Characterisation of social media conversations on syphilis: an unobtrusive observational study

Abby C. Dawson[A], Alyssa K. Fitzpatrick[B], Janet M. Matthews[A], Andrew A. K. Nguyen[A], Kelly Papanaoum[B] and Justine R. Smith[A,*] [iD]

For full list of author affiliations and declarations see end of paper

*Correspondence to:
Justine R. Smith
College of Medicine and Public Health, Flinders University, Adelaide, SA, Australia
Email: justine.smith@flinders.edu.au

Handling Editor:
Lei Zhang

ABSTRACT

Background. Conversations around disease conducted through social media provide a means for capturing public perspectives that may be useful in considering public health approaches. Syphilis is a sexually transmitted disease that is re-emerging. We sought to characterise online discourse on syphilis using data collected from the social media platform, Twitter. **Methods.** We extracted English-language tweets containing the word 'syphilis' posted on Twitter in 2019. Tweet identification number and URL, date and time of posting, number of retweets and likes, and the author's screen name, username and biographical statement were included in the dataset. A systematically sampled 10% subset of the data was subjected to qualitative analysis, involving categorisation on content. All tweets assigned to the category of medical resource were assessed for clinical accuracy. The engagement ratio for each category was calculated as (retweets + likes):tweets. **Results.** In 2019, 111,388 tweets mentioning syphilis were posted by 69,921 authors. The most frequent content category – totalling 5370 tweets (48%) – was a joke. Of 1762 tweets (16%) categorised as a medical resource, 1484 (84%) were medically correct and 240 (14%) were medically incorrect; for 38 (2%), medical accuracy could not be judged from the information posted. Tweets categorised as personal experiences had the highest engagement ratio at approximately 19:1. Medical resource tweets had an engagement ratio of approximately 7:1. **Conclusions.** We found medical information about syphilis was limited on Twitter. As tweets about personal experiences generate high engagement, coupling an experience with information may provide opportunity for public health education.

Keywords: medical resource, sexually transmitted disease, social media, syphilis, Twitter, X.

Introduction

Syphilis is a sexually transmitted disease caused by *Treponema pallidum*.[1] *T. pallidum* does not have an animal reservoir, and penicillin continues to be an effective treatment for syphilis.[2] These two factors have made syphilis a perfect candidate for disease elimination, and in 1999, the Centers for Disease Control and Prevention proposed the *Syphilis Elimination Effort* to reduce the number of individuals with primary and secondary syphilis in the USA to <1000.[3] In 2013, however, the initiative was suspended due to a significant increase in syphilis case numbers.[2] The incidence of syphilis is highest in low-income countries, but has been up-trending across multiple high- and medium-income countries.[4] A global prevalence estimate published in 2019 suggests that 0.5% of women and 0.5% of men worldwide have the disease.[5]

In recent years, researchers have looked to conversations on social media platforms, such as Twitter (https://twitter.com/?lang=en), as a means of capturing the public perception of different diseases.[6–10] One group of investigators reported an association between the level of syphilis-focused Twitter activity and numbers of reported cases across USA counties.[11] However, Twitter-based discussions about syphilis have not been analysed qualitatively. The content of engagement with user posts on social media could provide information about public understanding of syphilis. Measuring this knowledge may suggest strategies to address gaps that reduce the prevalence of syphilis. Our study characterises the posts – or tweets – written by Twitter users in 2019 related to syphilis.

Methods

The dataset was sourced on 8 September 2020 from Twitter (https://twitter.com/?lang=en) by TrackMyHashtag, a social media analytics company that uses proprietary software (https://www.trackmyhashtag.com). Every tweet posted in English between 1 January and 31 December 2019 that contained the word 'syphilis' was identified, including tweets containing the term '#syphilis'. Retweets were excluded. The tweet identification number and URL, the date and time of posting, and the number of retweets and likes were collected. The number of replies was not collected. To provide context to the tweets, the author's screen name, username and biographical statement, plus a link to their Twitter page, were also obtained.

This unobtrusive observational study involved analysis of publicly available data, and review by a human research ethics committee was not indicated.

Numbers of tweets for each month were compared by Poisson regression analysis, and considered to lie outside predicted if the P-value was <0.05. A 10% sample of the dataset was taken for qualitative analysis. This sample size was chosen for feasibility of analysis, and was comparable to similar published studies.[9,12–15] The sample was taken by systematic sampling, including every 10th tweet in the chronological list of all tweets, to avoid selection bias.

Sampled tweets were classified into the following categories based on content: medical resource, non-medical resource, personal experience, public conversation, joke, medical question, non-medical question, spam, testing advertisement and foreign language. This straight-forward classification system was modified from Chew *et al.*,[9] who undertook a content analysis of tweets related to the 2009 H1N1 influenza pandemic (Supplementary Table S1). One researcher (AD) classified all tweets taking a two-round approach to ensure precision. If needed for categorisation, the tweet was accessed online via hyperlink and read in the context of any available thread. A small number of tweets in a language other than English were collected; these tweets were removed from the content analysis.

All tweets assigned to the category of medical resource were reviewed by an infectious disease registrar (AF) to determine whether they were medically correct or incorrect. The content of the text, any associated still images or videos and links provided within each tweet were considered in the assessment. Engagement was calculated for each content category, as a ratio of the sum of retweets plus likes to the number of tweets within the category.

Results

In 2019, a total of 111,388 English-language tweets containing the term 'syphilis' were posted by 69,921 different authors. The number of tweets in each month was consistent overall, with a mean of 9282 and a median of 9079. A surge in tweets occurred during October, with an increase of 3096 from the mean number of monthly tweets, and there was a drop in tweets during the months of May and June, with a decrease of 1243 and 1239 from the mean, respectively ($P < 0.05$ for each difference).

Analysis of the content of the 10% sample of tweets ($n = 11{,}113$ after removal of 25 foreign language tweets) revealed that the most frequent category was joke ($n = 5370$, 48%). Other categories of tweets were less frequent, including: public conversation ($n = 2515$, 23%), medical resource ($n = 1762$, 16%), non-medical resource ($n = 622$, 6%), spam ($n = 462$, 4%), testing advertisement ($n = 161$, 1%), personal experience ($n = 112$, 1%), medical question ($n = 62$, 0.6%) and non-medical question ($n = 47$, 0.4%; Fig. 1*a*).

Of the 1762 tweets categorised as medical resource, 1484 tweets (84%) were judged to be medically correct, whereas 240 tweets (14%) were deemed to be medically incorrect; for 38 tweets (2%), it was not possible to determine medical accuracy (Fig. 1*b*). Although this shows the majority of tweets categorised as medical resource are often medically correct, in the context of the sample analysed, only 13% of all tweets presented medically accurate information. Ranking the authors according to the number of tweets posted, 'HIV Insight' posted the most tweets categorised as medical resource ($n = 34$; Table S2).

A calculation of the engagement ratio (retweets + likes:all tweets) for tweets in different content categories revealed the highest engagement for those categorised as personal experience (2109:112, approximately 19:1), whereas the category with the least engagement was non-medical question (39:47, approximately 1:1) (Fig. 1*c*). Medical resource tweets had an engagement ratio of approximately 7:1 (12,141:1762). The medical resources tweets attracting the highest engagement are presented in Table S3.

Discussion

A total of 111,388 tweets posted on Twitter in 2019 referenced syphilis. By analysing a 10% sample, we found that approximately 15% of tweets provided medical information, and the majority of these medical resource tweets presented correct information. We found that tweets on syphilis that presented a personal experience generated the most engagement, although medical resource tweets also generated some engagement.

Review of the literature identifies no prior studies of the prevalence of medical information on syphilis on social media. However, several groups have published studies analysing Twitter content related to other infectious diseases: Meadows *et al.*,[15] who focused on the 2015 Californian measles outbreak; Chew *et al.*,[9] who reported on the 2009 H1N1 influenza A pandemic; and Gabarron *et al.*,[16] who considered chlamydial and HIV infections. The annual
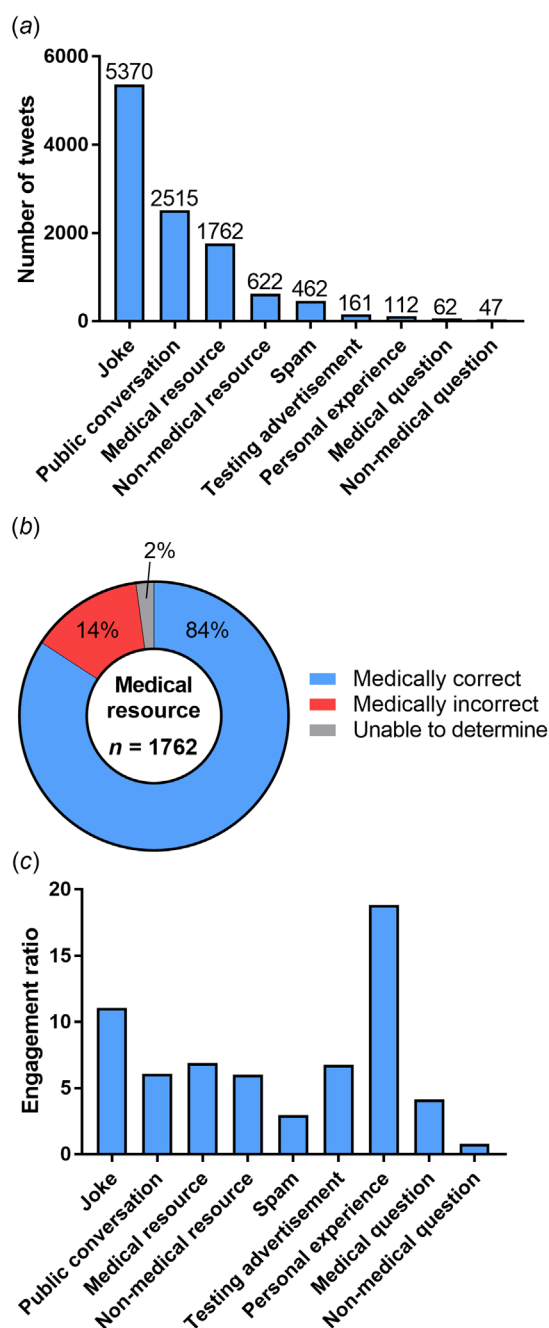
Fig. 1. Content of tweets containing the term 'syphilis' posted in 2019, determined by analysis of a 10% subset (*n* = 11,113 written in English). (*a*) Graph presenting number of tweets by content category, showed from most to least frequent. (*b*) Pie chart showing percentages of medically correct and incorrect tweets in the medical resource category (*n* = 1762). (*c*) Graph presenting engagement ratio, representing the ratio of retweets plus likes to all tweets for each category.

number of tweets on syphilis in our study far outweighed the time-corrected number on chlamydial and HIV sexually transmitted infections reported by Gabarron *et al.*[16]: 111,388 versus 15,340.

The other author groups found 41–80% of tweets were an information source,[9,15,16] up to five-times more than the percentage we measured. We observed considerably more tweets categorised as a joke (48%) than Chew *et al.*[9] and Gabarron *et al.*,[16] who reported percentages of 8% and 10%.[9,16] Oh *et al.*[17] studied engagement of HIV-related Twitter messaging, and found that tweets from personal accounts generated more engagement than those from institutional accounts. Consistently, our analysis of tweet engagement suggests those that convey a personal experience with the infection disease are of most interest to Twitter users.

Health misinformation is more common on Twitter than other social media platforms.[18] Incorrect medical information poses risks to the health of persons with syphilis, which may cause serious neurological and cardiovascular complications, and TORCH infection (toxoplasmosis, rubella, cytomegalovirus, herpes simplex and other organisms, including syphilis, parvovirus and Varicella zoster).[19] A total of 84% of tweets that represented a medical resource on syphilis were found to be medically correct. In a study of medical information posted on Twitter around the onset of the coronavirus 2019 (COVID-19) pandemic, albeit on a smaller scale than our work, Swetland *et al.*[20] described a similar percentage of correct medical resource tweets; that is, 74.4% of 358 tweets.

One limitation of our investigation into the syphilis content on social media is the focus on Twitter alone. Although Twitter has been identified as the most popular social media forum for sharing health care-related information,[21] future studies could expand on our findings by exploring conversations occurring on other social media platforms. As our research was unobtrusive, we cannot comment on author motivations for content type and correctness. Additionally, as only tweets written in English were analysed, our findings are representative of conversations about syphilis led by English-speaking authors. We observed temporal changes in tweet volume, suggesting value of research into correlations with public health events.

Research in the area of syphilis and social media may have value for guiding public health strategies that address the increasing rates of this sexually transmitted disease. Through understanding the type of information online and the types of messages that generate the most user engagement, social media could be used for effective public education. Our study has demonstrated that there is limited medical information about syphilis on Twitter. However, our results also suggest a potential solution to this health education gap could be framing resources with statements of personal experiences for maximum engagement.

Supplementary material

Supplementary material is available online.

References

1 Peeling RW, Mabey D, Kamb ML, Chen X-S, Radolf JD, Benzaken AS. Syphilis. *Nat Rev Dis Primers* 2017; 3: 17073. doi:10.1038/nrdp.2017.73

2 Willeford WG, Bachmann LH. Syphilis ascendant: a brief history and modern trends. *Trop Dis Travel Med Vaccines* 2016; 2: 20. doi:10.1186/s40794-016-0039-4

3 The Centers for Disease Control and Prevention. Syphilis elimination effort. 2013. Available at https://www.cdc.gov/stopsyphilis/default.htm

4 Read P, Fairley CK, Chow EPF. Increasing trends of syphilis among men who have sex with men in high income countries. *Sex Health* 2015; 12(2): 155–163. doi:10.1071/SH14153

5 Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, *et al.* Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. *Bull World Health Organ* 2019; 97(8): 548–562P. doi:10.2471/BLT.18.228486

6 Dowshen N, Lee S, Matty Lehman B, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. *AIDS Behav* 2015; 19: 106–111. doi:10.1007/s10461-014-0991-9

7 Coughlan E, Young H, Parkes C, Coshall M, Dickson N, Psutka R, *et al.* A novel response to an outbreak of infectious syphilis in Christchurch, New Zealand. *Sex Health* 2015; 12(2): 170–173. doi:10.1071/SH14140

8 Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg* 2012; 86(1): 39–45. doi:10.4269/ajtmh.2012.11-0597

9 Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* 2010; 5(11): e14118. doi:10.1371/journal.pone.0014118

10 Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. *J Med Syst* 2013; 37(3): 9942. doi:10.1007/s10916-013-9942-7

11 Young SD, Mercer N, Weiss RE, Torrone EA, Aral SO. Using social media as a tool to predict syphilis. *Prev Med* 2018; 109: 58–61. doi:10.1016/j.ypmed.2017.12.016

12 Krauss MJ, Sowles SJ, Moreno M, Zewdie K, Grucza RA, Bierut LJ, *et al.* Hookah-related Twitter chatter: a content analysis. *Prev Chronic Dis* 2015; 12: 150140. doi:10.5888/pcd12.150140

13 Sabus C, Johns B, Schultz N, Gagnon K. Exploration of content and reach of physical therapy-related discussion on Twitter. *Phys Ther* 2019; 99(8): 1048–1055. doi:10.1093/ptj/pzz063

14 Ahmed W, Vidal-Alaball J, Downing J, Lopez Segui F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *J Med Internet Res* 2020; 22(5): e19458. doi:10.2196/19458

15 Meadows CZ, Tang L, Liu W. Twitter message types, health beliefs, and vaccine attitudes during the 2015 measles outbreak in California. *Am J Infect Control* 2019; 47(11): 1314–1318. doi:10.1016/j.ajic.2019.05.007

16 Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. *J Med Internet Res* 2014; 16(10): e228. doi:10.2196/jmir.3259

17 Oh J, Bonett S, Kranzler EC, Saconi B, Stevens R. User- and message-level correlates of endorsement and engagement for HIV-related messages on Twitter: cross-sectional study. *JMIR Public Health Surveill* 2022; 8(6): e32718. doi:10.2196/32718

18 Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. *J Med Internet Res* 2021; 23(1): e17187. doi:10.2196/17187

19 Peeling RW, Mabey D, Chen X-S, Garcia PJ. Syphilis. *Lancet* 2023; 402(10398): 336–346. doi:10.1016/S0140-6736(22)02348-0

20 Swetland SB, Rothrock AN, Andris H, Davis B, Nguyen L, Davis P, *et al.* Accuracy of health-related information regarding COVID-19 on Twitter during a global pandemic. *World Med Health Policy* 2021; 13(3): 503–517. doi:10.1002/wmh3.468

21 Pershad Y, Hangge P, Albadawi H, Oklu R. Social medicine: Twitter in healthcare. *J Clin Med* 2018; 7(6): 121. doi:10.3390/jcm7060121

Author affiliations
[A]College of Medicine and Public Health, Flinders University, Adelaide, SA, Australia.
[B]Microbiology and Infectious Disease Unit, Southern Adelaide Local Health Network, Adelaide, SA, Australia.