Protocol

Insights From Twitter Conversations on Lupus and Reproductive Health: Protocol for a Content Analysis

Oleg Stens[1], MD; Michael H Weisman[2], MD; Julia Simard[3*], ScD; Katja Reuter[4,5*], PhD

[1]Department of Internal Medicine, Harbor-UCLA Medical Center, Torrance, CA, United States

[2]David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[3]Division of Epidemiology, Department of Health Research and Policy, Stanford University, Palo Alto, CA, United States

[4]Institute for Health Promotion and Disease Prevention Research, Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, United States

[5]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Katja Reuter, PhD
Southern California Clinical and Translational Science Institute
Keck School of Medicine of USC
University of Southern California
2250 Alcazar St
Los Angeles, CA, 90089
United States
Phone: 1 323 442 2046
Email: katja.reuter@gmail.com

Abstract

Background: Systemic lupus erythematosus (SLE) is the most common form of lupus. It is a chronic autoimmune disease that predominantly affects women of reproductive age, impacting contraception, fertility, and pregnancy. Although clinic-based studies have contributed to an increased understanding of reproductive health care needs of patients with SLE, misinformation abounds and perspectives on reproductive health issues among patients with lupus remain poorly understood. Social networks such as Twitter may serve as a data source for exploring how lupus patients communicate about their health issues, thus adding a dimension to enrich our understanding of communication regarding reproductive health in this unique patient population.

Objective: The objective of this study is to conduct a content analysis of Twitter data published by users in English in the United States from September 1, 2017, to October 31, 2018, in order to examine people's perspectives on reproductive health among patients with lupus.

Methods: This study will analyze user-generated posts that include keywords related to lupus and reproductive health from Twitter. To access public Twitter user data, we will use Symplur Signals, a health care social media analytics platform. Text classifiers will be used to identify topics in posts. Posts will be classified manually into the a priori and emergent categories. Based on the information available in a user's Twitter profile (ie, username, description, and profile image), we will further attempt to characterize the user who generated the post. We will use descriptive statistics to analyze the data and identify the most prevalent topics in the Twitter content among patients with lupus.

Results: This study has been funded by the National Center for Advancing Translational Science (NCATS) through their Clinical and Translational Science Awards program. The Institutional Review Board at the University of Southern California approved the study (HS-18-00912). Data extraction and cleaning are complete. We obtained 47,715 Twitter posts containing terms related to "lupus" from users in the United States, published in English between September 1, 2017, and October 31, 2018. We will include 40,885 posts in the analysis, which will be completed in fall 2020. This study was supported by funds from the has been funded by the National Center for Advancing Translational Science (NCATS) through their Clinical and Translational Science Awards program.

Conclusions: The findings from this study will provide pilot data on the use of Twitter among patients with lupus. Our findings will shed light on whether Twitter is a promising data source for learning about reproductive health issues expressed among

patients with lupus. The data will also help to determine whether Twitter can serve as a potential outreach platform for raising awareness of lupus and reproductive health and for implementing relevant health interventions.

Introduction

Background and Rationale

Lupus is a chronic autoimmune disease that can affect any part of the body (skin, joints, or vital organs) [1,2]. Estimates from recent population-based studies in the United States report the prevalence of systemic lupus erythematosus (SLE), the most common form of lupus, to be between 60 and 80 per 100,000, although this prevalence varies greatly by age, gender, race, and ethnicity. It is generally accepted that SLE is much more prevalent in women than men (up to 9 times higher prevalence) and that people of color have both higher prevalence rates and more severe manifestations of the disease compared to White populations. Rates as high as 196 per 100,000 have been reported in African American women [3,4].

SLE predominantly impacts women during the childbearing years, affecting contraception, fertility, and pregnancy, which are matters of importance to the patients and their family members. Providing care to pregnant patients with lupus is an important challenge for their families and the health care system. Although quite a few studies in the modern era have clarified the field of reproductive health care for SLE patients [5], misinformation abounds. Perspectives on reproductive health issues, especially those regarding medication risks and benefits, among patients with lupus and their family members remain poorly understood. In this study, we define the term "perspective" as an expression of thought, viewpoint, and attitude toward the reproductive health issues that have been identified in the literature, such as pregnancy prevention, pregnancy termination, pregnancy planning, conception, and concerns and management of childbirth [6]. A better understanding of the perspectives on reproductive health issues among patients with lupus can inform and improve the advocacy and education efforts to address the gaps in care, dispel misconceptions, and more effectively assist patients in making family planning decisions.

Social Media

Social media consists of web-based and mobile technologies that allow users to view, create, and share information online and participate in social networking [7-9]. Social media provides a unique source for data mining of health conditions and concerns, serving as a massive focus group [10-12]. A total of 72% of American adults use at least some type of social media [13], which provides an unprecedented opportunity for delivering information to reach large segments of the population [14] as well as hard-to-reach subpopulations [15,16]. Data from social networks such as Twitter, Instagram, and YouTube that allow users to discuss topics of their choice "unprimed by a researcher and without instrument bias" [10] can be used to capture and describe the social and environmental context in which individuals experience and describe their health conditions and concerns [17].

Twitter

Based on Pew Research data from 2019, nearly a quarter (22%) of adults in the United States use the social network Twitter; 40% of those are daily users [13]. Twitter allows users to post "tweets", short posts that are limited to 280 characters [18]. Users can search for any public tweet and engage with it through "like," "reply," and "retweet" (repost). Twitter is primarily public. Basic account information such as profile username, description, and location remains public. However, users can choose to keep their tweets protected to make them private or visible to subsets of users such as their followers or those they decided to follow [19,20]. Due to the more public nature of Twitter, previous research suggested that Twitter provides a "rich and promising avenue for exploring how patients conceptualize and communicate about their specific health issues" [21]. The increasing use of Twitter among the members of communities with disease is further evidenced by the abundance of disease-specific and health-related hashtags used in the tweets [22-24]. A hashtag is a word or phrase preceded by a hash or pound sign (#), which is used to identify tweets on a specific topic (eg, #lupus, #spoonies). These hashtags are used by users to assign their tweets to a topic and join ongoing conversations. Users can click on a hashtag and view all of the tweets that include the same hashtag; hence, discuss the same topic. This allows users to form online communities and share their health concerns, disease experience, and questions with other users [25]. However, there is little information about the use of social media among patients with lupus.

Previous Research on Social Media and Lupus

The emergence of social media has created new sources of analyzable data [12] and led to new research fields, such as infodemiology and infoveillance [11]. The data social media users generate through their online activities is referred to as their digital footprint [26] or social mediome [27].

Previous research examined user-generated content about lupus on Facebook [28]. Hale et al [28] looked at the representation of health conditions and found that lupus-related pages ranked the highest for patient support. Additionally, a patient commentary highlighted social media use (Twitter, in particular) by patients with lupus to find rheumatologists, specialist care, and peers and to build awareness of their health needs and experiences [29]. Health surveillance researchers have used

Twitter data to gain insights into the public perspectives on a variety of diseases and health topics such as influenza, autism, schizophrenia, smoking, and HIV/AIDS [30-35]. In some cases, social media user data demonstrated a correlation between the disease prevalence and frequency with which Twitter users discussed that disease [36]. To our knowledge, there are no studies that have leveraged Twitter to gain a better understanding of the perspectives of patients with lupus on reproductive health issues.

Study Objective and Research Questions

The objective of this study is to conduct a content analysis of tweets published in English by users in the United States during the period from September 1, 2017, to October 31, 2018, and to examine the perspectives of patients with lupus on reproductive health issues. We intend to answer the following research questions outlined in Textbox 1.

Our findings will shed light on whether Twitter is a promising data source for garnering insights about reproductive health concerns among the patients with lupus. The data will also help determine whether Twitter can serve as a potential outreach platform for raising awareness of lupus and reproductive health and for implementing relevant health interventions.

Textbox 1. Research questions.

- What is the volume of Twitter users who talk about lupus and reproductive health issues such as pregnancy prevention; pregnancy termination; and planning, conception, and management of pregnancy?

- How many of these users are patients with lupus?

- What are the perspectives, issues, and concerns that the patients with lupus express regarding their reproductive health?

- What are the demographics (ie, gender, race/ethnicity) of these patients with lupus on Twitter?

Methods

Data Collection

This qualitative study will analyze user-generated posts that include keywords related to lupus and fertility from the social network Twitter.

Data Source

To access public Twitter user data, we used Symplur Signals [37], a health care social media analytics company that maintains the largest publicly available database of health care– and disease-related conversations with the globally recognized Healthcare Hashtag Project. Symplur Signals extracts data from the Twitter representational state transfer (REST) application programming interface (API) and makes those available to researchers; those data are commonly used in peer-reviewed research [22,23,38-41]. We extracted data from Twitter using Symplur Signals user interface, searching for the relevant keywords and hashtags (Multimedia Appendix 1) from September 1, 2017, to October, 31, 2018. The data were provided in a spreadsheet, which we analyzed on local computers.

Search Filters

We utilized the framework suggested by Kim et al [42] for data collection, quality assessment, and reporting of standards. Twitter posts containing lupus-related terms were obtained for the period ranging from September 1, 2017, to October 31, 2018. The list of terms we used to collect the sample of tweets is shown in Multimedia Appendix 1. These terms can appear in the post or in an accompanying hashtag, for example, lupus or #LupusChat. LupusChat is a global health organization based in New York City, founded in 2012 by Tiffany Marie Peterson, a patient advocate who was diagnosed with SLE. The biweekly Twitter chat hosted by LupusChat is popular among patients with lupus to discuss related health concerns and the impact lupus has on their lives [43]. The selected keyword and hashtags are based on expert knowledge from clinicians and social media experts as well as on a systematic search of topic-related language using the Symplur Signals database. For each term, we viewed about 50 tweets to determine inactive as well as new keywords and hashtags that were being used in the lupus-related posts, particularly by patients. We will analyze the tweets from the patients with lupus to identify the issues and concerns they express regarding their reproductive health. Previous research has identified multiple challenges experienced by patients with SLE, for example, fertility preservation, optimal care during pregnancies, risks of adverse maternal or fetal outcomes, safety of contraceptive methods for women, and effects of dermatologic medications on male fertility [44-47].

Data Cleaning

The following types of posts were excluded: (1) non-English language tweets (which were identified using the methodology by Lui and Baldwin [48] and the language detection API of detectlanguage.com), (2) retweets that were originally composed/posted by other users, and (3) tweets that originated from outside the United States. We did not include retweets in the analysis dataset, as we intend to examine the patients' original perspectives on reproductive health issues. The locations of the users were determined using a mapped location filter as defined using "Profile Geo 2.0" algorithm (Gnip Inc) [49]. The algorithm uses a number of data points to determine a user's location, including the self-reported "Location" in the user profile and geotracking data, if available.

Furthermore, we relied on machine learning to recognize tweets by social bots or marketing-oriented accounts that could possibly influence the results and introduce bias [50,51]. Automated accounts on Twitter created by industry groups and private companies contribute to the corpus of Twitter data to influence discussions and promote specific ideas or products [28]. To identify those bias accounts, we identified a user account

XSL•FO

RenderX

responsible for each tweet collected in the dataset and analyzed its recent history, interactions, and metadata to determine the account was a social bot, a computer algorithm designed to automatically produce content and engage with humans on Twitter [50]. Tweets from these accounts "pollute social and health research data sets" [52]. They were identified and excluded from the dataset of tweets from patients with lupus. Bot accounts were identified using a system that analyzes the account's network (diffusion patterns), user (metadata), friends (account's contacts), temporal pattern (tweet rate), and sentiment (content of message), as previously described. The system detects bots with a 95% success rate [50].

Data Analysis

Coding

Two independent team members will be responsible for coding based on a set of a priori classifiers listed in Multimedia Appendices 2 and 3. We will use the profile information (ie, username, description, and profile image) of a Twitter account, which generated a relevant post, to characterize its user and determine if that user is a patient with lupus (Multimedia Appendix 3). Specifically, we will check if these users self-identify as patients with lupus in their profile description.

We will then code the tweets from patients with lupus (Multimedia Appendix 2). A tweet will be classified as the one by a patient with lupus—if that user has already been identified as such through examination of their Twitter profile or if the tweet describes lupus symptoms or lupus-related events in the first person (eg, My doctor had to change my medications today to the ones that are safe in pregnancy).

Additionally, we will code the person's gender and race/ethnicity if the profile contains sufficient information to do so. Cohen's kappa will be calculated for each code category to assess interrater reliability [53,54]. Once we establish concordance in the coder's classification with $\kappa > 0.8$ for each coding category, the remaining data will be divided between the 2 coders. Principal investigators of the project will help establish consensus in instances where coders disagree.

Statistical Analysis

The analysis will rely on public, anonymized data and will adhere to the terms and conditions, terms of use, and privacy policies of Twitter. This study will be conducted under the approval from the institutional review board of the authors' university. No tweets will be reported verbatim in the findings to protect the privacy of the users. Representative examples of tweets within each category will be selected to illustrate additional themes and will be shown as paraphrased quotes.

We will use descriptive statistics to identify the most prevalent topics in the Twitter content. Units of analysis will be unique terms in tweets, number of tweets, and number of users with lupus. For each analysis, we will present the findings in a confusion matrix, where diagonal lines would indicate the prevalence of a topic and off-diagonal lines, a topic overlap. The number of posts containing 2 or more topics would be found at the intersection of the matrix for these topics. We will further

describe the patient characteristics focusing on gender and race/ethnicity, as reported on Twitter.

Data Privacy and Confidentiality

Study data will be stored using the Research Electronic Data Capture (REDCap) system at the University of Southern California (USC). REDCap is a secure, web-based application designed to support data capture for research studies [55]. It provides (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for importing data from external sources. This database system facilitates the required provision of data to the USC Institutional Review Board, National Institutes of Health (NIH), and Food and Drug Administration (FDA).

Usernames will be initially available to the coders when they are examining the profiles to record the user demographics and determine whether a user is a patient with lupus. Profile usernames will then be redacted from the data file and replaced with unique numeric code identifiers before coders start examining the tweets. The link between the unique codes and the identifiable elements will be kept in a separate file. Thus, the coders will not be able to simultaneously view the identifiable elements of a Twitter profile and tweets made by that Twitter user. Additionally, any identifying and personal health information that the coders might find in the dataset of the tweets will be redacted by the coders. We will retain the data only for use in this project and destroy the identifiable information (tweet ID, tweet URL, thumbnail/URL of profile picture, username, and display name) prior to the data analysis. Given the sensitive nature of the topic "lupus and fertility," this step will be taken to protect the privacy of pregnant women whose tweets might be included in the data sample.

Risk Analysis

This research has minimal risk. We will use publicly available data from the social network Twitter. Identifiable information such as human subjects' names and Twitter usernames will not be included in the analysis dataset. We will further abide by the USC Institutional Review Board regulations and the USC Privacy of Personal Information policy. All data will be entered into a password-protected computer database. The data will be stored using appropriate secure computer software and encrypted computers.

Dissemination of Study Findings

The authors plan to publish the study findings in a peer-reviewed journal and present those at relevant conferences (to be determined at a later date). All the listed authors and contributors comply with the guidelines of the International Committee of Medical Journal Editors on author inclusion in a published work.

Results

Study approval was obtained from the Institutional Review Board at USC (Protocol HS-18-00912) (Multimedia Appendix 4). Data extraction and cleaning are complete. We obtained 47,715 tweets containing terms related to "lupus" from users

in the United states that were posted in English during the period September 1, 2017, to October 31, 2018. We will include 40,885 posts in the analysis. The detailed data extraction and cleaning flowchart is included in Multimedia Appendix 5. Data analysis will be completed in fall 2020.

Discussion

Limitations

This exploratory pilot study is limited to Twitter conversations from the patients of lupus who use the words lupus and SLE or the related hashtags in their tweets. As a result, tweets that share lupus-related experiences of patients without using the related terms and hashtags will be excluded from the study.

We recognize that this social media research and intervention favor those with the internet access and that this limitation could lead to potential bias in the research data. The generalizability of this study is also somewhat limited because the study excludes tweets from outside of the United States and tweets written in languages other than English. However, social media users "have grown more representative of the broader population." Twitter is used by 24% of Black Americans, 21% of White Americans, and 25% of Hispanic Americans. Twitter use is more common among younger (38% use among persons aged 18 to 29 years vs 7% use among those older than 65 years); educated (32% among college graduates vs 13% among those with a high school diploma or less); and urban (26% urban users vs 13% rural users) demographic [13].

Practical Significance

This pilot project will provide preliminary data and an insight into the application of publicly available Twitter data to gain a better understanding of the patients with lupus and their perspectives on reproductive health issues. If successful, our findings will shed light on whether Twitter provides a promising data source for garnering perspectives on reproductive health issues expressed by the patients with lupus. The data will also help to determine whether Twitter can be a potential outreach platform for raising awareness of lupus and reproductive health and for implementing the related health interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keywords and hashtags used for the Twitter search to assess Twitter conversations about lupus and reproductive health.
[PDF File (Adobe PDF File), 82 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Code categories to identify main themes in Twitter posts about lupus and reproductive health.
[PDF File (Adobe PDF File), 54 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Code categories to classify Twitter users.
[PDF File (Adobe PDF File), 43 KB-Multimedia Appendix 3]

Multimedia Appendix 4

IRB approval notice.
[PDF File (Adobe PDF File), 817 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Data extraction and cleaning flow diagram.
[PDF File (Adobe PDF File), 54 KB-Multimedia Appendix 5]

References

1. Lupus Basics. Centers for Disease Control and Prevention. 2018 Oct 17. URL: https://www.cdc.gov/lupus/basics/ [accessed 2019-07-09]

XSL•FO

RenderX

2. Updyke K, Urso B, Beg S, Solomon J. Developing a Continuous Quality Improvement Assessment Using a Patient-Centered Approach in Optimizing Systemic Lupus Erythematosus Disease Control. Cureus 2017;9(10):e1762. [doi: 10.7759/cureus.1762] [Medline: 29226052]

3. Lim SS, Drenkard C. The epidemiology of lupus. In: Wallace DJ, Hahn BH, editors. Dubois' Lupus Erythematosus and Related Syndromes. Edinburgh: Elsevier; 2019:23-43.

4. Lim SS, Bayakly AR, Helmick CG, Gordon C, Easley KA, Drenkard C. The Incidence and Prevalence of Systemic Lupus Erythematosus, 2002-2004: The Georgia Lupus Registry. Arthritis & Rheumatology 2014 Jan 27;66(2):357-368. [doi: 10.1002/art.38239] [Medline: 24504808]

5. Sammaritano LR, Chakravarty RF. Pregnancy and autoimmune disease, reproductive and hormonal issues. In: Wallace D, Hahn B, editors. Dubois' Lupus Erythematosus and Related Syndromes 9th Edition. United States: Elsevier; Oct 18, 2018:499-519.

6. Kartoz CR. Reproductive Health Concerns in Women with Systemic Lupus Erythematosus. MCN Am J Matern Child Nurs 2015;40(4):220-6; quiz E15. [doi: 10.1097/NMC.0000000000000145] [Medline: 25798747]

7. Dizon DS, Graham D, Thompson MA, Johnson LJ, Johnston C, Fisch MJ, et al. Practical guidance: the use of social media in oncology practice. J Oncol Pract 2012 Sep;8(5):e114-e124 [FREE Full text] [doi: 10.1200/JOP.2012.000610] [Medline: 23277774]

8. Obar JA, Wildman S. Social media definition and the governance challenge: An introduction to the special issue. Telecommunications Policy 2015 Oct;39(9):745-750. [doi: 10.1016/j.telpol.2015.07.014]

9. Lober WB, Flowers JL. Consumer empowerment in health care amid the internet and social media. Semin Oncol Nurs 2011 Aug;27(3):169-182. [doi: 10.1016/j.soncn.2011.04.002] [Medline: 21783008]

10. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 9;311(14):1399-1400. [doi: 10.1001/jama.2014.1505] [Medline: 24577162]

11. Zeraatkar K, Ahmadi M. Trends of infodemiology studies: a scoping review. Health Info Libr J 2018 Jun;35(2):91-120. [doi: 10.1111/hir.12216] [Medline: 29729073]

12. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Dec;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]

13. Social Media Fact Sheet. Pew Research Center. 2019 Jun 12. URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-12-24]

14. Carson KV, Ameer F, Sayehmiri K, Hnin K, van AJE, Sayehmiri F, et al. Mass media interventions for preventing smoking in young people. Cochrane Database Syst Rev 2017 Dec 02;6:CD001006. [doi: 10.1002/14651858.CD001006.pub3] [Medline: 28574573]

15. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]

16. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]

17. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]

18. Tsukayama H. Twitter is officially doubling the character limit to 280. The Washington Post website.: The Washington Post; 2017 Nov 07. URL: https://www.washingtonpost.com/news/the-switch/wp/2017/11/07/twitter-is-officially-doubling-the-character-limit-to-280/?utm_term=.eb2d65ecfe26 [accessed 2020-08-09]

19. Twitter privacy policy. Twitter. URL: https://twitter.com/en/privacy [accessed 2019-12-24]

20. Twitter terms of service. Twitter. URL: https://twitter.com/en/tos [accessed 2019-12-24]

21. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging Social Media to Promote Public Health Knowledge: Example of Cancer Awareness via Twitter. JMIR Public Health Surveill 2016 Apr;2(1):e17 [FREE Full text] [doi: 10.2196/publichealth.5205] [Medline: 27227152]

22. Katz M, Utengen A, Anderson P, Thompson M, Attai D, Johnston C, et al. Disease-Specific Hashtags for Online Communication About Cancer Care. JAMA Oncol 2016 Mar;2(3):392-394. [doi: 10.1001/jamaoncol.2015.3960] [Medline: 26539640]

23. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]

24. Katz MS, Anderson PF, Thompson MA, Salmi L, Freeman-Daily J, Utengen A, et al. Organizing Online Health Content: Developing Hashtag Collections for Healthier Internet-Based People and Communities. JCO Clin Cancer Inform 2019 Jun;3:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00124] [Medline: 31251658]

25. Moreno MA, D'Angelo J. Social Media Intervention Design: Applying an Affordances Framework. J Med Internet Res 2019 Mar 26;21(3):e11014 [FREE Full text] [doi: 10.2196/11014] [Medline: 30912754]

26. Zhang D, Guo B, Li B, Yu Z. Extracting social and community intelligence from digital footprints: an emerging research area. In: Yu Z, Liscano R, Chen G, Zhang D, Zhou X, editors. Ubiquitous intelligence and computing. Berlin: Springer; 2010:4-18.

27. Asch DA, Rader DJ, Merchant RM. Mining the social mediome. Trends Mol Med 2015 Sep;21(9):528-529 [FREE Full text] [doi: 10.1016/j.molmed.2015.06.004] [Medline: 26341614]

28. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014 Aug;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]

29. Greene A. Patient commentary: social media provides patients with support, information, and friendship. BMJ 2015 Feb 10;350:h256. [doi: 10.1136/bmj.h256] [Medline: 25670190]

30. Wagner M, Lampos V, Cox IJ, Pebody R. The added value of online user-generated content in traditional methods for influenza surveillance. Sci Rep 2018 Sep 18;8(1):13963 [FREE Full text] [doi: 10.1038/s41598-018-32029-6] [Medline: 30228285]

31. Hswen Y, Gopaluni A, Brownstein JS, Hawkins JB. Using Twitter to Detect Psychological Characteristics of Self-Identified Persons With Autism Spectrum Disorder: A Feasibility Study. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e12264 [FREE Full text] [doi: 10.2196/12264] [Medline: 30747718]

32. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study. JMIR Ment Health 2018 Dec 13;5(4):e11483 [FREE Full text] [doi: 10.2196/11483] [Medline: 30545811]

33. Malik A, Li Y, Karbasian H, Hamari J, Johri A. Live, Love, Juul: User and Content Analysis of Twitter Posts about Juul. Am J Health Behav 2019 Mar 01;43(2):326-336. [doi: 10.5993/AJHB.43.2.9] [Medline: 30808472]

34. Nielsen RC, Luengo-Oroz M, Mello MB, Paz J, Pantin C, Erkkola T. Social Media Monitoring of Discrimination and HIV Testing in Brazil, 2014-2015. AIDS Behav 2017 Jul;21(Suppl 1):114-120 [FREE Full text] [doi: 10.1007/s10461-017-1753-2] [Medline: 28349220]

35. Bychkov D, Young S. Social media as a tool to monitor adherence to HIV antiretroviral therapy. J Clin Transl Res 2018 Dec 17;3(Suppl 3):407-410 [FREE Full text] [Medline: 30873489]

36. Tufts C, Polsky D, Volpp KG, Groeneveld PW, Ungar L, Merchant RM, et al. Characterizing Tweet Volume and Content About Common Health Conditions Across Pennsylvania: Retrospective Analysis. JMIR Public Health Surveill 2018 Dec 06;4(4):e10834 [FREE Full text] [doi: 10.2196/10834] [Medline: 30522989]

37. Symplur Signals. Symplur. URL: https://www.symplur.com/product/symplur-signals/ [accessed 2020-08-09]

38. Utengen A, Rouholiman D, Gamble JG, Grajales FJ, Pradhan N, Staley AC, et al. Patient Participation at Health Care Conferences: Engaged Patients Increase Information Flow, Expand Propagation, and Deepen Engagement in the Conversation of Tweets Compared to Physicians or Researchers. J Med Internet Res 2017 Aug 17;19(8):e280 [FREE Full text] [doi: 10.2196/jmir.8049] [Medline: 28818821]

39. Parwani P, Choi AD, Lopez-Mattei J, Raza S, Chen T, Narang A, et al. Understanding Social Media: Opportunities for Cardiovascular Medicine. J Am Coll Cardiol 2019 Mar 12;73(9):1089-1093 [FREE Full text] [doi: 10.1016/j.jacc.2018.12.044] [Medline: 30846102]

40. Cumbraos-Sánchez MJ, Hermoso R, Iñiguez D, Paño-Pardo JR, Allende Bandres M, Latorre Martinez MP. Qualitative and quantitative evaluation of the use of Twitter as a tool of antimicrobial stewardship. Int J Med Inform 2019 Nov;131:103955. [doi: 10.1016/j.ijmedinf.2019.103955] [Medline: 31487575]

41. Venuturupalli RS, Sufka P, Bhana S. Digital Medicine in Rheumatology: Challenges and Opportunities. Rheum Dis Clin North Am 2019 Feb;45(1):113-126. [doi: 10.1016/j.rdc.2018.09.010] [Medline: 30447740]

42. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. J Med Internet Res 2016;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]

43. LupusChat. URL: https://www.lupuschat.org/ [accessed 2020-08-09]

44. McDonald EG, Bissonette L, Ensworth S, Dayan N, Clarke AE, Keeling S, et al. Monitoring of Systemic Lupus Erythematosus Pregnancies: A Systematic Literature Review. J Rheumatol 2018 Oct;45(10):1477-1490 [FREE Full text] [doi: 10.3899/jrheum.171023] [Medline: 30008450]

45. Andreoli L, Bertsias GK, Agmon-Levin N, Brown S, Cervera R, Costedoat-Chalumeau N, et al. EULAR recommendations for women's health and the management of family planning, assisted reproduction, pregnancy and menopause in patients with systemic lupus erythematosus and/or antiphospholipid syndrome. Ann Rheum Dis 2017 Mar;76(3):476-485 [FREE Full text] [doi: 10.1136/annrheumdis-2016-209770] [Medline: 27457513]

46. Culwell KR, Curtis KM, del Carmen Cravioto M. Safety of contraceptive method use among women with systemic lupus erythematosus: a systematic review. Obstet Gynecol 2009 Aug;114(2 Pt 1):341-353. [doi: 10.1097/AOG.0b013e3181ae9c64] [Medline: 19622996]

47. Zakhem GA, Motosko CC, Mu EW, Ho RS. Infertility and teratogenicity after paternal exposure to systemic dermatologic medications: A systematic review. J Am Acad Dermatol 2019 Apr;80(4):957-969. [doi: 10.1016/j.jaad.2018.09.031] [Medline: 30287313]

XSL·FO
RenderX

48. Lui M, Baldwin T. langid.py: An Off-the-shelf Language Identification Tool. In: Proceedings of the ACL 2012 System Demonstrations.: Association for Computational Linguistics; 2012 Jul Presented at: 50th Annual Meeting of the Association for Computational Linguistics; 2012; Jeju Island, Korea p. 25-30 URL: https://www.aclweb.org/anthology/P12-3005/

49. Profile Geo 2.0. GNIP. URL: http://support.gnip.com/enrichments/profile_geo.html [accessed 2020-08-09]

50. Davis C, Varol O, Ferrara E, Flammini A, Menczer F. Botornot: A system to evaluate social bots. In: WWW'16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web. Geneva: International World Wide Web Conferences Steering Committee; 2016 Apr Presented at: WWW'16: 25th International World Wide Web Conference; April 2016; Montréal Québec Canada p. 273-274 URL: https://dl.acm.org/doi/proceedings/10.5555/2872518 [doi: 10.1145/2872518.2889302]

51. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Communications of the ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]

52. Allem J, Ferrara E. The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors. J Med Internet Res 2016 Aug 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]

53. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]

54. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]

55. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

Abbreviations

API: Application Programming Interface
FDA: Food and Drug Administration
NCATS: National Center for Advancing Translational Science
NIH: National Institutes of Health
REDCap: Research Electronic Data Capture
REST: Representational State Transfer
SLE: systemic lupus erythematosus
USC: University of Southern California

XSL•FO
RenderX