Estimating effects of improved drinking water and sanitation on cholera

Andrew J. Leidner and Naveen C. Adusumilli

ABSTRACT

Demand for adequate provision of drinking-water and sanitation facilities to promote public health and economic growth is increasing in the rapidly urbanizing countries of the developing world. With a panel of data on Asia and Africa from 1990 to 2008, associations are estimated between the occurrence of cholera outbreaks, the case rates in given outbreaks, the mortality rates associated with cholera and two disease control mechanisms, drinking-water and sanitation services. A statistically significant and negative effect is found between drinking-water services and both cholera case rates as well as cholera-related mortality rates. A relatively weak statistical relationship is found between the occurrence of cholera outbreaks and sanitation services.

Key words | cholera, drinking water, Millennium Development Goals, sanitation

Andrew J. Leidner (corresponding author)
Formerly of the Department of Agricultural Economics,
Texas A&M University,
2124 TAMU, College Station,
TX 77843-2124,
USA
E-mail: *andrew.leidner@gmail.com*

Naveen C. Adusumilli
Water Economist Fellow, Natural Resources Defense Council,
1152 15th Street NW,
Washington, DC 20005,
USA

INTRODUCTION

Economists have long focused substantial effort to understand the relationship between an individual's health and wealth (Grossman 1972). In more recent years, public health and economic development researchers have focused on regional and national issues, to understand the mechanisms which contribute to regions possessing both substandard health and economic poverty (Anand & Ravallion 1993). Efforts have focused on health as it affects economic development (Mayer 2001; Sachs & Malaney 2002; Bloom *et al.* 2004), poverty as it affects health (Tol & Dowlatabadi 2001), and the two issues together as they simultaneously affect each other (Tol 2007; Bonds *et al.* 2009).

Among the illnesses that seem to pervade impoverished regions, the persistence of diarrheal disease poses a particularly tragic problem because well-known and well-executed methods exist in the developed world by which diarrheal disease can largely be prevented and by which many of those that become infected can be medically treated to achieve a full and relatively fast recovery. Diarrheal disease is the second leading cause of death among children less than five years old and can be caused by a wide range of viral, bacteriological, or parasitic organisms (WHO 2013). The morbidity associated with diarrheal illnesses leads to under-nutrition, dehydration and stunted growth among children (Nandy *et al.* 2005). This paper focuses on one particular pathogenic agent *Vibrio cholera* which causes the diarrheal disease more commonly known as cholera.

One of the objectives of the Millennium Development Goals (MDGs) is to halve the proportion of people without access to improved sanitation and drinking water (UN General Assembly 2000). This goal is in part motivated by the recognition that water and sanitation infrastructure plays a role in preventing disease, encouraging economic growth, and ultimately reducing poverty (Prüss-Üstün & Corvalán 2006). Furthermore, drinking water and sanitation improvements become more important public health considerations when viewed in the context of rapidly urbanizing human populations across the globe (Vlahov *et al.* 2007).

With these dynamics in mind, this paper seeks to better understand the relationships among the development of drinking water facilities, sanitation facilities, and the burden of the water-borne disease cholera. In this study, a cross-country and time-series dataset (i.e., a panel dataset) for nations in Asia and Africa has been assembled from several sources, integrating disease infection levels, water and sanitation infrastructure levels, economic wealth indicators,

doi: 10.2166/wh.2013.238

and environmental variables. To our knowledge, no such study exists that estimates the relationship among so many disease-relevant factors and across such a broad geographic range. Moreover, this study is timely and relevant in light of the ongoing efforts of the MDGs to reduce poverty and improve the lives of citizens worldwide. The progress and merits of the MDGs may be up for discussion long into the future, and this study provides an evaluation of one particular facet of the MDGs in the context of cholera, a well-known scourge of the developing world. This study builds a novel dataset to look at these questions. Empirical findings indicate a complex relationship between the varied aspects of cholera-related public health burdens, access to improved drinking water and sanitation systems, and environmental and economic conditions. The remainder of the study will proceed with the following sections: cholera background, methods and empirical approach, data description, results, and conclusion.

Relevant literature and background

Diarrheal disease can originate from a variety of microbiological agents (Jose & Bobadilla 1994) and results in the deaths of possibly more than 1.6 million children annually, mostly in developing countries (WHO & UNICEF 2006). Cholera is one such disease caused by the pathogen *Vibrio cholera*. Cholera case burden numbers between 100,000 and 300,000 recorded cases per year worldwide (WHO 2009a). More recent estimates of the total disease burden in areas of the globe where cholera is endemic suggest that between 1.4 and 4.2 million cases occur annually around the world and result in as many as 91,000 deaths annually (Ali *et al.* 2012). Cholera transmissions occur primarily as a consequence of contaminated shared resources, such as lakes (Bompangue *et al.* 2008) or coastal estuaries (Jutla *et al.* 2010). Other environmental and water-related factors influence cholera incidences, such as proximity to coast (Borroto & Martinez-Piedra 2000), seasonal weather conditions (Akanda *et al.* 2009), temperature (Lama *et al.* 2004), rainfall (Sasaki *et al.* 2009) and rainfall anomalies (de Magny *et al.* 2008). In the context of climate change, these environmental factors may yield profound implications for developing nations. Consider the case of Bangladesh, where substantial variation in cholera prevalence has been linked to ocean warming in conjunction with El Niño or the Southern Oscillation (ENSO) (Rodó *et al.* 2002).

Water treatment methods, whether they be piped water service or simple point-of-use systems, have been found to be effective at reducing infections of all types of diarrheal disease (Fewtrell *et al.* 2005; Zwane & Kremer 2007). However, evidence collection and evaluation for the effectiveness of water treatment and sanitation interventions remains ongoing. A review by Cairncross *et al.* (2010) considers study design and finds that the effectiveness of water treatment interventions in some random controlled trials are not reproduced when trial participants are blinded. Looking at treatment interventions that are specific to cholera, the implementation of simple filters to remove waterborne copepods, to which cholera bacterium are known to attach (Huo *et al.* 1996), has been shown to result in 48% reduction in cholera cases (Colwell *et al.* 2003). Other water-related infrastructure projects, such as drainage systems, have also been linked to cholera incidences (Sasaki *et al.* 2009). Well-designed evaluation studies of sanitation-based infrastructure interventions are few in number, which results in a dearth of high-quality evidence to support the effectiveness of these interventions in reducing the burden of cholera and other diarrheal illnesses (Cairncross *et al.* 2010). Beyond drinking water and sanitation-related control strategies, other personal health-based interventions are also acknowledged for reducing the cholera disease burden. These interventions include breastfeeding (Glass *et al.* 1983), vaccinations (Sanchez *et al.* 1994), simple hygiene such as hand washing prior to food preparation (Hutin *et al.* 2003), and bathing (Sack *et al.* 2003).

Therapeutic disease mitigation strategies have also demonstrated effectiveness at reducing the disease burden of diarrheal illnesses. The most common therapy for acute diarrhea regardless of the disease-causing agent (i.e., not specifically cholera-induced diarrhea) is oral-rehydration therapy (ORT), the consumption of large amounts of water with a balance of salts and sugars to replenish the essential nutrients lost due to diarrhea. From Jose & Bobadilla (1994), ORT reduces the duration of diarrhea as well as the mortality rate associated with diarrhea. The primary management strategy for most types of diarrheal diseases is to prevent and treat dehydration while maintaining nutrient intake levels throughout the course of the disease (Alam & Ashraf 2003). Specific to cholera, the deployment of ORT has reduced mortalities associated with cholera outbreaks from greater than 40% of reported cases, with tens of thousands of deaths, to less than 4% of reported cases (WHO 2009a).

A variety of other factors may also influence disease levels, including a region's economic and institutional characteristics. Gerolomo & Penna (2000) show the incidence of cholera is linked to lower household income levels. In Latin America, higher levels of gross national product per capita, above US$2,000, was negatively correlated with cholera incidence rates (Ackers *et al.* 1998). Talavera & Perez (2009) found that the number of cholera cases was highest in countries with low gross national income, strengthening the case that cholera, and sub-standard health in general, are linked to poverty.

The many factors associated with cholera outbreaks and disease-burden that are identified by previous literature inform the conceptual framework and the empirical strategy of this study. This study endeavors to construct a dataset that is designed to isolate the effects of the variables of interest: drinking water and sanitation infrastructure, while controlling for relevant environmental, demographic, and economic factors. Given the complexity of the relationships between cholera disease-burden and the various social, economic, technological/infrastructural, and behavioral attributes that contribute to cholera, assembling such a dataset provided a considerable challenge.

As is discussed in subsequent sections on this study's data imputation procedures, locating available data on even the more readily measurable of these contributing factors proves difficult. And in the case of the more difficult to measure factors, such as behavioral factors like hand washing and hygiene, data collection proves to be impossible while maintaining the same temporal and spatial scales as the other variables. The absence of explicitly behavioral and hygienic data constitutes a limitation of this article and highlights a worthy objective for future studies and data collection efforts. Assuming that public health awareness and hygiene practices are likely to increase with a nation's wealth, education, and economic development, the decision to include economic and geographic variables, i.e., gross domestic product (GDP) per capita and urban concentration, is partially motivated to control for behavioral changes which may also be correlated to these broader economic and geographic trends.

METHODS

Empirical approach

This study considers the factors that contribute to annually-reported cholera levels of countries in Africa and Asia. The analysis considers cholera events (or episodes) as consisting of three statistically-tractable components. Those components are: (1) the occurrence of an outbreak (a binary variable), (2) the number of cases per 100,000 population in a given outbreak, and (3) the number of disease-associated mortalities per cases in a given outbreak. Since negative values for the number of cholera infections in a given country at a given time cannot occur, the data for cholera cases and by extension the data for cholera rates (cases per 100,000 population) are lower bound at zero. Decomposing a given cholera event into the three components itemized above also allows for the selection of the three different estimation procedures, whichever is most appropriate given the type of data available for each component.

The three empirical models used to characterize a cholera outbreak and its severity are presented more explicitly below. The occurrence of a cholera outbreak is estimated with a probit model Equation (1) (Gujarati 2003; Wooldridge 2003). The number of cases (per 100,000 population) given in a cholera outbreak is represented by Equation (2). The intensity of cholera-related mortality rates given at least one cholera-related death is represented by Equation (3). Since the outcome variables of Equations (2) and (3) are continuous variables, they are estimated using ordinary least squares (OLS) (Wooldridge 2003).

$$P\left(y_{i,t}^{O}=1\right)=F\left(\beta^{O}X_{i,t}^{O}\right) \tag{1}$$

where $F(\cdot)$ is the normal cumulative density function

$$y_{i,t}^{C}\big|_{\left(y_{i,t}^{O}=1\right)}=\beta^{C}X_{i,t}^{C}+u^{C} \tag{2}$$

$$y^D_{i,t}\big|_{\left(y^O_{i,t}=1,\, y^D_{i,t}>0\right)} = \beta^D X^D_{i,t} + u^D \tag{3}$$

In Equation (1), the probability of an outbreak ($(P\left(y^O_{i,t}\right)$, where superscript O denotes 'outbreak') in country i at time t is a function of components in $X^O_{i,t}$. The components of $X^O_{i,t}$ include time-varying and country-specific levels of improved drinking water and sanitation services, i.e., the percentage of population served by improved drinking water and improved sanitation services, the occurrence of cholera outbreak in the previous year (or the lagged outbreak variable), a nation's wealth (measured as GDP per capita), the trend to capture time-correlated effects, and other geographic and environmental characteristics of a country that are fixed through time, such as latitude (a proxy for climate), and area of coastal and inland water bodies. The complete empirical specification of Equation (1) is presented below in Equation (4). All of the empirical variables are described in greater detail in the following section.

$$\begin{aligned}\text{Outbreak}_{i,t} &= \beta^C_1\text{Water}_{\ i,t} + \beta^C_3\text{Sanitation}_{\ i,t} + \beta^C_5\text{Outbreak}_{i,t-1}\\ &+ \beta^C_6\text{Urban}_{i,t} + \beta^C_7\text{ GDP per capita}_{i,t} + \beta^C_8\text{Year}_t\\ &+ \beta^C_9\text{Latitude}_i + \beta^C_{10}\text{Coast Land Ratio}_i\\ &+ \beta^C_{11}\text{Water Land Ratio}_i + \beta^C_{12}\text{Constant}\end{aligned} \tag{4}$$

The outcome variable in Equation (2) is the cholera case rate $y^C_{i,t}$, where superscript C denotes 'case rate' (number of reported infections per 100,000 population). This case rate is a function of water services, sanitation services, economic and environmental characteristics, much as the determinants of the binary outbreak variable in Equation (1) are a function of those same variables. One important distinction between the set of explanatory variables for outbreaks in Equation (1) and case rates in Equation (2) is that where Equation (1) implements a lagged outbreak variable $(y^O_{i,t-1} \in X^O_{i,t})$, Equation (2) implements a lagged case rate variable $(y^C_{i,t-1} \in X^C_{i,t})$. The lagged case rate variable, defined in a similar way, is the number of reported infections per 100,000 people that occurred in the previous period. The complete empirical specification of Equation (2) is presented below in Equation (5).

$$\begin{aligned}\text{Cholera rate}_{i,t} &= \beta^C_1\text{Water rural}_{i,t} + \beta^C_2\text{Water urban}_{i,t}\\ &+ \beta^C_3\text{Sanitation rural}_{i,t} + \beta^C_4\text{Sanitation urban}_{\ i,t}\\ &+ \beta^C_5\text{Cholera rate}_{i,t-1} + \beta^C_6\text{Urban}_{i,t} + \beta^C_7\text{ GDP per capita}_{i,t}\\ &+ \beta^C_8\text{Year}_t + \beta^C_9\text{Latitude}_i + \beta^C_{10}\text{Coast Land Ratio}_i\\ &+ \beta^C_{11}\text{Water Land Ratio}_i + \beta^C_{12}\text{Constant}\end{aligned} \tag{5}$$

Equation (3) is the model for the mortality rate associated with reported cholera cases where superscript D in the outcome variable $y^D_{i,t}$ denotes 'deaths'. Many of the same environmental and economic variables (i.e., urban proportion, GDP, latitude, ratios of coastal land and inland water area to land area) are used in this model of cholera-related mortality rates. One exception is that lagged (nor current-time) cholera cases ($y^C_{i,t}$) are not used as an explanatory variable in the mortality model of Equation (3). Since the mortality rate is by definition a function of current cases, by virtue of being the number of deaths divided by the number of cases, using current case rates as a right-hand-side covariate would introduce simultaneity into the model. One potential alternative is to model case rates and mortality rates in a simultaneous system, perhaps using lagged cholera cases in an instrumental variable approach, but that kind of analysis is beyond the current scope of this study. The complete empirical specification of Equation (3) is presented below in Equation (6).

$$\begin{aligned}\text{Mortality rate}_{i,t} &= \beta^C_1\text{Water rural}_{i,t} + \beta^C_2\text{Water urban}_{i,t}\\ &+ \beta^C_3\text{Sanitation rural}_{i,t} + \beta^C_4\text{Sanitation urban}_{\ i,t}\\ &+ \beta^C_5\text{Urban}_{i,t} + \beta^C_6\text{ GDP per capita}_{i,t} + \beta^C_7\text{Year}_t\\ &+ \beta^C_8\text{Latitude}_i + \beta^C_9\text{Coast Land Ratio}_i\\ &+ \beta^C_{10}\text{Water Land Ratio}_i + \beta^C_{11}\text{Constant}\end{aligned} \tag{6}$$

Data description

The dataset used for this study comes from several sources. The nations included in this dataset comprise the majority of nations located in Africa (including 43 nations of 54) and Asia (including 31 nations of 50). African nations included in the dataset are: Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Republic of the Congo, Côte d'Ivoire, Djibouti, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, South Africa, Swaziland, Tanzania, Togo, Uganda, Zambia, and Zimbabwe. Asian nations included in the dataset are: Armenia, Azerbaijan, Bangladesh, Bhutan, Cambodia, China, India, Indonesia, Iran, Iraq, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Laos, Lebanon, Malaysia, Mongolia, Nepal, Oman, Pakistan, Philippines, Qatar, Russia, South Korea, Sri Lanka, Thailand, Turkmenistan, United Arab Emirates, Vietnam, and Yemen.

The time-series of the dataset is from 1990 to 2008. The variable *Cases* is the count of reported cholera cases for a given country in a given year. Observations where no cholera cases are reported for a given country and year are initially assumed to be missing values. The *Cholera Rate* is a function of *Cases* and the level of *Population* reported to the World Bank. The *Death Rate* is the number of cholera-related mortalities per reported case. The variable *Deaths* is found alongside *Cases* in the World Health Organization (WHO) Epidemiological Record (2009b). The variables *Water* and *Sanitation* are the percentage rates of coverage for 'improved' drinking water and 'improved' sanitation services for each country. The source for these variables comes from World Bank (2011) and WHO & UNICEF (2010). Data for *Water* and *Sanitation* are available for urban and rural populations, hence the subscripts urban and rural are used, $Water_{urban}$, $Sanitation_{rural}$, etc. The raw data for *Water* and *Sanitation* are available for only a specific set of years: 1990, 1995, 2000, 2005, and 2008. The demographic and economic variables GDP and *Urban* are reported annually in World Bank (2011). Environmental variables are found in The World Factbook (2009) and are presumed to be fixed through time.

Imputed data on cholera outbreaks

One of the first data-related challenges this study encounters is the construction of the outbreak variable, which is $y_{i,t}^{O}$ from Equation (1). Data on the number of cholera cases in a given country and year are found in the archives of the Weekly Epidemiological Record (WER) (WHO 2009b). With a few exceptions, the WER does not report a '0' for countries that, presumably, either did not have any cholera cases (or deaths) or did not report their cases to the WHO. Rather, a country's name is simply omitted from the annual cholera report, which may or may not imply zero cases for that particular country. Zero cases are not necessarily implied because the situation could be that a country actually experienced a positive number of cholera cases, but, due to a variety of possible reasons, insufficient resources for monitoring and reporting being just one, such data were never transmitted to WHO and thereby not published in the WER.

To generate a binary variable amenable to probit analysis, specific instances of the outbreak variable are populated with a value (i.e., $y_{i,t}^{O} = 0$) indicating that no cholera outbreak occurred. Imputing $y_{i,t}^{O} = 0$ from the original data occurs in two ways: (1) for non-missing observations, defining an 'outbreak' as a country-year that reported cholera cases above an assumed threshold, and (2) for missing observations, selecting specific missing observations that, based on reasonable criteria, most likely represent non-occurrence of cholera.

An assumed number of cholera cases are established as an outbreak threshold to indicate a cholera 'outbreak' has occurred. This assumption is made for two reasons. First, very low case levels (as few one or two) in a given year in a country do not comport with the general conception of a disease outbreak. Second, and more importantly, small numbers of cases may be a consequence of international travel to nations that are both unknown to the analysis as well as potentially outside the African and Asian scope of this article. Since these cases are not generated within the drinking water, sanitation, environmental, and economic contexts of the country in which they are reported, using them as criteria to estimate the effectiveness of drinking water and sanitation coverage is unreasonable. This threshold is operationalized in this analysis as follows: if the threshold is set at 200 cases, then in a given year if more than 200 cholera cases is recorded, the outbreak variable for that year is set equal to one (i.e., $y_{i,t}^{O} = 1$), else it is zero. More formally, for a given country i and time t:

$$\text{if Cases}_{i,t} \geq \text{Cases}_{\text{threshold}} \text{ then } y_{i,t}^{O} = 1 \tag{7}$$

$$\text{if Cases}_{i,t} < \text{Cases}_{\text{threshold}} \text{ then } y_{i,t}^{O} = 0 \tag{8}$$

Due to the arbitrary nature of selecting a value for $Cases_{threshold}$, several values ranging from 50 cases in a year to 500 cases in a year are used in robustness checks. Results presented in this paper are largely consistent across the entire range of possible assumed threshold case levels (from 50 to 500). With no other satisfying guidance available for the selection of the most appropriate value for the outbreak threshold, we assumed the baseline value for the threshold for presentation in this article to be 200.

The second step in completing the 'outbreak' variable is to address the observations where the value of $Cases_{i,t}$ is not reported in the WER, initially becoming a missing value in the dataset. For a country to report its cholera level to the WHO (and to be then recorded in the WER), we conjecture that two criteria must be met. Those criteria are: (1) that the country is wealthy enough to monitor and distribute data on cholera incidences, and (2) that the country possesses the political desire to report those findings to the WHO. With these criteria in mind, we concluded a reasonable procedure is to assume that reporting cholera cases to the WHO in year t reveals (or demonstrates) both an ability (or sufficient wealth level) to report cholera cases and a willingness-to-report cholera cases for that country in all years going forward. Implicitly, this assumes countries that are wealthy enough to report cholera in year t are also wealthy enough to do so in year $t+1$. Similarly, this assumes a country's political willingness to report cholera is nondecreasing and/or not reversed through the years of the dataset. A more precise interpretation of these missing values would likely include country-by-country analyses of political stability and wealth over the years of the dataset, but such an effort is beyond the scope of the current study. To make this procedure more concrete, Table 1 contains an illustrative example using three countries in the dataset of the generation of the outbreak variable ($y^O_{i,t}$) over the years 1990 to 1999.

Imputed data on drinking water and sanitation services

Data on the proportion of people with access to 'improved' drinking water infrastructure and 'improved' sanitation service infrastructure are collected and published in conjunction with measuring progress towards the MDGs (WHO & UNICEF 2010). As far as we can tell, the published data for this particular goal occur for the following years: 1990, 1995, 2000, 2005, and 2008. These years are referred to as the reported years. Since cholera cases and most of the other covariates used in this study are reported annually, imputing values of drinking water and sanitation for the

Table 1 | An illustrative example of the imputation of the binary 'outbreak' variable for three countries (Rwanda, Mali, and India) over the years 1990 to 1999 in the context of cholera

	RWANDA			**MALI**			**INDIA**		
Year	**Cases ($Cases_{i,t}$)**	**Case rate ($y^c_{i,t}$)**	**Outbreak ($y^O_{i,t}$)**	**Cases ($Cases_{i,t}$)**	**Case rate ($y^c_{i,t}$)**	**Outbreak ($y^O_{i,t}$)**	**Cases ($Cases_{i,t}$)**	**Case rate ($y^c_{i,t}$)**	**Outbreak ($y^O_{i,t}$)**
1989	1	–	–		–	–	5,026	–	–
1990			0				3,583	0.410	1
1991	679	9.882	1				6,993	0.784	1
1992	530	8.214	1				6,911	0.759	1
1993	568	9.491	1				9,437	1.017	1
1994	10	0.177	0				4,973	0.526	1
1995	3	0.054	0				3,315	0.344	1
1996	106	1.826	0	5,723	56.663	1	4,396	0.447	1
1997	274	4.351	1	6	0.058	0	2,768	0.277	1
1998	3,220	46.370	1			0	7,151	0.701	1
1999	217	2.862	1	6	0.055	0	3,839	0.369	1

Note: Cases in year 1989 are not used in this analysis except to generate the outbreak variable ($y^O_{i,t}$) for year 1990.

years between the reported years seems a straightforward and reasonable way to increase the size of the dataset. The imputation strategy used is a simple linear interpolation of the unknown years, where the reported years (1990, 1995, 2000, 2005, and 2008) serve as endpoints. In this case, using linear interpolation is analogous to drawing a line between the known values in the reported years, where the drawn line represents the imputed values. To make this procedure concrete, an illustrative example of the two improved drinking water variables (reported and imputed), using three countries, is presented in Figure 1.

An important caveat, the WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation is charged with tabulating and reporting progress made on the Water and Sanitation MDG goal (WHO & UNICEF 2012). The methodology JMP uses to calculate water and sanitation coverage rates are described by WHO & UNICEF (2012, 2013). The levels of water and sanitation services which this article refers to as the reported data are themselves also a product of a data imputation process that occurs as JMP tabulates the water and sanitation progress reports. In brief, for each progress report, JMP accumulates survey data and updates a model that projects each nation's levels of water and sanitation services. Most importantly, their model is linear with respect to time.

The JMP data imputation process has two important consequences for the analysis done in this article. First, the efforts of JMP are ongoing. As additional surveys are collected for each nation, the JMP projection model is updated. More recent projections from JMP utilize newly collected or previously unavailable survey data. Since the linear model is used to project water and sanitation coverage at points in the past, the addition of new data may alter what this article refers to as the 'reported' levels of water and sanitation coverage estimates for past years. This issue introduces the possibility that analyses (such as this article) which are based on JMP water and sanitation coverage data can be repeated in the future with using updated coverage data from JMP and thereby yield different, potentially more reliable, results. Secondly, a model that is linear with respect to time is best suited for capturing processes with time-constant rates of change. For the build-out of water and sanitation coverage in conjunction with the MDGs, the case could be that water and sanitation coverage are increasing (or possibly decreasing) at an increasing rate. In particular, as the 2015 target year associated with the MDGs approaches, if efforts to achieve the MDGs have been increased, then any resulting acceleration in the rates of water and sanitation coverage may not be adequately captured by the linear projection model. More details on the JMP methodology can be found in WHO & UNICEF (2012).

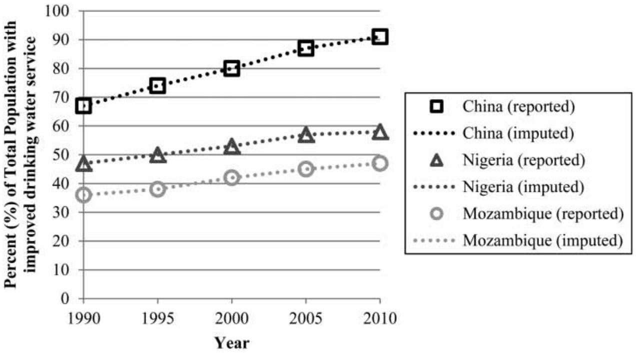

Figure 1 | Illustrative example of the two variants of the drinking water services variable, the reported values ($Water_{total}$), and imputed values ($Water_{total,imputed}$) for three countries (China, Nigeria, and Mozambique) over the years 1990 to 2010.

Summary statistics

Table 2 presents the summary statistics for the variables used in this analysis. Subscripts on *Water* and *Sanitation* include: *total*, *rural*, and *urban*, which indicates access to improved water and sanitation services of the population as a whole, the rural population, and the urban population, respectively. The subscript *imputed* differentiates the reported data (only the years 1990, 1995, 2000, 2005, and 2008) from imputed data that covers all the years from 1990 to 2008. The difference in years covered is reflected in the increase in sample size among the *imputed* variables. Overall, Asian countries seem to have fewer outbreaks, fewer cases per outbreak, fewer deaths per case, and more of both drinking water infrastructure and sanitation infrastructure.

RESULTS

Probit estimation of cholera outbreaks

Table 3 illustrates the difference between two models with regard to the imputation of improved drinking water and

Table 2 | Summary statistics for cholera case levels, water and sanitation services, and other variables for countries in Africa and Asia from 1990 to 2008

		Pooled		Africa		Asia	
Variable	**Details**	**N**	**Mean**	**N**	**Mean**	**N**	**Mean**
Cases	Number of cholera cases	693	4,109.889	488	5,178.207	205	1,566.771
Cholera rate	(Cases/Population) * 100,000	693	41.242	488	57.182	205	3.296
Outbreak	Binary variable =1, if cases >= 200	1220	0.401	695	0.545	525	0.210
Death rate	Deaths/Cases	644	0.038	472	0.046	172	0.014
$Water_{total}$	% total pop. with improved drinking water	283	72.159	163	63.215	120	84.308
$Water_{total,imputed}$	% total pop. with improved drinking water	1220	72.118	695	62.937	525	84.271
$Water_{rural,imputed}$	% rural pop. with improved drinking water	1220	63.023	695	52.016	525	77.595
$Water_{urban,imputed}$	% urban pop. with improved drinking water	1220	87.916	695	83.624	525	93.598
$Sanitation_{total}$	% total pop. with improved sanitation	283	48.385	163	34.037	120	67.875
$Sanitation_{total,imputed}$	% total pop. with improved sanitation	1220	48.304	695	33.256	525	68.225
$Sanitation_{rural,imputed}$	% rural pop. with improved sanitation	1220	40.908	695	25.387	525	61.454
$Sanitation_{urban,imputed}$	% urban pop. with improved sanitation	1220	60.318	695	45.231	525	80.291
Urban	% total pop. lives in urban area	1220	41.494	695	36.440	525	48.184
GDP per capita	Gross domestic product per capita (US$2,000)/1,000	1220	2.301	695	0.656	525	4.478
Latitude	Latitude (absolute value)	1220	18.634	695	12.259	525	27.074
Coast Land Ratio	Ratio of coastline to land area	1220	0.012	695	0.010	525	0.015
Water Land Ratio	Ratio of water area to land area	1220	0.035	695	0.039	525	0.029
Year	Year	1220	2,001.362	695	2,001.325	525	2,001.411

sanitation services. Model I uses the limited dataset (with no imputation of *Water* or *Sanitation*), whereas Model II uses the dataset where the years between original observations are linearly interpolated. For the most part, the results are consistent across models. The lagged outbreak variable is a strong predictor of current cholera outbreaks, with the largest marginal effect, which is statistically significant in both the models, presented in Table 3. The estimate on the lagged outbreak variable suggests that cholera outbreak in a previous time period is linked to an increase in the probability of cholera outbreak in the current time period. This could be capturing the effects of an outbreak event that extends across multiple years or mutiple months if the months include December and January.

In the expanded (imputed) dataset, the absolute value of latitude, a proxy for climate, is also significant and indicates that at higher latitudes (as climates become more seasonal and cooler), the likelihood of an outbreak falls. Interestingly, this table shows little evidence for the effectiveness of water and sanitation interventions on the occurrence of an outbreak. Only in the imputed dataset is a negative effect measured on sanitation services with statistical significance at 10%. These results are consistent across the range of outbreak threshold levels.

Observations increased from 287 in Model I to 1,236 in Model II. The explanatory power (as measured by the McFadden's R^2) of the two models are not dramatically different. Coefficient estimates, where statistically significant, also do not hold opposite relationships across the two models. Taken together, the comparison of estimated coefficient values and model statistics across Models I and II suggest that imputation of the water and sanitation variables for the unknown years increases the statistical power of the analysis, without altering (i.e., increasing or decreasing) any statistical bias that may be in either model. In particular, the imputations used in Model II do nothing to address any bias that may arise from the JMP methods used to project past drinking water and sanitation coverage levels that are discussed in the 'Data description' section.

Table 3 | Marginal effects and robust standard errors from probit estimations on the occurrence of a cholera outbreak (at least 200 documented cases) in African and Asian countries from 1990 to 2008 with no imputation (Model I) and with linear interpolation (Model II) of water and sanitation services

Model: **Dependent variable:**	**Original data** **I** **Outbreak**		**Imputed water and sanitation** **II** **Outbreak**	
	Marg. Eff.	**R.S.E.**	**Marg. Eff.**	**R.S.E.**
$\text{Water}_{\text{total}}$	0.0027	0.0029		
$\text{Water}_{\text{total,imputed}}$			0.0001	0.0020
$\text{Sanitation}_{\text{total}}$	– 0.0036	0.0023		
$\text{Sanitation}_{\text{total,imputed}}$			– 0.0023*	0.0013
Outbreak lag	0.5836***	0.0576	0.4752***	0.0396
Urban	– 0.0036	0.0025	– 0.0009	0.0015
GDP per capita	0.0056	0.0049	– 0.0045	0.0045
Year	0.0056	0.0057	– 0.0042	0.0031
Latitude	– 0.0034	0.0033	– 0.0090***	0.0024
Coast Land Ratio	– 0.0413	0.6477	– 0.6203	0.3857
Water Land Ratio	0.7046	0.8306	0.6196	0.5710
N	283		1220	
McFadden's R^2	0.353		0.339	
Log-likelihood	– 124.0		– 543.1	

Note: Robust standard errors are clustered at the country level. This pooled sample includes Africa and Asian countries.

*, ***indicates significance less than or equal to 10 and 1%, respectively.

OLS estimation of cholera rates

The OLS results for cholera rates, conditional on an outbreak, are presented in Table 4. This table presents three groups of estimates, based on the type of sample. Model III includes African and Asian nations pooled together. In Models IV and V, the African and Asian nations are separated into subsamples. As a factor that is associated with reduced intensity of an outbreak, water access in rural areas is significant and negative in the pooled and African samples. In the pooled sample, a one percent increase in access to improved water infrastructure in rural populations decreases the cholera rate by two cases per 100,000 population (Table 4). Interestingly, *Water*, rural or urban, is not significant in the Asian sample. In the Asian sample, the lagged cholera rate is the most important factor related to cholera infection rate in the current outbreak.

The estimate on the *Water Land Ratio* exhibits somewhat weak evidence (statistically significant at 10%) that more cholera cases can occur in countries with larger proportions of open water bodies. This result does comport with the idea that these water bodies are environmental reservoirs for the pathogen that causes cholera. However, interpretation of the results on the ratio of water area and coastline to land area warrants caution since these variables are crude proxies for a variety of factors related to a nation's population's risk from cholera and a nation's natural water systems. In particular, these variables do not directly account for the nation's population density near, nor its proximity to, surface water bodies or coastlines.

Perhaps the most interesting and relevant finding for disease-control policies is that the effects of *Water* are measurable with respect to cholera rates, but not statistically different from zero with respect to cholera outbreaks. By contrast, *Sanitation* exhibits evidence, albeit weak, as negatively correlated with outbreaks, but exhibits no statistical difference from zero with respect to cholera rates. These results illustrate that not only are intervention effects difficult to measure, especially at the national scale, but the effects of different interventions may have a heterogeneous influence on each of several aspects of a cholera event.

OLS estimation of cholera mortality rates

Table 5 is a presentation of the results of OLS estimation on the *Death rate* (scaled up by 1,000 for easier presentation of the coefficients). Similar to the results of the cholera case model, *Water* has an estimated effect that is negative and statistically significant with respect to cholera-related deaths, and no such effect is found for *Sanitation*. In the pooled sample, rural access to improved water is negative and significant at 5%, and, in the Asian sample, urban access to water is negative and significant at 10%. Sanitation, rural or urban, is not significant in any sample with respect to changes in cholera-related death rates. A point that was raised in the literature review is that in many cases those infected with cholera may be treated by consuming uncontaminated water (possibly augmented with ORT) and may achieve a full recovery. As such, access to improved drinking water infrastructure seems likely to promote disease remediation.

Table 4 | Coefficient values and robust standard errors from ordinary least squares regression on the cholera rates, conditional on at least 200 documented cases (*Outbreak* = 1), in African and Asian countries from 1990 to 2008

Sample:	**Pooled**		**Africa**		**Asia**	
Model:	**III**		**IV**		**V**	
Dependent variable:	**Cholera rate**		**Cholera rate**		**Cholera rate**	
	Coef.	**R.S.E.**	**Coef.**	**R.S.E.**	**Coef.**	**R.S.E.**
$\text{Water}_{\text{rural,imputed}}$	– 2.066**	0.877	– 2.218*	1.114	0.373	0.314
$\text{Water}_{\text{urban,imputed}}$	– 2.286	1.673	– 2.086	1.922	– 0.362	0.274
$\text{Sanitation}_{\text{rural,imputed}}$	– 1.094	1.330	– 2.514	2.259	0.153	0.205
$\text{Sanitation}_{\text{urban,imputed}}$	1.286	1.258	2.390	2.138	– 0.323	0.296
Cholera rate lag	– 0.091	0.229	– 0.268	0.282	0.231***	0.073
Urban	1.018	0.899	0.962	1.219	– 0.262	0.178
GDP per capita	1.099	10.443	70.787	49.824	– 0.151	0.164
Year	1.796	1.515	-0.915	1.827	– 0.700	0.627
Latitude	1.980	1.397	3.012	3.284	0.115	0.149
Coast Land Ratio	3,549.8	2,505.3	5,759.2*	3,240.5	– 50.2	32.143
Water Land Ratio	1,138.9*	672.5	1,514.3*	762.4	– 93.1	87.016
Constant	– 3,388.4	2,992.5	1,972.5	3,652.7	1,436.7	1,242.8
N	437		342		95	
R^2	0.203		0.341		0.297	

*,**,*** indicate significance less than or equal to 10, 5, and 1%, respectively.

Table 5 | Coefficient values and robust standard errors from ordinary least squares regression on the death rates associated with cholera, conditional on at least 200 documented cases (Outbreak =1) and at least one documented cholera-related death, in African and Asian countries from 1990 to 2008

Sample:	**Pooled**		**Africa**		**Asia**	
Model:	**VII**		**VIII**		**IX**	
Dependent variable:	**Death rate**		**Death rate**		**Death rate**	
	Coef.	**R.S.E.**	**Coef.**	**R.S.E.**	**Coef.**	**R.S.E.**
$\text{Water}_{\text{rural,imputed}}$	– 0.465**	0.190	– 0.239	0.232	– 0.339	0.347
$\text{Water}_{\text{urban,imputed}}$	0.103	0.210	0.240	0.236	– 0.676*	0.324
$\text{Sanitation}_{\text{rural,imputed}}$	0.083	0.177	0.038	0.285	0.179	0.266
$\text{Sanitation}_{\text{urban,imputed}}$	– 0.098	0.195	0.151	0.221	– 0.099	0.409
Urban	– 0.177	0.160	– 0.161	0.237	– 0.059	0.146
GDP per capita	– 0.431	0.714	– 7.473**	3.614	0.041	0.275
Year	– 1.523***	0.355	– 2.200***	0.345	– 0.297	0.294
Latitude	– 0.713**	0.326	– 0.526	0.525	– 0.073	0.133
Coast Land Ratio	– 67.5	56.6	– 104.6**	48.9	– 7.3	32.6
Water Land Ratio	– 117.8***	42.0	– 138.2***	46.2	– 41.6	61.7
Constant	3,120.4***	706.5	4,448.3***	688.1	698.7	581.6
N	455		364		91	
R^2	0.288		0.264		0.459	

* indicates significance less than or equal to 10%,**5%, and ***1%.

The significance of *Year* in the pooled and African sample suggests that time-correlated factors, not otherwise controlled for in the model, are at least reducing the mortality effects of cholera infections. Some of these time-correlated factors could include: broader distribution of human capital, such as knowledge about the effects and possible treatments of cholera, broader physical products, such as ORT salts and water purification devices, as well as faster and more effective responses of public health organizations that intervene when a cholera epidemic occurs. Another interesting result in this regression is, unlike the estimation of the case rate (Table 4), the coefficient on the relation of water area to land area is negative. Together, these results suggest countries with proportionally more water area are likely to have larger outbreaks in terms of case volumes but be subject to lower mortality rates per cholera case. This could be explained in a few ways. For example, countries with a higher ratio of water area to land area may have more experience dealing with cholera infections, or perhaps more intense outbreaks invoke greater public health responses. The significance of GDP per capita in the African sample agrees with conventional sense that wealthier nations are more able to care for the ill and prevent fewer infections from becoming fatalities.

SUMMARY AND CONCLUSIONS

This article presents an analysis of factors that are associated with cholera outbreaks, case rates, and mortality rates in countries of Africa and Asia. For this study, a new dataset is constructed from a variety of sources. Implementing data imputation for specific missing values in two variables allows for the development of a larger dataset. In the case of imputing water and sanitation levels for unknown years, the power of the model seems to increase with no apparent increase (or decrease) in any potential model bias.

Results indicate a mixture of geographic, environmental, and anthropogenic factors contribute to the three components of a cholera event. Latitudes nearer to the poles are associated with lower probabilities of outbreaks. Over the time frame captured by the dataset, the proportion of fatal cholera cases is falling in the pooled and African samples. Lagged outbreaks are a strong predictor of current outbreaks. In the Asian sample, lagged cholera case rates predict larger current case rates. These results are robust to assumptions about the threshold number of cases that define a cholera outbreak.

More relevant to policymakers and development researchers, this article tries to isolate and estimate the effects of actionable variables monitored as a part of efforts made towards achieving MDGs. In particular, the effects of national levels of water and sanitation, representing access to ‘improved’ drinking water and sanitation services, are examined in the context of cholera outbreaks, outbreak intensity or cholera case rates, and cholera-related mortality rates. In the pooled sample using imputed water and sanitation data, cholera outbreaks are negatively related to access to improved sanitation albeit with statistical significance at 10%. In the pooled and African samples, the cholera case rates (conditional on an outbreak) are statistically significant and negatively related to access to improved water systems among rural populations. In the context of cholera-related mortality, the number of deaths per case is negatively related to water access among rural populations in the pooled sample (significant at 5%) and among urban populations in the Asian sample (significant at 10%).

The most cohesive policy implication that follows from this study is to underscore the importance of high quality monitoring and reporting of both disease incidence and progress towards social goals such as the MDGs. Such data are not only necessary to monitor progress towards social objectives but are also useful in retrospect to evaluate the benefits and costs from such projects. Any other policy recommendations that follow are based on empirical findings and are therefore qualified by the context, specifically the limitations of our data, in which the findings are produced. The most convincing empirical result found in this analysis is that drinking water infrastructure is negative correlated with cholera case rates and cholera-related mortality rates in rural areas of the sample containing both African and Asian nations. Therefore, expanding water supply systems in rural areas seem to present a viable opportunity to reduce cholera burden in Africa and Asia.

In conclusion, this study takes one particular aspect of the MDGs, the water and sanitation program, and examines the potential effect on the issue of cholera outbreaks, which is just one of a myriad of water and sanitation-related public health issues confronted by economically developing regions. The empirical findings, being one important contribution of this study to the literature, are couched in the other main contribution of this study, which is the assimilation of a new dataset that captures a large spatial scale and a fairly large temporal scale, relative to other cholera studies. Interpretation of any results from this dataset should be qualified by obvious limitations on the availability and the quality of the data. The numbers of cholera cases and deaths are voluntarily reported to the WHO, published in the WER, and almost certainly do not represent the full extent of cholera infections across the geographic region of the sample. As mentioned previously, the reported data on the coverage levels of drinking water and sanitation infrastructure, which are generated from the JMP projection model, while based on survey data are not strictly empirical records.

ACKNOWLEDGEMENTS

Work on this paper was completed while the coauthors were housed in the Department of Agricultural Economics at Texas A&M University. Views, opinions, and results presented in the paper do not necessarily represent those held by their current or past employers. The authors are grateful for the thoughtful comments provided by Dr Piya Abeygunawardena, Nancy Leahy and two anonymous reviewers on earlier drafts of this paper.

REFERENCES

Ackers, M.-L., Quick, R. E., Drasbek, C. J., Hutwagner, L. & Tauxe, R. V. 1998 Are there national risk factors for epidemic cholera? the correlation between socio-economic and demographic indices and cholera incidence in Latin America. *International Journal of Epidemiology* **27**, 330–334.

Akanda, A. S., Jutla, A. S. & Islam, S. 2009 Dual peak cholera transmission in bengal delta: a Hydroclimatological explanation. *Geophysical Research Letters* **36**, L19401.

Alam, N. H. & Ashraf, H. 2003 Treatment of infectious diarrhoea in children. *Pediatric Drugs* **5**, 151–165.

Ali, M., Lopez, A. L., You, Y. A., Kim, Y. E., Sah, B., Maskery, B. & Clemens, J. 2012 The global burden of cholera. *Bulletin of the World Health Organization* **90**, 209–218.

Anand, S. & Ravallion, M. 1993 Human development in poor countries: on the role of private incomes and public services. *The Journal of Economic Perspectives* **7**, 133–150.

Bloom, D. E., Canning, D. & Sevilla, J. 2004 The effect of health on economic growth: a production function approach. *World Development* **32**, 1–13.

Bompangue, D., Giraudoux, P., Handschumacher, P., Piarroux, M., Sudre, B., Ekwanzala, M., Kebela, I. & Piarroux, R. 2008 Lakes as source of cholera outbreaks, Democratic Republic of Congo. *Emerging Infectious Diseases* **14**, 798–800.

Bonds, M. H., Keenan, D. C., Rohani, P. & Sachs, J. D. 2009 Poverty trap formed by the ecology of infectious diseases. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1185–1192.

Borroto, R. J. & Martinez-Piedra, R. 2000 Geographical patterns of cholera in Mexico, 1991–1996. *International Journal of Epidemiology* **29**, 764–772.

Cairncross, S., Hunt, C., Boisson, S., Bostoen, K., Curtis, V., Fung, I. C. H. & Schmidt, W.-P. 2010 Water, sanitation and hygiene for the prevention of diarrhoea. *International Journal of Epidemiology* **39**, i193–i205.

Colwell, R. R., Huq, A., Islam, M. S., Aziz, K. M. A., Yunus, Md., Huda Khan, N., Mahmud, A., Sack, R. B., Nair, G. B., Chakraborty, J., Sack, D. A. & Russek-Cohen, E. 2003 Reduction of cholera in Bangladeshi villages by simple filtration. *Proceedings of the National Academy of Sciences* **100**, 1051–1055.

de Magny, G. C., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown, C. W., Busalacchi, A. J., Yunus, M., Nair, G. B., Gil, A. I., Lanata, C. F., Calkins, J., Manna, B., Rajendran, K., Bhattacharya, M. K., Huq, A., Sack, R. B. & Colwell, R. R. 2008 Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences* **105**, 17676–17681.

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L. & Colford Jr, J. M. 2005 Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infectious Diseases* **5**, 42–52.

Gerolomo, M. & Penna, M. L. F. 2000 Cholera and living conditions. *Revista de Saúde Pública* **34**, 342–347.

Glass, R. I., Svennerhold, A-M., Stoll, B. J., Khan, M. R., Belayet Houssain, K. M., Imdadui Hug, M. & Holmgren, J. 1983 Protection against cholera in breast-fed children by antibodies in breast milk. *The New England Journal of Medicine* **308**, 1389–1392.

Grossman, M. 1972 On the concept of health capital and the demand for health. *The Journal of Political Economy* **80**, 223–225.

Gujarati, D. M. 2003 *Basic Econometrics*. McGraw-Hill, New York.

Huo, A., Xu, B., Chowdhury, M. A., Sirajul Islam, M., Montilla, R. & Colwell, R. R. 1996 A simple filtration method to remove plankton-associated *Vibrio Cholerae* in raw water supplies in developing countries. *Applied and Environmental Microbiology* **62**, 2508–2512.

Hutin, Y., Luby, S. & Paquet, C. 2003 A large cholera outbreak in Kano City, Nigeria: the importance of hand washing with soap and the danger of street-vended water. *Journal of Water and Health* **1**, 45–52.

Jose, M. V. & Bobadilla, J. R. 1994 Epidemiological model of diarrhoeal diseases and its application in prevention and control. *Vaccine* **12**, 109–116.

Jutla, A. S., Akanda, A. S. & Islam, S. 2010 Tracking cholera in coastal regions using satellite observations. *Journal of the American Water Resources Association* **46**, 651–662.

Lama, J. R., Seas, C. R., Leon-Barua, R., Gotuzzo, E. & Sack, R. B. 2004 Environmental temperature, cholera, and acute diarrhoea in adults in Lima, Peru. *Journal of Health Population and Nutrition* **22**, 399–403.

Mayer, D. 2001 The long-term impact of health on economic growth in Latin America. *World Development* **29**, 1025–1033.

Nandy, S., Irving, M., Gordon, D., Subramanian, S. V. & Smith, G. D. 2005 Poverty, child undernutrition and morbidity: new evidence from India. *Bulletin of the World Health Organization* **83**, 210–216.

Prüss-Üstün, A. & Corvalán, C. F. 2006 *Preventing Disease through Healthy Environments: Towards an Estimate of the Environmental Burden of Disease.* World Health Organization, Geneva, Switzerland.

Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. 2002 ENSO and cholera: A nonstationary link related to climate change? *Proceedings of the National Academy of Sciences* **99**, 12901–12906.

Sachs, J. & Malaney, P. 2002 The economic and social burden of malaria. *Nature* **415**, 680–685.

Sack, R. B., Siddique, A. K., Longini Jr, I. M., Nizam, A., Yunus, M., Islam, M. S., Morris Jr, J. G., Ali, A., Huq, A., Nair, G. B., Qadri, F., Faruque, S. M., Sack, D. A. & Colwell, R. R. 2003 A four-year study of the epidemiology of *Vibrio Cholerae* in four rural areas of bangladesh. *The Journal of Infectious Diseases* **187**, 96–101.

Sanchez, J. L., Vasquez, B., Begue, R. E., Meza, R., Castellares, G., Cabezas, C., Watts, D. M., Svennerhold, A-M., Sadoff, J. C. & Taylor, D. N. 1994 Protective efficacy of oral whole-cell/recombinant-B-subunit cholera vaccine in Peruvian military recruits. *The Lancet* **344**, 1273–1276.

Sasaki, S., Suzuki, H., Fujino, Y., Kimura, Y. & Cheelo, M. 2009 Impact of drainage networks on cholera outbreaks in Lusaka, Zambia. *American Journal of Public Health* **99**, 1982–1987.

Talavera, A. & Perez, E. M. 2009 Is cholera disease associated with poverty? *The Journal of Infection in Developing Countries* **3**, 408–411.

The World Factbook 2009 Central Intelligence Agency. Washington, DC. https://www.cia.gov/library/publications/the-world-factbook/index.html.

Tol, R. S. J. 2007 Climate, development and malaria: an application of FUND. *Climatic Change* **88**, 21–34.

Tol, R. S. J. & Dowlatabadi, H. 2001 Vector-borne diseases, development & climate change. *Integrated Assessment* **2**, 173–181.

United Nations General Assembly 2000 55/2 United Nations Millennium Declaration. Sept. 8, 2000. Available at: http://www.un.org/millennium/declaration/ares552e.htm (accessed 29 June 2013).

Vlahov, D., Freudenberg, N., Proietti, F., Ompad, D., Quinn, A., Nandi, V. & Galea, S. 2007 Urban as a determinant of health. *Journal of Urban Health* **84**, 16–26.

Wooldridge, J. M. 2003 *Introductory Econometrics: A Modern Approach*. South-western, New York.

World Bank, Development Data Group 2011 *World Development Indicators Datasets Published Online*. Available at: http://data.worldbank.org/indicator (accessed 10 June 2011).

World Health Organization 2009a *Diarrhoeal Diseases: Cholera. Initiative for Vaccine Research*. Available at: http://www.who.int/vaccine_research/diseases/diarrhoeal/en/index3.html (accessed 10 August 2010).

World Health Organization 2009b *Weekly Epidemiological Record: Cholera Articles*. Available at: http://www.who.int/topics/cholera/wer/en/index.html (accessed 12 June 12010).

World Health Organization 2013 *Diarrhoeal Disease Fact Sheet*. WHO Media Centre. Available at: http://www.who.int/mediacentre/factsheets/fs330/en/ (accessed 29 June 2013).

World Health Organization & UNICEF 2006 *Meeting the MDG Drinking Water and Sanitation Target: the Urban and Rural Challenge of the Decade*. WHO Press, Geneva, Switzerland.

World Health Organization & UNICEF 2010 *Progress on Sanitation and Drinking-Water: 2010 Update*. WHO Press, Geneva, Switzerland.

World Health Organization & UNICEF 2012 *Progress on Drinking Water and Sanitation: 2012 Update*. WHO Press, Geneva, Switzerland.

World Health Organization & UNICEF 2013 *Joint Monitoring Program's method – deriving Progress Estimates*. Available at: http://www.wssinfo.org/definitions-methods/method/ (accessed 29 June 2013).

Zwane, A. P. & Kremer, M. 2007 What works in fighting diarrhoeal diseases in developing countries? A critical review. *World Bank Research Observer* **22**, 1–24.

First received 31 August 2012; accepted in revised form 18 July 2013. Available online 23 August 2013