Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines

Heidi J Larson, David M D Smith, Pauline Paterson, Melissa Cumming, Elisabeth Eckersberger, Clark C Freifeld, Isaac Ghinai, Caitlin Jarrett, Louisa Paushter, John S Brownstein, Lawrence C Madoff

Summary

Background The intensity, spread, and effects of public opinion about vaccines are growing as new modes of communication speed up information sharing, contributing to vaccine hesitancy, refusals, and disease outbreaks. We aimed to develop a new application of existing surveillance systems to detect and characterise early signs of vaccine issues. We also aimed to develop a typology of concerns and a way to assess the priority of each concern.

Methods Following preliminary research by The Vaccine Confidence Project, media reports (eg, online articles, blogs, government reports) were obtained using the HealthMap automated data collection system, adapted to monitor online reports about vaccines, vaccination programmes, and vaccine-preventable diseases. Any reports that did not meet the inclusion criteria—any reference to a human vaccine or vaccination campaign or programme online reports about vaccines, vaccination programmes, and vaccine-preventable diseases. Any reports that did not meet the inclusion criteria—any reference to a human vaccine or vaccination campaign or programme that was accessible online—were removed from analysis. Reports were manually analysed for content and categorised by concerns, vaccine, disease, location, and source of report, and overall positive or negative sentiment towards vaccines. They were then given a priority level depending on the seriousness of the reported event and time of event occurrence. We used descriptive statistics to analyse the data collected during a period of 1 year, after refinements to the search terms and processes had been made.

Findings We analysed data from 10 380 reports (from 144 countries) obtained between May 1, 2011, and April 30, 2012. 7171 (69%) contained positive or neutral content and 3209 (31%) contained negative content. Of the negative reports, 1977 (24%) were associated with impacts on vaccine programmes and disease outbreaks; 1726 (21%) with beliefs, awareness, and perceptions; 1371 (16%) with vaccine safety; and 1336 (16%) with vaccine delivery programmes. We were able to disaggregate the data by country and vaccine type, and monitor evolution of events over time and location in specific regions where vaccine concerns were high.

Interpretation Real-time monitoring and analysis of vaccine concerns over time and location could help immunisation programmes to tailor more effective and timely strategies to address specific public concerns.

Funding Bill & Melinda Gates Foundation.

Introduction

Although immunisation has successfully reduced the global burden of illness and death, a range of concerns have converged to affect public confidence in vaccines. When confidence in vaccination breaks down, hesitancy can lead to delays and refusal, disrupting research and delivery programmes, and sometimes leading to disease outbreaks.1,2 The most serious example is the 2003–04 northern Nigeria boycott of polio vaccination, which set the global polio eradication initiative back substantially, cost millions of US dollars, and led to a resurgence of the disease.3,4 The boycott, driven mainly by politics and unfounded fears of vaccine-induced sterilisation, contributed to reinfecion in 20 previously polio-free countries, reaching as far as Indonesia. The fundamental breakdown in public trust still affects polio eradication efforts in Nigeria.5,6 Another example includes fear and refusal of the measles, mumps, and rubella (MMR) vaccine, initially ignited by Andrew Wakefield and colleagues’ now-retracted6,7 1998 study8 suggesting an association between the vaccine and autism. The study’s findings were amplified by the media, Wakefield’s own public appearances advocating his research, and networks of parents who felt that Wakefield finally had an answer to the cause of their child’s autism. The effect of the media and personal attention to the since-debunked research, officially retracted 4 years after it was published and deemed both unethical and fraudulent,9 resulted in a substantial decline in MMR vaccine coverage9,10 that has still not returned to the high of 92% before 1998.11 Meanwhile, MMR vaccine anxieties continue to circulate worldwide. In 2009–10, high non-acceptance rates of the pandemic influenza A H1N1 vaccine, including among health-care professionals, were another example of the potential effects of public distrust in vaccines.12

In such cases, the time between the prompting events and their effect on public health outcomes is important—eg, months or years can elapse, with extended periods of
vaccine hesitancy or uncertainty. Various factors amplify the spread of information, and misinformation, affecting perceptions and behaviours and creating what Kasper and colleagues term the “social amplification of risk”, by which they mean amplification of the spread of information via social, cultural, and institutional processes, and amplification of society’s response.

We postulated that media monitoring might provide important information about perceived problems with specific vaccines or immunisation programmes that might take longer to register through official channels. We therefore aimed to track the emergence and spread, geographically and temporally, of media and social media reports on vaccines by developing a new application for rumour surveillance systems—typically designed to detect signs of disease outbreaks—to detect and characterise early signs of public concern about vaccines. In doing so, we aimed to characterise and create a typology of the content and to develop an approach to assess the priority of each concern for further investigation and intervention.

In January, 2010, The Project to Monitor Public Confidence in Immunisation, now called The Vaccine Confidence Project, of which we are members, was launched to address three objectives. The first being to establish a global information surveillance system to detect emerging public concerns by monitoring media and social media and building a global key informant network. Second, to systematically code all reports to identify positive or neutral versus negative content and key areas of concern, to develop a typology of concerns. And third, to develop a diagnostic method to prioritise which reports need further investigation and intervention on the basis of patterns of clustering or persistence of reports, and when viewed against contextual and historical factors that have contributed to their amplification. This report focuses mainly on our first two objectives. Concurrent research and analysis of the third objective will be reported elsewhere.

**Methods**

**Data collection**

Early on in the project, an international advisory group consisting of experts in vaccine-preventable diseases, vaccine safety, risk and decision science, immunisation programme management, and public health—identified and discussed key examples of breakdowns in public confidence in vaccines or vaccine programmes that resulted in serious declines in vaccine acceptance. Detailed retrospective analyses then identified crucial factors that led to these breakdowns in acceptance, which were then used as criteria for report categorisation.

Data collection began in April, 2010. Refinements to search terms, collection, and coding were made periodically between April, 2010, and April, 2011. Most reports were collected with the HealthMap automated data collection system, which was adapted to monitor online reports about vaccines, vaccination programmes, and vaccine-preventable diseases. An example of the search criteria used is “intitle:vaccine OR intitle:rotavirus OR intitle:measles”. Reports included online news articles, blogs, website pages, public service announcements, government announcements, book reviews, and broadcast media. The three primary acquisition channels were Google News, Google Blog Search, and Moreover Public Health, a global news aggregator service. The Moreover content was culled from a broad range of RSS (rich site summary) news feeds by use of health-related keywords determined by Moreover. Foreign language reports were translated using Google Translate. The non-English languages that were translated were Chinese, Czech, Danish, French, German, Greek, Hindu, Indonesian, Italian, Japanese, Persian, Polish, Portuguese, Russian, Spanish, Swedish, Thai, and Vietnamese.

Reports were passed through an exclusion filter to remove any that were not vaccine-related or were related only to animal vaccines. Any report that included a reference to a human vaccine or vaccination campaign or programme that was accessible online was included in the analysis, which included articles about vaccine-preventable diseases whenever a vaccine was mentioned. Reports not meeting these inclusion criteria were excluded from analysis but logged and archived. The articles were presented on a web-based user interface system for rapid review, while the system automatically forwarded the reports to analysts from the London School of Hygiene & Tropical Medicine and the Program for Monitoring Emerging Diseases (ProMED).

**Data categorisation**

Reports were automatically classified according to the HealthMap algorithms every hour. Relevant reports were classified by location and disease and entered into the project database, with title, date, report source, URL, and plain text in full automatically added, taking less than 1 min per report. The entry date and the team member entering the report were also automatically recorded. Analysts then reviewed and amended the autopopulated fields, then allocated data categories, report source, vaccine-preventable disease, vaccine, and vaccine manufacturer, and disaggregated the automatically identified locations into three fields: country of origin of report, countries referred to, and country of origin of any additional sources referenced. Additionally, a one-line summary of the report content was created by the analyst. After populating these fields, an overall impression of the report tone (either positive or neutral, or negative), and report priority level (high, medium, or low) were decided on the basis of agreed criteria.

A report was coded as negative if it contained any indication of concern about a vaccine or vaccination programme, such as information about an adverse event that occurred after immunisation, a vaccine suspension, or any other factor that has a negative effect on a vaccine
programme. A report was coded as positive or neutral if it contained no indication of public concern about a vaccine or vaccination programme. A report that discussed an adverse event that occurred after immunisation could be coded as positive if, for instance, the article was about the investigation of a reported adverse event that had been confirmed to be unrelated to a vaccine. Other positive reports included topics such as a country announcing the introduction of a new vaccine or plans for a vaccination campaign.

We used the definition of adverse events as any untoward medical occurrence that follows immunisation that does not necessarily have a causal relation with the use of the vaccine, as described by the Council for International Organizations of Medical Sciences and WHO working group on vaccine pharmacovigilance.\(^{23}\)

On average it took 5 min per report for this stage of the data curation to be completed, although time taken varied substantially depending on the length, complexity, and clarity of the article. Within 24 h of receipt, each report was read and assessed by one of the team of analysts trained with a standardised database user guide developed by the research team. Periodic quality assessment was done to ensure consistency between reviewers (figure 1).

**Data prioritisation**

Report priority was classified as high if the report contained any reference to a serious adverse event that occurred after immunisation, such as death or hospitalisation, or any reference to a serious adverse event that occurred after immunisation, such as death or hospitalisation, or

potential to distinguish recent events from articles referencing historic events that were neither new nor recurring, but that needed monitoring (medium priority). Every report classified as high priority was automatically deleted, but those with different URLs were recorded as separate reports, recognising the fact that replicated reports show the spread of information, and are therefore important to the analyses.

**Statistical analysis**

Summary statistics given here comprise simple counts and percentages. All figures were generated in Matlab version 8.0.0.783 (R2012b).

**Role of the funding source**

The sponsor of the study had no role in research design, data collection, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Results**

We analysed only the data that were entered into the database between May 1, 2011, and April 30, 2012 (after all refinements to the search terms and processes were made), during which 10 380 reports originating from 144 countries were collected and coded. Of these reports, 7171 (69%) contained positive or neutral content about
vaccines and vaccination programmes and 3209 (31%) contained negative content (figure 2).

More than one data category could be allocated to one report. Of the 10 380 reports identified, 22 499 data categories were recorded (14 145 for the 7171 positive or neutral reports and 8354 for the 3209 negative reports). During the analysis of the data, related data categories were clustered into the following groups: beliefs, awareness, and perceptions (eg, religious beliefs, risk perceptions); contextual factors (eg, conflict or war); impacts (eg, vaccine suspension, vaccine refusals); vaccine delivery programme (eg, mass versus routine campaign, vaccine supply, costs); vaccine safety (eg, reports or concerns about adverse events occurring after immunisation, additives, or preservatives); vaccine development and introduction (eg, vaccine research or new products launched or approved); and recommendations about vaccines, disaggregated according to source (eg, national immunisation programme, religious leader, non-governmental health organisation; appendix).

Of the positive and neutral reports, 4632 (33%) data categories were associated with vaccine development and introduction; 4332 (31%) with vaccine delivery programmes; 2141 (15%) with vaccine recommendations; 1475 (10%) with contextual factors; 970 (7%) with beliefs, awareness, and perceptions of vaccines; 428 (3%) with vaccine safety; and 167 (1%) with impacts on vaccine programmes and disease outbreaks (figure 3). Of the negative reports, 1977 data categories (24%) were associated with impacts on vaccine programmes and disease outbreaks; 1726 (21%) with beliefs, awareness, and perceptions; 1371 (16%) with vaccine safety; 1336 (16%) with vaccine delivery programmes; 1119 (13%) with vaccine recommendations; 582 (7%) with contextual factors; and 243 (3%) with vaccine development and introduction (figure 3).

Our grouping of data categories included the group contextual factors, because some contextual factors can amplify negative sentiments (eg, marginalised populations, negative high-profile individual) and others attenuate them (eg, high vaccine acceptance rates, positive high-profile individual), which is important for the prioritisation of reported concerns. A more detailed breakdown of the data categories allocated to the 10 380 reports is shown in the appendix.

We categorised the reported data by vaccine type—worldwide and in five selected countries (China, Finland, France, Nigeria, and Pakistan; figure 4). We chose these countries because they had known or widely reported issues with specific vaccines before or during the study period, and we wanted to show that our data collection was consistent with established concerns. We note that the dominant vaccines discussed (either negatively or positively) in the media in the five countries are consistent with genuine issues that were reported concurrently by WHO, providing some validation of the reliability of the

Figure 2: Proportion of vaccine-related reports categorised as positive or neutral, by country
Based on analysis of all 10 380 reports. Of the 9655 reports (93%) that mentioned a country or countries, 11 535 countries were mentioned. Countries about which there were fewer than ten vaccine-related reports are shaded grey. The world proportion (69%) is shown by the arrow on scale bar. Country border data are from the Global Administrative Areas database.24
Vaccine development and introduction
Vaccine delivery programme
Recommendations
Contextual factors
Beliefs, awareness, and perceptions
Vaccine safety
Impacts

Figure 3: Number of data categories allocated to all reports, per data category group
Data is from all 10 380 reports, obtained between May, 2011, and April, 2012, split according to whether the report was classified as positive or neutral, or negative. Reports could be classified by more than one data category.

Figure 4: Proportionate number of times specific vaccine types were covered positively and negatively in the media, worldwide and in five selected countries
Vaccine types were reported positively (A) 9157 times worldwide, 45 times in France, 22 times in Finland, 171 times in China, 236 times in Nigeria, and 261 times in Pakistan. Vaccine types were reported negatively (B) 4900 times worldwide, 154 times in France, 29 times in Finland, 94 times in China, 193 times in Nigeria, and 311 times in Pakistan. Country border data are from the Global Administrative Areas database. HPV=human papillomavirus. JE=Japanese encephalitis.
Articles

For instance, polio vaccination dominates in Nigeria and Pakistan, two of the three remaining endemic countries, and the polio vaccination coverage in China is probably a result of the 2011 spread of the disease from Pakistan to China—the first cases of the disease in China since 1999. The media coverage of H1N1 influenza in Finland was due to the heightened concerns around the 2010–11 suspected links between narcolepsy and H1N1 influenza vaccination, and the attention to measles vaccination in France is consistent with the 2011 outbreak.

One of the important dimensions of the surveillance system is the ability to monitor the evolution of concerns over time and location. We mapped reports about human papillomavirus (HPV) vaccination from India, from May, 2011, to November, 2011, to monitor the effect of the April, 2010, government suspension of the HPV vaccine demonstration projects in Gujarat and Andhra Pradesh, caused by pressure from activists (figure 5). Although 11 (28%) of the 39 reports about HPV in India obtained during the study period (May 1, 2011, to April 30, 2012) were positive or neutral about HPV vaccines, most, 28 (72%), were negative—with some already posted on European and US anti-HPV vaccine websites—and were still focused on the issues surrounding the suspension of the HPV vaccine demonstration projects towards the end of our data collection period (April 30, 2012), 2 years after the government suspension of the project.

Discussion

33% of positive reports obtained worldwide were categorised as being about vaccine development and introduction, whereas only 3% of negative reports made reference to that topic; this percentage point difference was the largest noted between positive and negative reports out of all topic categories. The category of beliefs, awareness, and perceptions shows a similar large difference, albeit in the opposite direction, with 3% of positive reports making reference to this topic compared with 21% of negative reports. These reports include those discussing religious and philosophical beliefs about

Figure 5: Cumulative number of reports from India about human papillomavirus vaccination over time, by location and priority

Data consists of 48 location mentions from 31 reports obtained between May, 2011, and November, 2011. Country border data are from the Global Administrative Areas database.
vaccines, and perceptions about the motive behind vaccination (eg, government, industry). More negative reports discussed beliefs and perceptions than mentioned vaccine safety (16% of negative reports); negative reports discussing vaccine safety focused on adverse events and concerns about additives such as thiomersal.

Clearly, the proportions of positive and negative reports discussing each topic would change if examined at a more local level or by specific vaccine. Disaggregation of reports by positive or negative sentiment, country, and vaccine type could identify countries where specific vaccines might need more tailored engagement strategies than are currently in place.

The subnational analysis of reports about HPV vaccination from India (figure 5), is an example of an analysis that could help to inform delivery and engagement strategies. The HPV vaccine is available in India only through the private sector; it has not yet been included in India’s national immunisation schedule. When the HPV vaccine is considered for inclusion in the national immunisation schedule, this type of analysis, which identifies the clustering of reports and their sentiments in specific localities, might be useful to inform the design of engagement strategies, which will be particularly needed in Gujarat and Andhra Pradesh. When faced with several concerns and limited human and financial resources, this type of assessment could help to prioritise attention and allocation of resources.

Despite the strengths of our surveillance system, including the creation of a broad-spectrum and reflexive typology of concerns, it also has limitations. First, we used only English search terms to collect the initial data. Although some reports from non-English speaking countries were translated and included in the dataset, most were published in English, and included English language media from countries where the primary language is not English. However, the primary search terms are now being translated into five additional languages: Arabic, French, Mandarin, Russian, and Spanish.

Although human curation was important in the development and refinement of the system, it also restricted the volume of reports that could be processed. Efforts are focused on automating as much of the data collection and classification as possible, using the human-curated dataset as a training set. These changes will result in an increase in the geographical reach, volume, and speed at which reports can be classified and analysed. The changes will also help to enhance the typology of concerns, understand specific contextual factors, and develop the risk assessment method. Future iterations of the system will also expand the numbers of sources that are included. This expansion will be achieved by increasing the languages of our search terms and using other modes of social media (eg, Twitter).

The nature of public concerns about vaccines is complex and highly diverse. No single report can be relied upon to signal a genuine public concern. Instead, the strength and potential effects of a vaccine rumour or public concern are determined by clusters and persistent patterns of reports expressing similar concerns.

The surveillance system we present in this report allows systematic monitoring and assessment of media reports for vaccine sentiment, with the aim of detecting concerns as they emerge and evolve in real time. We created a typology of the scope of public concerns about vaccines and a way to monitor the ecology of those concerns as they develop temporally and spatially. Although this system has not been running long enough to show its long-term predictive value, it allows real-time characterisation of vaccine sentiments by topic, negative or positive content, location, time, and risk level. The third objective, to assess which reported concerns merit priority attention, needs longer-term study to validate the prioritisation methods.

Although international systems exist to monitor and investigate adverse events that occur after immunisation, and several local and vaccine-specific studies have investigated factors affecting vaccine acceptance over restricted periods of time (panel), no global systems are in place to routinely monitor and investigate emerging vaccine concerns that are not solely related to adverse events that occur after immunisation. Adverse event reporting systems are crucial to ensure vaccine safety, but, as shown here, they do not adequately capture the multitude of concerns that drive vaccine hesitancy and refusals.

The importance of listening to the public throughout the design and implementation of vaccine programmes and research trials cannot be understated. Even more important is acting on what is learned and acting early.

Panel: Research in context

Systematic review

We did a systematic review with the search terms “vaccin*”, “immunis*”, “immuniz*”, AND “trust”, “confidence”, and a set of additional related terms (eg, “hesitancy”, “anxiety”), with no date restrictions, of both mainstream and regional databases including Medline, Embase, PsycINFO, International Bibliography of the Social Sciences, Cochrane, Web of Science, Cumulative Index to Nursing and Allied Health Literature, Index Medicus for the Eastern Mediterranean Region, Literatura Latino Americana em Ciências da Saúde, and Africa-Wide Information. We did the latest search on Nov 12, 2012. Within the results of the broader search, we then searched for relevant research about online monitoring, and found four relevant studies: an assessment of vaccination sentiments in social media that analysed content and correlated the findings with vaccination rates estimated by the Centers for Disease Control and Prevention;30 an analysis of discussions about the measles, mumps, and rubella vaccine in an online chat forum for parents;31 a study that surveyed web users and assessed the popularity of influenza-vaccine-related sites during an influenza pandemic;15 None used worldwide data and none collected data about all types of vaccines.

Interpretation

This real-time information surveillance system uniquely captures both positive and negative vaccine reports and sentiments as reported in both traditional news and social media reports around the world. Because the system can categorise report content by type and monitor the evolution of concerns spatially and temporally, the system could be used to inform vaccine delivery and engagement strategies.
The retrospective analyses that informed this system’s development showed that early signs of concern have often been available well before their most serious effects occurred, but were not acted on, largely because the potential results were not expected. Although we show that vaccine concerns can vary geographically, evidence of global dissemination of vaccine concerns was apparent, enhanced by internet-based communication. Thus a global approach might help to detect confidence issues before they become established in other geographical areas.

We now have a growing body of evidence of the potential risks of the spread of unchecked rumours, and of failing to address legitimate questions and concerns. Even if concerns driven by misinformation or reported adverse events are investigated and confirmed as not being caused by vaccines, concerns and reputations in the public mind need to be addressed. Although 69% of the global reports were positive about vaccines, 31% were not. And, of these 31%, a large proportion of concerns are related to belief systems. The public health community should not underestimate the implications of the global burden of belief.

Contributors
All authors contributed to study design, data analysis and interpretation, and writing and approval of the manuscript. HJL was the principal investigator. DMDS did the statistical analysis and produced all figures.

Conflicts of interest
One representative from the Bill & Melinda Gates Foundation, which provided the funding, is a member of the project’s international advisory group, which provided feedback over the course of the research, but the research team had sole responsibility for all decisions about the conduct of the research and analysis of the findings. Since submitting this Article, Epidemicos has been formed by Boston Children’s Hospital to handle the licensing of HealthMap data for commercial companies. JSB holds an equity stake in Epidemico. All other authors declare that they have no conflicts of interest.

Acknowledgments
We thank the international advisory group of The Vaccine Confidence Project for their valuable feedback throughout this research.

References
5. Godlee F, Smith J, Marcovich H. Wakefield’s article linking MMR vaccine and autism was fraudulent. BMJ 2011; 342: c7452.