



Inferring the proportion of undetected cholera infections from serological and clinical surveillance in an immunologically naive population

Original Paper



Cite this article: Finger F, Lemaître J, Juin S, Jackson B, Funk S, Lessler J, Mintz E, Dely P, Boncy J and Azman AS (2024). Inferring the proportion of undetected cholera infections from serological and clinical surveillance in an immunologically naive population. *Epidemiology and Infection*, **152**, e149, 1–7 <https://doi.org/10.1017/S0950268824000888>

Received: 24 October 2023
Revised: 06 May 2024
Accepted: 27 May 2024

Corresponding author:

Flavio Finger;
Email: flavio.finger@epicentre.msf.org

*F.F. and Jo.L. contributed equally to the work.

Flavio Finger^{1,2*} , Joseph Lemaître^{3*}, Stanley Juin⁴, Brendan Jackson⁵, Sebastian Funk¹, Justin Lessler^{3,6,7}, Eric Mintz⁵, Patrick Dely⁸, Jacques Boncy⁸ and Andrew S Azman^{7,9,10} 

¹Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK; ²Epicentre, Paris, France; ³Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ⁴Center for Global Health, Massachusetts General Hospital, Boston, MA, USA; ⁵United States Centers for Disease Control and Prevention, Atlanta, GA, USA; ⁶Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ⁷Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ⁸Ministère de la Santé Publique et de la Population, Port au Prince, Haiti; ⁹Division of Tropical and Humanitarian Medicine, Geneva University Hospitals, Geneva, Switzerland and ¹⁰Center for Emerging Viral Diseases, Geneva University Hospitals, Geneva, Switzerland

Abstract

Most infections with pandemic *Vibrio cholerae* are thought to result in subclinical disease and are not captured by surveillance. Previous estimates of the ratio of infections to clinical cases have varied widely (2 to 100 infections per case). Understanding cholera epidemiology and immunity relies on the ability to translate between numbers of clinical cases and the underlying number of infections in the population. We estimated the infection incidence during the first months of an outbreak in a cholera-naive population using a Bayesian vibriocidal antibody titer decay model combining measurements from a representative serosurvey and clinical surveillance data. 3,880 suspected cases were reported in Grande Saline, Haiti, between 20 October 2010 and 6 April 2011 (clinical attack rate 18.4%). We found that more than 52.6% (95% Credible Interval (CrI) 49.4–55.7) of the population ≥ 2 years showed serologic evidence of infection, with a lower infection rate among children aged 2–4 years (35.5%; 95%CrI 24.2–51.6) compared with people ≥ 5 years (53.1%; 95%CrI 49.4–56.4). This estimated infection rate, nearly three times the clinical attack rate, with underdetection mainly seen in those ≥ 5 years, has likely impacted subsequent outbreak dynamics. Our findings show how seroincidence estimates improve understanding of links between cholera burden, transmission dynamics and immunity.

Key Results

- We combine serological and clinical cholera incidence in an outbreak in a naive population in Grande Saline, Haiti.
- The rate of infection with *Vibrio cholerae* was several times the clinical attack rate.
- Half of the population ≥ 2 years showed serological evidence of infection.
- For every reported clinical case ≥ 5 years old, 3.2 (95% CrI 3.0–3.4) people were infected.
- Children 2–4 years old had a lower infection rate, which was not significantly different from the clinical attack rate.

Introduction

In its global roadmap to eliminate cholera as a public health threat by 2030, the Global Task Force for Cholera Control (GTFCC) considers at least 47 countries as cholera affected [1], where 2.86 million cases (uncertainty range 1.3m to 4.0m) and 95,000 deaths (uncertainty range 21,000–143,000) occur each year [2]. In addition to endemic cholera, small and large-scale cholera outbreaks are common [3]. However, the true number of pandemic *V. cholerae* infections and cases are greatly obscured by the nature of cholera surveillance systems, care-seeking behavior and the natural history of pandemic *V. cholerae*.

In most cholera-affected countries, surveillance is primarily passive, health facility-based and reliant on a suspected case definition that has low specificity [4,5]. Laboratory confirmation of cholera requires shipping stool samples to central laboratories for culture or PCR, and is typically done only for a subset of cases during outbreaks [6,7]. Rapid Diagnostic Tests (RDT) have become available in recent years, but the scale-up of their routine use has only just started [6].

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

Biases in estimates of the clinical burden of cholera can result from non-systematic use of diagnostic testing and variability in the proportion of people with symptomatic cholera who seek care outside of facilities under surveillance [5,8,9].

Most cholera infections are thought to be asymptomatic or subclinical [10–12]. They may not be of clinical relevance but the infected may still transmit the bacteria and acquire immunity, which can have important impacts on the epidemiological dynamics of the current or future outbreaks [13–16]. Importantly, the proportion of the population with full or partial immunity can thus not be determined from clinical surveillance alone but is of crucial importance to gauge the future course of an outbreak or to plan vaccination campaigns.

There is growing interest in complementing clinical surveillance with serological data to provide a more complete picture of infections and burden [17–19]. Recent work has shown that cross-sectional serologic data contain sufficient information to estimate the incidence of recent infections [10,18,20]. This quantity is referred to as *seroincidence*. However, the relationship between medically attended clinical incidence and seroincidence, which is shaped by host, pathogen and surveillance systems, remains unclear. Furthermore, most evidence comes from hyperendemic settings like Bangladesh, where exposure to cholera is extremely common. Despite the scientific consensus that the number of asymptomatic or mild infections is many times higher than the number of clinically relevant infections, estimates have varied widely (factor 2 to 100), likely as a result of variable case definitions, individual infection histories, different *V. cholerae* strains and host factors (Supplementary Material) [13,15,21–28]. Being able to translate between the number of cases observed at health facilities and the true number of infections in the community, or vice-versa, could help improve the public health utility of serosurveillance data and strengthen our understanding of cholera epidemiology and immunity.

Data collected from a representative serosurvey in Grand Saline, Haiti, between March 22 and April 4, 2011, six months after the start of the first cholera outbreak ever reported there, provide a rare opportunity to gain new insights into the relationship between infections with pandemic *Vibrio cholerae* O1 and clinical disease in a population with no previous exposures to the bacterium [17]. In this study we aim to estimate the true incidence of infections with *V. cholerae* O1 and the infection to case ratio during the first 6 months of this unprecedented cholera epidemic by combining information from serological and clinical surveillance.

Methods

Data and study population

Serologic data were obtained from a previously published representative serosurvey conducted in the municipality of Grande Saline (population 21,131) between 22 March and 6 April 2011, roughly 6 months after the beginning of the outbreak [17,29]. This study included 2,622 participants aged between 2 and 90 years from 1,240 households, with 2,527 (93%) participants providing serum samples. Vibriocidal antibody titers were measured using an initial 10-fold dilution followed by serial 2-fold dilutions and expressed as 1:20, 1:40, 1:80, 1:160, ..., 1:40960. Reported titer values represent the highest dilution with a $\geq 50\%$ reduction in turbidity compared with a bacteria-only control, indicating inhibition of bacterial growth. The true titer thus lies in the interval between the indicated and the next higher dilution on the scale (interval censoring). The sampling and laboratory analyses have been described

previously in Jackson *et al.* (2013). We \log_2 -transformed the reciprocal titer values (a dilution of 1:20 was transformed to $\log_2(20) = 4.32$, 1:40 to 5.32, 1:80 to 6.32, ...) and assigned 0 to titer values below the limit of detection ($<1:20$). While the only known *V. cholerae* strain circulating in Haiti before the serosurvey was of serotype Ogawa [30], vibriocidal titers to serotypes Inaba and Ogawa were detected. Due to known cross reactivity, we use the maximum measured titer across both serotypes (Figure 1A).

In addition to collecting data on individual demographics and home location, study staff asked each participant about episodes of watery diarrhoea occurring since the beginning of the epidemic, if they had sought treatment at a health facility, if they had been hospitalized overnight and if they had been diagnosed with cholera by a health care worker.

We also used the daily incidence of suspected cholera in the commune of Grande Saline reported between 20 October 2010, when the commune reported its first case, and the end of the serosurvey (6 April 2011) from the national surveillance system (Figure 2A) [17,31]. The case definition for suspected cholera used in Haiti at the time was “acute watery diarrhoea, with or without vomiting, in persons of all ages residing in an area in which at least one case of *Vibrio cholerae* O1 infection had been confirmed by culture”. The cases were reported in two age groups: children less than 5 years old and individuals aged 5 years and older. In order to compare reported cases to the target population of the serosurvey, which excluded those less than 2 years old, we multiplied the number of reported cases under 5 years by 0.74, the fraction of 2–4 year olds in under 5 year old cases obtained from unpublished age distributions of cases in Haiti in 2010. We estimated the population of 2–4 year olds and of ≥ 5 year olds from the national age distribution [32] and performed our analysis for the entire population 2 years and older and separately for 2–4 year olds and for 5 year olds and older.

Logistic regression analysis

In the survey, participants were asked if they had had cholera symptoms since the start of the epidemic in Haiti. These types of self-reports can be insensitive and non-specific proxies of having true cholera. To understand the relationship between vibriocidal titers (a known marker of recent infection) and self-reported cholera symptoms, we used logistic regression models. Specifically, we used either self-reported cholera diagnosis or self-reported watery diarrhoea as dependent variable, and vibriocidal antibody titer (\log_2 -transformed) as predictor, adjusting for covariates in a step-wise procedure. Full details on the logistic regression models can be found in the Supplementary Materials. Regression analyses were performed with R statistical software [33].

Seroincidence estimation using a Bayesian vibriocidal titer decay model

Our primary estimates of seroincidence and infection rate are based on modeling the vibriocidal titer decay in the blood of participants enrolled in the study. Vibriocidal titer values correlate with recent *V. cholerae* infection and have been shown to be a non-mechanistic correlate of protection when paired samples are available [10,34]. They rise to their maximal value within 1–2 weeks after infection and then start decaying toward pre-infection levels [34]. We developed a hierarchical, Bayesian statistical model of the decay of the vibriocidal titer of participants, from the inferred date of infection to the date the serum sample was taken. As

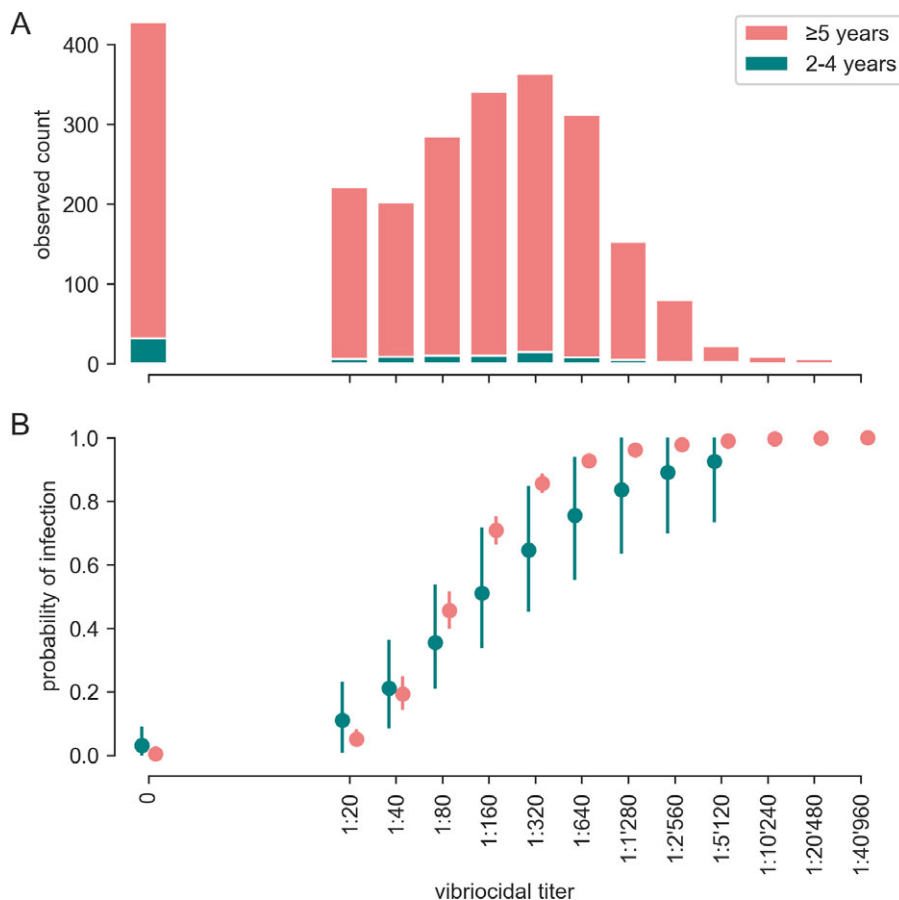


Figure 1. Measured titer values in the sample population (A), inferred individual probability of infection (with 90% prediction intervals) (B) by measured titer value (log₂-transformed). Note that no estimates of the infection probability for the three highest titer classes for children <5 are shown because no children included in the study had these titers.

described below, by modeling each individual’s probability of being infected, time of infection, titer decay and titer measurement process, we were able infer the infection rate (i.e., seroincidence) of cholera in Grande Saline from a snapshot of sampled vibriocidal titer values. We assumed that a constant proportion of infections was reported throughout the study period, equal incubation periods and equal reporting delays for all individuals, equal infection risk for all individuals (ignoring e.g. household structure or geography) and that infection led to seroconversion in all individuals. For each age group, we fitted the model to the number of individuals at each vibriocidal titer obtained from Jackson et al. [17]. As a comparison, two simpler methods to determine infection rate by classifying measured titer values according to fixed threshold or a mixture distribution are described in the [Supplementary Material](#).

Modeling infection

Our model assumes that the true number of infections each day was proportional to the reported clinical incidence (Figure 2A) lagged by the median incubation period and reporting delay (Table 1). The estimated number of individuals infected on day *t*, denoted *I(t)* is thus:

$$I(t) = \frac{1}{\rho} C(t - \tau_{I \rightarrow C})$$

where *C(t)* is the number of reported cases on day *t* (Figure 2A), ρ is the proportion of infections that are reported (a combination of the proportion of infections that are clinically apparent and of the proportion of clinically apparent infections that seek care at

reporting facilities) and $\tau_{I \rightarrow C}$ is the delay from infection to reporting. For each individual *i*, we modeled an infection indicator variable δ^i , which takes the value 1 if individual *i* was infected between the beginning of the outbreak and the time of sampling and 0 otherwise, as a Bernoulli process with probability of success equal to the infection rate:

$$\delta^i \sim \text{Bernoulli} \left(\frac{\sum_t C(t)}{\rho N} \right)$$

where *N* is the population size.

The time of infection T_i^i for each infected individual *i* is drawn from a distribution with the shape of the estimated infection incidence:

$$P(T_i^i = t) = \frac{I(t)}{\sum_{\tau} I(\tau)}$$

Modeling vibriocidal titer dynamics

Following previous studies, we modeled the vibriocidal titer dynamics of infected individuals by assuming an initial increase in titer *d* days after the presumed infection followed by an exponential decay [10,35]. Titers of non-infected individuals were assumed to remain at their baseline level. The modeled true titer value of individual *i* at the time of sampling T_s is thus:

$$\bar{A}_{T_s}^i = \begin{cases} \bar{\omega}^i & \text{if } \delta^i = 0 \\ \bar{\omega}^i + \lambda^i \cdot \exp[-r \cdot (T_s - (T_i^i + d))] & \text{if } \delta^i = 1 \text{ and } (T_s - d) > T_i^i \end{cases}$$

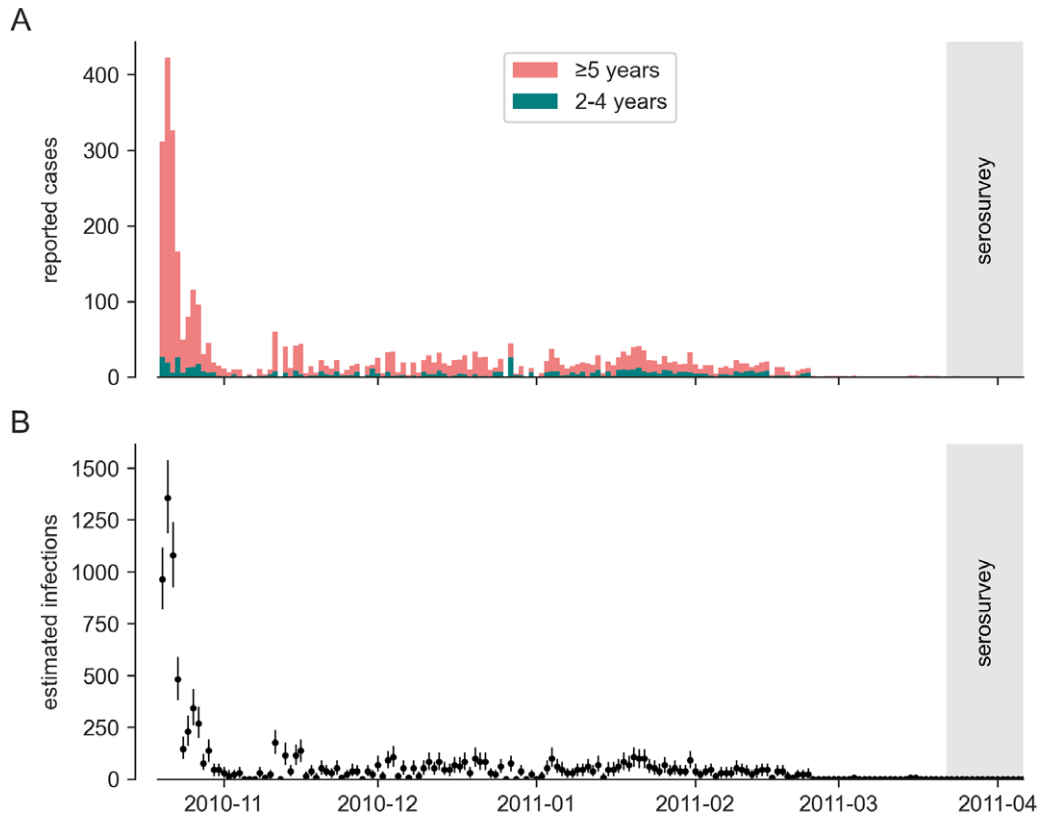


Figure 2. Reported daily clinical incidence of suspected cholera (A) and inferred median incidence of infections with 95% prediction intervals (B).

Table 1. Key parameters and priors of the vibriocidal titer decay model

Name	Description	Calibrated	Prior / Value
N_{pop}	population size	No	21,131 [29]
N_{2-4}	population size 2–4 years old	No	$1,588 = (N \cdot 0.1253 \cdot 3/5)$ (estimated from 12.53% under 5s in the 2010 national age distribution and assuming uniform age distribution in the population under 5 [29,32])
N_{5+}	population size, 5 years and older	No	19,403 (estimated from 2010 national age distribution: 87.47% of the population of over 5 [29,32])
N_{2+}	population size, 5 years and older	No	$N_{2-4} + N_{5+} = 20,991$
ρ_{2-4}	proportion of reported cases under 5 that are between 2 and 4 years old	No	73.94% derived from unpublished age structure of cases in Nov and Dec 2010
ρ	reporting fraction	Yes	Weakly informative prior $\rho \in [\sum_t C(t)/N, 1.5]$
σ	titer measurement standard deviation	Yes	$\mathcal{N}_{\geq 0}(0, 1)$
$\bar{\omega}^i$	individual baseline titer	Yes	$\begin{bmatrix} \bar{\omega}^i \\ \lambda^i \end{bmatrix} \sim \mathcal{MN}_{\geq 0} \left(\begin{bmatrix} 5, 31 \\ 5, 45 \end{bmatrix}, \begin{bmatrix} 3.96 & -4.02 \\ -4.02 & 8.10 \end{bmatrix} \right)$ [10]
λ^i	individual titer rise	Yes	
r	titer decay rate	Yes	$\mathcal{N}_{\geq 0}(0.0056, 0.0005)$ [day^{-1}]: mean half-life of 123 days [10]
δ^i	individual infection indicator	Yes	Bernoulli process with probability of success equal to the infection rate: $\delta^i \sim \text{Bernoulli} \left(\frac{\sum_t C(t)}{\rho N} \right)$
T_i^j	individual time of infection for infected individuals	Yes	Drawn from a categorical distribution with the shape of the estimated infection incidence. For each day t : $P(T_i^j = t) = \frac{I(t)}{\sum_{\tau} I(\tau)}$
d	time from infection to peak titer	No	11 days [10]
$T_{I \rightarrow C}$	time from infection to reporting	No	2 days: 1.5 days incubation [41] + 0.5 days of reporting delay

where $\bar{\omega}^i$ represents the baseline titer value of each individual i , λ^i is the individual temporary titer rise after infection and r the exponential decay rate. We do not explicitly model the dynamics of post-infection titer rise, justified by the absence of reported cases in the days immediately preceding the serosurvey (Figure 2A).

Measurement model

The measured vibriocidal titer $A_*^i = A_k \in \{0, \log_2(20), \log_2(40), \dots, \log_2(40960)\}$ lies on a discrete scale dictated by serial dilutions of the measurement process (see Data). Following Salje et al. (2018) we modeled the probability that the true (\log_2) vibriocidal titer \bar{A}^i for individual i is measured in bin A_k as:

$$P(A_*^i = A_k | \bar{A}^i) = \int_{A_k}^{A_{k+1}} \mathcal{N}(u | \bar{A}^i, \sigma) du,$$

where the width of the \log_2 titer measurement intervals $A_{k+1} - A_k$ is equal to 3.32 if $k=0$, corresponding to the initial 10-fold dilution, and 1 otherwise, corresponding to a 2-fold dilution. Measurement error was assumed to be normally distributed on the log-scale with standard deviation σ .

Parameters

We allowed parameters for titer rise upon infection λ^i and baseline titer $\bar{\omega}^i$ to vary between individuals, whereas the titer decay rate was assumed to be the same for the entire population (Table 1). We used a weakly informative prior for the fraction of reported infections ρ , ranging from $\sum_t C(t)/N = 0.17$ – if every inhabitant has been infected – to 1.5, an over-reporting of 50%. Prior distributions on decay parameters r , λ^i and $\bar{\omega}^i$ were derived from Jones et al. [10], we used a multivariate normal distribution for the latter two to account for correlation. In addition, we also inferred latent variables representing the infection history of each individual (T_i^j and δ^i). Sensitivity analyses on the prior distributions for the decay rate r and the baseline titer ω_i are presented in the Supplementary Material.

Inference

We implemented the model using the PyMC framework for probabilistic programming [36] and calibrated it with the No-U-Turn Sampler (NUTS) [37] with 4 chains of 20,000 tuning iterations and 2,000 posterior sampling iterations. Posterior predictive checks, posterior distributions and diagnostic checks are presented in the Supplementary Material. We ran separate inference for the entire population and for the different age groups and report mean and 95% credible intervals (CrI) of the resulting distributions.

Ethics

Participants in the original study provided written informed consent and the study protocol was approved by the Haitian National Ethics Committee IRB (Protocol #2011-FWD-CHOLERA-01) and the U.S. Centers for Disease Control and Prevention IRB (CDC Protocol #6038). This secondary analysis was approved by the Ministère de la Santé Publique et de la Population de l'Haiti and deemed to be exempt from human subjects research by the Johns Hopkins Bloomberg School of Public Health IRB.

Results

In the commune of Grande Saline a total of 3,880 suspected cholera cases (clinical attack rate 18.4%) were reported from the first reported case on 20 October 2010 to the end of the serosurvey on 6 April 2011 through the national cholera surveillance system.

Table 2. Clinical attack rate, attack rate of self reported cholera and watery diarrhoea, and estimates of infection rate with *V. cholerae* O1

Estimate	Value [95% CI]		
	All ≥ 2 years	2-4 years	≥ 5 years
Clinical attack rate (reported cholera incidence) ^a	18.2% [17.7–18.8]	39.5% [37.0–41.9]	16.4% [15.9–16.9]
Clinical attack rate (self-reported cholera diagnosis) ^a	17.8% [16.3–19.3]	18.5% [12.1–26.5]	17.8% [16.3–19.3]
Clinical attack rate (self-reported watery diarrhoea) ^a	20.6% [19.1–22.2]	25.0% [17.7–33.6]	20.4% [18.9–22.1]
Infection rate (vibriocidal decay model) ^b	52.6% [49.4–55.7]	35.5% [24.2–51.6]	53.1% [49.4–56.4]

^aExact binomial confidence intervals (Clopper-Pearson).

^bCredible Intervals are given for Bayesian estimates.

Age specific clinical attack rates were 16.4% in people 5 years and above 39.5% in children under 5. We estimated the clinical attack rate to be 39.5% in children 2-4 years old (Table 2).

Five months after the peak of reported cases in the first wave, 2,622 people 2 years and older were enrolled in a representative serosurvey [17], 4.7% (124) were 2-4 years of age and 59% (1,553) were female. 20.6% (541) of participants reported watery diarrhoea since the start of the outbreak in October 2010, and 17.8% (466) reported to have been diagnosed with cholera by a health care worker. 14.8% (388) of participants reported to have sought treatment at a health facility with half of them (50.5%, 196/388) having spent at least one night there (Table 2). Having a higher vibriocidal \log_2 -titer was a predictor of self-reported cholera diagnosis (odds ratio (OR) 1.12 (95%CI 1.08 to 1.16)) and self-reported watery diarrhoea (OR 1.10 (95%CI 1.07 to 1.14)) in multivariable logistic regression analysis when adjusting for age and location (Supplementary Material). Among the 2,526 participants with reported vibriocidal titer values, 18.2% had no detectable vibriocidal antibodies and 38.8% had a titer of ≥ 320 to either of the two serotypes.

We combined vibriocidal titers and reported clinical incidence in a Bayesian vibriocidal titer decay model to infer the infection to reported case ratio and the infection rate. We found a sigmoidal dose-response relationship between titer and the probability of having been infected. People with a titer of 40 had a probability of 18.5% (95% CrI 13.2–24.1) and those with a titer of 320 a probability of 85.8% (95% CrI 82.6–88.9) of having been infected. When stratifying by age, we found that the curves had similar shapes between children 2-4 years old and those 5 and older (Figure 1B).

Using the vibriocidal titer decay model, we estimated that 52.6% (95%CrI 49.4–55.7) of Grande Saline's population ≥ 2 years old was infected during the first 6 months of the cholera epidemic, including 35.5% (95%CrI 24.2–51.6) of children 2-4 years old and 53.1% (95%CrI 50.1–56.4) of those aged 5 and above (Table 2, Figure 2B and Figure 3A). We find that the implied infection to reported case ratio (parameter ρ) was far higher in older children and adults than among those 2-4 years old. On average we estimate that for every reported clinical case among those ≥ 5 years old, a total of 3.2 (95% CrI 3.0–3.4) people were infected. In contrast, for every reported clinical case among those 2-4 years old there were 0.9 (95% CrI 0.6–1.3) infections (Figure 3B).

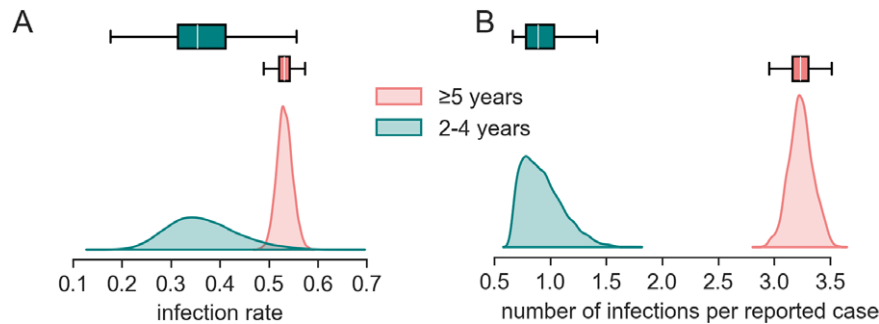


Figure 3. Distribution of the infection rate (A) and the infection (symptomatic and asymptomatic) to reported case ratio (B) in Grande Saline during the study period inferred from combining serological and incidence data and modeling vibriocidal decay.

Discussion

We estimated that more than 52.6% (95% CrI 49.4 to 55.7) of the population in Grand Saline, Haiti was infected by *V. cholerae* O1 during the first 6 months of the outbreak in 2010, with young children (2-4 years old) having a 32.5% (95% CrI 2.4 to 54.1) lower seroincidence than those 5 years and older, despite both age groups being immunologically naive at the start of the outbreak. Compared to the reported clinical incidence from the same community, we found that for every three infections among those five years and older in the community one clinical case was reported through the national surveillance system, though this ratio was closer to 1 for those 2-4 years old. As the global community looks for new ways to understand the burden and transmission of cholera, these estimates provide a unique perspective to help interpret clinical and seroincidence especially in immunologically naive populations.

Our estimates of the ratio of infections to reported clinical cholera cases are at the lower end of those previously reported, which likely reflects that this population had no previous exposure to the bacteria, but could also reflect differences in host factors (e.g., undernutrition) and exposure routes [13–15,21,22,25–28]. Our estimated number of infections among young children nearly matches the numbers of reported suspected cholera cases, although with wide confidence intervals, possibly arising from the small sample size limiting power. This result either implies that there were almost no infections in this age group that did not result in medically-attended disease, or more realistically that unreported infections were compensated by poor specificity of the suspected cholera case definition, especially in young children, where other etiologies frequently lead to similar symptoms. Given that other diarrhoeal pathogens may also have distinct seasonality and patterns of immunity (e.g., rotavirus [38]), our infection to reported case ratio may not be as generalizable across time and geographies as the estimate for adults.

Our study comes with a number of limitations. In reconstructing the infection incidence, our estimates of vibriocidal baseline titer, titer rise and decay rate among all infected individuals, across the spectrum of clinical severity and age, were based on priors informed from estimates from a cohort of severe, mostly adult, cholera cases in Bangladesh. In addition, to avoid identifiability issues, we assumed that the titer decay rate was the same for all individuals, and assigned fixed values to the delay from infection to peak titer and reporting delay. Moreover, there is known laboratory to laboratory variability in the results of the vibriocidal assay which could further limit the generalizability of the results from Bangladesh [39]. While to our knowledge there are no detailed kinetic data available across age and severity, especially among

immunologically naive, it may be that mild infections lead to a more blunted immune response, as seen with other pathogens [40], and that young children have different boosts and decay rates from adults. Our sensitivity analyses show that whereas our results in 2-4 year olds are robust to variations in the prior for decay rate, decreasing the prior distribution for the baseline titer could result in higher estimates of infection rate in both age groups (Supplementary Material). In estimating infection incidence we only relied on vibriocidal titers, with no other markers considered. Recent work has shown that inclusion of multiple serologic markers (i.e. antibodies against cholera toxin) can improve estimates of seroincidence from cross-sectional data [10,18], and inclusion of these in a multivariate decay model could have improved our precision. However, we decided not to include data on anti-cholera toxin B antibodies as they were measured using an assay for which no post-infection dataset to parametrize the kinetic model were available, in addition to the computational complexity of simultaneously modeling the decay of multiple antibodies. We assumed that the hazard of infection was proportional to the observed epidemic curve of suspected cholera cases, and that this ratio was constant in time. Although the agreement of attack rate estimates from clinical surveillance and self-reported cholera indicates that surveillance performed well, the surveillance system may have changed over the course of this six month period given the unprecedented context, and people's care seeking behaviors may have changed as awareness of cholera increased over the course of the epidemic. Simpler methods to determine infection by classifying measured titer values according to fixed thresholds or a mixture distribution exist (Supplementary Material), but given that they don't account for antibody dynamics they may be less accurate to reconstruct recent infection than the method employed in our main analysis.

Our estimates of the number of infections per reported suspected cholera case during an outbreak in an immunologically naive population suggests that only one in three infections with pandemic *V. cholerae* resulted in medically attended clinical disease that was reported via the surveillance system. More than half of the population of Grande Saline thus likely acquired at least partial immunity to cholera with unknown duration, a result which, given the similar reported clinical attack rates and similar surveillance system, can likely be generalized to large parts of Haiti. This has likely impacted the dynamics of the 2011 and subsequent epidemic waves. Although estimates may not be generalizable to populations with more frequent historical exposures to *V. cholerae*, this study shows that population immunity to cholera cannot be understood based on clinical surveillance data alone. While more work is needed to grasp the links between infections, immunity and its duration across different populations, our results highlight the

potential utility of combining clinical and serologic data in assessing cholera risk and making resource allocation decisions for cholera prevention and control.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0950268824000888>.

Data availability. The code and data are freely available on Github (github.com/HopkinsIDD/grande-saline-cholera-serosurvey) and archived on Zenodo (<https://zenodo.org/doi/10.5281/zenodo.10063169>). Note that the logistic regression analysis requires individual level data, which is not included in the repository to protect the privacy of the survey participants. Interested researchers are invited to contact the authors of the survey [17].

Acknowledgements. The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control (CDC). FF started to work on this analysis when at the London School of Hygiene and Tropical Medicine. He currently works for Epicentre.

Author contribution. Conceptualization: A.A., Jo.L., F.F.; Methodology: A.A., J.L., F.F.; Supervision: A.A.; Validation: A.A., S.F., J.L., E.M., B.J., J.B., J.L., P.D., S.J., F.F.; Writing – original draft: A.A., J.L., F.F.; Writing – review & editing: A.A., S.F., J.L., E.M., B.J., J.B., J.L., P.D., S.J., F.F.; Formal analysis: J.L.; Investigation: J.L., F.F.; Software: J.L., F.F.; Visualization: J.L., F.F.; Data curation: F.F.; Funding acquisition: F.F.; Project administration: F.F.

Financial support. F.F., Jo.L., A.A. acknowledges support from the Swiss National Science Foundation (P2ELP3_175079). JoL acknowledges support from the US National Institutes of Health (NIH 5R01AI102939). ASA acknowledges support from the US National Institutes of Health (R01 AI135115).

Competing interest. The authors declare no conflicts of interest.

References

- [1] **Global Task Force on Cholera Control.** *Ending cholera. A global roadmap to 2030.* 2017, p. 32–32.
- [2] **Ali M,** et al. Updated Global Burden of Cholera in Endemic Countries. *PLOS Neglected Tropical Diseases* 2015; **9**: e0003832.
- [3] **Zheng Q,** et al. Cholera outbreaks in sub-Saharan Africa during 2010–2019: a descriptive analysis. *International Journal of Infectious Diseases* 2022; **122**: 215–221.
- [4] **Nadri J,** et al. Sensitivity, Specificity, and Public-Health Utility of Clinical Case Definitions Based on the Signs and Symptoms of Cholera in Africa. *The American Journal of Tropical Medicine and Hygiene* 2018; **98**: 1021–1030.
- [5] **Wiens KE,** et al. Estimating the proportion of clinically suspected cholera cases that are true *Vibrio cholerae* infections: A systematic review and meta-analysis. *PLOS Medicine* 2023; **20**: e1004286.
- [6] **GTFCC Surveillance Working Group.** *Public health surveillance for cholera.* Global Task Force on Cholera Control, 2023 Feb.
- [7] **Ratnayake R,** et al. Early detection of cholera epidemics to support control in fragile states: estimation of delays and potential epidemic sizes. *BMC Medicine* 2020; **18**: 397.
- [8] **Ganesan D, Gupta SS, Legros D.** Cholera surveillance and estimation of burden of cholera. *Vaccine* 2020; **38**: A13–A17.
- [9] **Summers A,** et al. Notes from the Field: Diarrhea and Acute Respiratory Infection, Oral Cholera Vaccination Coverage, and Care-Seeking Behaviors of Rohingya Refugees — Cox's Bazar, Bangladesh, October–November 2017. *Morbidity and Mortality Weekly Report* 2018; **67**: 533–535.
- [10] **Jones FK,** et al. *Identifying Recent Cholera Infections Using a Multiplex Bead Serological Assay.* mBio American Society for Microbiology, 2022; **0**: e01900–22.
- [11] **Chen WH,** et al. Single-dose Live Oral Cholera Vaccine CVD 103-HgR Protects Against Human Experimental Infection With *Vibrio cholerae* O1 El Tor. *Clinical Infectious Diseases* 2016; **62**: 1329–1335.
- [12] **Maluda MCM,** et al. The incidence, and spatial trends of cholera in Sabah over 15 years: Repeated outbreaks in coastal areas. *PLOS Global Public Health* 2024; **4**: e0002861.
- [13] **Glass RI, Black RE.** *The Epidemiology of Cholera.* In: Barua D, Greenough WB, eds. *Cholera.* Boston, MA: Springer US, 1992, p. 129–154.
- [14] **Kaper JB, Morris JG, Levine MM.** Cholera. *Clinical Microbiology Reviews* 1995; **8**: 48–86.
- [15] **King AA,** et al. Inapparent infections and cholera dynamics. *Nature* 2008; **454**: 877–880.
- [16] **Nelson EJ,** et al. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Reviews Microbiology* 2009; **7**: 693–702.
- [17] **Jackson BR,** et al. Seroepidemiologic Survey of Epidemic Cholera in Haiti to Assess Spectrum of Illness and Risk Factors for Severe Disease. *The American Journal of Tropical Medicine and Hygiene* 2013; **89**: 654–664.
- [18] **Azman AS,** et al. Estimating cholera incidence with cross-sectional serology. *Science Translational Medicine* 2019; **11**: eaau6242.
- [19] **Azman AS,** et al. *Vibrio cholerae* O1 transmission in Bangladesh: insights from a nationally representative serosurvey. *The Lancet Microbe* 2020; **1**: e336–e343.
- [20] **Harris JB,** et al. Susceptibility to *Vibrio cholerae* Infection in a Cohort of Household Contacts of Patients with Cholera in Bangladesh. *PLOS Neglected Tropical Diseases* 2008; **2**: e221.
- [21] **Bart KJ,** et al. Seroepidemiologic studies during a simultaneous epidemic of infection with El Tor ogawa and classical inaba *vibrio cholerae*. *Journal of Infectious Diseases* Oxford University Press, 1970; **121**: S17–S24.
- [22] **Blake PA,** et al. Cholera—a possible endemic focus in the United States. *The New England Journal of Medicine* 1980; **302**: 305–309.
- [23] **Feachem RG.** Environmental aspects of cholera epidemiology. III. Transmission and control. *Tropical Diseases Bulletin* 1982; **79**: 1–47.
- [24] **Gangarosa EJ.** The epidemiologic basis of cholera control. *Bulletin of the Pan American Health Organization (PAHO)* 1974; **8**: 189–197.
- [25] **Harris JR,** et al. Impact of Epidemic Cholera in a Previously Uninfected Island Population: Evaluation of a New Seroepidemiologic Method. *American Journal of Epidemiology* 1986; **123**: 424–430.
- [26] **Johnston JM,** et al. Cholera on a Gulf Coast Oil Rig. *New England Journal of Medicine* 1983; **309**: 523–526.
- [27] **Mccormack WM, Islam MS, Fahimuddin M.** A community study of inapparent cholera infections. *American Journal of Epidemiology* 1969; **89**: 658–664.
- [28] **Van de Linde PAM, Forbes GI.** Observations on the spread of cholera in {Hong Kong}, 1961–63. *Bulletin of the World Health Organization* 1965; **32**: 515.
- [29] **Institut Haïtien de Statistique et d'Informatique (IHSI).** *Projection de la population par la section communale de 2009.* 2009.
- [30] **Centers for Disease Control and Prevention (CDC).** Notes from the field: Identification of *Vibrio cholerae* serogroup O1, serotype Inaba, biotype El Tor strain - Haiti, March 2012. *MMWR. Morbidity and mortality weekly report* 2012; **61**: 309.
- [31] **Barzilay EJ,** et al. Cholera Surveillance during the Haiti Epidemic — The First 2 Years. *New England Journal of Medicine* 2013; **368**: 599–609.
- [32] **United Nations.** 2011 *United Nations Demographic Yearbook.* United Nations Publication, 2012.
- [33] **R Core Team.** *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing, 2017.
- [34] **Clements ML,** et al. Magnitude, kinetics, and duration of vibriocidal antibody responses in North Americans after ingestion of *Vibrio cholerae*. *The Journal of infectious diseases* 1982; **145**: 465–473.
- [35] **Salje H,** et al. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature* 2018; **557**: 719–723.
- [36] **Salvatier J, Wiecki TV, Fonnesbeck C.** *Probabilistic programming in Python using PyMC3.* PeerJ Computer Science PeerJ Inc., 2016; **2**: e55.
- [37] **Hoffman MD, Gelman A.** The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research* 2014; **15**: 1593–1623.
- [38] **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**: 1487–1496.
- [39] **Son MS, Taylor RK.** Vibriocidal assays to determine the antibody titer of patient sera samples. *Current Protocols in Microbiology* 2011; Chapter **6**: Unit6A.3.
- [40] **Takahashi S, Greenhouse B, Rodríguez-Barrquer I.** Are Seroprevalence Estimates for Severe Acute Respiratory Syndrome Coronavirus 2 Biased? *The Journal of Infectious Diseases* 2020; **222**: 1772–1775.
- [41] **Azman AS,** et al. The incubation period of cholera: A systematic review. *Journal of Infection* 2013; **66**: 432–438.