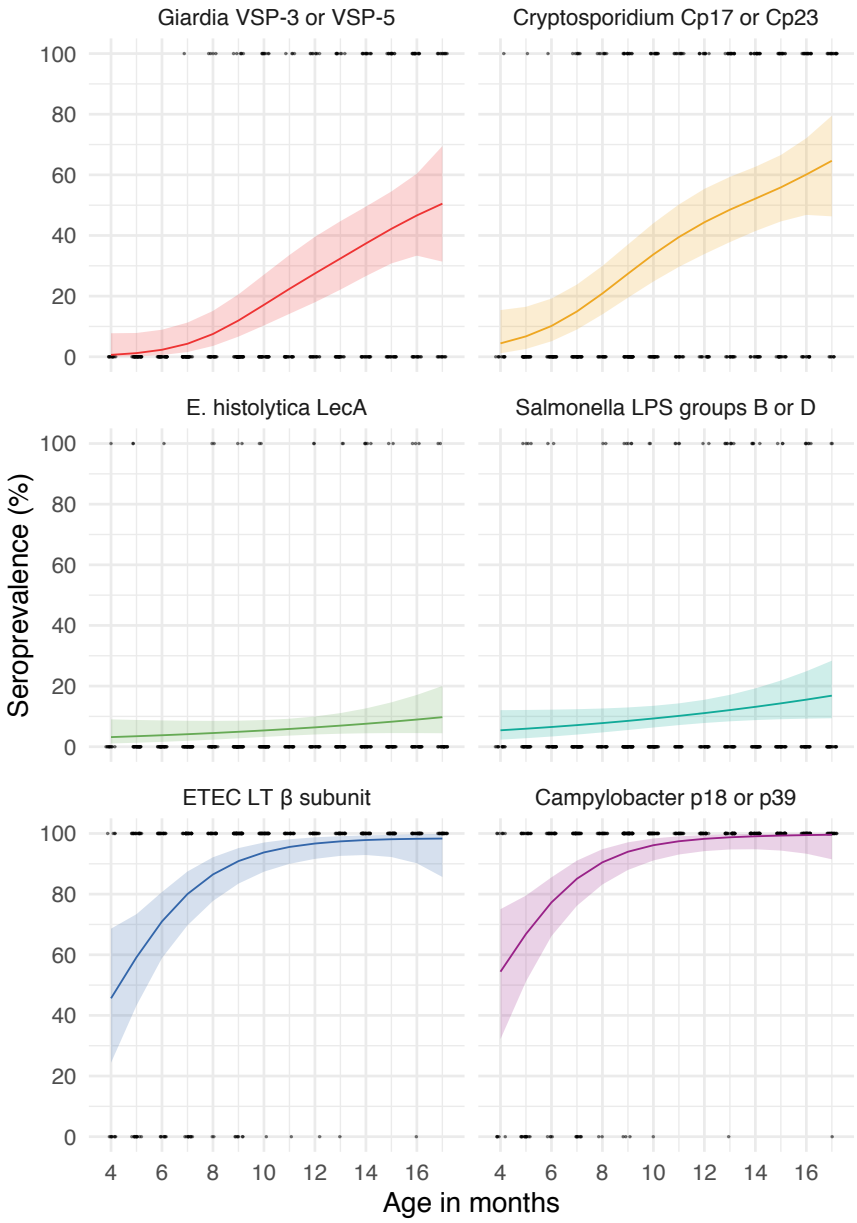

Figure 4 - supplement 1. Age dependent seroprevalence curves fit with cubic splines among children ages 4 to 17 months in Asembo, Kenya, 2013. Shaded bands are approximate, simultaneous 95% credible intervals. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform (N=439 measurements from 240 children). Points at 0 and 100 mark individual outcome measurements. Ages were measured in months completed (rounded) so points in the figure are jittered. Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/2q7zg`).

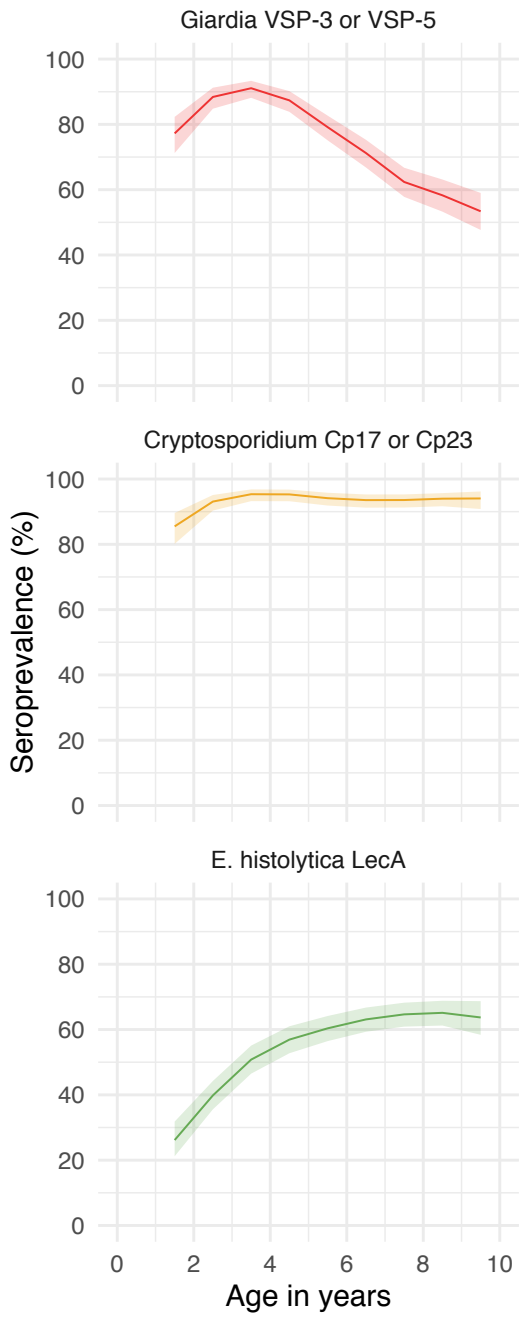

Figure 4 - supplement 2. Age dependent seroprevalence curves fit with cubic splines among children ages 1 to 9 years in Kongwa, Tanzania, 2012-2015. Shaded bands are approximate, simultaneous 95% credible intervals. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform in 4,989 specimens. Seropositivity cutoffs could not be estimated for *Salmonella spp.*, ETEC, and *Campylobacter spp.* in this study so seroprevalence curves are not shown for those pathogens. Created with notebook (`https://osf.io/pyhxe`) and data (`https://osf.io/kv4d3`).

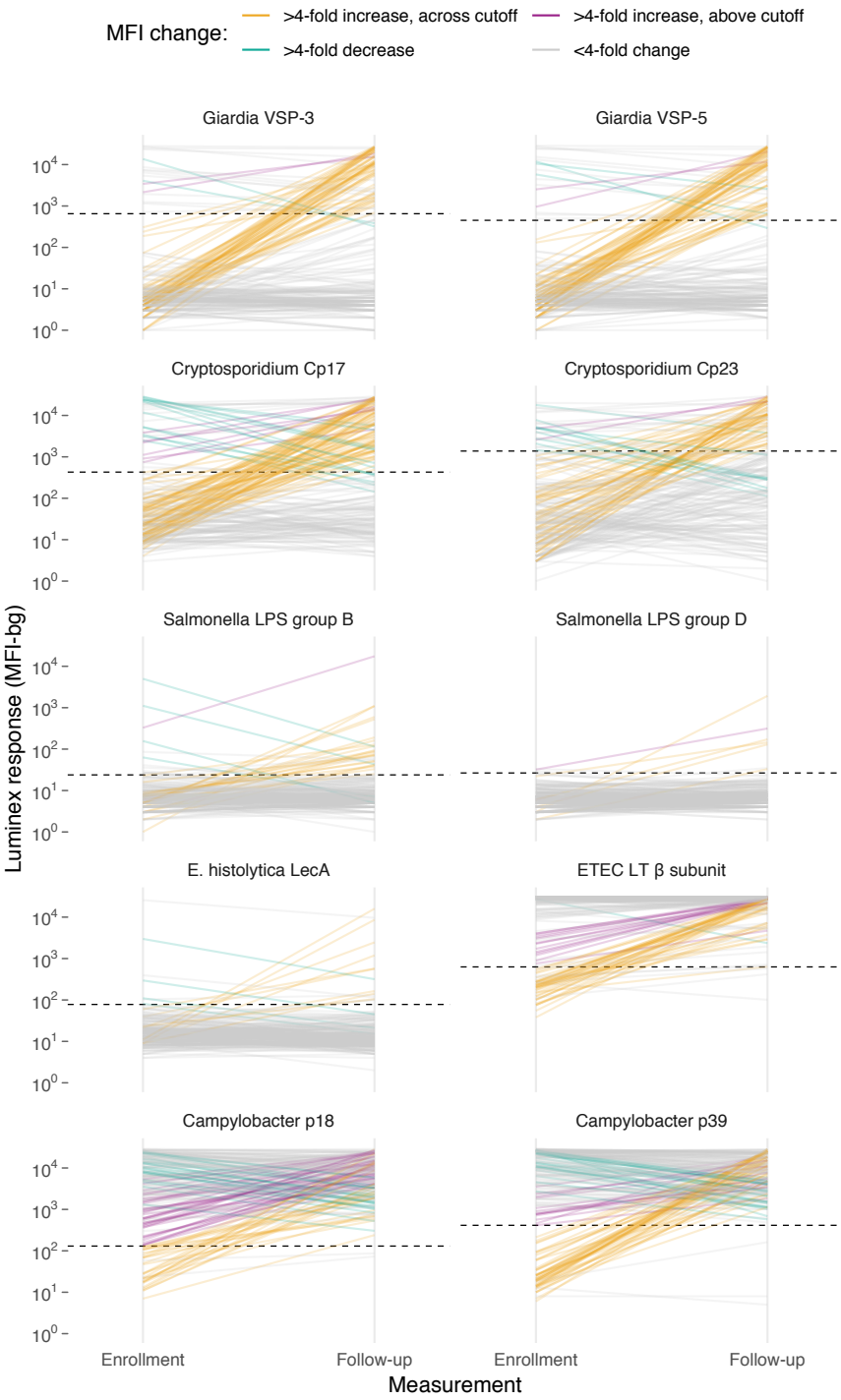

Figure 5. Longitudinal changes in IgG response between enrollment and follow-up among 199 children ages 4-17 months in Asembo, Kenya, 2013. Horizontal dashed lines mark seropositivity cutoffs for each antibody. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. Created with notebook (`https://osf.io/5kaht`), which includes additional visualizations, and data (`https://osf.io/2q7zg`). Supplement 1 includes individual antibody trajectories in the Haiti cohort.

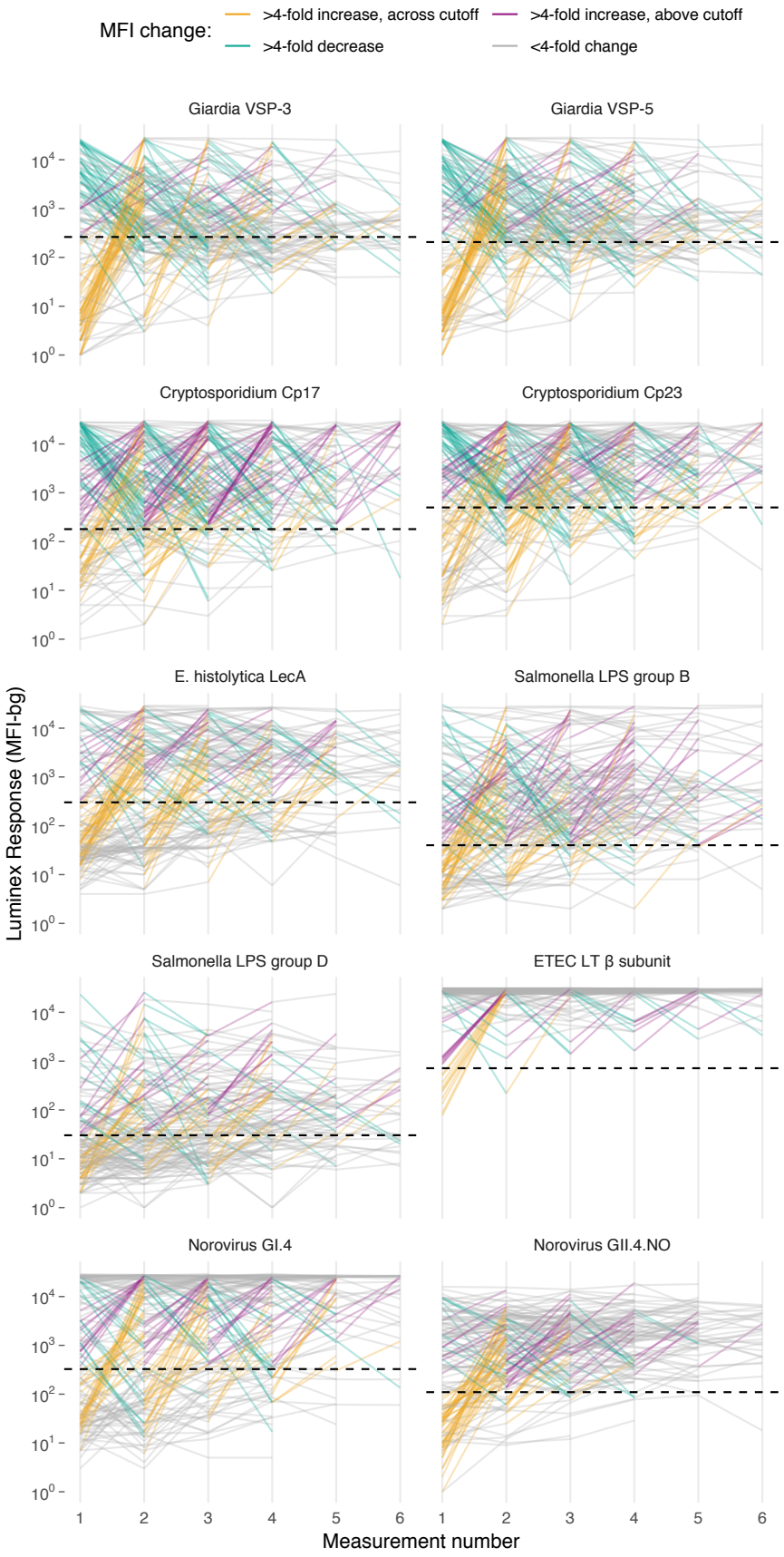
MFI change:
>4-fold increase, across cutoff
>4-fold increase, above cutoff
>4-fold decrease
<4-fold change
Giardia VSP-3
Giardia VSP-5
Cryptosporidium Cp17
Cryptosporidium Cp23
E. histolytica LecA
Salmonella LPS group B
Salmonella LPS group D
ETEC LT β subunit
Norovirus GI.4
Norovirus GII.4.NO
Luminex Response (MFI-bg)
Measurement number

Figure 5 - supplement 1. Longitudinal changes in IgG response over six repeated measurements among 142 children ages 0-11 years in Leogane, Haiti, 1990-1999. Measurements were spaced by approximately 1 year (median spacing = 1, IQR = 0.7, 1.3). Horizontal dashed lines mark seropositivity cutoffs for each antibody. The number of children measured at each visit was: n_1=142, n_2=142, n_3=140, n_4=131, n_5=111, n_6=66); 29 children had >6 measurements that are not shown. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform. Created with notebook (`https://osf.io/wnt3r`), which includes additional visualizations, and data (`https://osf.io/3nv98`).

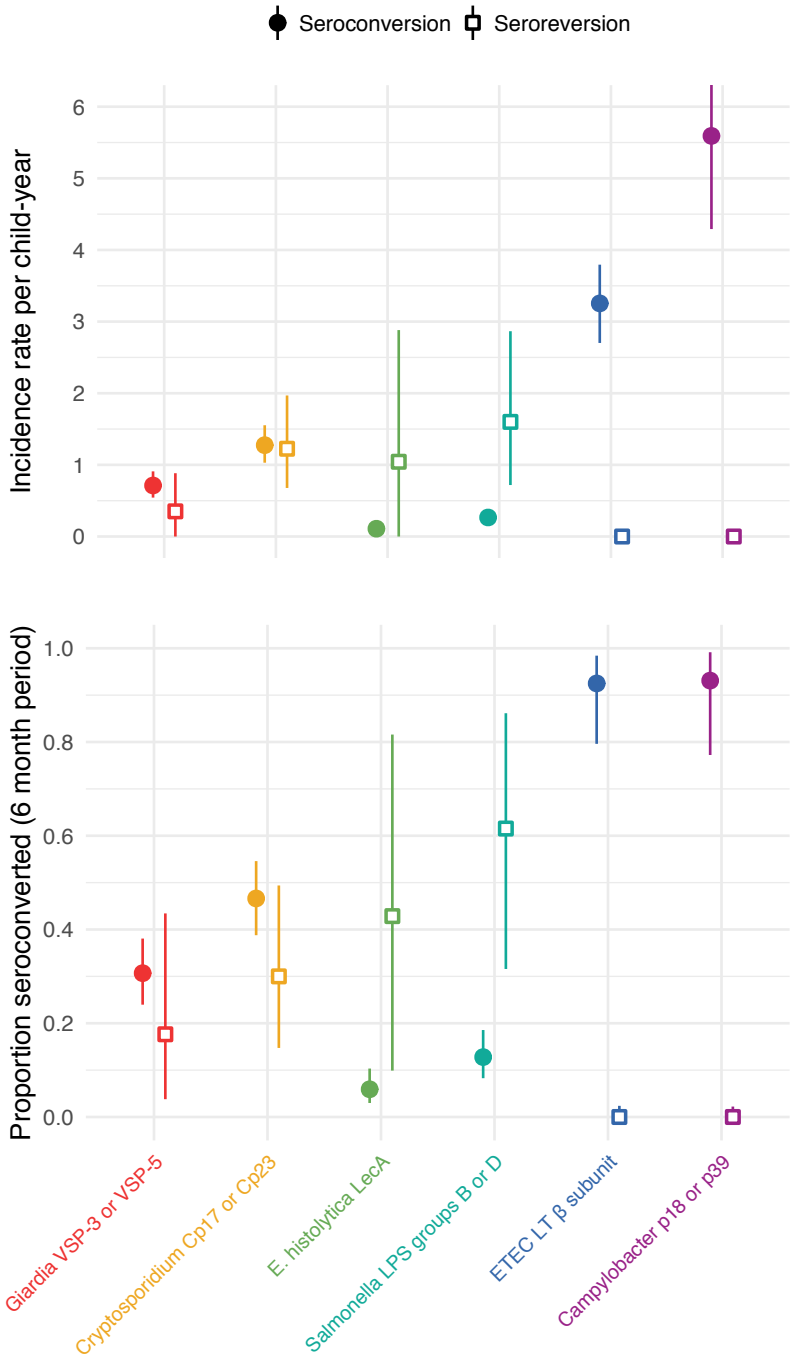

Figure 6. Enteric pathogen seroconversion and seroreversion incidence rates and proportion that seroconverted among 199 children ages 4 to 17 months measured longitudinally in Asembo, Kenya, 2013. Vertical lines mark bootstrap 95% confidence intervals. Children who seroconverted were those who were seronegative at enrollment and seropositive at follow-up 6-7 months later; seroreversion was the opposite. The upper bound of the confidence interval for *Campylobacter* seroconversion rate extends to 7.5 and is truncated in the figure. IgG response measured in multiplex using median fluorescence units minus background (MFI-bg) on the Luminex platform (N=398 measurements from 199 children). Created with notebook (`https://osf.io/sqvj7`) and data (`https://osf.io/2q7zg`). Supplement 1 compares prospective seroconversion rates with estimates from age-seroprevalence curves.

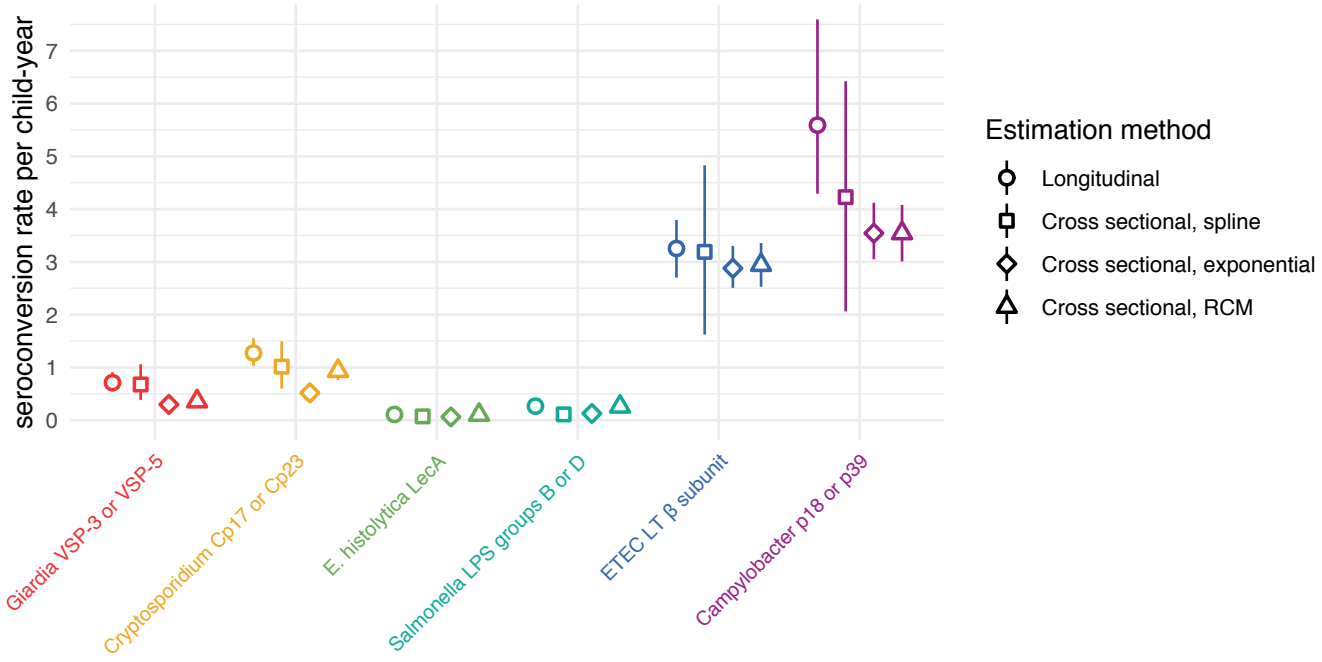

Figure 6 - supplement 1. Enteric pathogen seroconversion rate estimates from longitudinal and cross-sectional analyses among children ages 4 to 17 months in Asembo, Kenya, 2013. Cross-sectional estimates of incidence were derived from age-specific seroprevalence curves using nonparametric cubic splines (splines), a parametric constant rate survival model (exponential), and a reversible catalytic model (RCM) that assumed a constant seroconversion rate and fixed seroreversion rate (estimated from the prospective analysis). Vertical lines mark 95% confidence intervals. (N=398 measurements from 199 children). Created with notebook (`https://osf.io/3kczj`) and data (`https://osf.io/2q7zg`).

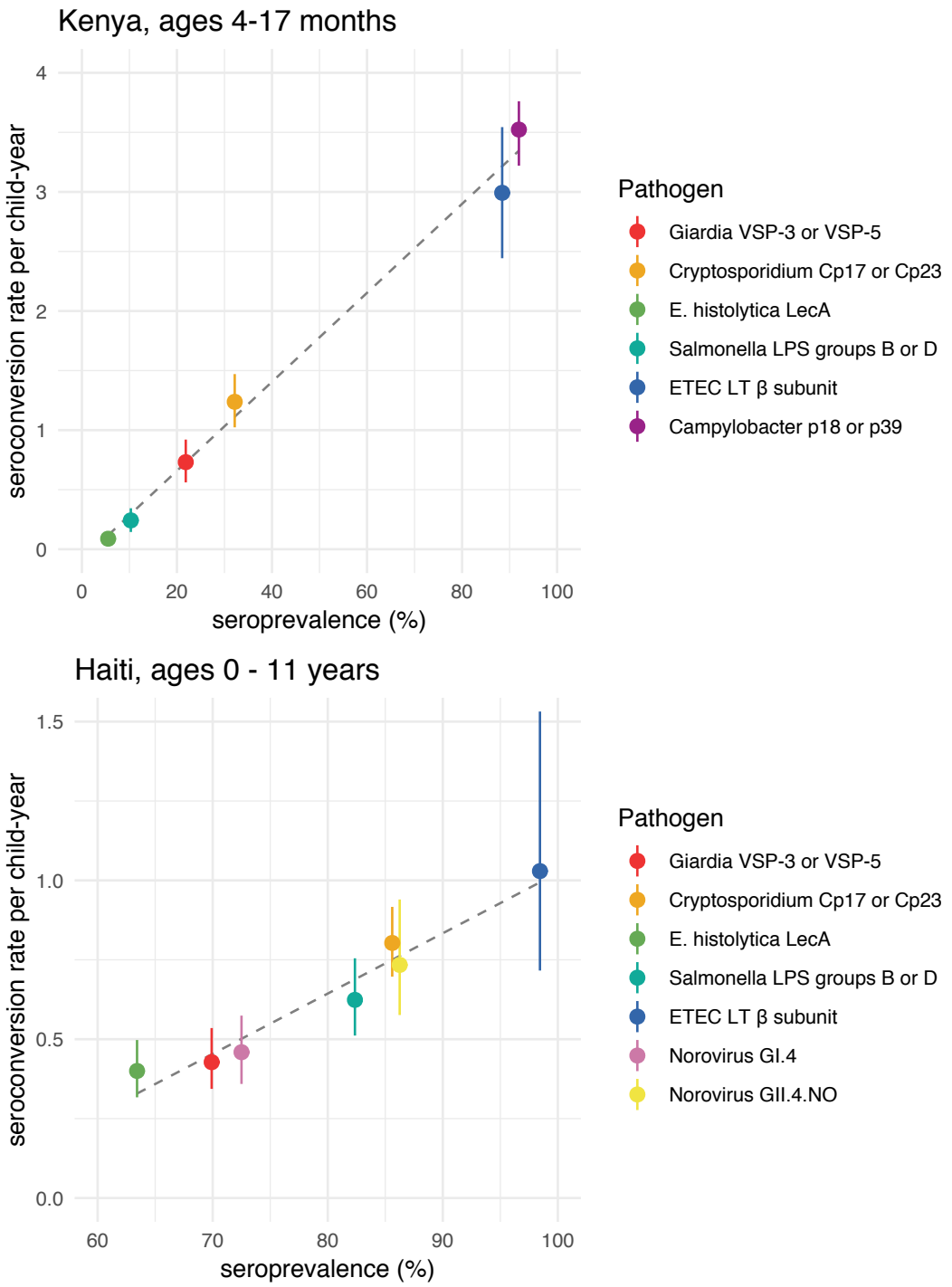

Figure 7. Prospective seroconversion rates versus seroprevalence among cohorts in Kenya and Haiti. Dashed lines show linear fits across pathogens. Note that axis scales differ between panels to best-illustrate the estimates. Created with notebook (`https://osf.io/pz8bd`) and data (`https://osf.io/2q7zg`, `https://osf.io/3nv98`).