

Towards estimating true cholera burden: a systematic review and meta-analysis of *Vibrio cholerae* positivity

Supplementary Materials

| | |
|-------------------------------------------------------------------------------------------------------------------------|----|
| Supplementary Methods | 1 |
| Systematic review search terms | 1 |
| Statistical model to estimate <i>V. cholerae</i> positivity | 2 |
| Supplementary Figures | 3 |
| Figure S1. Data coverage by geography | 3 |
| Figure S2. Data coverage over time | 4 |
| Figure S3. <i>Vibrio cholerae</i> positivity by incidence and suspected case characteristics | 5 |
| Figure S4. Posterior distributions of <i>V. cholerae</i> positivity | 6 |
| Supplementary Tables | 7 |
| Table S1. Priors used in the latent class meta-analysis to estimate sensitivity and specificity of each diagnostic test | 7 |
| Table S2. Characteristics of suspected cholera cases reported by each study | 8 |
| Table S3. Estimated sensitivity and specificity of each diagnostic test | 8 |
| Table S4. Estimated underlying <i>V. cholerae</i> positivity | 9 |
| Table S5. Odds of <i>V. cholerae</i> positivity by age and outbreak context | 9 |
| Table S6. Odds of <i>V. cholerae</i> positivity by categorical age in case definition | 10 |
| Table S7. Sensitivity analysis of the odds of <i>V. cholerae</i> positivity by age and outbreak context | 10 |
| References | 11 |

Supplementary Methods

Systematic review search terms

We searched PubMed, Embase, Scopus, Google Scholar and Medrxiv on October 16, 2021 using the following search terms:

Pubmed (4034 results):

(efficacy[Title/Abstract] OR effectiveness[Title/Abstract] OR surveillance[Title/Abstract] OR outbreak*[Title/Abstract] OR diagnos*[Title/Abstract] OR confirm*[Title/Abstract] OR incidence[Title/Abstract] OR PCR[Title/Abstract] OR "rapid diagnostic"[Title/Abstract] OR "rapid detection"[Title/Abstract] OR "rapid test*" [Title/Abstract] OR RDT[Title/Abstract] OR RDTs OR culture[Title/Abstract]) AND (cholera*[Title/Abstract] OR cholera[MeSH]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])

Embase (4653 results):

(effectiveness:ab,ti OR efficacy:ab,ti OR surveillance:ab,ti OR outbreak*:ab,ti OR diagnos*:ab,ti OR confirm*:ab,ti OR incidence:ab,ti OR pcr:ab,ti OR 'rapid diagnostic':ab,ti OR 'rapid detection':ab,ti OR 'rapid test*':ab,ti OR rdt:ab,ti OR rdts:ab,ti OR culture:ab,ti) AND cholera*:ab,ti AND [2000-2021]/py NOT ('animal'/exp NOT 'human'/exp)

Scopus (2415 results):

TITLE-ABS(effectiveness OR efficacy OR surveillance OR outbreak* OR diagnos* OR confirm* OR incidence OR pcr OR "rapid diagnostic" OR "rapid detection" OR "rapid test*" OR rdt OR culture) AND TITLE-ABS (cholera*) AND PUBYEAR > 2000 AND NOT INDEX (embase)

Google scholar (2180 results exported):

(cholera OR cholerae) AND (effectiveness OR efficacy OR surveillance OR outbreak OR outbreaks OR confirm OR confirmed OR confirmation OR diagnose OR diagnosed OR diagnosis OR incidence OR PCR OR culture OR "rapid test" OR "rapid tests" OR "rapid diagnostic" OR "rapid detection" OR RDT OR RDTs)

Medrxiv (13 results exported):

Simply searched "cholera" and reviewed all the 287 results, 13 were exported to endnote for possible inclusion.

Statistical model to estimate *V. cholerae* positivity

To estimate the pooled proportion of suspected cases that are true *V. cholerae* infections, we performed a hierarchical meta-analysis using CmdStanR version 0.5.2 as an interface to Stan for R [1,2].

In this model, $p_{obs}(i, j)$, probability of observing a positive test result by test j in study i is a function of the true percent positive $p_{pos}(i)$ of study i and test sensitivity θ^+ and specificity θ^- of test j :

$$p_{obs}(i, j) = p_{pos}(i) \theta^+(j) + (1 - p_{pos}(i)) (1 - \theta^-(j))$$
$$n_{obs}(i, j) \sim \text{Binomial}(N_{obs}(i, j), p_{obs}(i, j)),$$

where $N_{obs}(i, j)$ is the number of suspected cases tested and $n_{obs}(i, j)$ is the number that tested positive. The underlying pooled percent positive p_{pos} across all studies was modeled as a function of covariates X (i.e. sampling strategy for test, age constraint in case definition) and a random effect by study, $\varepsilon(i)$:

$$\text{logit}(p_{pos}(i)) = X(i) \beta + \varepsilon(i)$$

We used a $Normal(0,1)$ prior on both the covariate coefficients and the random effect and used the following priors to pool positivity:

$$\text{logit}(p_{pos}(i)) \sim \text{Normal}(\mu_{\text{logit}(p_{pos})}, \sigma_{\text{logit}(p_{pos})})$$
$$\mu_{\text{logit}(p_{pos})} \sim \text{Normal}(0,1)$$
$$\sigma_{\text{logit}(p_{pos})} \sim \text{Normal}(0,1)$$

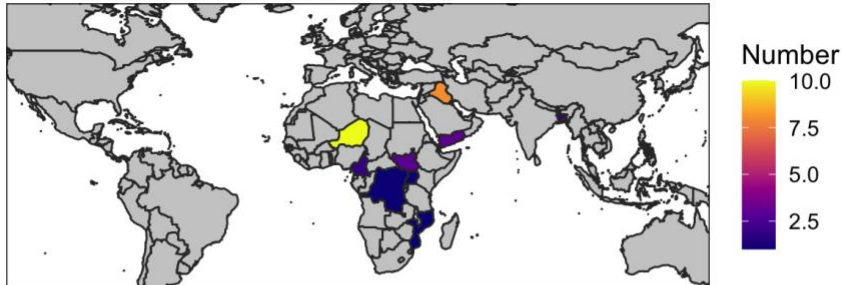
We used 1000 draws from the posterior distribution of sensitivity and specificity in the JAGS model output to incorporate uncertainty in test performance. In order to examine the effects of the priors, we increased their variance and used $Normal(0,2)$ priors in sensitivity analysis. We ran 8000 total iterations for all models implemented with Stan, including 4 chains each with 2000 sampling iterations and 2000 warm-up iterations. Convergence of all models was assessed by R-hat values and sampling iterations were increased to 4000 per chain for models that did not initially converge.

Supplementary Figures

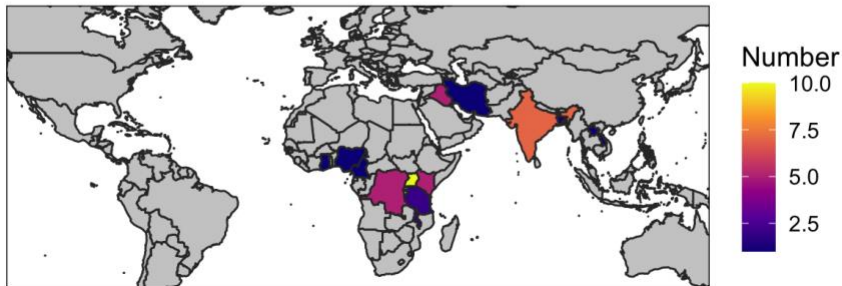
Figure S1. Data coverage by geography

Number of study-periods in the primary dataset at each administrative level as defined by GADM (<https://gadm.org/>) by country. An individual study has >1 "study-periods" if it reports cholera positivity for >1 time frame. Countries with >10 study-periods displayed as 10. Haiti includes 1 study at the national level, 2 at the first administrative division, and 4 at the second.

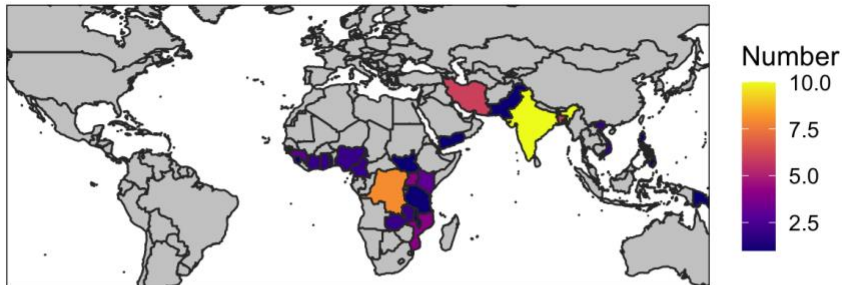
A Admin0



B Admin1



C Admin2



D Admin3

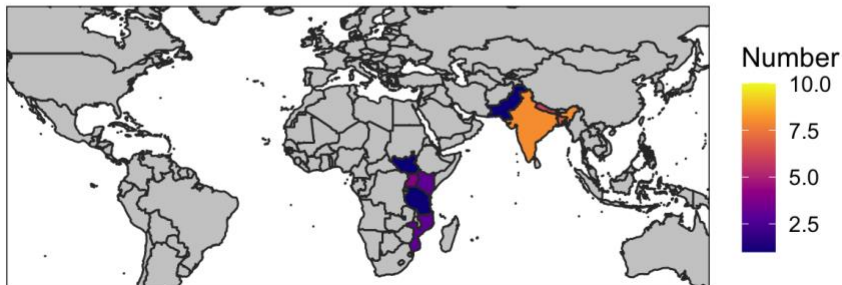
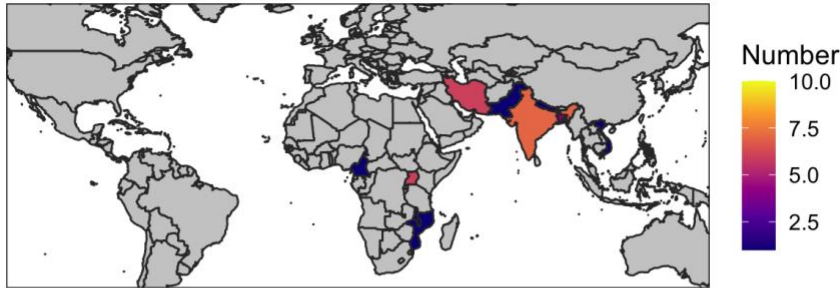


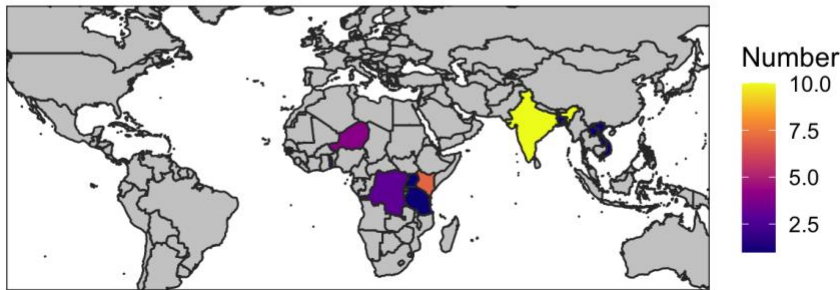
Figure S2. Data coverage over time

Number of study-location-periods in the primary dataset at each administrative level as defined by GADM (<https://gadm.org/>) by country within different time periods. An individual study has >1 "study-location-periods" if it reports cholera positivity for >1 time frame and/or for >1 administrative level. Countries with >10 study-location-periods displayed as 10. Year represents the year sampling was completed. Excludes 8 studies missing a study end date. Haiti includes 3 studies during 2010-2014 and 4 during 2015-2019.

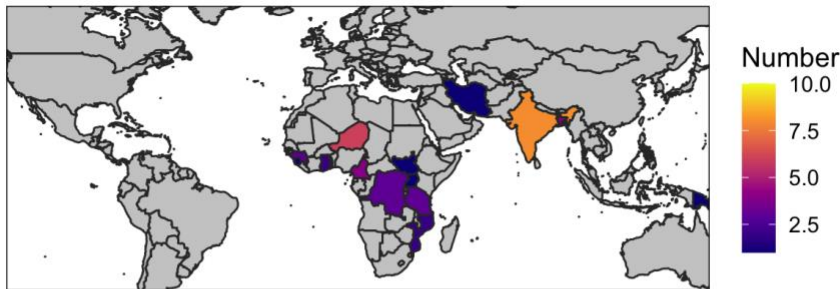
A 1997~2004



B 2005~2009



C 2010~2014



D 2015~2019

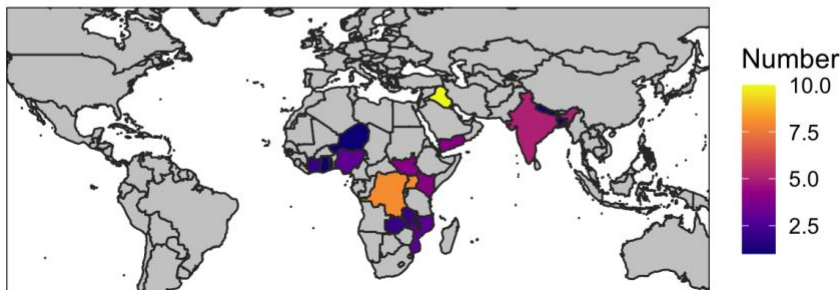


Figure S3. *Vibrio cholerae* positivity by incidence and suspected case characteristics
 Relationship between reported *V. cholerae* positivity and **A)** proportion of suspected cholera cases tests that were under 5 years of age, **B)** suspected cholera incidence rate per 10,000 at each study site in Africa, **C)** proportion of suspected cases severely dehydrated, and **D)** proportion of suspected cases on antibiotics prior to testing. Size of the points is proportional to the number of cases tested, and shapes indicate which diagnostic test was used to confirm *V. cholerae* infection. Confidence intervals for Spearman rank correlation coefficients estimated using bootstrapping (nrep=1000). Smoothing method is loess (without weights). In **B)** estimated suspected cholera incidence rates in Africa for 2010-2016 [3] were aggregated to the administrative division that best represented each study's catchment area by dividing the total estimated cholera cases in each area by its estimated population.

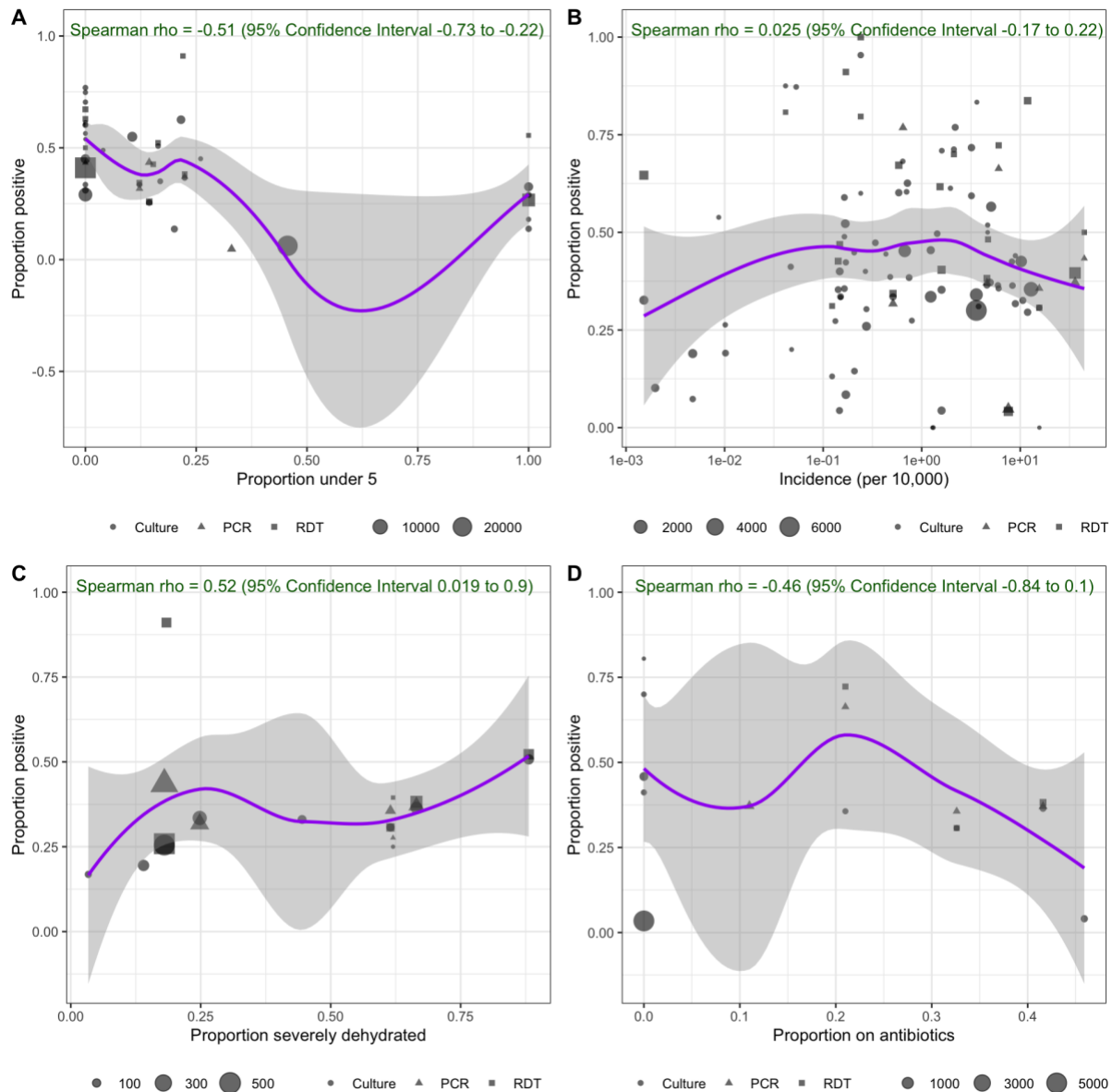
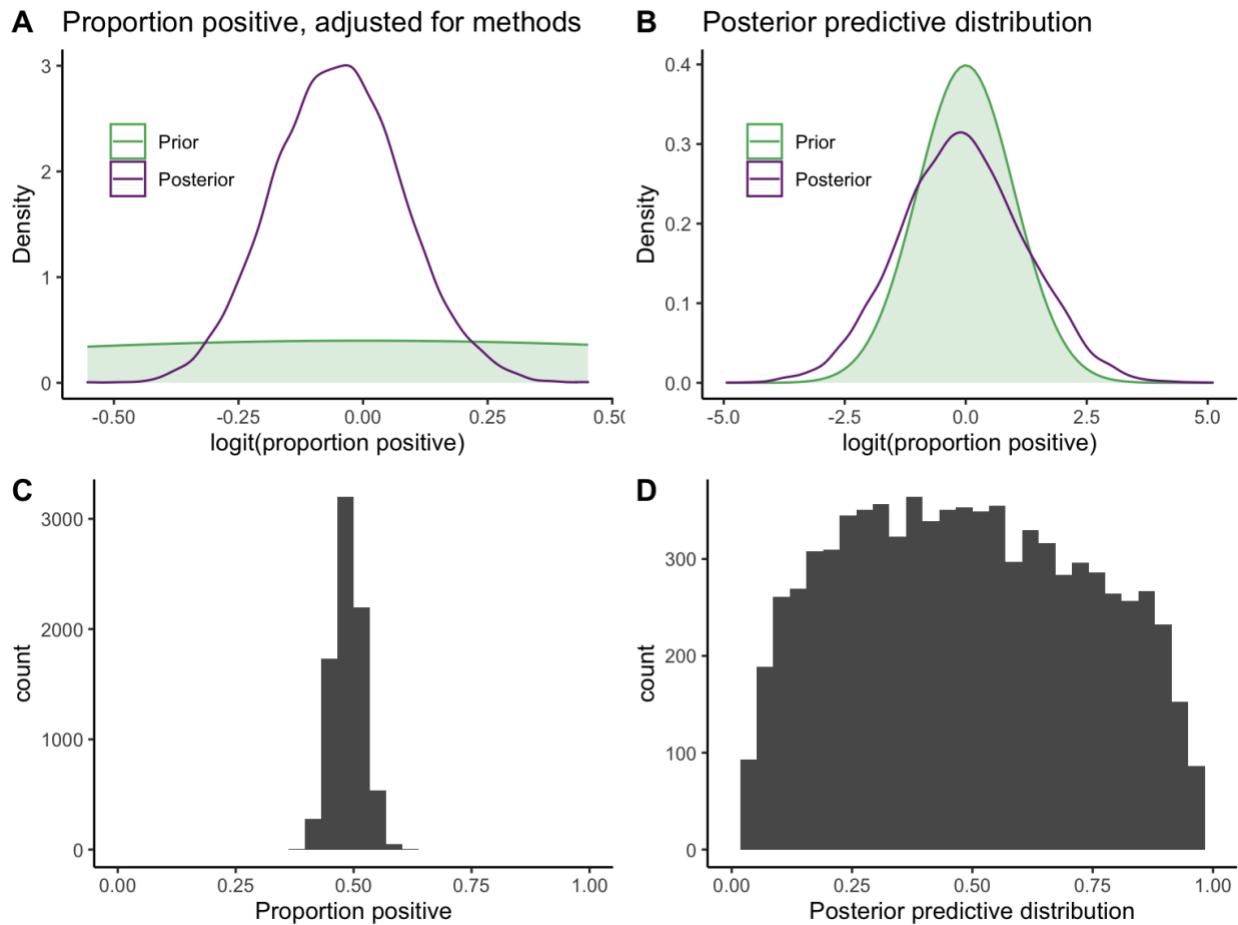


Figure S4. Posterior distributions of *V. cholerae* positivity

Posterior distributions of *V. cholerae* positivity estimated using the random-effects model adjusted for diagnostic test sensitivity and specificity and study methods (i.e., sampling quality and age minimum in case definition), corresponding to results in Table 2. **A)** Prior (green) and posterior (purple) distributions of positivity in logit space. **B)** Prior (green) and posterior predictive (purple) distributions of positivity in logit space. **C)** Histogram of estimated positivity, or the true underlying proportion positive, based on studies in the meta-analysis. **D)** Histogram of the posterior predictive distribution of positivity, or what we predict the underlying proportion positive would be in a new hypothetical study site.



Supplementary Tables

Table S1. Priors used in the latent class meta-analysis to estimate sensitivity and specificity of each diagnostic test

| | Prior distribution in JAGS model | Lower bound of truncation | Upper bound of truncation |
|------------------------------------------------------------|-----------------------------------------------------------|---------------------------|---------------------------|
| Latent prevalence of cholera infection π | $\pi \sim \text{beta}(1,1)$ | none | none |
| Sensitivity of culture $\mu_{\text{sen}(\text{culture})}$ | $\mu_{\text{sen}(\text{culture})} \sim \text{beta}(1,1)$ | 0.3 | 1 |
| Sensitivity of PCR $\mu_{\text{sen}(\text{PCR})}$ | $\mu_{\text{sen}(\text{PCR})} \sim \text{beta}(1,1)$ | 0.5 | 1 |
| Sensitivity of RDT $\mu_{\text{sen}(\text{RDT})}$ | $\mu_{\text{sen}(\text{RDT})} \sim \text{beta}(1,1)$ | 0.5 | 1 |
| Specificity of culture $\mu_{\text{spec}(\text{culture})}$ | $\mu_{\text{spec}(\text{culture})} \sim \text{beta}(1,1)$ | 0.8 | 1 |
| Specificity of PCR $\mu_{\text{spec}(\text{PCR})}$ | $\mu_{\text{spec}(\text{PCR})} \sim \text{beta}(1,1)$ | 0.8 | 1 |
| Specificity of PCR $\mu_{\text{spec}(\text{RDT})}$ | $\mu_{\text{spec}(\text{RDT})} \sim \text{beta}(1,1)$ | 0.5 | 1 |

Table S2. Characteristics of suspected cholera cases reported by each study

| Characteristic | Reported | Total studies | Percent |
|-----------------------------------------------------------------------------------------|----------|---------------|---------|
| Proportion vaccinated | No | 115 | 95.8 |
| | Yes | 5 | 4.2 |
| Proportion under 5 years old | No | 86 | 71.7 |
| | Yes | 34 | 28.3 |
| Proportion under 15 years old | No | 99 | 82.5 |
| | Yes | 21 | 17.5 |
| Proportion severely dehydrated | No | 110 | 91.7 |
| | Yes | 10 | 8.3 |
| Proportion moderately dehydrated | No | 111 | 92.5 |
| | Yes | 9 | 7.5 |
| Proportion female | No | 93 | 77.5 |
| | Yes | 27 | 22.5 |
| Proportion that received antibiotics prior to confirmation test | No | 110 | 91.7 |
| | Yes | 10 | 8.3 |
| Proportion under 5 years old & proportion severely dehydrated | No | 115 | 95.8 |
| | Yes | 5 | 4.2 |
| Proportion under 5 years old, severely dehydrated, & received antibiotics prior to test | No | 119 | 99.2 |
| | Yes | 1 | 0.8 |

Table S3. Estimated sensitivity and specificity of each diagnostic test

Estimate is median sensitivity and specificity pooled across four studies that reported results for all diagnostics tests, as described in Methods. Parentheses show 95% Credible Interval.

| Measure | Test | Estimate (%) |
|-------------|---------|--------------------|
| Sensitivity | Culture | 82.0 (37.5 - 98.7) |
| | PCR | 85.1 (53.6 - 98.9) |
| | RDT | 90.4 (55.2 - 99.5) |
| Specificity | Culture | 94.3 (81.5 - 99.6) |
| | PCR | 94.2 (81.1 - 99.7) |
| | RDT | 88.9 (54.8 - 99.4) |

Table S4. Estimated underlying *V. cholerae* positivity

"Unadjusted" is mean percent positive (95% confidence interval) from random effects meta-analysis implemented using R package meta without any adjustments. "Adjusted" is the underlying true percent positive (95% credible interval) estimated using the Bayesian hierarchical model, adjusted for sensitivity/specificity of the tests, age in case definition, sampling quality, and whether surveillance was initiated in response to an outbreak.

| Version | Positivity (%) |
|------------------------------------------------------------------------------------------|----------------|
| Unadjusted study data | 41 (37 - 46) |
| Adjusted for test performance | 48 (42 - 54) |
| Adjusted for test performance & methods | 49 (43 - 55) |
| Adjusted for test performance, methods, & outbreak setting | 49 (43 - 56) |
| Adjusted for test performance, methods, & outbreak setting, increased variance on priors | 49 (43 - 56) |

Table S5. Odds of *V. cholerae* positivity by age and outbreak context

Odds that a suspected cholera case seeking testing or care has a true *V. cholerae* O1/O139 infection with low sampling quality compared to high sampling quality, with increasing minimum age set in suspected case definition, and with surveillance initiated in response to an outbreak compared to non-outbreak surveillance (i.e., routine or post-vaccination surveillance).

| Variable | Odds ratio |
|--------------------------------|--------------------|
| Low sampling quality | 1.48 (0.92 - 2.38) |
| Minimum age in case definition | 1.16 (1.01 - 1.34) |
| Outbreak surveillance | 1.64 (1.06 - 2.52) |

Table S6. Odds of *V. cholerae* positivity by categorical age in case definition

Odds that a suspected cholera case seeking care has a true *V. cholerae* infection by minimum age set in suspected case definition. Coefficients correspond to random effects models run with age as a categorical variable (reference group = no age minimum age explicitly set in case definition). All other model settings and covariates are the same as in Table 3.

| Variable | Odds ratio |
|--------------------------------|---------------------|
| Low sampling quality | 1.47 (0.91 - 2.47) |
| Min age in case definition: 1 | 1.19 (0.48 - 2.98) |
| Min age in case definition: 2 | 1.23 (0.66 - 2.28) |
| Min age in case definition: 5 | 1.46 (0.64 - 3.19) |
| Min age in case definition: 10 | 2.57 (0.52 - 12.75) |
| Outbreak surveillance | 1.7 (1.09 - 2.65) |

Table S7. Sensitivity analysis of the odds of *V. cholerae* positivity by age and outbreak context

Odds that a suspected cholera case seeking testing or care has a true *V. cholerae* O1/O139 infection with low sampling quality compared to high sampling quality, with increasing minimum age set in suspected case definition, and with surveillance initiated in response to an outbreak compared to non-outbreak surveillance (i.e., routine or post-vaccination surveillance) in models where variance was increased on all priors.

| Variable | Odds ratio |
|--------------------------------|--------------------|
| Low sampling quality | 1.48 (0.92 - 2.38) |
| Minimum age in case definition | 1.16 (1.01 - 1.34) |
| Outbreak surveillance | 1.64 (1.06 - 2.52) |

References

1. Gabry J, Češnovar R, Bales B, et al. R interface to CmdStan. 2022. Available at: <https://mc-stan.org/cmdstanr/>. Accessed 17 August 2022.
2. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020; Available at: <https://www.R-project.org/>.
3. Lessler J, Moore SM, Luquero FJ, et al. Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. *The Lancet* **2018**; 391:1908–1915.