

Cholera at the Crossroads: The Association Between Endemic Cholera and National Access to Improved Water Sources and Sanitation

Benjamin L. Nygren,* Anna J. Blackstock, and Eric D. Mintz

Division of Foodborne, Waterborne and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract. We evaluated World Health Organization (WHO) national water and sanitation coverage levels and the infant mortality rate as predictors of endemic cholera in the 5-year period following water and sanitation coverage estimates using logistic regression, receiver operator characteristic curves, and different definitions of endemicity. Each was a significant predictor of endemic cholera at $P < 0.001$. Using a value of 250 for annual cases reported in 3 of 5 years, a national water access level of 71% has 65% sensitivity and 65% specificity in predicting endemic cholera, a sanitation access level of 39% has 63% sensitivity and 62% specificity, and an infant mortality rate of 65/1,000 has 67% sensitivity and 69% specificity. Our findings reveal the tradeoff between sensitivity and specificity for these predictors of endemic cholera and highlight the substantial uncertainty in the data. More accurate global surveillance data will enable more precise characterization of the benefits of improved water and sanitation.

INTRODUCTION

Cholera is an important, recurring source of morbidity and mortality in many developing countries. Illness is caused by infection with toxigenic *Vibrio cholerae* O1 or O139 bacteria, most often acquired through ingestion of fecally contaminated water or food. Symptoms include nausea, vomiting, and profuse watery diarrhea. Severe disease causes rapid dehydration, is marked by loss of skin turgor and sunken eyes, and can result in death within hours if untreated. Proper treatment with rehydration therapy is simple, inexpensive, and highly effective in preventing mortality. However, preventing new infections is difficult in settings with poor water, sanitation, and hygiene conditions. Ongoing transmission can lead to large outbreaks requiring an extensive clinical response and often straining limited resources. Furthermore, many infections are asymptomatic, or result in mild or moderate disease, but still serve to reintroduce the agent into the environment where it can infect others. Effective and enduring prevention at the national level requires costly improvements in water and sanitation infrastructure to ensure broad access to safe drinking water and proper disposal of human excreta. Cholera therefore persists in many developing countries lacking the resources for such major investments. Other areas, such as Haiti in 2010 and parts of Latin America in 1991, were historically isolated from cholera-endemic areas as a result of geography and other factors, only to experience large, and in some instances, persistent epidemics when toxigenic *Vibrio cholerae* O1 was introduced to immunologically naive populations with limited access to safe water and sanitation. In contrast, most developed countries have been free of endemic and epidemic cholera for many decades after major improvements to water and sanitation infrastructure, despite the frequent occurrence of imported, travel-associated cases.

The Millennium Development Goals (MDGs) include reducing by 50% the proportion of the world's population without sustainable access to safe drinking water and basic sanitation by 2015.¹ Although the global goal related to water was achieved 5 years ahead of schedule, global progress on

improving sanitation has been considerably slower.² As these important improvements are made, it is timely to consider their impact on cholera endemicity. To attempt to better define the levels of infrastructure improvements associated with endemic cholera, we investigated the association between national coverage levels for improved water and sanitation and the presence or absence of endemic cholera. We also evaluated childhood mortality as a predictor of cholera because it offers a strong indicator of national development and has been associated with cholera endemicity in the literature.³

MATERIALS AND METHODS

Data sources and time period. We used publicly available data from the World Health Organization (WHO) on national estimates of water and sanitation coverage, childhood mortality rates, and cholera case surveillance.⁴ Final data were downloaded on April 1, 2014, and analyzed with SAS 9.3 (SAS Institute Inc., Cary, NC). Our predictors were national access to improved water and sanitation and infant mortality rates. Our outcome was endemic cholera, based on WHO national cholera surveillance data. The analysis time period was 1990–2010 with predictor data for 1990, 1995, 2000, and 2005 and annual national cholera case counts from 1991 to 2010. Additional cases were added based on reported cases in The Bulletin of the World Health Organization.⁵ The relationship between predictor variables and cholera incidence was investigated by modeling cholera over a 5-year period versus the predictor variable from the preceding year (e.g., the sanitation estimate from 1995 was used to predict endemic cholera during the period from 1996 to 2010).

Predictors. An improved water source is considered protected from fecal and other contamination. The WHO/United Nations Children's Fund (UNICEF) Joint Monitoring Program (JMP) for Water Supply and Sanitation defines improved water sources as piped water into dwelling, yard, or plot; public tap or standpipe; tubewell or borehole; protected spring; protected dug well or rainwater collection.^{2,6} Unimproved sources include an unprotected dug well or spring; cart with tank or drum; tanker truck; surface water (river dam, lake, pond, stream, canal, or irrigation channel) or bottled water in most cases. Improved sanitation includes the use of flush or pour facilities to a piped sewer system, septic tank or

*Address correspondence to Benjamin L. Nygren, Centers for Disease Control and Prevention, 1600 Clifton Road NE, MS-C09, Atlanta, GA 30329. E-mail: bnygren@cdc.gov

pit latrine; ventilated improved pit latrine; pit latrine with a slab or a composting toilet.^{2,7} Examples of unimproved sanitation include flush or pour to any other type of containment, pit latrine without a slab, open pit, bucket, hanging latrine, any type of shared or public facility, or open defecation. Methods of measurement include nationally representative surveys and are further described by WHO. Point estimates are adjusted as follows: “For each country, survey and census data are plotted on a time series: 1980 to present. A linear trend line, based on the least-squares method, is drawn through these data points to estimate coverage for 1990, 1995, 2000, 2005 . . . The total coverage estimates are based on the aggregate of the population-weighted average of urban and rural coverage numbers.”

Infant mortality is the probability of dying between birth and age 1 per 1,000 live births, defined as “the number of deaths divided by the number of population at risk during a certain period of time . . . a probability of death derived from a life table and expressed as rate per 1,000 live births”.⁸

Outcome. We used national cholera surveillance data case counts to assess the presence of endemic cholera as a binary classification. Although these data are available back to 1949 for certain countries, we restricted our analysis of cholera surveillance to 1991 to 2010 to correspond with the availability of predictor variables, beginning in 1990. The cholera surveillance data include annual counts of cases and deaths. In these data, a case is defined as “Confirmed cholera cases, including those confirmed clinically, epidemiologically, or by laboratory investigation,” and a death is defined simply as “Reported cholera deaths.”^{9,10} We used case data exclusively except for instances in which no values for case counts were submitted, but values for deaths were submitted, and for the one instance in which the number of deaths exceeded the case count. In this instance, we used the value for deaths as the value for case count. Imported cases were omitted from the overall case totals used in this analysis. Where counts of imported cases among the country case total exceeded the country case total, a value of zero was used for case counts. As our focus was endemic cholera, we used case count values for determining whether cholera was present as a binary outcome rather than using overall count as the outcome.

The WHO defines endemic cholera as the occurrence of fecal-culture confirmed cholera diarrhea in a population in at least 3 years of a 5-year period.^{11,12} We used this outcome and an annual case count threshold of 10, 50, 100, 250, 500, and 1,000 for defining endemic cholera to assess how different thresholds affect the predictive ability of national development estimates. By using these different case threshold outcomes, we hoped to better understand any potential association of water, sanitation, and childhood mortality indicators on the general epidemiological picture of cholera in a country.

Analysis. We limited our analysis to countries with a population > 100,000 in 2010 because many smaller nations do not regularly report cholera, lack national estimates of access to improved water sources or sanitation, or feature city-state geography or other unique characteristics that may make them unsuitable for comparison to larger countries. We did not include the WHO Europe region because of the lack of indigenous cholera. Some countries, regardless of population, fail to report cholera cases to WHO, and many did not report zeroes despite regularly reporting periodic, small numbers of cases in other years. Based on information in The WHO Bul-

letin that further describes reported cholera, additional cholera cases were added for Bangladesh, Indonesia, North Korea, Laos, Myanmar, Namibia, Pakistan, Viet Nam, and Yemen.⁵ We treated missing data points as zeroes, unless a country did not report data at all, in which case it was excluded from the analysis. Because the absence of cholera was rarely actively reported, requiring complete annual observations from each country for inclusion in the analysis would have resulted in a data set limited almost entirely to cholera-reporting countries, which might not have allowed us to compare the absence of cholera with varying degrees of incidence and endemicity.

To statistically evaluate water and sanitation coverage levels as predictors of endemic cholera in the 5-year period following water and sanitation coverage estimates, we performed logistic regression and calculated sensitivity and specificity estimates. We also produced receiver operator characteristic (ROC) curves, or plots of sensitivity versus one minus specificity, and calculated values for the area under the ROC curves (AUC) to assess the predictive ability of each predictor. The AUC measures the ability of the test to correctly predict a binary outcome. A higher AUC indicates that the test is better overall at predicting the outcome; a perfect predictive test would have 100% sensitivity, 100% specificity, and an AUC of 1, whereas a test with no predictive ability would have an AUC of 0.5. In practical applications, there is a tradeoff between sensitivity and specificity. We calculated, using each value of the predictor as the cutoff, the sensitivity (the percentage of countries for which the coverage level cut-off correctly predicted that endemic cholera would be present) and specificity (the percentage of countries for which the coverage level cut-off correctly predicted that endemic cholera would not be present). We then compared ROC curves for each predictor using different case count threshold definitions of endemic cholera. To account for repeated measures—that is, each country contributes values of predictors for up to four different time points, and values for each of these time points are not independent of the other time points for the same country—we incorporated a correlation structure using generalized estimating equations. We also compared these results to results from logistic regression without a correlation structure to gauge the effect of repeated measures. After using AUC to evaluate national access to improved water sources, access to improved sanitation, and infant mortality rate, we evaluated each of the three predictors in a logistic model.

RESULTS

Our analysis included 97 countries. Each country contributed up to four observations for each of the predictor variables; six (6%) countries had three time point estimates of water coverage data available, and one (1%) had two values. Twelve (12%) countries had three sanitation estimates available, 15 (15%) had two estimates, one (1%) had one estimate, and one (1%) did not report sanitation estimates. All had infant mortality estimates available for each of the four time points. Cholera incidence data were included for 97 countries, each with up to 20 annual estimates from 1991 (or later year of national independence, if applicable) to 2010. These data are less complete; data were reported for 1,053 (54%) of 1937 possible country-year time points. The absence of cholera was rarely reported; in 74 (4%) instances, zeroes were

TABLE 1

Median values of predictors, by year, and proportion of countries with endemic cholera under two definitions over the following 5-year period

Predictor	1990	1995	2000	2005
Water	70%	74%	79%	79%
Sanitation	40%	40%	47%	43%
Infant Mortality*	64	62	59	49
Proportion of nations considered endemic	1991–1995 N (%)	1996–2000 N (%)	2001–2005 N (%)	2006–2010 N (%)
WHO (1)	55 (57)	54 (56)	44 (45)	41 (42)
Modified (250)	37 (38)	37 (38)	23 (24)	24 (25)

*Deaths under age 1 per 1,000 persons.

reported, compared with 979 (51%) instances where no data were reported.

In the subset of data we used, differences in predictors over time followed the results described in the United Nations (UN) and JMP progress reports, including increases in access to improved water sources, slow increases in sanitation access, and overall decreases in infant mortality (Table 1).^{13,14} Cholera was endemic under the WHO definition in 57% of countries in 1991–1995 and decreased over the 5-year time periods to 42% of countries in 2006–2010. Under a definition of 250 cases per year in 3 of 5 years, cholera was endemic in 38% of countries in 1991–1995 and decreased to 25% in 2006–2010.

National water and sanitation coverage estimates and the infant mortality rate were each a significant predictor of endemic cholera under all definitions at $P < 0.001$. Increasing the case count used to define endemic cholera for values up to 250 resulted in increased area under the ROC curve, above

which little or no increase in AUC was seen. The AUC ranged from 0.65 (95% confidence interval [CI] = 0.59–0.70) for water and 0.62 (0.56–0.68) for sanitation using the WHO definition of any reported cholera three or more times in a 5-year period, to 0.72 (0.67–0.77) for water and 0.70 (0.64–0.75) for sanitation using 250 cases per year in 3 or more of 5 consecutive years. Using a value of 250 for annual cases reported, a national water access level of 71% has 65% sensitivity and 65% specificity in predicting endemic cholera, and a sanitation access level of 39% has 63% sensitivity and 62% specificity (Figures 1–3, Table 2). Infant mortality had the largest AUC, ranging from 0.68 (0.62–0.73) for the WHO definition of endemic to 0.76 (0.72–0.81) in predicting endemic cholera with a case threshold of 250. An infant mortality value of 65 has 67% sensitivity and 69% specificity. We used the endemic threshold of 250 for univariable logistic regression as it provided the highest AUC values. All three predictors were significant at $P < 0.0001$.

DISCUSSION

National estimates of water and sanitation access and, in particular, infant mortality are strong individual statistical predictors of the persistence of cholera. There are, however, several reasons why it is challenging to confidently and precisely estimate development levels sufficient to prevent recurring cholera using these data. Limitations of the data include potential overestimation of the population served by improved water sources and of the microbiologic safety of water from sources classified as improved.¹³ In addition, the varying methods by which water and sanitation coverage estimates are derived and the effect of fitting predictor data to a trend line to generate

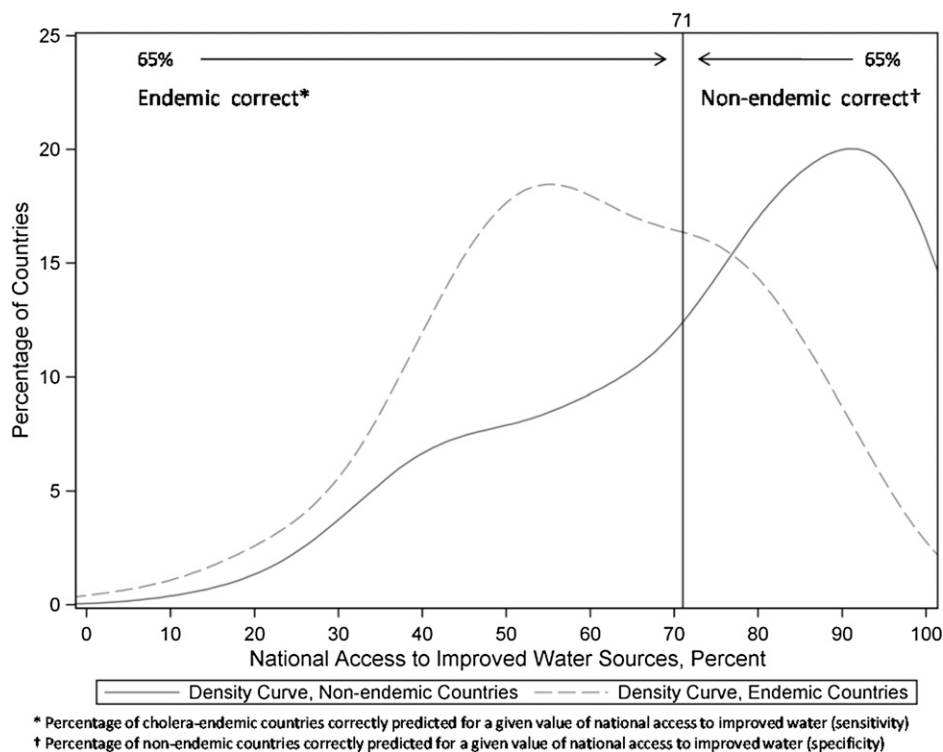


FIGURE 1. Tradeoff in sensitivity and specificity for access to improved water sources as predictor of cholera endemicity.

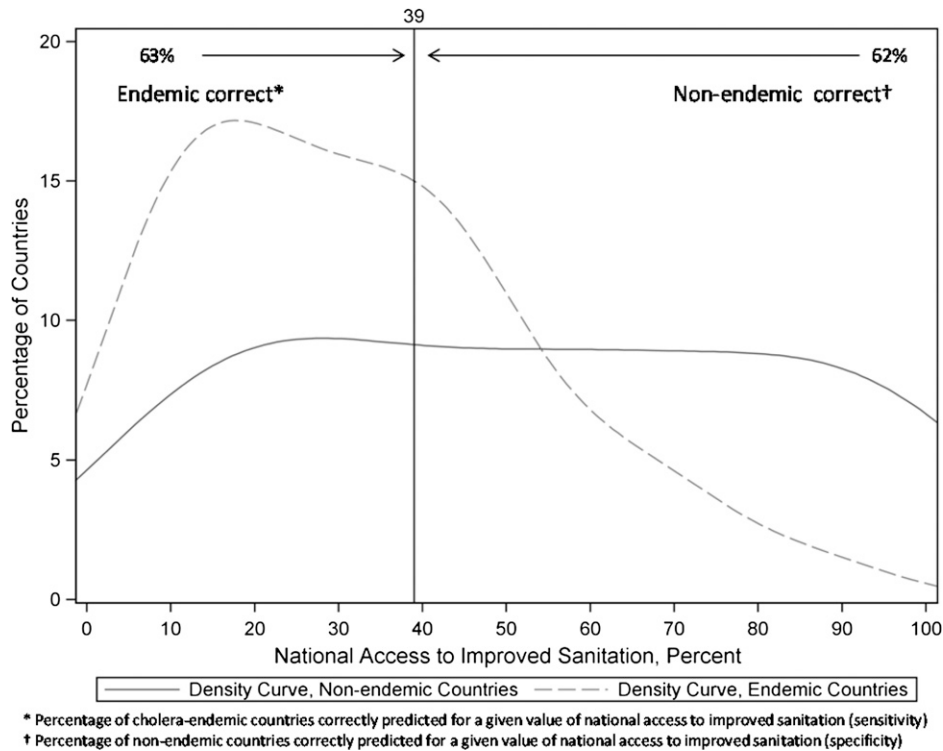


FIGURE 2. Tradeoff in sensitivity and specificity for access to improved sanitation as predictor of cholera endemicity.

estimates may introduce further imprecision and potential biases. Furthermore, the predictor data we required for a global focus and national-level comparison by time points lack stratification by important population characteristics such as socioeconomic classification. Singular national estimates of

access to improved water sources and sanitation may increase overall, but improvements might not have reached the groups most susceptible to endemic cholera within a country as a result of geographic, economic, or other reasons. Water and sanitation data stratified by socioeconomic status and other

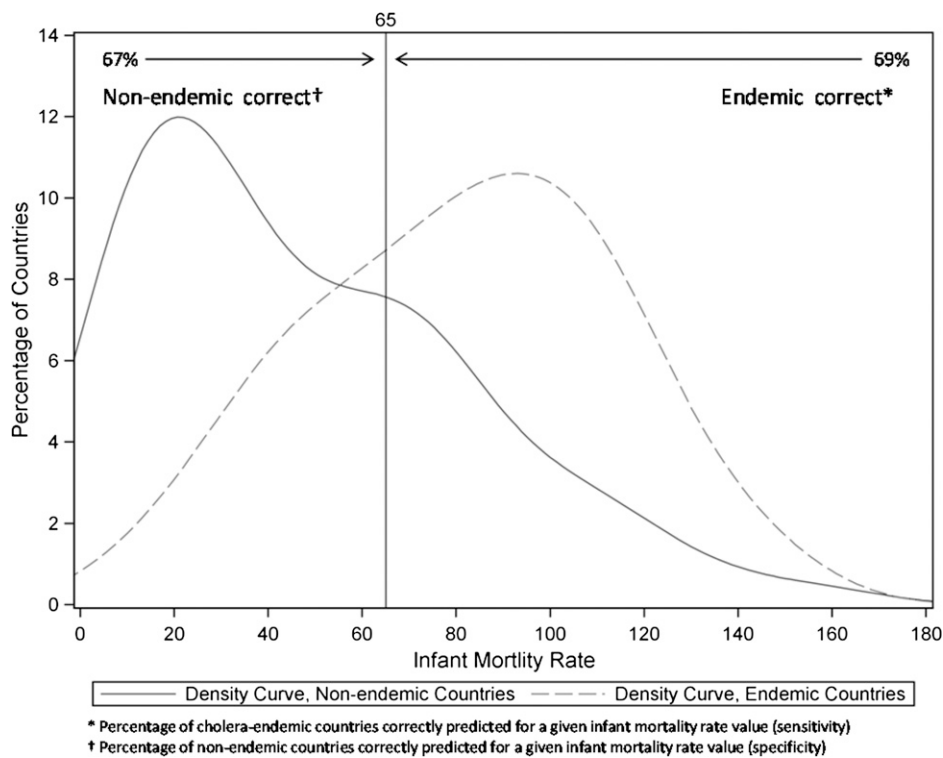


FIGURE 3. Tradeoff in sensitivity and specificity for infant mortality rate as predictor of cholera endemicity.

TABLE 2
Sensitivity and specificity values for different predictor value thresholds

Predictor	% Of cholera-free countries correctly classified as cholera-free (specificity)	% Of cholera-free countries incorrectly classified as cholera-endemic (1-specificity)	% Of cholera-endemic countries correctly classified as cholera-endemic (sensitivity)	% Of cholera-endemic countries incorrectly classified as cholera-free (1-sensitivity)
Water = 93%	28	72	100	0
Water = 81%	46	54	87	13
Water = 71%	65	35	65	35
Water = 61%	75	25	50	50
Water = 14%	100	0	1	99
Sanitation = 91%	17	83	100	0
Sanitation = 49%	53	47	83	17
Sanitation = 39%	62	38	63	38
Sanitation = 29%	68	32	51	49
Sanitation = 2%	100	0	0	100
Infant mortality = 11*	13	87	100	0
Infant mortality = 65*	69	31	67	33
Infant mortality = 161*	100	0	0	100

*Deaths under age 1 per 1,000 persons.

characteristics are available for some countries, but there is not enough for a global analysis.

Although infant mortality rate data may be higher quality because of better methods of estimation, it is also a function of other factors less directly related to cholera persistence than population access to safe water and sanitation. Thus, the causal relationship between national infant mortality rate and endemic cholera is less clear compared with that of water and sanitation access and endemic cholera, even though infant mortality is a stronger predictor based on AUC.

In the cholera case counts outcome data, data completeness and poor reporting of the absence of cholera is a notable limitation. The precision of case counts is another concern, including misattribution of watery diarrhea, but our analysis does suggest that using a slightly higher threshold than any cases in 3 of 5 years notably improves the predictive ability of national water and sanitation coverage estimates. And although WHO cautions generally that “case numbers are generally a poor indicator of the burden of disease,” case counts in the hundreds suggest a more significant public health problem than counts in the single digits.⁹ Where few cases are reported, it is important to discern between imported and indigenous cases because imported cases are not a function of the host country’s water and sanitation infrastructure. Logistical challenges in making this determination may affect the completeness of data reported on imported cases. The definition of endemic cholera could perhaps be modified to specify reporting of cases acquired in the country in 3 of 5 years. Substantial underreporting of cholera cases to the WHO has been noted as a result of political, economic, or logistical reasons, and WHO estimates that the reported case counts are about 5–10% of the true total.^{15–20} More comprehensive cholera case count data could be formed by augmenting the data officially reported to WHO with additional national-level data sources, and refinements to counts from other global active and passive infectious disease surveillance systems, media sources, and outbreak reports in the literature. Although we updated our data for several countries (e.g., Bangladesh) where clear documentation is available on the occurrence of cholera and cases were not reported in the original WHO data, we were not able to perform a complete literature review and media search of global cholera incidence

for this analysis. These and other data limitations suggest using caution in attempting to estimate a threshold value for national access to safe water or improved sanitation above which endemic cholera is no longer likely to be seen. For instance, a water access figure of 93% correctly predicted all of the cholera endemic countries, but only 28% of non-endemic countries had a figure this high (Table 2). Finally, we did not take into account geographic or environmental factors, such as temperature, rainfall, and the presence of suitable lacustrine, estuarine, or marine reservoirs for long-term survival of *V. cholerae* outside the host, which must also contribute to the risk of cholera endemicity. Refining and enhancing data on cholera incidence and on national and sub-national water and sanitation coverage, and including variables for additional risk factors for cholera in modeling, will improve the usefulness of estimates of coverage rates above which endemic cholera is unlikely to occur. Nevertheless, the results of these preliminary analyses may potentially be helpful in prioritizing cholera prevention or response efforts, for example in focusing and improving the potential value of a pre-emptive cholera vaccine campaign, or identifying high-priority national or sub-national targets for urgent investment in improved water and sanitation access.

In conducting this analysis, we sought to better understand the relationship between these national development indicators and the presence of endemic cholera. Our findings confirm the strong relationship between these development indicators and endemic cholera and detail the tradeoff between sensitivity and specificity when they are used as predictors of endemic cholera. However, they also reveal a substantial area of uncertainty. More accurate and comprehensive global surveillance data will enable more precise characterization of the benefits of access to improved water and sanitation, including more accurate estimates of the threshold values for endemic cholera, that will better inform policy and development decisions.

Received May 29, 2014. Accepted for publication August 6, 2014.

Published online September 8, 2014.

Acknowledgments: We thank the World Health Organization and national Ministries of Health for providing these data to the public.

Financial support: This work was supported in part by the United States Agency for International Development.

Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention or the United States Agency for International Development.

Authors' addresses: Benjamin L. Nygren, Anna J. Blackstock, and Eric D. Mintz, Division of Foodborne, Waterborne and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, GA, E-mails: bnygren@cdc.gov, hyp9@cdc.gov, and emintz@cdc.gov.

REFERENCES

1. United Nations Department of Public Information UNDoP, 2013. Millennium Development Goals and Beyond 2015: Fact Sheet, 1–2.
2. WHO, 2013. Progress on Sanitation and Drinking-Water - 2013 update, 3–13.
3. Ackers ML, Quick RE, Drasbek CJ, Hutwagner L, Tauxe RV, 1998. Are there national risk factors for epidemic cholera? The correlation between socioeconomic and demographic indices and cholera incidence in Latin America. *Int J Epidemiol* 27: 330–334.
4. WHO, 2013. *Global Health Observatory (GHO) Data Repository*. Available at: <http://www.who.int/gho/database/en/>. Accessed April 1, 2014.
5. Ali M, Lopez AL, You YA, Kim YE, Sah B, Maskery B, Clemens J, 2012. The global burden of cholera. *Bull World Health Organ* 90: 209–218A.
6. WHO. *Indicator Definition: Population Using Improved Drinking-Water Sources (%)*. Available at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=8. Accessed April 1, 2014.
7. WHO. *Indicator Definition: Population Using Improved Sanitation Facilities (%)*. Available at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=9. Accessed April 1, 2014.
8. WHO. *Indicator Definition: Infant Mortality Rate*. Available at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=1. Accessed April 1, 2014.
9. WHO. *Indicator Definition: Number of Reported Cases of Cholera*. Available at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=42. Accessed April 1, 2014.
10. WHO. *Indicator Definition: Number of Reported Deaths of Cholera*. Available at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=2446. Accessed April 1, 2014.
11. WHO, 2009. Meeting of the Strategic Advisory Group of Experts on immunization, October 2009: conclusions and recommendations. *Wkly Epidemiol Rec* 84: 526–528.
12. WHO, 2010. Cholera vaccines: WHO position paper. *Wkly Epidemiol Rec* 85: 127.
13. WHO, 2013. Progress on Sanitation and Drinking-Water - 2012 update, 4.
14. UNICEF, 2011. *United Nations Inter-Agency Group for Child Mortality Estimation: Levels and Trends in Child Mortality*. New York: UNICEF.
15. Bwire G, Malimbo M, Maskery B, Kim YE, Mogasale V, Levin A, 2013. The burden of cholera in Uganda. *PLoS Negl Trop Dis* 7: e2545.
16. Sarkar BL, Kanungo S, Nair GB, 2012. How endemic is cholera in India? *Indian J Med Res* 135: 246–248.
17. Griffith DC, Kelly-Hope LA, Miller MA, 2006. Review of reported cholera outbreaks worldwide, 1995–2005. *Am J Trop Med Hyg* 75: 973–977.
18. WHO, 2000. WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases.
19. WHO, 2009. Cholera: global surveillance summary, 2008. *Wkly Epidemiol Rec* 84: 309–324.
20. WHO, 2014. *Cholera Surveillance and Number of Cases*. Available at: <http://www.who.int/topics/cholera/surveillance/en/>. Accessed April 1, 2014.