

Original Paper

Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study

Keisuke Nakagawa^{1,2}, MD; Nuen Tsang Yang³, MS; Mabelle Wilson³, PhD; Peter Yellowlees¹, MBBS, MD

¹Department of Psychiatry and Behavioral Sciences, University of California, Davis School of Medicine, Sacramento, CA, United States

²Digital CoLab, Innovation Technology, University of California, Davis Health, Sacramento, CA, United States

³Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, Sacramento, CA, United States

Corresponding Author:

Keisuke Nakagawa, MD

Department of Psychiatry and Behavioral Sciences

University of California, Davis School of Medicine

2230 Stockton Blvd

Sacramento, CA, 95817

United States

Phone: 1 916 397 8941

Email: drknakagawa@ucdavis.edu

Abstract

Background: Physicians are increasingly using Twitter as a channel for communicating with colleagues and the public. Identifying physicians on Twitter is difficult due to the varied and imprecise ways that people self-identify themselves on the social media platform. This is the first study to describe a reliable, repeatable methodology for identifying physicians on Twitter. By using this approach, we characterized the longitudinal activity of US physicians on Twitter.

Objective: We aimed to develop a reliable and repeatable methodology for identifying US physicians on Twitter and to characterize their activity on Twitter over 5 years by activity, tweeted topic, and account type.

Methods: In this study, 5 years of Twitter data (2016-2020) were mined for physician accounts. US physicians on Twitter were identified by using a custom-built algorithm to screen for physician identifiers in the Twitter handles, user profiles, and tweeted content. The number of tweets by physician accounts from the 5-year period were counted and analyzed. The top 100 hashtags were identified, categorized into topics, and analyzed.

Results: Approximately 1 trillion tweets were mined to identify 6,399,146 (<0.001%) tweets originating from 39,084 US physician accounts. Over the 5-year period, the number of US physicians tweeting more than doubled (ie, increased by 112%). Across all 5 years, the most popular themes were general health, medical education, and mental health, and in specific years, the number of tweets related to elections (2016 and 2020), Black Lives Matter (2020), and COVID-19 (2020) increased.

Conclusions: Twitter has become an increasingly popular social media platform for US physicians over the past 5 years, and their use of Twitter has evolved to cover a broad range of topics, including science, politics, social activism, and COVID-19. We have developed an accurate, repeatable methodology for identifying US physicians on Twitter and have characterized their activity.

(*J Med Internet Res* 2022;24(9):e37752) doi: [10.2196/37752](https://doi.org/10.2196/37752)

KEYWORDS

social media; internet; health informatics; internet use; public health; Twitter; physician

Introduction

Twitter has become a popular communication tool, with almost 200 million daily active users and more than 500 million tweets being sent out every day [1]. Around the world, people have used Twitter as a platform for sharing information, expressing opinions, and engaging in social movements. It has democratized

communication, allowing everyday citizens to have a voice at the same potential scale as the voices of global leaders, politicians, and celebrities. The speed at which information is shared on Twitter has transformed the way information spreads and how communities create social movements. Studies have shown that Twitter is one of the best sources of real-time

information, outpacing traditional media outlets for reporting natural disasters, crimes, and major events [2-6].

Over the past 15 years, physicians have started to use Twitter as a platform for sharing scientific information, frontline experiences, and opinions on various topics [7-12]. Studies have shown that scientific papers obtain more visibility when shared on Twitter, and impact metrics have started to incorporate social media data as part of their calculations [13]. During the COVID-19 pandemic, Twitter has given unprecedented visibility to clinicians working on the front lines. Clinicians have engaged in public discourse by sharing data and scientific information on Twitter [14-17]. They have also engaged in dialogue with each other, often sharing encouragement, empathy, and compassion during challenging times [18,19].

Studies have already demonstrated the use of Twitter in building early warning systems for adverse drug reactions, understanding discussions about various diseases, and characterizing public perceptions regarding COVID-19 [20-24]. One of the earliest uses of Twitter among physicians was tweeting during conferences, and a number of studies have analyzed what factors make tweets more successful or more likely to be disseminated during medical conferences [9,25]. Recent studies have demonstrated that physician activity on Twitter is a potential predictor of COVID-19 surges [26]. As more physicians join Twitter and engage in dialogue across a wide range of topics, health systems and academic medical centers that train and often employ these physicians need to be aware of these trends to teach them about professionally appropriate use and potential repercussions. Physicians should be aware of the positive and negative consequences of engaging in public discourse on platforms such as Twitter, where content is instantly made public and is often permanent and irreversible [23,27].

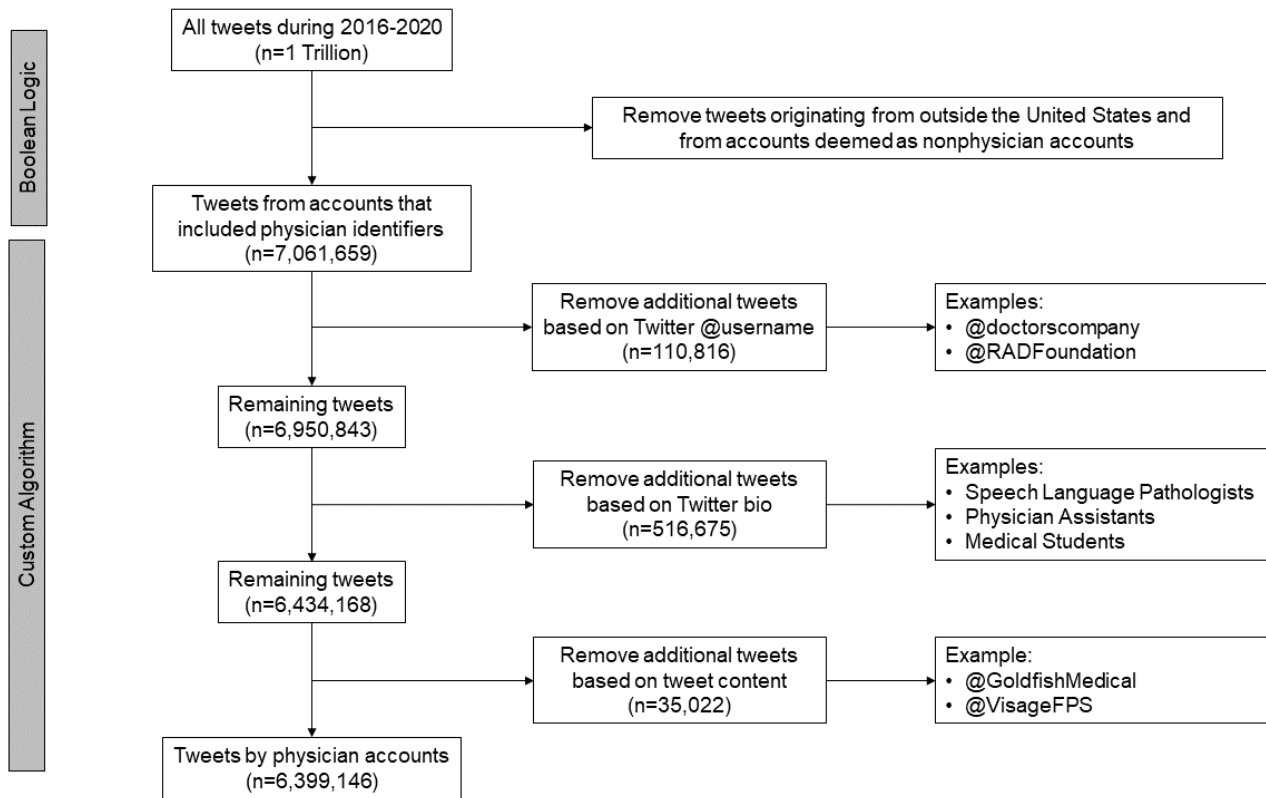
Although some studies have investigated the use of Twitter among physicians in narrow use cases, such as a specific specialty, no study has characterized the overall activity of US physicians on Twitter across multiple years [11]. Part of this can be attributed to the difficulty of identifying physicians on Twitter. Physicians self-report their professional status in varied, inconsistent, and sometimes vague ways, and there are limited methods for verifying and validating the identity of a user based on their profile and tweet content. Furthermore, the challenges of combing through billions of tweets and bios to identify physician accounts can be computationally and logistically intensive. It is for these reasons that a process for identifying US physicians and characterizing their activity has not been published, to our knowledge. Given the widespread use of Twitter by the general public and among physicians, it is becoming increasingly important to understand the use of social media among physicians, particularly with the backdrop of the COVID-19 pandemic. The aims of this study are to describe a reliable and repeatable methodology for identifying US physicians on Twitter and to characterize their use of Twitter and longitudinal trends across multiple years.

Methods

Process for Identifying US Physicians on Twitter

We conducted an *in silico* analysis of physicians' Twitter usage over a 5-year period. We identified physicians by using a 3-stage process (Figure 1) comprising a geographic filter for excluding accounts from outside of the United States, Boolean logic for further identifying US physician accounts based on their usernames and bios, and a custom algorithm for filtering out false-positive accounts. The details of the algorithm developed and the iterative refinement process are described in the following sections.

Figure 1. Identification and selection of physician tweets.



Algorithm Development and Refinement for Physician Identification

Initially, a Boolean logic was developed (Multimedia Appendix 1) to scan all Twitter usernames (ie, @username), names (eg, Jane Doe, M.D.), and bios for every tweet that was posted during the 5-year study period (2016-2020). The Boolean logic identified all tweets that included some indication of a Twitter user’s status as a physician (eg, mentioning physician in the bio or including a medical degree, such as M.D. or D.O., in the name or username). The Boolean logic included a corpus that contained every specialty and subspecialty listed by the

Accreditation Council for Graduate Medical Education and the American Osteopathic Association. A third-party software (Meltwater Explore; Meltwater, Inc) was used to review the user accounts and tweets generated from the Boolean algorithm to ensure the accuracy of the Boolean logic expression. With each iteration, the first 1000 tweets were manually reviewed for any tweets that did not originate from a physician-owned account. Appropriate logic was added to filter out false positives. Examples of false positives are provided in Table 1. This manual review process was repeated until the false-positive rate decreased to less than 5% in a 1000-tweet sample.

Table 1. Examples of false positives and the associated Boolean logic.

Description	Examples	Boolean logic
<i>MD</i> is used in the username, in the profile, or in both to indicate something other than a medical degree.	<ul style="list-style-type: none"> • <i>MD</i> in “@MDZeki” stands for “Muhammad” • <i>MD</i> in “Student living in MD” and “@MDJobs” stands for “Maryland” 	<ul style="list-style-type: none"> • Only identify tweets that are in English and from the United States • Only include usernames that have “,M.D.,” “,MD,” or “, MD” • Remove all accounts that include <i>jobs</i> in the username or profile
<i>DO</i> is used in the username, in the profile, or in both to indicate something other than a medical degree.	<ul style="list-style-type: none"> • <i>DO</i> in “@JUSTDOIT” does not indicate a DO medical degree 	<ul style="list-style-type: none"> • Only include names that have “,D.O.”
<i>Doctor</i> is used broadly without specifying an MD or DO degree.	<ul style="list-style-type: none"> • @DoctorJones • @aprokodoctor • @DoctorsEMres 	<ul style="list-style-type: none"> • The Boolean logic did not select handles with <i>Doctor</i>; the user would have had to indicate their status as a physician through more specific ways
Premedical and medical students who include <i>physician</i> or <i>doctor</i> in their username or profile.	<ul style="list-style-type: none"> • Profile includes “Aspiring Infectious Disease Physician” • Profile includes “aspiring physician working in epi & vaccine research” • Profile includes “future physician” or “Future Physician Assistant” 	<ul style="list-style-type: none"> • Remove all accounts that include <i>aspiring</i> or <i>future</i> in the profile

Data Acquisition

Once the Boolean logic was finalized, all tweets that were written in English; originated from the United States; were posted between January 1, 2016, to December 31, 2020; and met the Boolean criteria were acquired from a third-party vendor (Meltwater, Inc, San Francisco, California).

Algorithm Refinement

The data were imported into RStudio (Version 1.2.5033; RStudio, PBC), where additional rules were programmed in R (R Foundation for Statistical Computing) and applied to filter out nonphysician Twitter accounts. The next phase of algorithm refinement involved filtering out more nuanced cases that the Boolean logic could not catch. Accounts that used *MD* to indicate the state of Maryland; accounts of physician recruiting services that used variations of *MD* and *jobs* in their usernames and profiles; and premedical and medical students who used *aspiring MD*, *aspiring physician*, *future physician*, or similar terms in their profile were filtered out by using logic coded in R.

The algorithm was further refined by identifying patterns in the tweet text itself and then applying filtering logic to the accounts of those tweets. For example, to filter out more physician recruitment accounts, tweets that used the hashtags #*job* or #*jobs* in the text were identified. Among these tweets, any accounts that included the word *medical* or *surgery* in the account name were eliminated (eg, *Goldfish Medical* or *Visage Facial Plastic Surgery*). This 2-step process of screening accounts—analyzing the tweet text first and then identifying false-positive usernames and names—was used to filter out the final batch of nonphysician accounts.

Algorithm Accuracy and Validation

We used Twitter’s verified accounts (ie, those with a blue badge) as the validation set to test the accuracy of the algorithm. Twitter

uses a rigorous verification process to confirm the identities of verified account users, which includes checking for an official government-issued identification document or an official email address [28].

All of the tweets that originated from a verified Twitter account were separated from the larger 5-year data set. A team manually reviewed each unverified account by visiting the associated Twitter profiles and searching the internet to confirm the details of a user’s status as a physician (eg, checking the website of their stated institution for their profile). A false-positive error rate was calculated based on the number of verified accounts that belonged to Twitter users who were not actually physicians. In order to calculate the false-positive rate of unverified accounts, around 5% (1978/43,328, 4.6%) of the unverified accounts in the data set were randomly selected. The accounts were manually reviewed by the team to calculate a false-positive rate for the unverified physician accounts identified by the algorithm.

Hashtag and Topic Analysis

The top 100 hashtags used by US physicians were identified, tallied, and ranked for each year. Each tweet was tokenized, stemmed, and normalized, and stop words were removed. This was done by using the following R packages: *stringi*, *stringr*, and *tokenizers*. All words that followed a pound sign were identified and separately stored within a data frame in RStudio. For every tweet, each hashtag was counted and tallied across all tweets for each year (2016-2020). Researchers (KN and NTY) reviewed the list and grouped any terms that were related. For example, the COVID-19 topic was comprised of hashtags such as #*covid19*, #*coronavirus*, #*sarscov2*, #*covid19coronavirus*, and other related hashtags. Further, both the #*blacklivesmatter* and #*blm* hashtags were used for the Black Lives Matter topic’s tally and rank.

We analyzed hashtags in two ways. First, we created word clouds for a visual representation of Twitter activity. A word cloud was generated for each year of the study period; words were color-coded based on their frequency of use. Word clouds were generated by using the following R packages: *wordcloud*, *wordcloud2*, and *ggplot2*. Second, we identified the top 100 hashtags and categorized them into broad thematic areas. In order to analyze thematic trends between years, hashtags were organized into the following themes: general health care, COVID-19 and public health, politics, social activism, mental health and well-being, health technology, conferences, patient groups, specialties and subspecialties, medical conditions and procedures, and medical education. For the general health care theme, tweets that included generic health care-oriented hashtags like *#healthcare*, *#doctor*, *#medtwitter*, and *#medicine* were combined and tallied, and the proportion of such tweets was calculated as a percentage of all tweets for each year. For the COVID-19 and public health theme, hashtags like *#covid19*, as well as hashtags that conveyed public health messages like *#maskup*, *#socialdistancing*, and *#flattenthecurve*, were included. For the politics theme, any hashtag related to politics and elections, such as *#trump*, *#vote*, *#election2020*, *#tcot* (ie, top conservatives on Twitter), and *#newsmax*, were used [29,30]. Only the top 100 hashtags for each year were considered.

Ethical Considerations

No application for an ethics review board assessment was submitted, since this study involved third-party data sets with no experimental activities. Therefore, this study was deemed a quality assurance/quality improvement activity [31].

Table 2. Twitter activity among US physicians from 2016 to 2020.

Twitter activity	2016	2017	2018	2019	2020
US physicians on Twitter					
Tweets per year, n	1,461,753	1,344,911	1,205,053	1,049,279	1,338,150
Unique accounts tweeting, n	12,675	15,633	17,934	20,584	26,897
New accounts tweeting, n	1711	2170	1917	1760	2215
Number of tweets per account, mean	115.3	86.0	67.2	51.0	49.8
Verified Twitter accounts^a					
Tweets per year, n	57,255	56,082	49,743	51,776	77,609
Unique accounts tweeting, n	107	132	155	171	208
New accounts tweeting, n	2	6	3	8	2
Number of tweets per account, mean	535.1	424.9	320.9	302.8	373.1

^aVerified accounts have a blue verified badge next to the Twitter user's name to let people know that an account is authentic and has undergone rigorous verification by Twitter. More information is available on the Twitter Help Center [28].

Hashtag and Topic Analysis

Among US physicians using hashtags in their tweets, *#health*, *#healthcare*, *#doctor*, *#medicine*, *#meded*, *#mentalhealth*, and *#wellness* consistently ranked among the most frequently used hashtags across all 5 years of the study period (Multimedia Appendix 2). In 2016, *#digitalhealth* and *#realdonaldtrump* were trending hashtags, while *#trump*, *#backfiretrump*, *#diversity*, and *#inclusion* were trending in 2018 and 2019.

Results

Algorithm Accuracy

Of the 216 Twitter-verified accounts that the Boolean algorithm identified as physician accounts, 12 were not physician accounts, resulting in an accuracy rate of 94.4% (204/216). Of the 43,328 unverified accounts in the data set (ie, regular Twitter accounts), 1978 accounts were randomly selected for manual review as part of the 5% validation sample. Of the 1978 randomly-selected unverified accounts, 204 were not physician accounts, resulting in an accuracy rate of 89.7% (1774/1978; 95% CI 88.3%-91%).

Twitter Activity

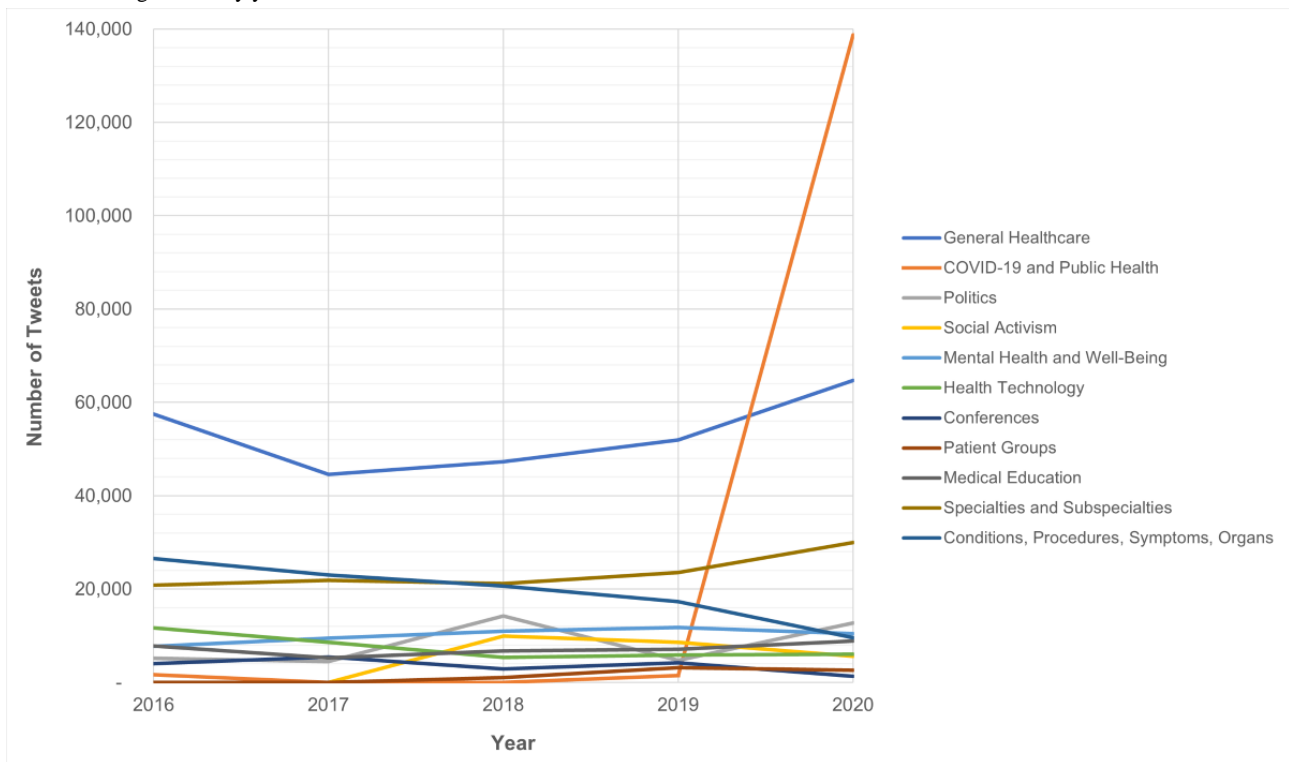
The number of US physicians who were active on Twitter increased from 2016 to 2020. The number of unique physician accounts tweeting increased from 12,675 in 2016 to 26,897 in 2020—a 112.2% increase in the number of active physician users over the 5-year period (Table 2). The number of new physician accounts per year also increased from 1711 new accounts in 2016 to 2215 new accounts in 2020—a 19.7% year-over-year increase in the number of new physician accounts created. Although the total number of US physician accounts increased, the total number of tweets varied across the 5-year period, with 1,461,753 total physician tweets in 2016 and 1,338,150 total physician tweets in 2020. The average number of tweets per account decreased from 115.3 tweets per physician account in 2016 to 49.8 tweets per physician account in 2020. These trends were also reflected in a subgroup analysis of verified Twitter accounts.

Hashtags related to *#blacklivesmatter* and *#covid19* represented 0.3% (3750/1,291,567) and 10.6% (137,107/1,291,567) of all hashtags used in 2020, respectively. Donald Trump-related hashtags were consistently ranked among the top 50 hashtags from 2016 to 2019, and in 2020, US physicians used COVID-19-related hashtags in 137,107 tweets, making them the most used hashtags during the entire 5-year study period (Multimedia Appendix 2), with *#health* being the second most used hashtag across all 5 years (used in 64,092 tweets).

The word clouds provide a visual representation of the frequency and distribution of the hashtags used each year. The frequency of words was color-coded; gray represents the most frequently used hashtags, green represents the least used hashtags, and other colors (light green, pink, purple, and orange) represent the different levels of frequency in between the two extremes. [Multimedia Appendix 3](#) shows the word clouds for tweets from 2016 to 2019. The word clouds for 2016 to 2019 show a broad distribution of words, as demonstrated by the wide variety of colors and the varying sizes of words in the word clouds ([Multimedia Appendix 3](#)). The 2020 word cloud shifts drastically, showing just 1 hashtag in gray (*#COVID-19*) and 1 hashtag in orange (*#medtwitter*); all other hashtags are in green (ie, the least used hashtags).

Once hashtags were organized into themes, the results showed that the general health care topic was the most consistently discussed topic across all 5 years and represented 5% (64,701/1,291,567;2020) of the total tweets in 2020 ([Figure 2](#)). In approximately 2% (29,964/1,291,567;2020) of tweets, US physicians discussed specialties and subspecialties, and 0.5% (8830; 1,291,567;2020) of tweets discussed medical education. Public health topics were discussed in less than 0.5% (1478/1,186,835; 2019) of the tweets from 2016 to 2019, but in 2020, US physicians discussed COVID-19 and related public health messages—the most dominant theme across all 5 years of the study period—in over 10% of all tweets.

Figure 2. Trending themes by year.



Discussion

Study Overview

This is the first longitudinal report of physician activity on Twitter. Physician activity on Twitter has been increasing over the past 5 years, with more physicians creating accounts and joining the platform every year. The diverse range of topics that are discussed on the platform, from social activism and politics to memes, demonstrate that Twitter is becoming a mainstream communication tool for everyday physicians.

The trends in the last 5 years likely indicate that the early adopters phase of Twitter use among physicians is sunsetting, and Twitter is becoming a mainstream platform for physicians [32]. This is supported by the decreasing average number of tweets per user, which shows that the fraction of power users, that is, users with a typically high number of tweets per year, is decreasing in the population of US physician users while use by mainstream physicians, who tend to have a lower number

of tweets per year, is increasing. The best example of this occurred in 2020, which was when the content topics drastically shifted toward the topic of COVID-19—a mainstream issue of interest among physicians and the public. Similar trends were seen during the election years (2016 and 2020) and around the time of the Black Lives Matter movement in 2020. If Twitter use among physicians was still in the early adopters phase, we would expect to see less volatile shifts in topics from year to year, since the interests of a niche group would be more stable and consistent.

We also want to acknowledge the role of misinformation and disinformation on the internet. Although this was not the focus of this study, this topic is inextricably linked to the roles and responsibilities that physicians have on Twitter, social media, and the internet. The study period included the 2016 US presidential election and the COVID-19 pandemic, of which both involved the sharing of misinformation and disinformation related to these events [33-36]. Physicians who are active on social media have a disproportionately large influence on the

public discourse around science, social movements, and politics, which is supported by the findings in this study. We anticipate that this trend will only continue to increase, placing more importance on studying and understanding the role that physicians play on social media in shaping society's perceptions and opinions.

Principal Findings

Physicians discuss a broad range of topics on Twitter and are not shy about opining on topics that are relevant to society beyond strictly medical topics. Physicians are unlikely to find an equivalent platform (traditional or social media) where ordinary clinicians' voices can be shared and heard globally in such a public, transparent way. This was illustrated by the outsized influence that physicians like Bob Wachter (@Bob_Wachter), physician influencers like Kevin Pho (@kevinmd), and ordinary clinicians have had by using Twitter during the COVID-19 pandemic. Among the hundreds of millions of tweets shared in 2020 [1], the 137,107 (<0.001%) physicians tweeting about COVID-19 may have had a disproportionately outsized impact on public dialogues and disease perceptions.

US physicians are increasingly using Twitter to teach and share medical education materials. Hashtags like #meded and #foamed (ie, free open-access medicine) were consistently ranked in the top 10 and top 100 hashtags, respectively (Multimedia Appendix 2). This demonstrates a growing shift toward the democratization of medical education through the sharing of materials, study sheets, and illustrations. This shift toward democratizing medical education is likely to continue, with Twitter and other social media platforms, such as Instagram, playing an increasing role in the future of medical education.

Tweeting during medical conferences continues to be a mainstay of use cases among US physicians, but it has become less prominent relative to more mainstream topics. The American Academy of Pediatrics, American Society of Clinical Oncology, and American College of Cardiology conference hashtags (eg, #aap18, #acc19, and #asco20, respectively) were ranked in the top 100 hashtags across the majority of the 5-year study period (Multimedia Appendix 2). Social media continues to be a powerful tool for disseminating scientific information through medical conferences and scientific journals.

Another major trend in Twitter activity was its growing use for marketing purposes. Hashtags like #healthcaremarketing, #medicalmarketing, and #plasticsurgery were consistently ranked in the top 50 hashtags (Multimedia Appendix 2). During the filtering process, a large group of accounts had to be removed that were associated with physician recruitment services using hashtags like #jobs in the profile, in the tweet text, or in both. We also observed individual physicians in certain specialties, such as plastic surgery, cosmetic surgery, and dermatology, using Twitter to market their practice, further reinforcing the fact that Twitter is now accepted as a mainstream platform that physicians can use to communicate directly with patient populations.

Finally, mental health and well-being were consistently ranked in the top 50 topics that US physicians discussed on Twitter.

Across a multitude of hashtags, including #mentalhealth, #wellness, #depression, and #mindfulness, US physicians used Twitter to engage in dialogue about wellness. These trends may be particularly useful for monitoring physicians' well-being at a local, regional, or national level, since the mental health of clinicians is notoriously challenging to measure on a regular basis. We hypothesize that during stressful events, such as the COVID-19 pandemic, Twitter may offer unprecedented insight into the overall mental health status of physicians, and we are following up this study with one involving a sentiment analysis of all physician tweets that have been posted during the pandemic.

Challenges, Strengths, and Limitations

During the planning of this study, it was clear to us how challenging it would be to identify physicians on Twitter, hence the lack of publications on this topic in the literature. Twitter was designed to be fast and easy with regard to sign-up and use, and the consequence of this focus on user experience is that there are very few denominators for identifying subgroups within Twitter. The most challenging task of this study was identifying US physician tweets (n=6,399,146) from the trillions of tweets generated worldwide.

Although numerous studies have analyzed the Twitter activity of the entire Twitter population, the complexity of this task increases significantly once subgroups need to be identified and analyzed. We believe that the methodology outlined in this study offers a robust, scalable, and repeatable method of identifying physician accounts on Twitter. We also recognize that this approach is not perfect and hope that this study serves as a foundation for future work to further improve the identification of physician accounts for future analyses.

The computational requirements for mining approximately 1 trillion tweets and analyzing millions of tweets should not be underestimated. Once the data set reaches beyond a few months of Twitter data, the data set is too large for manual assessment. As such, algorithmic approaches to filtering and analyzing the data are recommended. Analyzing this amount of data often exceeds the computational limits of a typical laptop or desktop, and more high-performance graphics processing units or high-performance computing clusters in the cloud are recommended to analyze the data in a reasonable amount of time, particularly if a real-time analysis is desired.

This study had 3 main limitations. First, all physicians who were identified as active users on Twitter were self-reported physicians. One's status as a physician can be very difficult to validate unless this information is cross-referenced against state licensing data. Physicians with verified accounts on Twitter (ie, accounts with blue badges) are the most reliable, since Twitter has a rigorous verification process that involves checking credible sources, such as health system websites and media sources, and using other means for verification [28]. Since activity trends were similar between the nonverified and verified accounts, we are reasonably confident that our approach produced a reliable cohort of physicians. Second, while Twitter's official verification process via the blue badge offers the most reliable method for validating physician accounts within pragmatic means, this approach has its limitations. Twitter has

accidentally verified fake accounts, has temporarily halted verifications multiple times to improve their internal processes, and has overall been slow to verify physicians—an issue Twitter has recently addressed during the COVID-19 pandemic [37-40]. Third, it is very likely that our report underestimates the number of active US physicians on Twitter because some physicians may not indicate their status as a clinician. Although we explored various methodologies for matching Twitter accounts, such as using state licensing data, these approaches were not practical for this study, and they will be considered for future investigations.

Future Directions

There are many opportunities to build on this study. We are planning future studies to focus on specific events, such as the COVID-19 pandemic. For example, by segmenting the physician tweets by state and correlating them with epidemiological data, we plan to study whether there was increased physician activity during pandemic surges. Similarly, machine learning and statistical analyses could be used to explore whether physicians' Twitter activity can be used as a predictor for public health alerts in the future. Physician activity could also be used as a potential real-time barometer for physician well-being at a regional or national level; by applying natural language processing techniques, such as sentiment analysis, such a data set could provide a baseline measurement of physician morale

at any given point in time. We are exploring more in-depth analyses of Twitter bios; topic modeling; and the correlation of self-reported identities with validated physician registries, such as medical licensing boards. Finally, we feel that our methodology and this first comprehensive study of physician activity on Twitter offer a foundation to building more accurate and precise algorithms for identifying physician accounts nationally and globally. We also provide a more complete characterization of the social media activity of US physicians than what has been possible in the past.

Conclusion

More physicians are using Twitter and covering a wider range of topics in their tweets. Twitter is entering a new phase of mainstream use among physicians, even though physician activity still represents a tiny fraction of total Twitter activity. These trends are evidence that Twitter can be a valuable source of data that can be used to understand social trends, interests, and the perceptions of the physician community. Unlike the more stringent settings of clinical practice, physicians are opening up on social media and showing a different side of themselves that is typically protected in the professional clinical setting. Physicians on Twitter may be offering a unique, more comprehensive view into the physician psyche, making Twitter an intriguing platform for further study and exploration.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Boolean logic for identifying physician accounts on Twitter.

[\[TXT File , 3 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Top 100 tweeted hashtags and themes between 2016 and 2020.

[\[PNG File , 62 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Word clouds for hashtags from 2016 to 2020.

[\[PNG File , 231 KB-Multimedia Appendix 3\]](#)

References

1. Sayce D. The number of tweets per day in 2020. David Sayce. URL: <https://www.dsayce.com/social-media/tweets-day/> [accessed 2021-04-23]
2. Hasan M, Orgun MA, Schwitter R. A survey on real-time event detection from the Twitter data stream. *J Inf Sci* 2017 Mar 17;44(4):443-463. [doi: [10.1177/0165551517698564](https://doi.org/10.1177/0165551517698564)]
3. Earle PS, Bowden DC, Guy M. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics* 2011;54(6):708-715 [FREE Full text] [doi: [10.4401/ag-5364](https://doi.org/10.4401/ag-5364)]
4. Sakaki T, Okazaki M, Matsuo Y. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Trans Knowl Data Eng* 2013 Apr;25(4):919-931. [doi: [10.1109/tkde.2012.29](https://doi.org/10.1109/tkde.2012.29)]
5. Li R, Lei KH, Khadiwala R, Chang KCC. TEDAS: A Twitter-based event detection and analysis system. 2012 Presented at: 2012 IEEE 28th International Conference on Data Engineering; April 1-5, 2012; Arlington, VA, USA p. 1273-1276. [doi: [10.1109/icde.2012.125](https://doi.org/10.1109/icde.2012.125)]
6. Wang X, Gerber MS, Brown DE. Automatic crime prediction using events extracted from Twitter posts. 2012 Presented at: SBP 2012: Social Computing, Behavioral-Cultural Modeling and Prediction; April 3-5, 2012; College Park, MD, USA p. 231-238. [doi: [10.1007/978-3-642-29047-3_28](https://doi.org/10.1007/978-3-642-29047-3_28)]

7. Daneshjou R, Adamson AS. Twitter journal clubs: Medical education in the era of social media. *JAMA Dermatol* 2020 Jul 01;156(7):729-730. [doi: [10.1001/jamadermatol.2020.0315](https://doi.org/10.1001/jamadermatol.2020.0315)] [Medline: [32186655](https://pubmed.ncbi.nlm.nih.gov/32186655/)]
8. Ghosh P, Schwartz G, Narouze S. Twitter as a powerful tool for communication between pain physicians during COVID-19 pandemic. *Reg Anesth Pain Med* 2021 Feb;46(2):187-188. [doi: [10.1136/rapm-2020-101530](https://doi.org/10.1136/rapm-2020-101530)] [Medline: [32321859](https://pubmed.ncbi.nlm.nih.gov/32321859/)]
9. Schwenk ES, Jaremko KM, Park BH, Stiegler MA, Gamble JG, Chu LF, et al. I tweet, therefore I learn: An analysis of Twitter use across anesthesiology conferences. *Anesth Analg* 2020 Feb;130(2):333-340. [doi: [10.1213/ANE.0000000000004036](https://doi.org/10.1213/ANE.0000000000004036)] [Medline: [31124801](https://pubmed.ncbi.nlm.nih.gov/31124801/)]
10. Chisolm MS. Social media in medicine: The volume that Twitter built. *Int Rev Psychiatry* 2015 Apr;27(2):83-84. [doi: [10.3109/09540261.2015.1026239](https://doi.org/10.3109/09540261.2015.1026239)] [Medline: [25828741](https://pubmed.ncbi.nlm.nih.gov/25828741/)]
11. Chretien KC, Azar J, Kind T. Physicians on Twitter. *JAMA* 2011 Feb 09;305(6):566-568. [doi: [10.1001/jama.2011.68](https://doi.org/10.1001/jama.2011.68)] [Medline: [21304081](https://pubmed.ncbi.nlm.nih.gov/21304081/)]
12. Riddell J, Brown A, Kovic I, Jauregui J. Who are the most influential emergency physicians on Twitter? *West J Emerg Med* 2017 Feb;18(2):281-287 [FREE Full text] [doi: [10.5811/westjem.2016.11.31299](https://doi.org/10.5811/westjem.2016.11.31299)] [Medline: [28210365](https://pubmed.ncbi.nlm.nih.gov/28210365/)]
13. Sharp SP, Mackenzie DG, Ong DSY, Mountziaris PM, Logghe HJ, Ferrada P, et al. Factors influencing the dissemination of tweets at the American College of Surgeons Clinical Congress 2018. *Am Surg* 2021 Apr;87(4):520-526. [doi: [10.1177/0003134820950680](https://doi.org/10.1177/0003134820950680)] [Medline: [33106021](https://pubmed.ncbi.nlm.nih.gov/33106021/)]
14. @Bob_Wachter. 1/ Covid (@UCSF) Chronicles, Day 80 Local update: @UCSFHospitals, just 8 pts, 3 vented, no change from yesterday. In SF, cases still steady at ~20-30/d; and still no new deaths in a week (cumulative cases in Fig). 43 pts in hospitals, up 4 from yest; ICU pts 16, down from 19. 2020 Jun 05. URL: https://twitter.com/Bob_Wachter/status/1269096800096407552 [accessed 2022-06-05]
15. @Bob_Wachter. SF folks: it's now a big-time surge. No longer just cases (~500/d reported, so true # >2K, or ~250/100K/d). Also major uptick in hospitalizations: 98 in SF (vs 18 six wks ago). @UCSFHospitals 41 pts in hospital, twice April #. If you're trying to stay well, time to up your game. 2022 May 30. URL: https://twitter.com/Bob_Wachter/status/1531315837264482304 [accessed 2022-06-05]
16. @EricTopol. It's not over, by any means, but the chances of dying from Covid keep declining. 2022 Jun 05. URL: <https://twitter.com/EricTopol/status/1533485528720416768> [accessed 2022-06-05]
17. @Craig_A_Spencer. Hospitalizations are increasing as Omicron cases reach record highs in NYC and all across the country. But as I noticed on my recent ER shifts, take a look at who needs admission to the hospital for severe Covid? TL;DR It's the unvaccinated. 2021 Dec 30. URL: https://twitter.com/Craig_A_Spencer/status/1476772672994488328 [accessed 2022-06-05]
18. Liu L, Woo BKP. Twitter as a mental health support system for students and professionals in the medical field. *JMIR Med Educ* 2021 Jan 19;7(1):e17598 [FREE Full text] [doi: [10.2196/17598](https://doi.org/10.2196/17598)] [Medline: [33464210](https://pubmed.ncbi.nlm.nih.gov/33464210/)]
19. Leibowitz MK, Scudder MR, McCabe M, Chan JL, Klein MR, Trueger NS, et al. Emergency medicine influencers' Twitter use during the COVID-19 pandemic: A mixed-methods analysis. *West J Emerg Med* 2021 Mar 22;22(3):710-718 [FREE Full text] [doi: [10.5811/westjem.2020.12.49213](https://doi.org/10.5811/westjem.2020.12.49213)] [Medline: [34125051](https://pubmed.ncbi.nlm.nih.gov/34125051/)]
20. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. *AMIA Annu Symp Proc* 2014 Nov 14;2014:924-933 [FREE Full text] [Medline: [25954400](https://pubmed.ncbi.nlm.nih.gov/25954400/)]
21. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: A sentiment analysis. *J Diabetes Sci Technol* 2019 May;13(3):439-444 [FREE Full text] [doi: [10.1177/1932296818811679](https://doi.org/10.1177/1932296818811679)] [Medline: [30453762](https://pubmed.ncbi.nlm.nih.gov/30453762/)]
22. Doogan C, Buntine W, Linger H, Brunt S. Public perceptions and attitudes toward COVID-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of Twitter data. *J Med Internet Res* 2020 Sep 03;22(9):e21419 [FREE Full text] [doi: [10.2196/21419](https://doi.org/10.2196/21419)] [Medline: [32784190](https://pubmed.ncbi.nlm.nih.gov/32784190/)]
23. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. *JMIR Public Health Surveill* 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: [10.2196/21978](https://doi.org/10.2196/21978)] [Medline: [33108310](https://pubmed.ncbi.nlm.nih.gov/33108310/)]
24. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health Surveill* 2020 May 22;6(2):e19447 [FREE Full text] [doi: [10.2196/19447](https://doi.org/10.2196/19447)] [Medline: [32412418](https://pubmed.ncbi.nlm.nih.gov/32412418/)]
25. Cevik M, Ong DSY, Mackenzie G. How scientists and physicians use Twitter during a medical congress. *Clin Microbiol Infect* 2019 Dec;25(12):1561.e7-1561.e12 [FREE Full text] [doi: [10.1016/j.cmi.2019.04.030](https://doi.org/10.1016/j.cmi.2019.04.030)] [Medline: [31102781](https://pubmed.ncbi.nlm.nih.gov/31102781/)]
26. Margus C, Brown N, Hertelendy AJ, Safferman MR, Hart A, Ciottone GR. Emergency physician Twitter use in the COVID-19 pandemic as a potential predictor of impending surge: Retrospective observational study. *J Med Internet Res* 2021 Jul 14;23(7):e28615 [FREE Full text] [doi: [10.2196/28615](https://doi.org/10.2196/28615)] [Medline: [34081612](https://pubmed.ncbi.nlm.nih.gov/34081612/)]
27. Kilic Y, Chauhan D, Avery P, Horwood N, Nakov R, Disney B, et al. The public's attitude towards doctors' use of Twitter and perceived professionalism: an exploratory study. *Clin Med (Lond)* 2021 Sep;21(5):e475-e479 [FREE Full text] [doi: [10.7861/clinmed.2021-0357](https://doi.org/10.7861/clinmed.2021-0357)] [Medline: [34507932](https://pubmed.ncbi.nlm.nih.gov/34507932/)]
28. Twitter verification requirements - how to get the blue check. Twitter. URL: <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts> [accessed 2021-06-28]
29. Urban Dictionary: TCOT. Urban Dictionary. URL: <https://www.urbandictionary.com/define.php?term=TCOT> [accessed 2021-07-04]

30. Benjamin Y. The secret Twitter war for America's independents: #tcot vs #p2. SFGate. 2012 Nov 19. URL: <https://web.archive.org/web/20121120014027/https://blog.sfgate.com/ybenjamin/2010/07/27/the-secret-twitter-war-for-americas-independents-tcot-vs-p2/> [accessed 2021-07-24]
31. Does my project need review by the IRB. UC Davis Office of Research. URL: <https://research.ucdavis.edu/policiescompliance/irb-admin/researchers/irb-submissions/new-projects/do-i-need-irb-review/> [accessed 2022-08-11]
32. Rogers EM, Singhal A, Quinlan MM. Diffusion of innovations. In: Stacks DW, Salwen MB, editors. An Integrated Approach to Communication Theory and Research, Second Edition. New York, NY: Routledge; 2009.
33. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. *Science* 2019 Jan 25;363(6425):374-378. [doi: [10.1126/science.aau2706](https://doi.org/10.1126/science.aau2706)] [Medline: [30679368](https://pubmed.ncbi.nlm.nih.gov/30679368/)]
34. Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *Am Polit Sci Rev* 2021 May 06;115(3):999-1015. [doi: [10.1017/s0003055421000290](https://doi.org/10.1017/s0003055421000290)]
35. Memon SA, Carley KM. Characterizing COVID-19 misinformation communities using a novel Twitter dataset. arXiv Preprint posted online on September 19, 2020 [FREE Full text] [doi: [10.48550/arXiv.2008.00791](https://doi.org/10.48550/arXiv.2008.00791)]
36. Huang B, Carley KM. Disinformation and misinformation on Twitter during the novel coronavirus outbreak. arXiv Preprint posted online on June 7, 2020 [FREE Full text] [doi: [10.48550/arXiv.2006.04278](https://doi.org/10.48550/arXiv.2006.04278)]
37. Porter J. Twitter admits it verified several fake accounts. *The Verge*. 2021 Jul 13. URL: <https://www.theverge.com/2021/7/13/22575025/twitter-verified-fake-accounts-botnet-researcher> [accessed 2022-06-05]
38. @verified. We've temporarily hit pause on rolling out access to apply for Verification so we can make improvements to the application and review process. For those who have been waiting, we know this may be disappointing. We want to get things right, and appreciate your patience. 2021 Aug 13. URL: <https://twitter.com/verified/status/1426214732403466241> [accessed 2022-06-05]
39. Clark M. Twitter pauses its verification program rollout after giving fake accounts blue checks. *The Verge*. 2021 Aug 13. URL: <https://www.theverge.com/2021/8/13/22623714/twitter-verification-program-pause-fake-accounts-review> [accessed 2022-06-05]
40. Neporent L. Coronavirus social: Twitter hustles to verify physicians. *Medscape*. 2020 May 12. URL: <http://www.medscape.com/viewarticle/930360> [accessed 2022-06-05]

Edited by T Leung; submitted 05.03.22; peer-reviewed by J Luo, J Li, D Pessoa, S El kefi; comments to author 25.03.22; revised version received 05.06.22; accepted 12.07.22; published 06.09.22

Please cite as:

Nakagawa K, Yang NT, Wilson M, Yellowlees P

Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study

J Med Internet Res 2022;24(9):e37752

URL: <https://www.jmir.org/2022/9/e37752>

doi: [10.2196/37752](https://doi.org/10.2196/37752)

PMID: [36066939](https://pubmed.ncbi.nlm.nih.gov/36066939/)

©Keisuke Nakagawa, Nuen Tsang Yang, Machel Wilson, Peter Yellowlees. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.