

A New Approach to Monitoring Dengue Activity

Lawrence C. Madoff^{1,2}, David N. Fisman^{3,4,5*}, Taha Kass-Hout⁶

1 Division of Infectious Disease and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, **2** Division of Epidemiology and Immunization, Massachusetts Department of Public Health, Boston, Massachusetts, United States of America, **3** Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada, **4** Departments of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada, **5** Department of Medicine, University of Toronto, Toronto, Ontario, Canada, **6** Public Health Surveillance Program Office, Office of Surveillance, Epidemiology & Laboratory Services, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

Among the world's most vexing emerging infectious diseases, dengue continues to spread, and in its many endemic areas is a major public health problem [1–3]. There is no vaccine available, and the immunology of dengue, whereby immunological “priming” can result in extremely severe manifestations (e.g., dengue hemorrhagic fever) complicates vaccine development [4]. Thus dengue control is dependent on controlling the mosquito vector, and resistance to insecticides, environmental and social disruption, climate change, and global movement of goods and people (and incidentally, vectors) provide ongoing hurdles to effective vector control [5,6].

Accurate risk analysis and allocation of resources for dengue control depends on disease surveillance. Dengue surveillance is similarly complex and depends in most areas on formal surveillance systems that capture case counts (or via syndromic surveillance at sentinel sites) [7,8]. Laboratory reporting of serology can confirm not only individual cases but identify the viral serotypes found in a given area at a point in time [9]. Analysis of mosquito populations can also confirm dengue circulation and provide information on viral types [10–12]. Formal surveillance has many advantages: precise counts of case numbers, good geographic localization and the potential to identify precise disease etiology among them.

However, formal infectious disease surveillance systems have important limitations, including lags between case occurrence and reporting. Sentinel sites may report cases only periodically or fail to report altogether for a variety of reasons. Delays in reporting may occur when governmental organizations charged with surveillance aren't able to adequately collect and analyze data or publish reports in a timely manner. These problems may be particularly daunting in developing countries with limited resources to devote to strengthening surveillance systems: robust formal public health surveillance is expensive, requiring major investments in trained personnel, communications, buildings and equipment. Indeed, the economic conditions that prevent development of robust surveillance systems may also be those that potentiate dengue transmission: for example, a seroprevalence study performed in a city straddling the Texas-Mexico border found marked differences in dengue seroprevalence on the Mexican side of the border in association with economic disadvantage [13].

The hierarchical nature of formal public health surveillance also poses challenges to surveillance. Hierarchical reporting structures can lose data at any point of interaction, for example when a regional authority fails to report to a national one. Finally, in some situations there can be short-term disincentives for the timely and transparent reporting of disease activity: governments may fear that surges in disease activity may chase away tourists or visitors, or may undermine government credibility [14].

To address these drawbacks, a complementary system of informal surveillance tools have been developed, some by governmental agencies, but many by non-governmental organizations and/or

researchers. Event-based surveillance systems such as ProMED, GPHIN HealthMap and BioCaster rely on unofficial reports of disease, for example from clinicians or web-based health-related news media, to report on disease outbreaks [15–17]. Such systems have proven reliable and timely and informal sources of information were even recognized in the 2005 revision of the International Health Regulations as important sources of epidemic intelligence [18,19]. The rapid and accelerating growth of the Internet has improved the usefulness and sensitivity of these systems and they have likely improved the timeliness of outbreak reporting [18], and the ever-expanding availability of electronic information has also led to the discovery of other types of analyses that detect disease outbreaks. “Web-crawlers” (software programs that search internet sites for specific terms, and then use these search terms to generate reports or maps of disease activity) can provide important information on disease outbreaks that may be published on nongovernmental websites, in online newspapers, and in blogs, and this approach powers the widely-used HealthMap system mentioned above [20]. In the context of the recent cholera outbreak in Haiti, there were inconsistencies in initial accounts of regional disease activity, but information from HealthMap proved useful in the construction of a mathematical model that predicted disease spread on the island [21].

The analysis of real time search queries—the so-called “searchstream”—has been shown to be a sensitive and timely means of evaluating geographically-specific trends in influenza; both Yahoo and Google search engines have proven to be powerful tools for influenza surveillance [22,23]. More recently, evaluation of data from the microblogging website Twitter has been shown to provide useful information about *both* disease activity and disease concern related to the 2009 influenza pandemic [15]. Finally, the widespread availability of smartphone technologies makes it possible to interact with population members to elicit information on illness (so-called “crowdsourcing”), and also (by using cellphone or smartphone network data) to evaluate the movement of populations, which may be a key predictor of how epidemics spread [24–26].

Citation: Madoff LC, Fisman DN, Kass-Hout T (2011) A New Approach to Monitoring Dengue Activity. *PLoS Negl Trop Dis* 5(5): e1215. doi:10.1371/journal.pntd.0001215

Editor: Serap Aksoy, Yale School of Public Health, United States of America

Published: May 31, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The authors received no funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: david.fisman@utoronto.ca

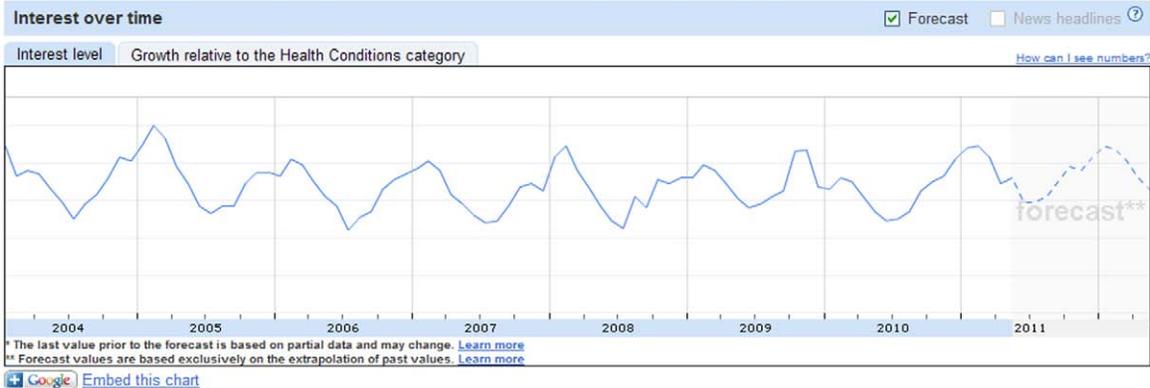
Compare by <input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	Search terms <small>Tip: Use a comma as shorthand to add comparison items. (tennis, squash)</small> <input type="text" value="pneumonia"/> + Add search term	Filter Web Search Worldwide 2004 - present Health Conditions	<input type="button" value="Search"/>
--	--	---	---------------------------------------

Web Search Interest: pneumonia

Worldwide, 2004 - present

[All Categories](#) > [Health](#) > Health Conditions

Subcategories: [Respiratory Conditions \(75-100%\)](#), [Infectious Diseases \(0-10%\)](#), [Cold & Flu \(0-10%\)](#), [more...](#)



Search terms

Top searches		Rising searches	
1. pneumonia symptoms	██	1. is pneumonia contagious	+250%
2. walking pneumonia	████████████████████████████████████	2. signs of pneumonia	+250%
3. pneumonia contagious	██████████████████████████████████	3. pneumonia contagious	+180%
4. bronchitis	██████████████████████████████████	4. síntomas de pneumonia	+180%
5. bronchitis pneumonia	██████████████████████████████████	5. pneumonia sintomas	+160%
6. pneumonia treatment	██████████████████████████████████	6. pneumonia in children	+90%
7. symptoms of pneumonia	██████████████████████████████████	7. pneumonia symptoms	+60%
8. is pneumonia contagious	██████████████████████████████████	8. walking pneumonia symptoms	+60%
9. pneumonia sintomas	██████████████████████████████████	9. what is pneumonia	+50%
10. pneumonia children	██████████████████████████████████	10. pneumonia children	+40%

[+ Google](#) [Embed this table](#)

Insights for Search aims to provide insights into broad search patterns. Several approximations are used to compute these results. The Insights for Search map is intended for general analysis of volume patterns. Borders are an approximation and may not be accurate.
 ©2011 Google - [Terms of Use](#) - [Privacy Policy](#) - [Insights for Search Help Center](#)

Figure 1. Screenshot of search performed on the term “pneumonia” using the Google Insights for Search tool (<http://www.google.com/insights/search/#>). The expected wintertime seasonality of pneumonia incidence is mirrored in seasonal surges in search volumes.
 doi:10.1371/journal.pntd.0001215.g001

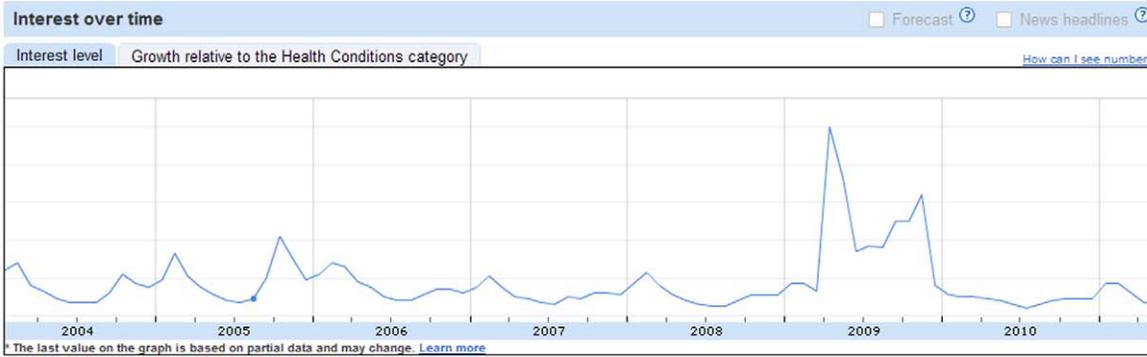
Compare by	Search terms	Filter
<input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	Tip: Use a comma as shorthand to add comparison items. (tennis, squash) <input type="text" value="influenza"/> + Add search term	Web Search Worldwide 2004 - present Health Conditions
		<input type="button" value="Search"/>

Web Search Interest: influenza

Worldwide, 2004 - present

[All Categories](#) > [Health](#) > Health Conditions

Subcategories: [Infectious Diseases \(50-75%\)](#), [Cold & Flu \(10-25%\)](#), [AIDS & HIV \(0-10%\)](#), [more...](#)



[Google](#) [Embed this chart](#)

Regional interest [Region](#) [City](#)

- [Mexico](#)
- [Italy](#)
- [Denmark](#)
- [Hungary](#)
- [Philippines](#)
- [Chile](#)
- [New Zealand](#)
- [South Korea](#)
- [Australia](#)
- [Hong Kong](#)



[View change over time](#)

Search terms

Top searches

1.	influenza a	
2.	h1n1	
3.	h1n1 influenza	
4.	influenza virus	
5.	flu	
6.	h1n1 influenza a	
7.	influenza sintomi	
8.	influenza vaccine	
9.	avian influenza	
10.	la influenza	

[Google](#) [Embed this table](#)

Rising searches

1.	h1n1	Breakout
2.	h1n1 influenza	Breakout
3.	h1n1 influenza a	Breakout
4.	influenza 2009	Breakout
5.	influenza mexico	Breakout
6.	influenza suina	Breakout
7.	influenza suina sintomi	Breakout
8.	swine influenza	+1,100%
9.	influenza sintomi	+450%
10.	la influenza	+400%

[Google](#) [Embed this table](#)

Insights for Search aims to provide insights into broad search patterns. Several approximations are used to compute these results. The Insights for Search map is intended for general analysis of volume patterns. Borders are an approximation and may not be accurate.

©2011 Google - [Terms of Use](#) - [Privacy Policy](#) - [Insights for Search Help Center](#)

Figure 2. Screenshot of search performed on the term “influenza” using the Google Insights for Search tool (<http://www.google.com/insights/search/#>). Although influenza searches would expect to display similar wintertime seasonality to pneumonia searches, depicted in Figure 1, the public concern and interest generated by the 2009 influenza pandemic generated a large spike in searches in that year, which obscures seasonal oscillation in other years.
doi:10.1371/journal.pntd.0001215.g002

Readers unfamiliar with these approaches may wish to try a simple experiment using the Google Insights for Search tool, which provides a graphical depiction of both search term volumes and online media reports of disease (<http://www.google.com/insights/search/#>). Searches on terms such as “norovirus” or “pneumonia” produce seasonally oscillating patterns of searches as one might anticipate in diseases with strong wintertime seasonality (Figure 1), and which is presumably generated by individuals who have, or know someone who has, this diagnosis seeking to learn more about it online. However, the pitfalls of this approach can be appreciated in a similar manner: a search on the term “influenza” produces a graph with a tremendous spike in 2009 (Figure 2); indeed a spike so large that it obscures influenza activity in all other years. This reflects the difficulties that searchstream-based surveillance methods may encounter when evaluating diseases that generate extreme public concern or media attention.

Chan et al., in this edition of *PLoS Neglected Tropical Diseases* [27], apply searchstream surveillance techniques to the monitoring of dengue. In this case, search queries appear to closely track (rather than lead) dengue activity as measured by traditional systems. The authors have limited their model to certain locations defined in part by the extent of Internet use in these areas (Bolivia, Brazil, India, Indonesia and Singapore). Their findings are exciting: when evaluated in a “testing set” of data not used to derive initial models, they found extremely strong correlation between dengue-related query volumes and case counts reported by traditional surveillance systems, but their approach has the advantage of both timeliness and transparency (including the availability of the system on the Google.org website).

As with any prediction-oriented surveillance tool, a major concern relates to model “over-fitting” such that the prediction model performs well in the dataset that was used in its creation but fails to work well in the “real world”. Reassuringly the authors divided their data into a derivation set and a testing set (or “holdout set” as they call it), with the former used for model construction. As can be seen in the table and figure they present, their derived models perform extremely well in both sets in all countries, in the derivation set as expected, but also in the testing set. Perhaps less straightforward is the authors’ decision to “smooth out” unusual spikes in search volumes in candidate queries; as demonstrated by the influenza example above, extreme surges in public interest in a disease can cause surges in query volumes, as can surges in interest related particular subject that is unrelated to the disease under surveillance but shares attributes that would be the subject of searches. By smoothing search volumes, the authors may have incorporated into their models terms that have the potential to “misbehave” in the future. For example, one imagines that if a novel (and frightening) new hemorrhagic fever unrelated to dengue emerges in one of these countries in coming years, one would imagine that the correlation between the search term “haemorrhagic fever” and dengue volumes would decline. As we don’t have access to the precise

query terms that were included in each country-specific model, it is difficult to know whether or not the terms included in the model would be vulnerable to such effects. The authors note that the expanding range of a clinically similar illness (Chikungunya) may confound the utility as well [28].

It would also be helpful to see to what extent there is overlap in components of models across countries, as this may help us understand whether these models can be applied to other jurisdictions or whether they are applicable only in the country for which they were constructed. As dengue is a disease whose range may change under the influence of climate change, it is important to know whether such an approach is applicable in the face of novel emergence of dengue in a new region or jurisdiction, or whether it is only applicable in countries like these in which dengue is currently endemic.

Perhaps the greatest challenge for the use of the approach described here is the same that applies across surveillance modalities: the same geographic locations that lack public health resources to control dengue, and to perform traditional surveillance, are likely to lag in access to the Internet as well. Nonetheless, the application of web-query based monitoring to a major and growing health threat in the developing world represents an important step forward. The ability to inexpensively and reliably maintain situational awareness of dengue activity will be welcomed by those charged with the public health response.

Does the development of web-based surveillance tools represent a revolution in how we conceptualize surveillance? We think not: current high-quality public health surveillance already utilizes multiple sources of information to gain a more complete picture of the incidence and distribution of disease. For example, influenza surveillance may include laboratory-based virological surveillance, sentinel syndromic surveillance (e.g., school-based absenteeism reports) and evaluation of mortality trends for pneumonia and influenza, which taken together may provide a more complete picture of disease risk and impacts. Searchterm-based surveillance and other modalities mentioned above thus provide an additional tool in the surveillance toolbox, which has advantages over traditional surveillance as well as limitations. It should be noted, however, that limitations such as those described above are not absent from traditional surveillance systems either: estimates of incidence can change markedly with changing case definitions, incidence of laboratory-confirmed disease can change markedly with augmentation or restriction of clinical testing or changes in diagnostic test methodologies, and syndromic surveillance systems can be subject to poor specificity and frequent false alarms. Thus supplementary information derived using methods such as the one developed by Chan and colleagues should be welcomed by public health professionals. The transparency of such systems may also help demonstrate the value of openness in disease reporting, which may have “spillover effects” on traditional surveillance systems.

References

1. Guzman MG, Kouri G (2002) Dengue: an update. *Lancet Infect Dis* 2: 33–42.
2. Messer WB, Gubler DJ, Harris E, Sivananthan K, de Silva AM (2003) Emergence and global spread of a dengue serotype 3, subtype III virus. *Emerg Infect Dis* 9: 800–809.
3. Gubler DJ, Clark GG (1995) Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerg Infect Dis* 1: 55–57.
4. Webster DP, Farrar J, Rowland-Jones S (2009) Progress towards a dengue vaccine. *Lancet Infect Dis* 9: 678–687.

5. McElroy KL, Santiago GA, Lennon NJ, Birren BW, Henn MR, et al. (2011) Endurance, refuge, and reemergence of dengue virus type 2, Puerto Rico, 1986–2007. *Emerg Infect Dis* 17: 64–71.
6. Hotez PJ, Remme JH, Buss P, Alleyne G, Morel C, et al. (2004) Combating tropical infectious diseases: report of the Disease Control Priorities in Developing Countries Project. *Clin Infect Dis* 38: 871–878.
7. Runge-Ranzinger S, Horstick O, Marx M, Kroeger A (2008) What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Trop Med Int Health* 13: 1022–1041.
8. Meynard JB, Chaudet H, Texier G, Ardillon V, Ravachol F, et al. (2008) Value of syndromic surveillance within the Armed Forces for early warning during a dengue fever outbreak in French Guiana in 2006. *BMC Med Inform Decis Mak* 8: 29.
9. van Panhuis WG, Gibbons RV, Endy TP, Rothman AL, Srikiatkachorn A, et al. (2010) Inferring the serotype associated with dengue virus infections on the basis of pre- and postinfection neutralizing antibody titers. *J Infect Dis* 202: 1002–1010.
10. Sharma K, Angel B, Singh H, Purohit A, Joshi V (2008) Entomological studies for surveillance and prevention of dengue in arid and semi-arid districts of Rajasthan, India. *J Vector Borne Dis* 45: 124–132.
11. Victor TJ (2009) Detection of dengue viral infections in *Aedes* mosquitoes: an essential tool for epidemiological surveillance. *Indian J Med Res* 129: 634–636.
12. Seng CM, Seta T, Nealon J, Socheat D (2009) Pupal sampling for *Aedes aegypti* (L.) surveillance and potential stratification of dengue high-risk areas in Cambodia. *Trop Med Int Health* 14: 1233–1240.
13. Brunkard JM, Robles Lopez JL, Ramirez J, Cifuentes E, Rothenberg SJ, et al. (2007) Dengue fever seroprevalence and risk factors, Texas-Mexico border, 2004. *Emerg Infect Dis* 13: 1477–1483.
14. Wilder-Smith A (2006) The severe acute respiratory syndrome: impact on travel and tourism. *Travel Med Infect Dis* 4: 53–60.
15. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, et al. (2008) BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* 24: 2940–2941.
16. Brownstein JS, Freifeld CC, Madoff LC (2009) Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med* 360: 2153–2157.
17. Brownstein JS, Freifeld CC, Reis BY, Mandl KD (2008) Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 5: e151. doi:10.1371/journal.pmed.0050151.
18. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, et al. (2010) Global capacity for emerging infectious disease detection. *Proc Natl Acad Sci U S A* 107: 21701–21706.
19. Fidler DP, Gostin LO (2006) The new International Health Regulations: an historic development for international law and public health. *J Law Med Ethics* 34: 85–94.
20. Freifeld CC, Mandl KD, Reis BY, Brownstein JS (2008) HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 15: 150–157.
21. Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, et al. (2011) Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154: 593–601.
22. Polgreen PM, Chen Y, Pennock DM, Nelson FD (2008) Using internet searches for influenza surveillance. *Clin Infect Dis* 47: 1443–1448.
23. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014.
24. Freifeld CC, Chunara R, Mekar SR, Chan EH, Kass-Hout T, et al. (2010) Participatory epidemiology: use of mobile phones for community-based health reporting. *PLoS Med* 7: e1000376. doi:10.1371/journal.pmed.1000376.
25. Epstein JM, Parker J, Cummings D, Hammond RA (2008) Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PLoS One* 3: e3955. doi:10.1371/journal.pone.0003955.
26. Wang P, Gonzalez MC, Hidalgo CA, Barabasi AL (2009) Understanding the spreading patterns of mobile phone viruses. *Science* 324: 1071–1076.
27. Chan EH, Sahai V, Conrad C, Brownstein JS (2011) Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 5(5): e1206. doi:10.1371/journal.pntd.0001206.
28. Pialoux G, Gauzere BA, Jaureguierry S, Strobel M (2007) Chikungunya, an epidemic arbovirolos. *Lancet Infect Dis* 7: 319–327.