

## Twitter Data Offers Opportunities for Public Health Professionals

Chris Okugami<sup>1</sup>, Ross Sparks<sup>1\*</sup> and Sam Woolford<sup>2</sup>

<sup>1</sup>CSIRO Computational Informatics Locked Bag 17, Bentley University, Australia

<sup>2</sup>Bentley University, 175 Forest St, Waltham, MA 02452 USA

### Introduction

Twitter, an online social media network, offers a new vehicle for tracking and managing public health risks. Twitter allows users to keep friends, family and co-workers up-to-date on their status in real time using short messages of under 140 characters, called 'tweets', sent through computers and mobile phones. In addition, tweets carry information regarding the user and the user's location making it a potential data source for the detection and tracking of major events. Sakaki, Okazaki and Masuo [1] have used Twitter data to detect the occurrence and location of major events such as earthquakes and typhoons in Japan within minutes of the event. We propose that Twitter data can provide an opportunity to detect and manage public health 'events'.

Current estimates place the number of Twitter users as high as 650 million [2]. By 2011 there were 14 countries (e.g. India, Indonesia, USA, UK, Japan, and Germany) where 25% of the population used Twitter at least once in a 6 months period. In the Netherlands the percentage was 33%. Twitter use is rapidly growing in the Middle East (32% of the population in Saudi Arabia are Twitter users [3]) and in African countries such as Nigeria and South Africa. China has just less than 20% of its internet population visiting twitter at least once per month.

The broad penetration of Twitter users along with the real time nature of Twitter data and the associated information about the networks of users and their locations, suggests that Twitter data could be a valuable data source to detect, monitor and manage public health events and disease outbreaks. Moreover, the detailed information in tweets may provide information on the extent and severity of the event more quickly than emergency departments, health departments, and Google Trends. This information could then be used by public health officials to manage the event more effectively by helping them identify and interact with networks of sufferers who would benefit from interventions.

This article suggests that Twitter data offers advantages over other on-line data available through, e.g., Google Trends or electronic health data from hospital emergency departments. In addition, due to the accessibility of individual Twitter data, Twitter data may have advantages over other forms of social media such as blogs or Facebook [2,3] for detecting and tracking health risks.

### Benefits of Twitter for Disease Surveillance and Management

Twitter users can be thought of as noisy sensors whose tweets provide information regarding the appearance and severity of symptoms associated with different diseases. Within their tweets is information regarding the time of the tweet and the location of the user (the availability of GPS data for tweets is increasing with the spread of smartphones and can be supplemented by the user account registration address). Longitudinal tracking of the network of individuals who tweet about a disease or public health risk may provide additional information on the progression of any associated symptoms. This allows for a spatial as well as temporal tracking of high risk areas in real time and allows public health entities to provide warnings of local health risks on a timelier basis than might be available using emergency room data or other on-line data from sources like Google Trends.

Twitter is becoming a popular way of building communities and can provide public health professionals with a micro blogging capability that can reach people who suffer particular health problems and help them manage their disease burden during peak suffering periods. Since the complete text of tweets is available and an individual's tweets can be followed over time, multiple symptoms and other information on how unwell an individual is with a disease can be tracked over time. In addition to using this information for the detection and tracking of disease, these same Twitter networks can also be targeted by public health officials with more specific information, risk warnings and other interventions in order to more effectively manage the health risk.

To illustrate these ideas, we considered asthma sufferers in New South Wales and used Twitter data to detect the health risk to asthma sufferers due to an outbreak of bushfires which occurred in mid-October 2013. Asthma is a chronic condition that has several well documented risk factors such as smoking, allergens (atopy) and obesity. However, there are other real time risk factors such as bushfire smoke, air pollution [4] and other common allergens that might be detected and their risk effectively managed through Twitter data. Twitter data could be used to identify geographical regions at the greatest risk due to these real time risk factors. Since the best way of managing asthma is the timely administration of medication, Twitter messages from health departments during high risk periods could be used to help those sufferers and their care givers in the high risk areas to recognize the risk earlier and undertake appropriate interventions. Ideally such timely information would help sufferers avoid the need for hospital or emergency care. The tweets from sufferers who accidentally get caught in the high risk areas could be used to understand the consequences of exposure to the hazard and therefore to better manage sufferers caught unexpectedly in similar high risk situations.

Figure 1 below shows the time series plot of the number of tweets per hour which contain the keyword asthma in a region of New South Wales Australia during the period of the brush fire event. Below are some examples of asthma related tweets during this bushfire event:

- Very hazy and strong smell of smoke in Sydney this morning ... not good for asthma sufferers
- I do not envy asthmatics with this sh\*\*. It's annoying my healthy lungs....
- The smoke has done something to my asthma...idk how I'm going to do the exam to-day

\*Corresponding author: Ross Sparks, CSIRO Computational Informatics Locked Bag 17, Bentley University, 175 Forest St, Waltham, MA 02452 USA, Tel: (02)93253262; E-mail: [Ross.Sparks@csiro.au](mailto:Ross.Sparks@csiro.au)

Received January 20, 2014; Accepted January 22, 2014; Published January 28, 2014

Citation: Okugami C, Sparks R, Woolford S (2014) Twitter Data Offers Opportunities for Public Health Professionals. J Health Med Informat. 5: e123. doi:10.4172/2157-7420.1000e123

Copyright: © 2014 Okugami C, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

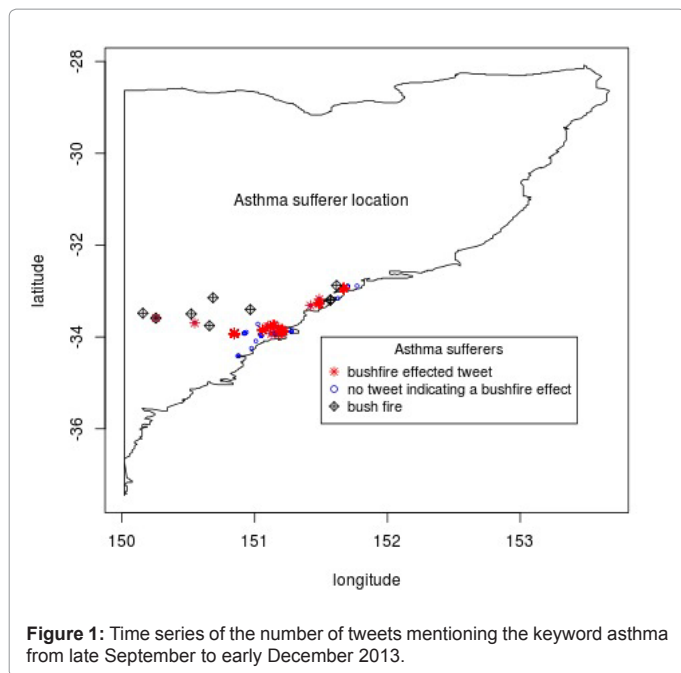


Figure 1: Time series of the number of tweets mentioning the keyword asthma from late September to early December 2013.

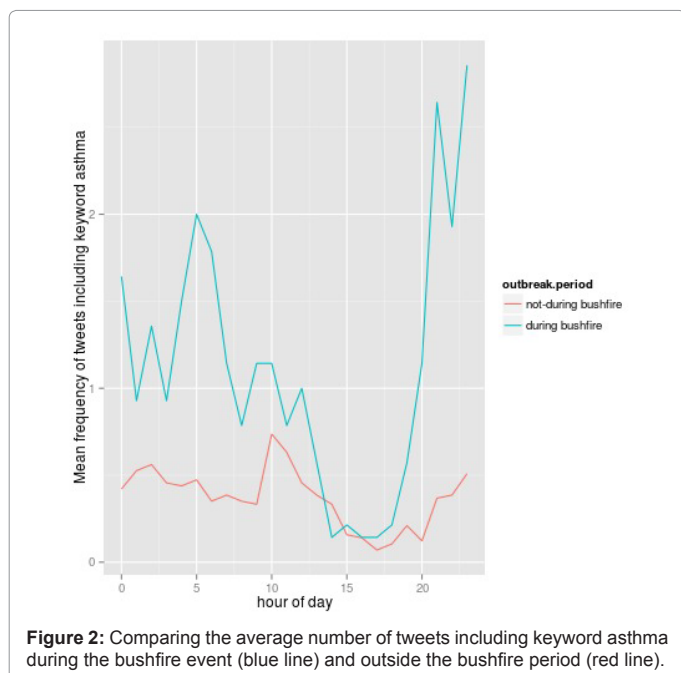


Figure 2: Comparing the average number of tweets including keyword asthma during the bushfire event (blue line) and outside the bushfire period (red line).

• ...smoke/hazy in the air, this is really affecting my asthma

The plot clearly indicates a spike in tweets at the same time the bushfire started to spread allowing for early detection of a possible health threat. Analysis of the content of the tweets provides additional information on the nature of the risk. Figure 2 shows the average number of tweets containing the word asthma during the 24 hours periods of the day for the duration of the fire outbreak and during the same 24 hours periods when there were no bushfires. Peaks in the tweet frequency during the midnight to early morning hours give some indication of the level of severity of the risk as it appears that sufferers may not be sleeping. Moreover, the peak around 9 am to 10 am may indicate that sufferers are staying at home rather than risk going

to work. Finally, Figure 3 provides a map indicating the locations of known asthma sufferers identified through their prior tweets. The blue dots in Figure 3 represent locations of asthmatics who did not tweet that they were suffering from the bushfire smoke, while red stars represent locations of asthmatics who tweeted they were affected. The location of the bushfires is also indicated in Figure 3. Although the wind direction varied over the duration of the bushfire event, it was the westerly winds that affected the Sydney and north Sydney areas. Accumulating records for several of these events could build the information needed to establish the likelihood that asthmatics in different regions would suffer due to bushfire smoke, and thus highlight the people who need to be alerted about specific bushfire events.

The longitudinal analysis of asthma sufferers' tweets during hazardous events provides public health officials with information on the severity of the hazardous event. This information can be useful to inform the response required to manage the health risk allowing health officials to more effectively and efficiently utilize their scarce resources in responding to the risk. The longitudinal analysis of the tweets also would be helpful in determining the pathways that symptoms take for given hazardous events. Knowledge of such common pathways could be helpful in the early detection and management of similar such hazardous events in the future.

Finally, public health agencies can define an asthma social network or interest group made up of Twitter users that could be valuable in several ways including:

- Providing a support group for those who suffer similar hazardous events at the same time, allowing them to leverage off the collective experience of the individuals within the network.
- Broadcasting hazardous events (location and risk) in real-time.
- Identifying the tweet population that would benefit from early detection of asthma outbreaks.

Defining a cohort of people that could benefit from other outreach

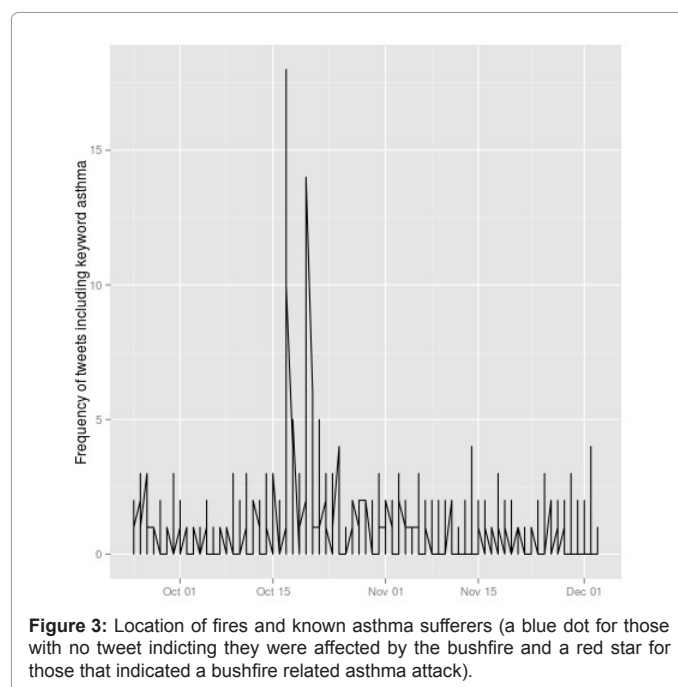


Figure 3: Location of fires and known asthma sufferers (a blue dot for those with no tweet indicating they were affected by the bushfire and a red star for those that indicated a bushfire related asthma attack).

efforts to help diagnose and control Asthma (e.g. from the Asthma Foundation).

## Conclusion

Twitter, and its associated social networks, offers significant potential for public health disease management, not only in terms of early detection and better control, but also in terms of providing information on how to better manage the health risk. Tweets are likely to occur soon after symptoms present which suggests that tweets offer early information on most diseases that start with certain symptoms and that ramp up slowly in severity. Temporal and spatial tracking through Twitter could be useful in tracking infectious diseases and otherwise managing health risks. The ability to extract information regarding disease severity and longitudinal disease pathways from tweets is helpful to better understand the disease burden and may also be helpful to further enhance the early detection of disease outbreaks. Twitter data can be a source for user feedback regarding the efficacy and adverse effects of specific drugs and other interventions in managing the disease risk.

Given the expansion in the use of Twitter, there are numerous possibilities for future research into the use of Twitter data for public health applications. In particular, there is clearly a 'selection bias' issue with who uses Twitter which needs to be better understood. The

information extracted from tweets is noisy and at times incomplete leading to issues with false discovery rates which need to be assessed and addressed. In terms of disease surveillance we don't suggest that tweets replace emergency department data or Google trend data for early detection of disease outbreaks. Each of these data sources has different strengths and weaknesses. However, the use of Twitter data in conjunction with Google search trends and hospital ED data could be used to provide an early warning system of unexpected incidences of disease. Jointly these data could also reduce the false discovery rates for disease outbreaks to an acceptable level however this is still an open issue.

## References

1. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real time event detection by social sensors. *Proceedings of the 19th International Conference on the World Wide Web* 851-860.
2. Hay SI, George DB, Moyes CL, Brownstein JS (2013) Big Data Opportunities for Global Infectious Disease Surveillance. *PLoS Med* 10: e1001413.
3. Schmidt CW (2012) Trending Now: Using Social Media to Predict and Track Disease Outbreaks. *Environ Health Perspect* 120: a30-a33.
4. YoussefAgha AH, Jayawardene WP, Lohrmann DK, El Afandi GS (2012) Air pollution indicators predict outbreaks of asthma exacerbations among elementary school children: integration of daily environmental and school health surveillance systems in Pennsylvania. *J Environ Monit* 14: 3202-3210.