

# Evaluation of online media reports for global infectious disease intelligence

John S. Brownstein, Ph.D., Clark C. Freifeld, Ben Y. Reis, Ph.D., Kenneth D. Mandl, M.D,  
*Children's Hospital Boston Informatics Program, Harvard-MIT HST, Harvard Medical School*

## OBJECTIVE

Internet-based resources such as discussion sites and online news sources have become invaluable sources for a new wave of surveillance systems. The WHO relies on these informal sources for about 65% of their outbreak investigations.(1) Despite widespread use of unstructured information there has been little, if any, data evaluation.

## BACKGROUND

While traditional means of surveillance by governments, multi-national agencies, and institutional networks assist in reporting and confirming infectious disease outbreaks, these formal sources of information are limited by their geographic coverage and timeliness of information flow. In contrast, rapid global reach of electronic communication has resulted in the advent of informal sources of information on outbreaks. Informal resources include discussion sites, online news media, individual and organization reports and even individual search records. The earliest descriptions of the severe acute respiratory syndrome (SARS) outbreak in Guangdong Province, south China came from informal reports. However, system development to date has been geared toward knowledge management and strategies for interpreting these data are underdeveloped. There is a need to move from simple knowledge reorganization to an analytic approach for disseminating timely yet specific signals.

## METHODS

For multi-stream surveillance to be effective, basic characteristics such as sensitivity, specificity and timeliness of different news source types need to be quantified. In our evaluation, we used officially confirmed outbreaks obtained from WHO Outbreak News, available in the public domain, as a "gold standard" indicator of an infectious disease outbreak.(2) We measured key detection characteristics of news reports for outbreaks over the 9-month period (October 1, 2006-June 30, 2007) in both English and Spanish. We apply standard evaluation metrics (volume, geography covered, diseases captured, timeliness, sensitivity and specificity). Our analysis is informed by evaluation of the performance of HealthMap, a freely accessible, automated system for real-time monitoring of online information about emerging diseases.(3)



Figure 1. Screenshot of [www.HealthMap.org](http://www.HealthMap.org)

## RESULTS

Over the evaluation period, HealthMap found 11,194 news reports of infectious disease outbreaks (a mean of 38.6 per day, 95% CI, 33.1 to 44.1) covering a 105 pathogens and 160 countries. Mean timeliness for news sources, defined as the time between detection by the surveillance source and report by the WHO, was 12 days. However, actual timeliness varied widely from 102 days earlier to 59 days after the WHO report. For instance, a diarrheal outbreak in Ethiopia was detected by the media almost three weeks before reporting by the WHO. In contrast, a plague outbreak in the Democratic Republic of the Congo and a Chikungunya outbreak in India were only reported in the media once the official WHO report was released. Sensitivity, defined as the proportion of WHO alerts detected by news data, was moderate, with 58% of the alerts reported in the news. We identified 962 unique alerts (country-disease pairs) from news sources, compared to only 24 released by WHO and 623 found on the ProMED mail system. News sources are shown to be especially valuable for monitoring spatial and temporal patterns of larger scale epidemics, especially seasonal or endemic diseases.

## CONCLUSIONS

Overall, we find that internet-based alert mapping offers a promising tool for surveillance and that there is value in the integration of distributed electronic resources for public health communication and intervention. Future work should be directed at modeling and data integration, including improving risk assessment

## REFERENCES

- 1.D. L. Heymann, G. R. Rodier, *Lancet Infect Dis* 1, 345-53 (Dec, 2001).
2. World Health Organization Epidemic Alert and Response, (Accessed April 4, 2007.).
- 3.C. Holden, *Science* 314, 1363d (December 1st., 2006).