Forecasting High-Priority Infectious Disease Surveillance Regions: A Socioeconomic Model

Emily H. Chan,1,2 David A. Scales,1,2 Timothy F. Brewer,4,6,7 Lawrence C. Madoff,4,5 Marjorie P. Pollack,4 Anne G. Hoen,8 Tenzin Choden,1,2 and John S. Brownstein1,2,3,6,7

1Children’s Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, 2Division of Emergency Medicine, Children’s Hospital Boston, 3Department of Pediatrics, Harvard Medical School, Boston, 4ProMED-Mail, International Society for Infectious Diseases, Brookline, and 5Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts; Departments of 6Medicine and 7Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; and 8Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

(See the Editorial Commentary by Polgreen and Polgreen on pages 525–6.)

Background. Few researchers have assessed the relationships between socioeconomic inequality and infectious disease outbreaks at the population level globally. We use a socioeconomic model to forecast national annual rates of infectious disease outbreaks.

Methods. We constructed a multivariate mixed-effects Poisson model of the number of times a given country was the origin of an outbreak in a given year. The dataset included 389 outbreaks of international concern reported in the World Health Organization’s Disease Outbreak News from 1996 to 2008. The initial full model included 9 socioeconomic variables related to education, poverty, population health, urbanization, health infrastructure, gender equality, communication, transportation, and democracy, and 1 composite index. Population, latitude, and elevation were included as potential confounders. The initial model was pared down to a final model by a backwards elimination procedure. The dependent and independent variables were lagged by 2 years to allow for forecasting future rates.

Results. Among the socioeconomic variables tested, the final model included child measles immunization rate and telephone line density. The Democratic Republic of Congo, China, and Brazil were predicted to be at the highest risk for outbreaks in 2010, and Colombia and Indonesia were predicted to have the highest percentage of increase in their risk compared to their average over 1996–2008.

Conclusions. Understanding socioeconomic factors could help improve the understanding of outbreak risk. The inclusion of the measles immunization variable suggests that there is a fundamental basis in ensuring adequate public health capacity. Increased vigilance and expanding public health capacity should be prioritized in the projected high-risk regions.

Keywords. communicable diseases; disease outbreaks; socioeconomic factors; public health; epidemiology.
plentiful, but calls for more research from a population-level perspective in the 1990s highlighted the scarcity of work at the distal range of the spectrum [4, 5]. Since then, a rich literature has built up connecting macro-level socioeconomic factors to population health outcomes such as infant mortality [3] and life expectancy [6]. Broadly speaking, these outcomes have been shown to improve as a country becomes increasingly “developed,” with variation being partly attributed to externalities related to inequality [7] or social cohesion [8]. It would be expected that these trends extend to infectious disease outcomes as well, although in fact the role of population-level socioeconomic factors in infectious disease outcomes has received relatively little attention [9].

In this study, we examined the association between national-level socioeconomic variables and national-level risk for infectious disease outbreaks, as approximated by the number of outbreaks reported in the World Health Organization’s (WHO) Disease Outbreak News reports for which a country was the first to have cases. Efforts toward a better understanding of how various factors drive emergence and spread are critical for formulating policies that optimize strategies for the prevention, detection, and management of outbreaks. Understanding geographic risk is also important to determine whether global resources are being properly allocated, especially as efforts for infectious disease surveillance and research are skewed toward a limited set of wealthier countries [10].

METHODS
Data
Outbreak Data
In a previous study [12], a historical global database of infectious disease outbreaks reported in the WHO’s Disease Outbreak News reports (http://www.who.int/csr/don/en/) was assembled. These reports describe confirmed public health events deemed of international concern. Certain types of outbreaks, such as those of endemic diseases, were excluded from the database (see Appendix for details). The country listed for each outbreak is generally the first country to have cases, or if unclear, the country with the most cases.

We counted the number of times each country (of 213 study countries or territories) appeared in the outbreak database each year (0 if none) from 1996 to 2008 (we excluded 2009 because much of the 2009 socioeconomic data were not yet available at the time of the study).

Socioeconomic Data
Focusing on the distal level of a hierarchical socioeconomic model of disease outbreaks (Figure 1), we selected a single variable to represent each of 9 categories: education, financial welfare, population health, urban development, health infrastructure, gender equality, communication infrastructure, transportation infrastructure, and political health (Table 1). Our choice of variables is based on literature on social and environmental determinants of health and infectious disease, as well as on our impression of data quality and availability. For example, numerous studies have linked education, wealth, and gender equality to better health outcomes [6, 12, 13]. The more educated are more likely to be familiar with healthy practices such as immunization and personal hygiene [14], or more fundamentally, basic literacy and numeracy skills allow one to continue enriching their health knowledge throughout life [15]. Higher incomes also enable access to better nutrition, safe water, sanitation, and medical care [8], while gender equality has positive outcomes for both women’s health [13] and the health of their children [15]. In addition, communication and transportation infrastructure could represent accessibility (for disease spread as well as flow of information and medical care). On the other hand, infrastructure development is often an outcome of urbanization, and as hubs for travel and high-density populations, cities facilitate rapid person-to-person spread of disease [14]. However, in developed countries, cities also tend to have wealthier and more-educated populations, but calls for more research from a population-level perspective in the 1990s highlighted the scarcity of work at the distal range of the spectrum [4, 5]. Since then, a rich literature has built up connecting macro-level socioeconomic factors to population health outcomes such as infant mortality [3] and life expectancy [6]. Broadly speaking, these outcomes have been shown to improve as a country becomes increasingly “developed,” with variation being partly attributed to externalities related to inequality [7] or social cohesion [8]. It would be expected that these trends extend to infectious disease outcomes as well, although in fact the role of population-level socioeconomic factors in infectious disease outcomes has received relatively little attention [9].

In this study, we examined the association between national-level socioeconomic variables and national-level risk for infectious disease outbreaks, as approximated by the number of outbreaks reported in the World Health Organization’s (WHO) Disease Outbreak News reports for which a country was the first to have cases. Efforts toward a better understanding of how various factors drive emergence and spread are critical for formulating policies that optimize strategies for the prevention, detection, and management of outbreaks. Understanding geographic risk is also important to determine whether global resources are being properly allocated, especially as efforts for infectious disease surveillance and research are skewed toward a limited set of wealthier countries [10].

METHODS
Data
Outbreak Data
In a previous study [12], a historical global database of infectious disease outbreaks reported in the WHO’s Disease Outbreak News reports (http://www.who.int/csr/don/en/) was assembled. These reports describe confirmed public health events deemed of international concern. Certain types of outbreaks, such as those of endemic diseases, were excluded from the database (see Appendix for details). The country listed for each outbreak is generally the first country to have cases, or if unclear, the country with the most cases.

We counted the number of times each country (of 213 study countries or territories) appeared in the outbreak database each year (0 if none) from 1996 to 2008 (we excluded 2009 because much of the 2009 socioeconomic data were not yet available at the time of the study).

Socioeconomic Data
Focusing on the distal level of a hierarchical socioeconomic model of disease outbreaks (Figure 1), we selected a single variable to represent each of 9 categories: education, financial welfare, population health, urban development, health infrastructure, gender equality, communication infrastructure, transportation infrastructure, and political health (Table 1). Our choice of variables is based on literature on social and environmental determinants of health and infectious disease, as well as on our impression of data quality and availability. For example, numerous studies have linked education, wealth, and gender equality to better health outcomes [6, 12, 13]. The more educated are more likely to be familiar with healthy practices such as immunization and personal hygiene [14], or more fundamentally, basic literacy and numeracy skills allow one to continue enriching their health knowledge throughout life [15]. Higher incomes also enable access to better nutrition, safe water, sanitation, and medical care [8], while gender equality has positive outcomes for both women’s health [13] and the health of their children [15]. In addition, communication and transportation infrastructure could represent accessibility (for disease spread as well as flow of information and medical care). On the other hand, infrastructure development is often an outcome of urbanization, and as hubs for travel and high-density populations, cities facilitate rapid person-to-person spread of disease [14]. However, in developed countries, cities also tend to have wealthier and more-educated populations, but calls for more research from a population-level perspective in the 1990s highlighted the scarcity of work at the distal range of the spectrum [4, 5]. Since then, a rich literature has built up connecting macro-level socioeconomic factors to population health outcomes such as infant mortality [3] and life expectancy [6]. Broadly speaking, these outcomes have been shown to improve as a country becomes increasingly “developed,” with variation being partly attributed to externalities related to inequality [7] or social cohesion [8]. It would be expected that these trends extend to infectious disease outcomes as well, although in fact the role of population-level socioeconomic factors in infectious disease outcomes has received relatively little attention [9].

In this study, we examined the association between national-level socioeconomic variables and national-level risk for infectious disease outbreaks, as approximated by the number of outbreaks reported in the World Health Organization’s (WHO) Disease Outbreak News reports for which a country was the first to have cases. Efforts toward a better understanding of how various factors drive emergence and spread are critical for formulating policies that optimize strategies for the prevention, detection, and management of outbreaks. Understanding geographic risk is also important to determine whether global resources are being properly allocated, especially as efforts for infectious disease surveillance and research are skewed toward a limited set of wealthier countries [10].

METHODS
Data
Outbreak Data
In a previous study [12], a historical global database of infectious disease outbreaks reported in the WHO’s Disease Outbreak News reports (http://www.who.int/csr/don/en/) was assembled. These reports describe confirmed public health events deemed of international concern. Certain types of outbreaks, such as those of endemic diseases, were excluded from the database (see Appendix for details). The country listed for each outbreak is generally the first country to have cases, or if unclear, the country with the most cases.

We counted the number of times each country (of 213 study countries or territories) appeared in the outbreak database each year (0 if none) from 1996 to 2008 (we excluded 2009 because much of the 2009 socioeconomic data were not yet available at the time of the study).

Socioeconomic Data
Focusing on the distal level of a hierarchical socioeconomic model of disease outbreaks (Figure 1), we selected a single variable to represent each of 9 categories: education, financial welfare, population health, urban development, health infrastructure, gender equality, communication infrastructure, transportation infrastructure, and political health (Table 1). Our choice of variables is based on literature on social and environmental determinants of health and infectious disease, as well as on our impression of data quality and availability. For example, numerous studies have linked education, wealth, and gender equality to better health outcomes [6, 12, 13]. The more educated are more likely to be familiar with healthy practices such as immunization and personal hygiene [14], or more fundamentally, basic literacy and numeracy skills allow one to continue enriching their health knowledge throughout life [15]. Higher incomes also enable access to better nutrition, safe water, sanitation, and medical care [8], while gender equality has positive outcomes for both women’s health [13] and the health of their children [15]. In addition, communication and transportation infrastructure could represent accessibility (for disease spread as well as flow of information and medical care). On the other hand, infrastructure development is often an outcome of urbanization, and as hubs for travel and high-density populations, cities facilitate rapid person-to-person spread of disease [14]. However, in developed countries, cities also tend to have wealthier and more-educated populations, but calls for more research from a population-level perspective in the 1990s highlighted the scarcity of work at the distal range of the spectrum [4, 5]. Since then, a rich literature has built up connecting macro-level socioeconomic factors to population health outcomes such as infant mortality [3] and life expectancy [6]. Broadly speaking, these outcomes have been shown to improve as a country becomes increasingly “developed,” with variation being partly attributed to externalities related to inequality [7] or social cohesion [8]. It would be expected that these trends extend to infectious disease outcomes as well, although in fact the role of population-level socioeconomic factors in infectious disease outcomes has received relatively little attention [9].

In this study, we examined the association between national-level socioeconomic variables and national-level risk for infectious disease outbreaks, as approximated by the number of outbreaks reported in the World Health Organization’s (WHO) Disease Outbreak News reports for which a country was the first to have cases. Efforts toward a better understanding of how various factors drive emergence and spread are critical for formulating policies that optimize strategies for the prevention, detection, and management of outbreaks. Understanding geographic risk is also important to determine whether global resources are being properly allocated, especially as efforts for infectious disease surveillance and research are skewed toward a limited set of wealthier countries [10].
residents with access to better public health infrastructure. Studies have also found positive effects of democracy on population health, while political corruption could lead to mismanagement of government spending [16]. We used The Economist’s Democracy Index [17] to represent political health, while all of our other socioeconomic variables came from the World Development Indicators (WDI) databank (http://data.worldbank.org), December 2010 edition, which covered the 213 study countries. Finally, we decided to also test the United Nations Development Programme’s Human Development Index, which is defined as a “composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, access to knowledge and a decent standard of living” [18].

Confounders (Demographic and Geographic Data)

Previous modeling work has suggested that population and host density [19, 20] and climate (with absolute latitude often used as a proxy measure) [21] are related to both pathogen and disease prevalence and spread. In addition, we believed there would be an elevation gradient for the density of human and disease vector populations. Therefore, we included total population, absolute latitude of a country’s population centroid, and the proportion of a country’s...
surface area below an elevation of 500 meters above mean sea level in our model to control for demographic and geographic factors that potentially elevate the baseline risk for outbreaks.

**Missing Values**

There was a large proportion of missing values in our socioeconomic data, as expected, as not all countries collect and/or report these data or do so every year. Therefore we employed a multiple imputation method using the Amelia II version 1.5 package in R [22]. The method involves filling in n estimated values for each missing value, resulting in n “complete” datasets. Analyses are then carried out on each dataset separately, and the results are combined at the end. Certain transformations (eg, logarithmic) were applied to the data before the missing values were imputed (see Appendix for details).

**Statistical Analysis**

A multivariate mixed-effects Poisson regression model was constructed to predict the number of outbreaks in a given country in a given year from national-level socioeconomic indicators. There was no overdispersion in the outcome variable. The model included fixed-effects terms for the various socioeconomic indicators, and for population, latitude, and elevation as potential confounders, and a random-effects term (intercept only) to account for country-level clustering. Autocorrelation function plots showed no significant temporal autocorrelation. Because the release of the socioeconomic data seemed to generally be delayed by at least 2 years, we lagged the independent and dependent variables by 2 years.

Fixed-effects regressors were chosen for the multivariate model via a backward elimination procedure using the Akaike information criterion. We performed this backward elimination procedure separately on each of the 5 imputed datasets, then identified the variables that remained at the end of at least 3 of the 5 runs. Models were reestimated for each of the 5 imputed datasets including just the identified variables, then a final model was obtained by combining the results of the 5 sets using the procedure described by King and Scheve [23]. Models were constructed using the Zelig package in R [24].

To assess model fit, a Monte Carlo–based version of the Pearson $\chi^2$ goodness of fit test [25] with 2000 replicates was conducted. This simulation-based approach was chosen to avoid having to pool categories, as would be required for the standard Pearson $\chi^2$ test owing to the many cells with an expected count of <5. To assess prediction efficiency and thereby perform model validation, leave-one-out cross-validation was conducted, and average prediction errors between actual and predicted values were computed.

The final model was then used to forecast the expected number of outbreaks in 2010 for each country using socioeconomic data for 2008 (2-year lag). Analyses were performed in R version 2.12.1 (R Project for Statistical Computing, Vienna, Austria).

### RESULTS

Our database included 389 outbreaks during 1996–2008. The highest outbreak rate occurred in the Democratic Republic of Congo (DR Congo; average of 1.77 outbreaks per year), followed by Sudan (0.92) and China, Brazil, India, and Guinea (0.77 each). Of the 213 study countries, 108 had no outbreaks during the entire study period (see Supplementary Table A2 for list), and the majority of the other countries had no reported outbreaks in most years. A summary of the socioeconomic data (minimum and maximum values for countries with no outbreaks as compared with countries with at least 1 outbreak during the study period) is given in Table 1.

In preliminary models (data not shown), all socioeconomic variables except for health expenditure in the public sector were on their own significantly associated with outbreak rate. The final multivariate Poisson mixed-effects regression model predicting the number of outbreaks in each country in each year included 2 socioeconomic variables: the log of the number of telephone lines (per 100 people; coefficient = −0.59, $P$ value = .02) and the log of measles immunization (percentage of children aged 12–23 months; coefficient = −0.20, $P$ value <.001); see Table 2. In addition, the population and latitude

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (P Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-11.59 (&lt;.001)$</td>
</tr>
<tr>
<td>Log(Total population)</td>
<td>0.47 (&lt;.001)</td>
</tr>
<tr>
<td>Absolute latitude</td>
<td>$-0.03 (&lt;.001)$</td>
</tr>
<tr>
<td>Elevation (proportion of country &lt;500 m)</td>
<td>...</td>
</tr>
<tr>
<td>Logit(Measles immunization)</td>
<td>$-0.59 (.02)$</td>
</tr>
<tr>
<td>Log(Telephones)</td>
<td>$-0.20 (&lt;.001)$</td>
</tr>
<tr>
<td>Logit(Primary school enrollment)</td>
<td>...</td>
</tr>
<tr>
<td>Logit(Human Development Index)</td>
<td>...</td>
</tr>
<tr>
<td>Logit(Poverty headcount ratio)</td>
<td>...</td>
</tr>
<tr>
<td>Logit(Urban population)</td>
<td>...</td>
</tr>
<tr>
<td>Democracy Index</td>
<td>...</td>
</tr>
<tr>
<td>Logit(Health expenditure, public)</td>
<td>...</td>
</tr>
<tr>
<td>Log(Paved road density)</td>
<td>...</td>
</tr>
<tr>
<td>Ratio of girls to boys in primary + secondary education</td>
<td>...</td>
</tr>
</tbody>
</table>

A dash indicates that an initially included variable was not included in the final model after an Akaike information criterion backwards elimination procedure was conducted. The estimated coefficients for the fixed effects terms are shown. Random-effects terms (intercept only) were also included in the model to account for country-level clustering but are not shown.
confounder variables remained. There was a high correlation between many of the socioeconomic variables themselves as well as with the confounder variables (Supplementary Table A2). Model diagnostic tests indicated an adequate model; the Monte Carlo–based $\chi^2$ goodness of fit test did not demonstrate any significant lack of fit ($P$ value = .72), while a leave-one-out cross-validation indicated that the error between predicted and observed values differed by 0.19 outbreaks on average.

Using the model to forecast the expected number of outbreak for 2010, the top 5 high-risk geographic areas were predicted to be DR Congo (0.77 outbreaks), China (0.59), Brazil (0.54), India (0.48), and Sudan (0.43); see Figure 2A and Table 3. Figure 2B depicts for each country the difference between the forecasted rate for 2010 and the average annual rate during 1996–2008. For most countries, the model generally predicted fewer outbreaks in 2010 compared to their historical rates, consistent with the general decline in the number of outbreaks during the period 1996–2008. Dividing countries into risk-level categories based on their historical average annual outbreak rate during 1996–2008, Table 4 lists for each category the countries with the greatest percentage of decrease between their historical average annual outbreak rate and their model-forecasted number of outbreaks for 2010. Contrarily, several low-risk countries became moderately higher risk. Colombia’s 2010 forecast was an 89% increase (from 0.08 to 0.15), and Indonesia’s was 22% (from 0.23 to 0.28). Togo and Haiti also had a projected increased risk that was 13% and 12% higher, respectively, although the absolute change (0.01 each) was minimal. Countries that originally had zero outbreaks in our database for 1996–2008 were excluded from these percentage of change calculations as the denominator cannot be zero. Among these countries, Myanmar, Mexico, and Papua New Guinea were forecasted to have 0.16, 0.12, and 0.11 outbreaks, respectively.

**DISCUSSION**

Of the various socioeconomic factors related to education, financial welfare, population health, urban development, health infrastructure, gender equality, communication infrastructure, transportation infrastructure, and political health that were included in the initial full multivariate model to explain the number of outbreaks in a country in a given year, only the
proportion of children aged 12–23 months vaccinated against measles (representing population health) and the density of telephone lines (representing communication infrastructure) remained in the final model. These results nonetheless demonstrate that national-level socioeconomic indicators, in conjunction with other known correlates such as population and latitude, could be useful for model-based projections for surveillance purposes.

Because of the high correlation between the socioeconomic variables themselves as well as with the confounder variables, it is not possible to tease apart the individual contribution of each factor to a country’s outbreak rate or to infer causality, but the model results as a whole do provide a basis for a link between socioeconomic well-being at the population level and national risk for infectious disease outbreaks.

Both governments and nongovernmental development assistance agencies spend billions of dollars on public health and medical care on the principle that disease control can reduce prevalence. Previous studies have shown strong negative associations between healthcare spending and pathogen prevalence [26] and the speed of infectious disease spread [27]. Although we looked at outbreak rates and not disease prevalence, the absence of the variable for the proportion of health expenditure in the public sector in our final model was still a surprising result. This finding may be related to the “displacement effect,” whereby countries receiving foreign aid for health programs use the aid in place of (and thereby reduce) domestic health spending [15, 28]. Furthermore, in most countries, government spending ends up being concentrated in cities, though city dwellers tend to be better off already [15].

Another related health variable, the proportion of children aged 12–23 months vaccinated against measles, did consistently remain in our model across most of our imputation sets. Measles vaccination coverage is often used as a proxy indicator for access to child health services [29], and in contrast to the health expenditure variable is perhaps a better measure of the public health state as it represents actual implementation and efficacy; designating funds for health expenditures is less meaningful if committed funds are not ultimately dispensed, or if dispensed funds are not managed effectively. Publicly funded immunization programs and strong public health education campaigns enable access and encourage vaccine uptake. Therefore, the inclusion of the measles immunization variable in our final model suggests a fundamental basis in adequate public health capacity and efficacy. A 1992 Institute of Medicine report also identified breakdown of public health measures as 1 of 6 categories of factors responsible for the emergence of infectious diseases [30].

Density of telephone lines, included to represent communication infrastructure, was the only other socioeconomic variable that remained in our final model. Its connection with outbreak rate is less obvious, but one theory is that it may be representative of the flow of information in and out of the community. A community that is better informed may be better equipped for public health threats. However, it may merely be acting as a proxy of development, which is associated with better healthcare and public health infrastructure.

Other studies have described the global geographic historical distribution of the origins of emerging infectious disease events [20] and of human pathogen richness [26]. We analogously identified some potential future high-risk geographic areas for propensity of outbreaks. Some of these countries, such as the DR Congo, which took the top spot, have historically been at high risk for outbreaks and were predicted to continue to sit high in the risk ranking in 2010. We also looked at countries with a high increase, proportionately, between their historic average annual outbreak rate during

Table 4. Top 5 Countries, for Each of 4 Risk-Level Categories, With the Greatest Projected Percentage of Decrease Between the Average Number of Outbreaks per Year During 1996–2008 and the Model-Forecasted Number of Outbreaks for 2010

<table>
<thead>
<tr>
<th>Country</th>
<th>Projected No. Outbreaks, 2010</th>
<th>Average No. Outbreaks per Year, 1996–2008</th>
<th>Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk (&gt;0.75 outbreaks/y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinea</td>
<td>0.31</td>
<td>0.77</td>
<td>−69</td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>0.77</td>
<td>1.77</td>
<td>−57</td>
</tr>
<tr>
<td>Sudan</td>
<td>0.43</td>
<td>0.92</td>
<td>−54</td>
</tr>
<tr>
<td>India</td>
<td>0.48</td>
<td>0.77</td>
<td>−37</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.54</td>
<td>0.77</td>
<td>−30</td>
</tr>
<tr>
<td>Upper-intermediate risk (0.5–0.75 outbreaks/y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozambique</td>
<td>0.20</td>
<td>0.62</td>
<td>−67</td>
</tr>
<tr>
<td>Liberia</td>
<td>0.24</td>
<td>0.69</td>
<td>−66</td>
</tr>
<tr>
<td>Angola</td>
<td>0.26</td>
<td>0.69</td>
<td>−63</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>0.21</td>
<td>0.54</td>
<td>−62</td>
</tr>
<tr>
<td>Zambia</td>
<td>0.22</td>
<td>0.54</td>
<td>−60</td>
</tr>
<tr>
<td>Lower-intermediate risk (0.25–0.5 outbreaks/y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabon</td>
<td>0.08</td>
<td>0.31</td>
<td>−74</td>
</tr>
<tr>
<td>Congo Republic</td>
<td>0.10</td>
<td>0.38</td>
<td>−73</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>0.12</td>
<td>0.38</td>
<td>−68</td>
</tr>
<tr>
<td>Cambodia</td>
<td>0.15</td>
<td>0.38</td>
<td>−62</td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.19</td>
<td>0.46</td>
<td>−59</td>
</tr>
<tr>
<td>Low risk (&lt;0.25 outbreaks/y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>0.01</td>
<td>0.08</td>
<td>−86</td>
</tr>
<tr>
<td>Mongolia</td>
<td>0.01</td>
<td>0.08</td>
<td>−83</td>
</tr>
<tr>
<td>Norway</td>
<td>0.01</td>
<td>0.08</td>
<td>−83</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>0.03</td>
<td>0.15</td>
<td>−82</td>
</tr>
<tr>
<td>Albania</td>
<td>0.02</td>
<td>0.08</td>
<td>−80</td>
</tr>
</tbody>
</table>

Countries are categorized into risk-level categories based on their historical 1996–2008 average outbreak rate.
1996–2008 and their projected number of outbreaks in 2010. In particular, the model results suggest that Colombia and Indonesia could potentially require increased vigilance and outbreak prevention initiatives as they were predicted to have a large percentage of increase in risk. Myanmar, Mexico, and Papua New Guinea, which had no outbreaks listed in our historical database during 1996–2008 but had the highest predicted absolute increase in outbreak rate for 2010, are perhaps additional target countries.

We decided to not compare the expected number of outbreaks for 2010 as predicted by the model to the observed number of outbreaks in 2010. Because the model essentially represents a smoothed overall trend, the comparison would be a difficult one owing to the instability of data when looking at any given year.

Limitations of this study include reporting and selection biases in the WHO’s Disease Outbreak News reports, from which we drew our dataset. The problem of missing values is also an inherent issue in socioeconomic data and can result in biases when data is not “missing at random.” Another limitation is that all outbreaks, whether of high morbidity and far-reaching or ultimately relatively limited, were weighted equally. However, the problem is lessened by the fact that the WHO’s Disease Outbreak News reports focus on outbreaks that are large or dangerous enough to be of international significance.

The high correlation between many of the socioeconomic variables is a main concern as it makes it difficult to distinguish their individual effects when analyzed together in a multivariate analysis. While we believe ecological studies such as this one do have value [31], one must be cautious to not imply causation in an ecological study, as some of these socioeconomic factors, as indicators of development status and financial capacity, may merely be proxy measures for other underlying causal pathways [32]. The aggressive implementation of health programs in areas at need could further complicate the picture by creating an impression that areas with health programs are associated with poorer health outcomes [33].

Our results demonstrate that socioeconomic epidemiology should play an important role in research efforts aimed at understanding the drivers of infectious disease outbreaks. Although there are individual-level mechanisms that biologically dictate how infection proceeds, it is important to understand population-level dynamics as well, since outbreaks of infectious disease are ultimately driven at the population level, and it is this gap in our understanding at the population level that we have aimed to help fill in deciding to conduct this study as an ecological one. From the perspective of actionable strategies, it may also be easier to enact policy changes than it is to try to control biological mechanisms or enforce individual behaviors. Others have also highlighted the need for identifying population-level risk factors so as to not miss opportunities to adopt societal-level interventions [4] and to revitalize the concept of public health [5]. Understanding these components in finer detail could help inform how to best strategically allocate funds to specific public health measures and high-priority geographic regions.

**Supplementary Data**

**Supplementary materials** are available at *Clinical Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/cid/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyrighted. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

**Notes**

**Acknowledgments.** We are grateful to Alexei Zelenev (Yale University) for his statistical feedback.

**Financial support.** This work was supported by research grants from Google.org and the National Library of Medicine at the National Institutes of Health (1R01LM01812–01, S08LM009776–02).

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**