Published in final edited form as:

Vaccine. 2015 May 15; 33(21): 2395–2398. doi:10.1016/j.vaccine.2015.03.100.

Internet Activity as a Proxy for Vaccination Compliance

Yuval Barak-Corren and

Predictive Medicine Group, Children's Hospital Boston; Technion Israeli Institute of Technology, Boston Children's Hospital, 1 Autumn St. Room 540.1, Boston, MA 02115, Phone: +1 (617) 299 6484

Ben Y. Reis¹

Predictive Medicine Group, Children's Hospital Boston and Harvard Medical School, Boston Children's Hospital, 1 Autumn St. Room 540.1, Boston, MA 02115, Phone: +1 (857) 218 4561

Yuval Barak-Corren: yuval.barakcorren@childrens.harvard.edu; Ben Y. Reis: Ben.Reis@childrens.harvard.edu

Abstract

Tracking the progress of vaccination campaigns is a challenging and important public health need. Examining a recent Polio outbreak in the Middle East, we show that novel methods utilizing online search trends have great potential to provide a real-time, reliable proxy for vaccination rates over space and time.

Keywords

Internet/utilization; Polio/immunology; Polio Infections/prevention & control; Vaccination/statistics & numerical data; Sentinel Surveillance

Introduction

Population-scale vaccination programs are essential for controlling disease outbreaks, yet health officials often lack the real-time information on compliance necessary for guiding ongoing efforts and assessing the success of vaccination campaigns.[1,2] Without this knowledge, precious time is often lost as high-priority populations fail to be sufficiently vaccinated while the disease spreads.

Current methods for tracking vaccinations typically rely on manual public health reporting or ad-hoc surveys,[3] requiring the aggregation of different data sources and involving significant delays. For example, tracking of flu vaccination rates in the United States relies

© 2015 Published by Elsevier Ltd.

Correspondence to: Yuval Barak-Corren, yuval.barakcorren@childrens.harvard.edu.

¹Funded in part by National Library of Medicine, 5R01 LM009879-04.

Conflict of Interest

Mr. Barak-Corren has nothing to disclose; Dr. Reis reports grants from NIH, during the conduct of the study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

on the integrated evaluation of six different surveys administered nationally,[4] and can lag behind by as much as two to four years.[5] New approaches are needed to complement existing reporting methods and enable a more effective utilization of limited healthcare resources.[6]

One potential new approach involves the use of online data from Internet Query Share (IQS) statistics.[7] IQS data measure the popularity of different search terms over different geographical regions and time periods using web resources such as Google Trends, Wikipedia, Twitter, and others. While the use of IQS data has gained popularity in healthcare studies in recent years, its utility for tracking vaccination campaigns has yet to be measured. The geographical resolution of IQS data can help pinpoint specific areas with low compliance in vaccination campaigns, while the timeliness of IQS data allows rapid response to changes in compliance. IQS data are freely available and can thus minimize the costs required for tracking population-level vaccination compliance, especially in areas where public Internet access is available but health-reporting systems are not electronically integrated.

A recent Polio outbreak in Israel serves as an ideal natural experiment for evaluating the use of such new approaches for vaccination monitoring.[8] In June 2013, poliovirus was detected in the sewage system in southern Israel, and tracked to Bedouin residents living in the region. In response, Oral Polio Vaccination (OPV) was immediately administered to nearby communities. Two months later, in August 2013, a nation-wide vaccination campaign named “Two Drops” was launched. The aim of the campaign was to vaccinate all children who were born after January 1, 2004 (zero to nine years old) with the live attenuated Polio vaccine (OPV). During the “Two Drops” campaign, a total of 976,756 children were vaccinated out of a total of 1,383,296 children targeted (64.8% vaccination coverage). The vaccinations were administered at 1,015 Family Health Centers across the nation, operated by either the Israeli Ministry of Health, one of the four national Health Maintenance Organizations (HMOs), or one of the local municipalities.[9]

This Israeli outbreak allows us to access both highly detailed official reporting as well as IQS data, enabling the validation of IQS-based methods for tracking vaccination campaigns. It is a test case that can be of great relevance for other settings where real-time vaccination compliance information is not available from traditional sources in real time, but IQS data are available. In such cases IQS-based approach could complement and augment traditional reporting methods currently in use.

Methods

Vaccination data based on on-site reporting was obtained from the Israeli Ministry of Health (MOH). The data included weekly statistics on the number of vaccinations that were administered, stratified by district (Southern, Central, Haifa, Tel Aviv, and Jerusalem) for the period of August 5th to November 28th, 2013. We supplemented this data with additional reports released by the MOH, which provided two-day temporal resolution at the national level. We imputed these data by averaging to fill in the missing days and create a national

daily time series (i.e. if 10% were vaccinated on Sunday and 30% were vaccinated on Tuesday, we assumed that 20% were vaccinated on Monday).

Google Trends IQS statistics were obtained both at the national and district levels, for the period covering January through December 2013 to evaluate Internet searching behavior for both outbreak and non-outbreak time periods. These data are provided as normalized values between 0–100, where 100 represents the peak in searches for the specified search term over the selected time period and geographic region.

All IQS studies rely on the selection of appropriate search terms. In keeping with the approach described in previous studies [7,10–12] we began with three basic terms – ‘polio’, ‘vaccination’, and ‘Family Health Center’ (where the vast majority of vaccinations were administered), and tested related terms suggested by Google Correlate. We found that the additional search terms suggested by Google Correlate did not add significantly to the performance of the three initial search terms and thus decided to exclude them. We then added alternate forms (e.g. plural form, common misspellings, etc.) to each of the three terms, as well as translations into other common languages used in the region, including Arabic, English, and Russian. We tested each term separately (i.e. any form of a given term) and all terms combined – “combined search” (i.e. any form of any term).

Having both the IQS data from Google and the official reports from the Israeli MOH, we examined different correlations between the two datasets. We first analyzed weekly district-level data for correlations between MOH and IQS statistics within specific districts. We then queried a more specific time period encompassing August–October (including) and analyzed the national daily data for correlations between MOH statistics and next-day, same-day, previous-day, and two-day-prior IQS data.

Results

At the district level, IQS statistics and MOH vaccination rates were highly correlated overall ($R=0.79$, ranges from [0.61 at the Southern district to 0.96 at the Haifa district]). With a simple linear transformation we were able to predict the vaccination rate with an average absolute error of only 2.7–3.9%. As can be seen in Figure 1, web searches for polio-associated terms closely followed vaccination trends within each district, in many cases slightly preceding them. Examining the cumulative performance of the simple linear transformation over the course of the vaccination campaign, we see that throughout the campaign and combining all districts, IQS predictions were 11.41% ($n=76,698$) higher than the MOH statistics.

Day-by-day correlations were calculated between each of the aforementioned IQS terms and the data collected by the MOH using Spearman and Pearson coefficients. We tested whether official MOH vaccination rates were associated with next-day, same day, previous-day, and two-days-prior Internet search activity. As can be seen in Table 1, the strongest overall correlations were between MOH data and previous-day IQS data, while the weakest correlations were between MOH data and next-day IQS data.

Discussion

In the setting described in this study, IQS data was highly correlated with official manually-reported vaccination compliance rates. Given their timeliness and low costs, these data can potentially serve as a valuable complementary data source for real-time vaccination compliance rates with high temporal and geographic resolution. The results also suggest that IQS data has the ability to serve as a valid predictor for next-day vaccination rates.

We believe that the suggested system could be most useful in nations where Internet access is available to the general public including minority groups, but official tracking either lags in time, has partial coverage, or is not integrated nationally due to a divided healthcare system. In such scenarios, IQS data can be calibrated to historically available official data. Once IQS data is calibrated, it could provide a valuable complementary perspective on the progress of outbreaks and vaccination campaigns. In settings where Internet access is available to the population but no official vaccination tracking is performed, IQS data could still be used as a proxy for vaccination rates, by interpreting the data relative to historical baseline IQS rates. This situational awareness can still be very valuable when no other data sources are available.

Limitations of the current study are that its utility depends on the degree of Internet penetration in the population which is currently limited in some nations susceptible to outbreaks, though Internet reach is rapidly growing in many developing nations.[13] Furthermore, complete Internet penetration is not necessary, so long as the Internet user population is generally representative of the broader population. Second, it was carried out retrospectively and is only correlational and not causal. Also, due to the small population size of Israel, we were limited to use Google Trends' with less specific terms (e.g. "Family Health Center") and broader time scales (e.g. weekly data for by-district analysis) as well as omit the 'Northern' district due to insufficient search volume. Future work will further validate this approach prospectively and in other settings as well.

The results of this study suggest that an increase in web searches for outbreak-relevant terms, as a result of vaccination campaigns or media coverage, can potentially reflect current vaccination coverage for different districts, as well as indicate near-term future demand for vaccinations. Such capability can prove especially useful for countries that have public Internet access, but do not have a centralized healthcare information reporting infrastructure. IQS offers a widely available, free and timely data source that can be used today to complement and improve traditional surveillance methods currently in use.

Acknowledgments

Funding

This work was supported by the National Library of Medicine grant number 5R01LM009879-04.

References

1. Murray CJ, Shengelia B, Gupta N, Moussavi S, Tandon A, Thieren M. Validity of reported vaccination coverage in 45 countries. *lancet*. 2003; 362:1022–1027. [PubMed: 14522532]

2. Hanquet G, Van Damme P, Brasseur D, et al. Lessons learnt from pandemic A(H1N1) 2009 influenza vaccination. Highlights of a European workshop in Brussels (22 March 2010). *vaccine*. 2011; 29:370–377. [PubMed: 21075164]
3. Walter D, Böhmer MM, an der Heiden M, Reiter S, Krause G, Wichmann O. Monitoring pandemic influenza A(H1N1) vaccination coverage in Germany 2009/10 – Results from thirteen consecutive cross-sectional surveys. *vaccine*. 2011; 29:4008–4012. [PubMed: 21463683]
4. CDC. [Accessed 15 April 2014] Influenza Vaccination Coverage | FluVaxView | Seasonal Influenza (Flu). Available at: <http://www.cdc.gov/flu/fluview/>
5. Schuck-Paim C, Taylor R, Lindley D, Klugman KP, Simonsen L. Use of near-real-time medical claims data to generate timely vaccine coverage estimates in the US: The dynamics of PCV13 vaccine uptake. *vaccine*. 2013; 31:5983–5988. [PubMed: 24144470]
6. Harris KM, Schonlau M, Lurie N. Surveying a nationally representative internet-based panel to obtain timely estimates of influenza vaccination rates. *vaccine*. 2009; 27:815– 818. [PubMed: 19059447]
7. Desai R, Lopman BA, Shimshoni Y, Harris JP, Patel MM, Parashar UD. Use of Internet search data to monitor impact of rotavirus vaccination in the United States. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2012; 54:e115–118.
8. Kopel E, Kaliner E, Grotto I. Lessons from a Public Health Emergency — Importation of Wild Poliovirus to Israel. *N Engl J Med*. 2014; 371:981–983. [PubMed: 25207764]
9. Gruto, I. Polio Vaccination Campaign Notice. Israeli Ministry of Health; 2013. Available at: http://www.health.gov.il/hozer/bz18_2013.pdf
10. 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 Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3104029/>.
11. Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from National and Californian Internet searches and social media mentions, 2012. *Bmj Open*. 2014; 4:e004417.
12. Fazeli Dehkordy S, Carlos RC, Hall KS, Dalton VK. Novel Data Sources for Women’s Health Research: Mapping Breast Screening Online Information Seeking Through Google Trends. *Acad Radiol*. 2014
13. Devlin, K. [Accessed 3 September 2014] Emerging Nations Embrace Internet, Mobile Technology. Pew Res Centers Glob Attitudes Proj. 2014. Available at: <http://www.pewglobal.org/2014/02/13/emerging-nations-embrace-internet-mobile-technology/>

Highlights

- In Summer 2013 Poliovirus was detected in Israel's sewage system. In response the Israeli Ministry of Health (MOH) carried out a nation-wide immunization campaign.
- We analyze Internet search statistics by district for Polio-related terms for the time period of this campaign.
- We compare Internet search statistics with official reporting obtained from the MOH.
- Internet searches were highly correlated with same district MOH reported vaccination rates ($R=0.786$).
- These findings suggest a novel method for monitoring vaccination campaigns.

Figure 1. District level data. Comparing IQS results for “polio”, “vaccination(s)”, “Family Health Center(s)” (FHC) and a combined query of any of these terms, with data obtained from the MOH. Northern district not shown due to insufficient IQS data.

Table 1

Spearman coefficients for the temporal correlation between IQS data for different search terms and national vaccination compliance. Correlations are calculated between MOH data and IQS data for two days prior, one day prior, same day, and as a control next day IQS data, There were no differences between Spearman & Pearson coefficients.

	“Polio”	“Vaccine”	“FHC”	Combined Search
MOH-IQS two days before	0.73	0.74	0.59	0.70
MOH-IQS previous day	0.77	0.75	0.64	0.76
MOH-IQS same day	0.75	0.73	0.68	0.75
MOH-IQS next day	0.68	0.67	0.57	0.66