Impact of Twitter intensity, time, and location on message lapse of bluebird’s pursuit of fleas in Madagascar

Omar B. Da’ar a,*, Faisal Yunus b, Nassif Md. Hossain c, Mowafa Househ a

a College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
b Institute of Public Health, College of Medicine and Health Sciences, UAE University, Al-Ain, United Arab Emirates
c King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

Received 6 January 2016; received in revised form 25 May 2016; accepted 24 June 2016

KEYWORDS
Twitter; Social media; Bubonic plague; Madagascar; Outbreak; Surveillance; Public health

Abstract

Background: The recent outbreak of bubonic plague in Madagascar reminds us of the continuing public health challenges posed by such deadly diseases in various parts of the world years after their eradication. This study examines the role of Twitter in public health disease surveillance with special focus on how Twitter intensity, time, and location issues explain Twitter plague message delay.

Method: We retrospectively analyzed the Twitter feeds of the 2014 bubonic plague outbreak in Madagascar. The analyses are based on the plague-related data available in the public domain between November 19th and 27th 2014. The data were compiled in March 2015. We calculated the time differential between the tweets and retweets, and analyzed various characteristics of the Tweets including Twitter intensity of the users.

Results: A total of 6873 Twitter users were included in the study, of which 52% tweeted plague-related information during the morning hours (before mid-day), and 87% of the tweets came from the west of the epicenter of the plague. More importantly, while session of tweet lease and relative location had effect on message lapse, absolute location did not. Additionally, we found no evidence of differential effect of location on message lapse based on relative location i.e. tweets from west or east nor number of following. However, there is evidence that more intense
Introduction

The history, origin, cause, transmission, and death rate of bubonic plague is well documented [1]. However, bubonic plague continues to recur around the world years after it was eradicated. The number of cases reported globally in recent times is between 1000 and 2000, with most cases emerging out of Africa [2]. The recurrence of plague is due to many factors including absence of effective antibiotics and resistance of plague causing fleas to the insecticides. A 2014 outbreak of plague in Madagascar started from a single case and rapidly spread to 16 districts in the country due to high population density. The disease was confirmed in 119 persons and proved fatal in 40 of them [3]. Hence, it is increasingly essential to monitor the transmission of diseases and restrict their spread in a timely manner to prevent public health catastrophes [4]. Many public health agencies across the globe are adopting innovative ways to improve the detection of disease outbreaks. The electronic monitoring systems, such as EpiSPIDER, HealthMap, BioCaster, and the Global Public Health Intelligence Network mine the websites, like Twitter, for any reports on flagged diseases and provide a real-time tracking of disease progression [5–7].

Twitter is a publicly accessible social media platform where the users express a diverse range of views (tweets) within the allowable 140 characters. Most of the tweets are accessible via the Twitter Application Programming Interface (API), and have been used by the public health researchers in disease surveillance [7]. Studies have shown a strong similarity between the information extracted from the formal disease surveillance systems and that from the tweet mining [7]. For example, the information on cholera incidents in Haiti gathered from the HealthMap and the tweets collected via surveillance systems operated by the Haitian Ministry of Public Health indicated significant positive correlation [8]. Unlike traditional means of disease surveillance, which may take up to a week to confirm existence of an outbreak after its inception, social media can alert the public health agencies of an outbreak situation almost instantaneously [9,10]. In spite of this potential, it is not uncommon for the social media messages to face problems in their relay. The amount of time it takes for the first message about a disease outbreak hitting the public domain and its subsequent dissemination via social media is crucial to contain the spread of an outbreak. To best of our knowledge, the dependence of Twitter message relay on time and place issues and Twitter intensity has not been extensively explored in the context of disease outbreak situations. We have, therefore, examined the role of Twitter in the recent bubonic plague outbreak in Madagascar in this study with special focus on whether time of the day, the place of the message origin, and Twitter intensity can affect the relay of messages in emergency situations.

Methods

Fig. 1 presents a schematic framework that we have used as a basis for analysis and interpretation of the study findings.

Data source and timeframe

We retrospectively analyzed a cross-section of Twitter feeds during the 2014 bubonic plague outbreak in Madagascar. The analyses are based on the plague-related data available in the public domain between 19th November and 27th November 2014. The data were compiled in March 2015.

Operational definitions

We analyzed various characteristics of the tweets and the Twitter users. For example, we divided the tweets into the time of the day (session) when the
Impact of Twitter intensity, time, and location on message lapse of bluebird’s pursuit

message was sent (morning versus afternoon), the relative geographical location of the Twitter user (east versus west of the plague epicenter), and the absolute location of the Twitter users based on their global positioning system (GPS) coordinates. From these GPS coordinates, we calculated the users’ absolute distance from the plague epicenter. In addition, we observed Twitter intensity of the users in terms of the number of tweets, followers, following, and retweets. Retweeting is a common activity on Twitter and reflects the popularity of individual tweets. Follower or following are the terms used when individuals are subscribed to other persons’ tweets.

We calculated plague-related message relay as a measure of time interval (time differential) between a starting point (when the message first hit the public domain) and the current point in time (when a message is subsequently tweeted) to spread information about the plague.

**Statistical analysis**

We performed the statistical analyses using SPSS 21.0 software (IBM, NY, USA). Number of observations (n) and percentages (%) were reported for all categorial data. Correlation analysis was performed to find associations between message relay time and its likely predictors. Multiple regression analysis was used to predict the plague-related message relay time based on time of the day when the message was released, the absolute location of message origin, relative geographical position, and the social media intensity such as the number of tweets, retweets, followers, and following. All statistical tests were 2-tailed and the level of significance was set at \( p = 0.05 \). The variables are defined in Table 1.

**Results**

**Characteristics of the tweets**

A total of 6873 Twitter users were included in the study, of which 52% tweeted plague-related information during the morning hours of the day, and 87% of the tweets came from the west of the

---

**Table 1  Variables and definitions.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Message delay/lapse</td>
<td>A measure of time interval (time differential) between a starting point (when the message first hit the public domain) and the current point in time (when a message is subsequently tweeted) to spread information about the plague</td>
</tr>
<tr>
<td>Am</td>
<td>A dummy/binary variable taking a value of one if a tweet was released in a morning session, and zero if tweeted otherwise</td>
</tr>
<tr>
<td>West</td>
<td>A dummy/binary variable that depicts the relative geographical location of the Twitter user and takes a value of one if a tweet comes from west of the epicenter of the plague, and zero if tweet comes from east of epicenter</td>
</tr>
<tr>
<td>Absolute location</td>
<td>Twitter users’ absolute distance from the plague center calculated from GPS coordinates</td>
</tr>
<tr>
<td>Number of tweets</td>
<td>Measures of Twitter intensity</td>
</tr>
<tr>
<td>Number of retweets</td>
<td>Measures of Twitter intensity</td>
</tr>
<tr>
<td>Number of followers</td>
<td>Measures of Twitter intensity</td>
</tr>
<tr>
<td>Number of following</td>
<td>Measures of Twitter intensity</td>
</tr>
<tr>
<td>Tweets squared and Retweets squared</td>
<td>Depicts the effect of more intensive number of tweets and retweets</td>
</tr>
</tbody>
</table>

---

Please cite this article in press as: Da’ar OB, et al. Impact of Twitter intensity, time, and location on message lapse of bluebird’s pursuit of fleas in Madagascar. J Infect Public Health (2016), http://dx.doi.org/10.1016/j.jiph.2016.06.011
epicenter of the plague as shown in Table 2. A higher proportion of tweets were sent from the west of the epicenter irrespective of the session of the day.

**Effect of time and place on tweet delay**

The effect of time and space variables on tweet message time lapse is shown in Table 3 (Specification 1). The results indicate that other things being equal, tweeting during morning session (AM) resulted in more message lapse or delay compared to tweeting in afternoon session (p < 0.01). However, while tweeting from a location west of the epicenter of the plague increased message delay compared to tweeting from the east, this result was not statistically significant (p > 0.05). In addition, the absolute geographic location of the Twitter user did not statistically influence the message delay (p > 0.05).

**Effect of Twitter intensity on tweet time lapse**

As highlighted in Table 3 (Specification 2), unlike time and space variables, Twitter intensity characteristics significantly explained variations in tweet message time lapse. Other factors held constant, an increase in the number of tweets of a user and tweet followers minimized delay of the plague message while an increase in the number of following and retweets of a user resulted in more message delay (p < 0.01, for all analyses). However,

### Table 2: Characteristics of Twitter users and Tweets for the plague outbreak in Madagascar (total users = 6873).

<table>
<thead>
<tr>
<th>Twitter Intensity</th>
<th>Number (% of total users)</th>
</tr>
</thead>
</table>
| Tweets sent by session of the day | Morning: 3548 (51.6)  
 | Afternoon: 3325 (48.4) |
| Tweets by users’ geographical location | East: 897 (13.1)  
 | West: 5976 (86.9) |

<table>
<thead>
<tr>
<th>Tweets by Session of the day and Geographical location</th>
<th>Number (% of users within each category)</th>
</tr>
</thead>
</table>
| Session of the day — Morning | East: 465 (13.1)  
 | West: 3083 (86.9) |
| Session of the day — Afternoon | East: 432 (13)  
 | West: 2893 (87) |

### Table 3: The impact of Twitter intensity and time-space on plague-related message delays — coefficients and statistical significance.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of day (AM)</td>
<td>0.757***</td>
<td>0.785***</td>
<td>0.831***</td>
</tr>
<tr>
<td>Location (West of epicenter)</td>
<td>0.115*</td>
<td>0.0476</td>
<td>0.0741</td>
</tr>
<tr>
<td>Tweet following</td>
<td>0.0138***</td>
<td>0.0126***</td>
<td>0.0119</td>
</tr>
<tr>
<td>Retweets</td>
<td>0.00569***</td>
<td>0.0123***</td>
<td>0.0123***</td>
</tr>
<tr>
<td>Tweets</td>
<td>−0.00189***</td>
<td>−0.00398***</td>
<td>−0.00398***</td>
</tr>
<tr>
<td>Tweet followers</td>
<td>−0.000339***</td>
<td>−0.000406***</td>
<td>−0.000406***</td>
</tr>
<tr>
<td>Absolute location</td>
<td>0.000138</td>
<td>0.000175</td>
<td>0.000173</td>
</tr>
<tr>
<td>Tweets squared</td>
<td>2.93e−06***</td>
<td>2.93e−06***</td>
<td>2.93e−06***</td>
</tr>
<tr>
<td>Retweets squared</td>
<td>−1.49e−05***</td>
<td>−1.48e−05***</td>
<td>−1.48e−05***</td>
</tr>
<tr>
<td>AM × West</td>
<td>−0.0540</td>
<td>0.000703</td>
<td></td>
</tr>
<tr>
<td>Tweet following × West</td>
<td>4.150***</td>
<td>4.237***</td>
<td>4.214***</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.410</td>
<td>0.421</td>
<td>0.421</td>
</tr>
</tbody>
</table>

Specification 1 is baseline estimation, Specifications 2 and 3 include various interactions. Coefficient for each independent variable gives the size of the effect that variable is having on the dependent variable (message delay), and the sign on the coefficient (positive or negative) gives the direction of the effect. × implies interaction.

*p < 0.1, **p < 0.05, ***p < 0.01.

Please cite this article in press as: Da’ar OB, et al. Impact of Twitter intensity, time, and location on message lapse of bluebird’s pursuit of fleas in Madagascar. J Infect Public Health (2016), http://dx.doi.org/10.1016/j.jiph.2016.06.011
the results suggest that more intense tweets and retweets (tweets and retweets squared) resulted in more message delay ($p < 0.01$).

**Effect of interaction of time, space and Twitter intensity on message delay**

Based on the extended analysis as depicted in Table 3 (Specification 3), the results further revealed that there was no statistical evidence to support the hypothesis that tweeting from the west or east of the plague center sufficed to express the effect of when a tweet was sent (session during tweeting, AM/PM) on message delay of plague message ($p > 0.05$). Similarly, there was no statistical evidence to support the hypothesis that west or east positioning sufficed to express the effect of the number of Twitter following on message delay of plague message ($p > 0.05$). Although not statistically significant, the direction of interaction effect appears to suggest washing out the main effects of Twitter intensity such as the number of following and retweets on message delay ($p > 0.05$).

**Discussion**

This study is based on the premise that health informatics can greatly and positively help to protect the global population from harm and ill-health in a timely fashion. By the year 2017, the number of corporations and individuals using social media platforms will amount to 4.8 billion [11], suggesting the need for more research and development in areas of real-time disease surveillance. We examine whether Twitter media intensity and the time and place of origin of tweets can affect bubonic plague Twitter message relay.

In our study’s context, time and place are proxied by when the message was released, the absolute positioning of the Twitter user and their relative geographical location with respect to the plague epicentre. The number of Twitter followers, following, and number of tweets and retweets collectively measured Twitter intensity. We analyzed the associations between various variables by regression models. We further estimated a nonlinear variant to allow for interaction between some of the predictors that can lead to plague message delay. The results demonstrate that the time of the day when the message is sent and the geographical location of the users have no impact on message delay even in emergency situations. We also found a statistically significant relationship between Twitter intensity and the plague-related message delay, where a higher number of tweets and followers minimize message delay. These results are intuitive and consistent with the notion of “liquid modernity” in which when a node distance decreases, communication becomes more liquid, and digital geography between the two people, thoughts, ideas, or groups is instantly traversed [12].

The social media channels are increasingly being used in this day and age to supplement the work of traditional media [13], and have the potential to increase the reach and efficiency of essential public health services, such as surveillance, research, and communication [14,15]. In addition, social media can also be used to quickly and directly reach out to the public in situations that require prompt action such as food recalls, disease outbreaks and weather emergencies [16].

This study shows how communicating disaster or plague related information can be effectively and quickly relayed to disease surveillance officials irrespective of the distance and time, which have both shown to be statistically insignificant. This proves especially useful in the surveillance and detection of disease strains for all infectious diseases. This is because any new disease strain would carry the potential to trigger pandemics and epidemics. It should be kept in mind that the most effective use of disease surveillance of this kind is best complemented by strong laboratory facilities. These laboratory facilities could harbor the right mix of resources (both financial and labor) in conjunction with a strong sense of quality control in battling disease outbreaks once their spread and location has been detected via surveillance [17].

Our findings also shed light upon how the social media inclusive of Twitter may prove useful in collating data on various other variables which may precipitate the progression of plague like diseases. This holds promise to be a key feature in the future of evidence-based practice of disease surveillance. Climate variables, such as temperature, humidity and precipitation can change the dynamics of effects and progression of disease in geographical terms from the epicenter, thereby potentially having effect on message time delay [18].

Since its inception, the very use of Twitter has also changed over the years. Users are not only using Twitter as a "simple message posting service" but also to schedule events comprised of Twitter followers. These events may be in the form of "free flowing discussions", Q&A sessions and critical appraisal of disease related information via tweets [19].

Other upcoming features such as the "Twitter Town Hall" and "Live tweeting" will further develop the ways in which messages regarding disease outbreak spread. Twitter Town Halls are
scheduled forums which allow users to convey questions regarding a topic. Public health officials may very well regard this avenue when relaying information of public health safety and advice in case of a crisis [19]. Live Tweeting would also improve emphasis on disease outbreak news by highlighting important points of a live event, which would change how followers engage and utilize information both gathered and relayed using tweets. This information can then be evaluated using classifier performance markers such as the 10-fold cross-validation method to measure performance accuracy [20].

Further temporal and text analysis of tweets in detail may prove useful in future studies of how tweet volume related data may be used to infer characteristics of disease related tweets and how this may change over time [21]. By use of in-depth text analysis we may also begin to infer any patterns in mention of symptoms, treatments and exacerbating disease circumstance through tweets made regarding outbreaks [22].

Though our study illustrates the benefits of Twitter use, there are a few inherently natured limitations in using social media for health policy measures. One such limitation is "coverage", this is the rate at which extracted data from Twitter users can in real-life be utilized to formulate estimates of disease prevalence for use in disease prevention policy making [23]. Another limitation involves how countries which do not use English as an official language are disadvantaged from social media systems which only cater for tweets written in English [23]. Automated translation of key words may be a useful consideration in the global development of disease surveillance systems used by a widely diverse social media audience. With regard to social media, veracity of data is an issue. A single individual could severally tweet using multiple accounts about the same message. This why even important analysis carried out, results revealed ought to be conservatively interpreted. Finally, while the regression we use considered all the assumptions of the classical ordinary least square model, the naivety of the model cannot be ruled out. A further analysis is, therefore, in order that considers the issues of measurement errors and relationship between included independent variables and those variables excluded and therefore assumed to be in the residual.

Our study alongside other published works point towards an expanding scope in the field of disease surveillance via use of "social media analytics" [24]. This refers to a finer and more evidence-based grade of disease surveillance which addresses the real-time threats and precautions taken with regards to outbreaks. Further development of effective disease prediction models and protocols [25] catered for varying geographical locations, languages and cultures would maximize overall benefits derived from such surveillance systems, as well as their subsequent use by the public health officials.

Conclusion

This study affirms that Twitter intensity does matter while time and space issues do not affect the plague-related message delay. This study also affirms that Twitter intensity can play an important role in ongoing disease surveillance and the timely dissemination of information during public health emergencies independent of the time and space restrictions.

Further ways should be explored to embed social media channels in routine public health practice. These may include the adoption of disease surveillance modeling techniques in accurately tracing disease progression amongst the population via use of extracted data available on social media. Further emerging technologies coupled with social media allows the connection of thousands and millions of users and their voices. This would allow the effective and systematic spread of health and safety related information inclusive of their extent of impact on our communities.

More research can be done on the variables of climate and migration of disease to help refine allocation of scarce surveillance resources by governing public health bodies. This would help in improving the equity in the distribution of disease prevention resources and public empowerment offered by social media can promote a safer and healthier atmosphere for us all.

Funding

No funding sources.

Competing interests

None declared.

Ethical approval

Not required.
References


[19] The health communicator’s social media toolkit. Centers for Disease Control and Prevention, Ofﬁce of the Associate Director for Communication; 2011, July.


