
Electronic Event-based Surveillance for Monitoring Dengue, Latin America

Anne G. Hoen,¹ Mikaela Keller, Aman D. Verma,
David L. Buckeridge, and John S. Brownstein

The current dengue epidemic in Latin America represents a major threat to health. However, surveillance of affected regions lacks timeliness and precision. We investigated real-time electronic sources for monitoring spread of dengue into new regions. This approach could provide timely estimates of changes in distribution of dengue, a critical component of prevention and control efforts.

Dengue, a potentially fatal viral disease, has been recognized for >200 years (1). Once sporadic and limited geographically, dengue viruses (DENVs) and their mosquito vectors have spread globally, putting an estimated 2.5 billion persons at risk throughout the tropical and subtropical regions of the world (2). Factors contributing to the dramatic expansion of DENV activity include demographic changes such as population growth, urbanization, and globalization, and reductions in vector control and other public health measures (3). Because of the nature of passive surveillance, conventional systems have limited ability in identifying new epidemics quickly (1), thus suggesting a role for alternative information sources.

Free or low-cost sources of unstructured information, such as Internet news outlets, health expert mailing lists, social media sites, and queries to online search engines, when computationally filtered and mined, can provide detailed local and near real-time data on potential or confirmed disease outbreaks (4). For dengue in particular, our group and others recently reported on a set of Google search terms that parallel temporal trends in official

dengue case counts (5,6). These event-based data sources can provide insight into new and ongoing public health challenges in areas of the world with limited public health reporting infrastructure.

Few studies have investigated the value of unofficial sources for monitoring recent geographic expansion of infectious disease risk. Using dengue as a case study, we report on the utility of electronic outbreak surveillance for real-time monitoring of recent infectious disease spread.

The Study

We focused on the geographic range of DENVs in Latin America and the Caribbean, where dengue is widespread and expanding in range. We attempted to identify areas contiguous with previously known dengue-endemic zones where new DENV transmission is occurring by using reports of recent outbreaks.

Known dengue-endemic areas were defined as dengue risk areas identified by the US Centers for Disease Control and Prevention (Atlanta, GA, USA) Health Information for International Travel (commonly referred to as the Yellow Book), 2010 (7) and 2012 (8) editions. Each edition of this book reflects the known distribution of dengue risk in the prior 2 years. To characterize spread according to the Yellow Book, we identified areas that were classified as no known dengue risk in 2010 but were changed to risk areas in the 2012 edition (hereafter referred to as new dengue-endemic areas).

Outbreak data for December 1, 2009–March 18, 2011, were collected from HealthMap (<http://www.healthmap.org/en/an>), an open access online infectious disease outbreak monitoring system (9,10). HealthMap integrates outbreak-related data from >30,000 electronic sources, including the news media, ProMED-mail, and other electronic public health reporting sources, by using algorithms to classify the diseases and locations associated with each report. Because we wanted to identify spread into new dengue-endemic zones, we limited our analyses to areas that were identified as having no known dengue risk in the 2010 Yellow Book but that were contiguous with ≥ 1 risk areas in the 2010 Yellow Book. We identified 53 dengue outbreaks distributed in 60 of these areas.

We fitted a bivariate Gaussian mixture model to the extracted HealthMap alerts to model a continuous surface of outbreak density (online Technical Appendix, wwwnc.cdc.gov/EID/pdfs/12-0055-Techapp.pdf). This modeled outbreak probability density surface represents a risk map of recent DENV spread into areas of previously unknown dengue endemicity according to the 2010 Yellow Book (Figure 1). We compared our map with the geographic

Author affiliations: Children's Hospital Boston, Boston, Massachusetts, USA (A.G. Hoen, J.S. Brownstein); Harvard Medical School, Boston (A.G. Hoen, J.S. Brownstein); Institut de Recherche en Informatique et Automatique, Lille, France (M. Keller); Université de Lille, Lille (M. Keller); McGill University, Montreal, Quebec, Canada (A.G. Hoen, A.D. Verma, D.L. Buckeridge, J.S. Brownstein); and Agence de la Santé et des Services Sociaux de Montreal, Montreal (D.L. Buckeridge)

DOI: <http://dx.doi.org/10.3201/eid1808.120055>

¹Current affiliation: The Geisel School of Medicine at Dartmouth College, Lebanon, New Hampshire, USA.

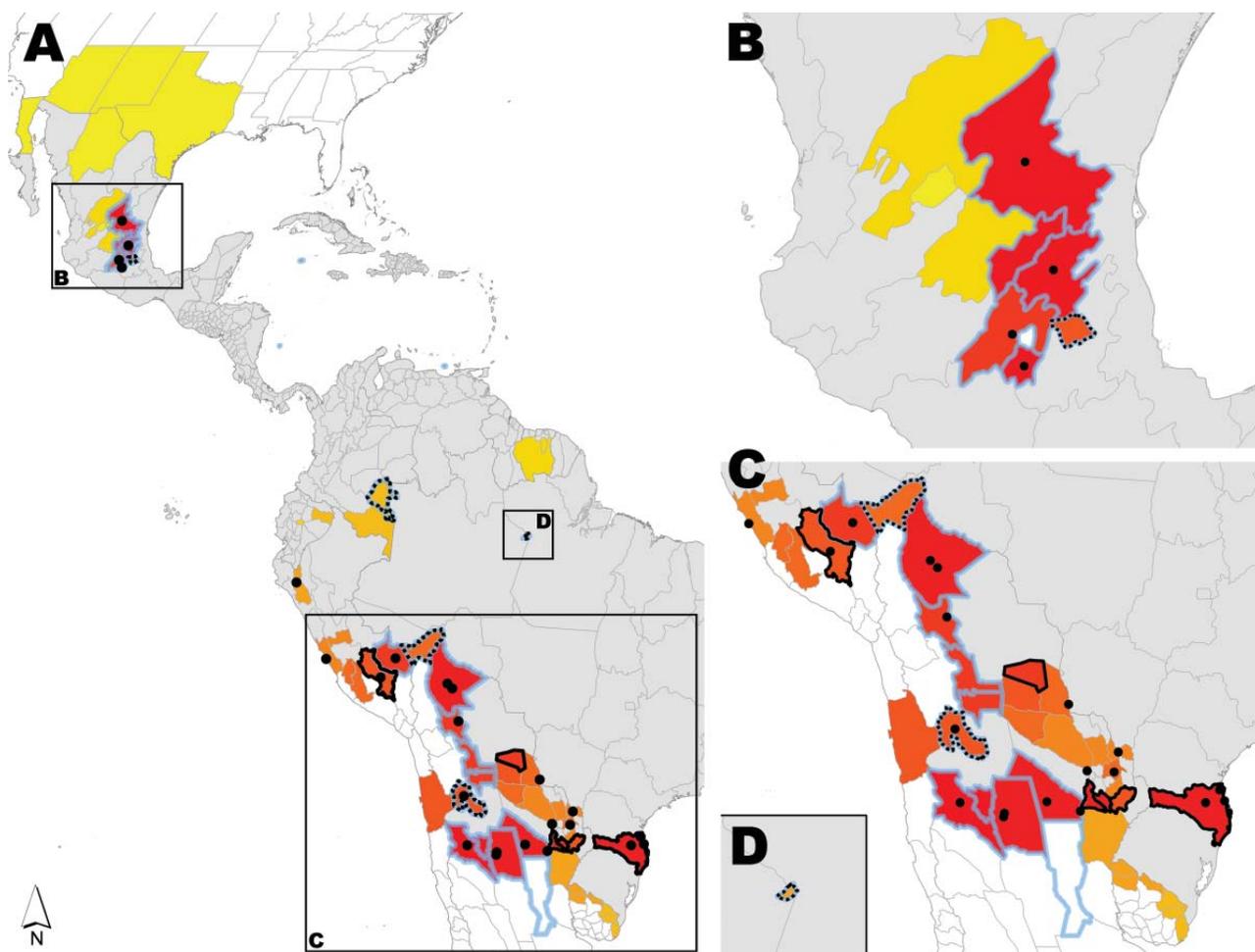


Figure 1. A) Regions in Latin America analyzed for dengue. B) Central Mexico; C) central South America; D) eastern Amazonas, Brazil. Thin gray lines indicate boundaries of first-level administrative units. Areas to which dengue was identified in the 2010 Yellow Book are shaded in gray. New dengue-endemic areas identified in the 2012 Yellow Book are outlined in blue. Dots indicate HealthMap dengue-related alerts. Modeled HealthMap alert probability density surface is shown in a gradient from yellow to red with yellow areas predicted as having lower alert densities and red areas predicted as having higher alert densities according to the model. Areas outlined with heavy black solid lines were classified as high HealthMap alert density but were not identified in either Yellow Book edition as dengue risk areas. Areas outlined with heavy black dashed lines were classified as low HealthMap alert density but were identified in the 2012 Yellow Book as areas at risk for dengue.

distribution of new dengue-endemic areas identified in the 2012 Yellow Book. Details of the datasets, models, and statistical methods are available in the online Technical Appendix.

Figure 1 shows that high dengue outbreak activity occurred adjacent to previously recognized dengue-endemic zones in 6 states in central Mexico and in parts of northern Argentina, southern Brazil, Bolivia, and Paraguay. We used receiver-operating characteristic analysis with cross-validation (Figure 2) to set a threshold dengue report density that best identifies new dengue-endemic areas (Figure 1; online Technical Appendix). Of the 19 new dengue-endemic areas reported in the 2012 Yellow Book, this threshold identified 14 (74%) as being

at elevated risk of endemicity, according to the dengue outbreak probability density estimated by our model. Of the 41 areas that remained unidentified as dengue-endemic areas in the 2012 Yellow Book, our model classified 35 (85%) as having reduced risk of endemicity.

When compared with the Yellow Book, our model incorrectly classified 6 areas as at elevated risk (Figure 1). All alerts in these areas described outbreaks of cases acquired in a nearby known dengue-endemic region of the country. One alert also warned of the recent discovery of dengue vector mosquito larvae by the local surveillance program. The model also classified 5 Yellow Book dengue-endemic areas as at reduced risk (Figure 1). Four of these areas were rural or isolated. Although other explanations

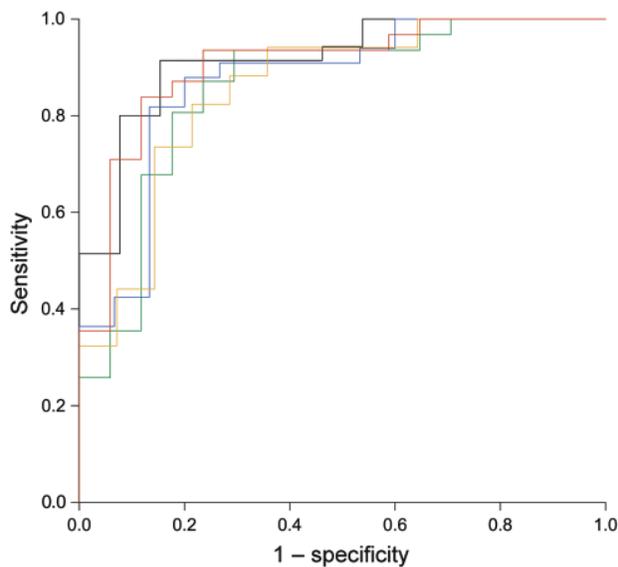


Figure 2. Receiver-operating characteristic plot of 5-fold cross-validated HealthMap alert density-based classification with new dengue-endemic areas identified by the 2012 Yellow Book as the standard.

likely exist, the low observed sensitivity in these areas illustrates certain limits of any system that relies on Internet-based information flow for monitoring disease spread.

Conclusions

Electronic event-based surveillance systems such as HealthMap and others are frequently used by public health authorities, travelers, physicians and patients, to gain a real-time understanding of global outbreak activity. The HealthMap dengue feed, DengueMap, is currently part of the online dengue information resource of the Centers for Disease Control and Prevention (<http://www.cdc.gov/dengue/>). Used in combination with traditional case reporting, HealthMap and other electronic surveillance systems have proven value for enhancing the timeliness of outbreak discovery and information dissemination (11). However, these information sources may also provide added value for monitoring ongoing spread.

Although the signal of DENV activity detected by HealthMap is relatively robust, it has certain limitations. First, the signal tends to be sparse in areas with limited reporting because of low population density or incomplete coverage by the news or social media. Second, the signal can be surrounded by background noise because separating reports caused by cases in travelers from true autochthonous transmission is difficult with automated methods. By limiting our analysis to areas contiguous with known dengue-endemic areas and smoothing outbreak alerts into an outbreak-density surface, we were able to

identify a reliable signal of dengue spread. Although this analysis was performed retrospectively, the timeliness of this signal far outperforms any traditional surveillance data stream. Passive case report-based surveillance systems typically operate at a delay of weeks to months, which limits their value for providing a picture of geographic spread, especially on an international scale where surveillance delays may be even more prolonged.

We have demonstrated a novel approach to real-time monitoring of recent expansion of DENV activity in Latin America. Using outbreak reports captured by HealthMap, we identified a signal of geographic expansion of dengue activity that would precede official reports of the geographic distribution of dengue-endemic areas. Currently, no reliable surveillance system is in widespread use that reports the distribution of DENV activity on an ongoing basis and enables near real-time monitoring of trends in geographic expansion. Such a system should enhance the ability of regional and global public health authorities to dynamically allocate resources within a time frame that might effectively avert a full-blown epidemic. Like other large-scale surveillance data sources, our results must be interpreted cautiously. However, when used in conjunction with traditional surveillance methods, our approach has the potential to provide a timely estimate of changes in the geographic distribution of dengue, a critical component of targeted prevention and control efforts.

Acknowledgments

We thank Michael Johansson for critically revising the manuscript and providing Yellow Book data; and Katia Charland, Laura Pinheiro, Emily Chan, Susan Aman, Clark Freifeld, Katelynn O'Brien and Sumiko Mekaru for helpful discussions and database support.

This study was supported by National Institutes of Health grants R01 LM010812 and G08 LM009776.

Dr Hoen is an infectious disease epidemiologist in the Department of Community and Family Medicine at The Geisel School of Medicine at Dartmouth College. Her research interests include informatics-based approaches to understanding infectious disease risk, emergence and spread.

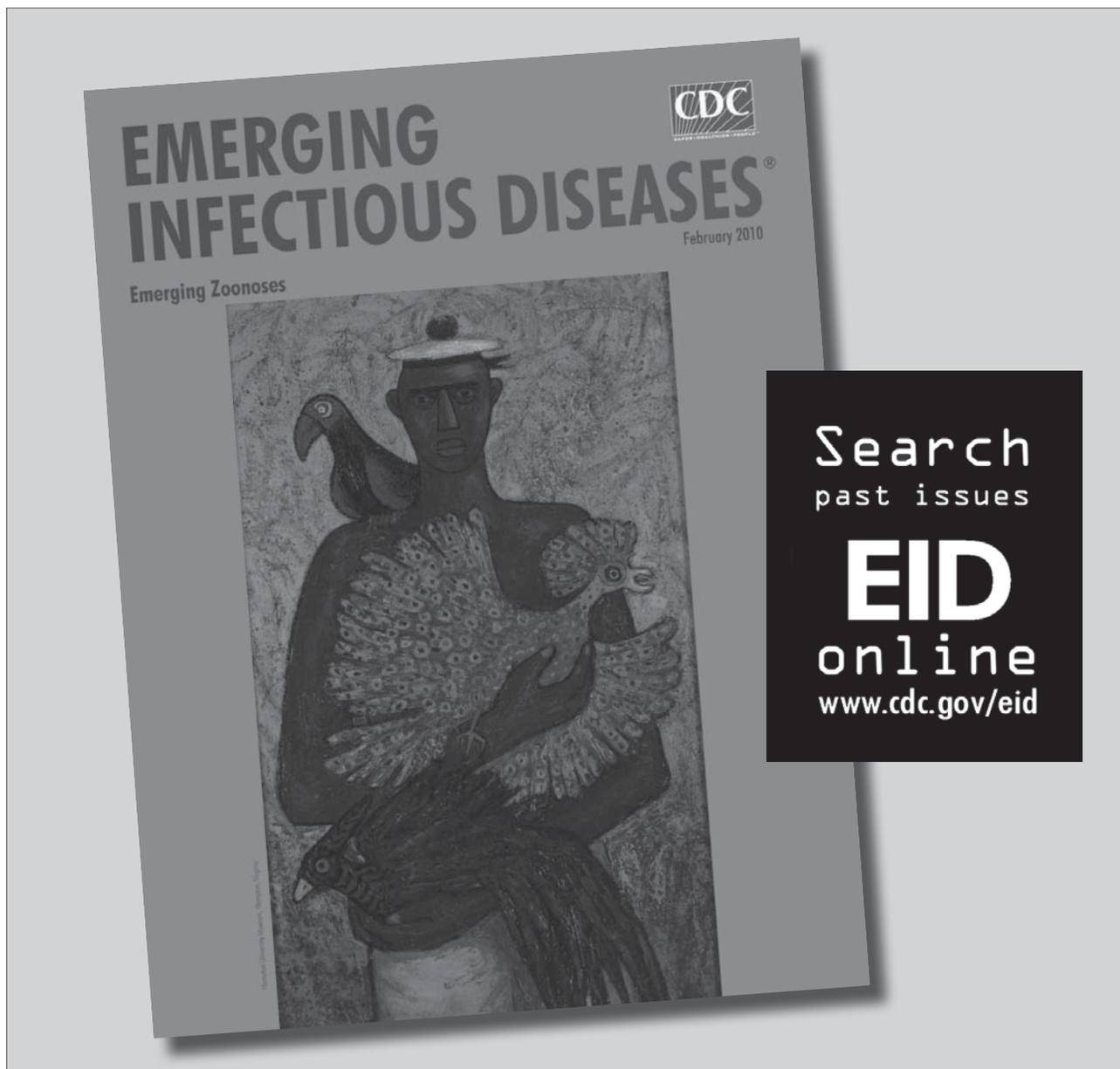
References

- Gubler DJ. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerg Infect Dis.* 1995;1:55-7. <http://dx.doi.org/10.3201/eid0102.950204>
- World Health Organization. Fact sheet no. 117: dengue and dengue haemorrhagic fever. Geneva: The Organization; 2009.
- Gubler DJ. The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res.* 2002;33:330-42. [http://dx.doi.org/10.1016/S0188-4409\(02\)00378-8](http://dx.doi.org/10.1016/S0188-4409(02)00378-8)

4. Morse SS. Global infectious disease surveillance and health intelligence. *Health Aff (Millwood)*. 2007;26:1069–77. <http://dx.doi.org/10.1377/hlthaff.26.4.1069>
5. Amarasinghe A, Kuritsky JN, Letson GW, Margolis HS. Dengue virus infection in Africa. *Emerg Infect Dis*. 2011;17:1349–54.
6. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis*. 2011;5:e1206. <http://dx.doi.org/10.1371/journal.pntd.0001206>
7. Centers for Disease Control and Prevention. CDC health information for international travel 2010. Atlanta: The Centers; 2010.
8. Centers for Disease Control and Prevention. CDC health information for international travel 2012. Atlanta: The Centers; 2012.
9. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc*. 2008;15:150–7. <http://dx.doi.org/10.1197/jamia.M2544>
10. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontiers: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*. 2008;5:e151. <http://dx.doi.org/10.1371/journal.pmed.0050151>
11. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. *Proc Natl Acad Sci U S A*. 2010;107:21701–6. <http://dx.doi.org/10.1073/pnas.1006219107>

Address for correspondence: Anne G. Hoen, Dartmouth-Hitchcock Medical Center, 1 Medical Center Dr, HB 7937, Lebanon, NH 03756, USA; email: anne.g.hoen@dartmouth.edu

Use of trade names is for identification only and does not imply endorsement by the Public Health Service or by the US Department of Health and Human Services.



Electronic Event–based Surveillance for Monitoring Dengue, Latin America

Technical Appendix

Bivariate Gaussian mixture model applied to extracted HealthMap alerts to model a continuous surface of outbreak density

Datasets

CDC Yellow Book

We obtained the geographic boundaries of dengue risk areas reported in the U.S. Centers for Disease Control and Prevention’s Health Information for International Travel (commonly referred to as the Yellow Book), 2010 and 2012 editions (1,2). The Yellow Book is published every 2 years as a reference for health care providers who advise international travelers on health risks. The Yellow Book classifies regions of the world into dengue risk areas and areas with no known dengue risk on the basis of whether dengue is considered to be endemic in these areas. This classification relies on expert-reviewed reports and peer-reviewed publications, as well as communications with subject-matter authorities. Areas are drawn at the scale of first-level administrative units (subnational regions such as state or province). For our study, we used only areas that had been labeled as no known risk but were adjacent to and contiguous with risk areas in the 2010 Yellow Book.

Because the Yellow Book is intended as information for clinicians and travelers and does not provide detailed information on the precise criteria that are used in making its classifications, we are limited in our ability to make inferences regarding the exact meanings of the 2 classifications. However, to our knowledge, the Yellow Book provides the most geographically comprehensive and most frequently updated dengue risk map currently available, which motivated its use as our reference map for this study.

HealthMap

HealthMap has been described (3). In brief, HealthMap is an online infectious disease monitoring system that captures online information about outbreaks with automated hourly scans of >30,000 electronic sources in 9 languages and then uses a text processing algorithm to classify items by disease and geographic location. HealthMap sources include the news media through Google News and other news aggregators; the moderated listserv ProMED Mail; the GeoSentinel Surveillance Network's database of illness among travelers; and formal reports from the Food and Agriculture Organization and World Health Organization, among others. Approximately 80% of the volume of HealthMap alerts reflects information captured directly from the news media.

We extracted all HealthMap alerts in the Americas that had been geolocated to first-level administrative areas contiguous with 2010 Yellow Book dengue-positive areas. Our analysis was limited to HealthMap alerts issued from December 1, 2009, when HealthMap began curating Spanish and Portuguese language reports, through March 18, 2011, when we extracted our dataset for this analysis.

Spatial Modeling

We fit a bivariate Gaussian mixture model to our dataset of HealthMap alerts. This is a statistical model of a probability density function made up of a weighted sum of Gaussian densities. While a common application of mixture modeling is cluster detection, mixture models are also used for density estimation. In a spatial context, this enables estimation of the underlying and unobservable continuous density of a set of observed points. Our candidate spatial models included the covariates latitude and longitude represented by Mollweide equal area-projected x- and y-coordinates. Therefore, the computed models are weighted sums of component bivariate Gaussian distributions that represent the probability distribution of HealthMap observations in the study area. We fit bivariate Gaussian mixture models with 1, 2, ..., 10 components and diagonal or spherical covariance functions and then selected the best fit model according to the Bayesian information criterion. We ensured that the best-fit model did not fall at the minimum or maximum number of components considered. Using the best-fit model, we then estimated the mean probability density for each first-level administrative unit contiguous with 2010 Yellow Book dengue-positive areas. This value that can be thought of as a model-based estimate of the intensity of HealthMap alerts for each subnational area.

Statistical Methods

Of included administrative areas, we identified all that had been changed from an area with no known dengue risk in the 2010 Yellow Book edition to dengue risk area in the 2012 Yellow Book edition. We hypothesized that these were areas into which dengue had spread between editions of the Yellow Book, and that the areas with the strongest HealthMap alert density would correspond to these areas of recent expansion. To test this hypothesis, we plotted receiver operating characteristic curves to evaluate the sensitivity and specificity of a range of threshold HealthMap alert density values for predicting the occurrence of new dengue risk areas in the 2012 Yellow Book and selected optimally predictive density threshold cutoffs by using the Youden statistic (4). To avoid over fitting, we performed receiver operating characteristics analysis with 5-fold cross validation.

Mapping was performed by using ArcGIS Desktop version 10 (ESRI, Redlands, CA, USA). Statistical analyses were performed by using statistical software version 2.12.2 with the packages MCLUST and ROCR (<http://cran.r-project.org/bin/windows/base/old.2.12.2>).

References

1. Centers for Disease Control and Prevention. CDC health information for international travel 2010. Atlanta: The Centers; 2010.
2. Centers for Disease Control and Prevention. CDC health information for international travel 2012. Atlanta: The Centers; 2012.
3. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc.* 2008;15:150–7. [PubMed http://dx.doi.org/10.1197/jamia.M2544](http://dx.doi.org/10.1197/jamia.M2544)
4. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J.* 2005;47:458–72. [PubMed http://dx.doi.org/10.1002/bimj.200410135](http://dx.doi.org/10.1002/bimj.200410135)