

Original Paper

Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data

Sunny Jung Kim^{1,2}, PhD; Lisa A Marsch^{1,2}, PhD; Jeffrey T Hancock³, PhD; Amarendra K Das⁴, MD, PhD

¹Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Dartmouth College, Lebanon, NH, United States

²Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States

³Department of Communication, Stanford University, Stanford, CA, United States

⁴Healthcare Effectiveness Research, IBM, Cambridge, MA, United States

Corresponding Author:

Sunny Jung Kim, PhD

Department of Biomedical Data Science

Geisel School of Medicine at Dartmouth

Dartmouth College

46 Centerra Parkway

Lebanon, NH,

United States

Phone: 1 603 646 7041

Fax: 1 603 646 7068

Email: sunny.j.kim@dartmouth.edu

Abstract

Background: Substance use–related communication for drug use promotion and its prevention is widely prevalent on social media. Social media big data involve naturally occurring communication phenomena that are observable through social media platforms, which can be used in computational or scalable solutions to generate data-driven inferences. Despite the promising potential to utilize social media big data to monitor and treat substance use problems, the characteristics, mechanisms, and outcomes of substance use–related communications on social media are largely unknown. Understanding these aspects can help researchers effectively leverage social media big data and platforms for observation and health communication outreach for people with substance use problems.

Objective: The objective of this critical review was to determine how social media big data can be used to understand communication and behavioral patterns of problematic use of prescription drugs. We elaborate on theoretical applications, ethical challenges and methodological considerations when using social media big data for research on drug abuse and addiction. Based on a critical review process, we propose a typology with key initiatives to address the knowledge gap in the use of social media for research on prescription drug abuse and addiction.

Methods: First, we provided a narrative summary of the literature on drug use–related communication on social media. We also examined ethical considerations in the research processes of (1) social media big data mining, (2) subgroup or follow-up investigation, and (3) dissemination of social media data-driven findings. To develop a critical review-based typology, we searched the PubMed database and the entire e-collection theme of “infodemiology and infoveillance” in the Journal of Medical Internet Research / JMIR Publications. Studies that met our inclusion criteria (eg, use of social media data concerning non-medical use of prescription drugs, data informatics-driven findings) were reviewed for knowledge synthesis. User characteristics, communication characteristics, mechanisms and predictors of such communications, and the psychological and behavioral outcomes of social media use for problematic drug use–related communications are the dimensions of our typology. In addition to ethical practices and considerations, we also reviewed the methodological and computational approaches used in each study to develop our typology.

Results: We developed a typology to better understand non-medical, problematic use of prescription drugs through the lens of social media big data. Highly relevant studies that met our inclusion criteria were reviewed for knowledge synthesis. The characteristics of users who shared problematic substance use–related communications on social media were reported by general group terms, such as adolescents, Twitter users, and Instagram users. All reviewed studies examined the communication characteristics, such as linguistic properties, and social networks of problematic drug use–related communications on social media. The mechanisms and predictors of such social media communications were not directly examined or empirically identified

in the reviewed studies. The psychological or behavioral consequence (eg, increased behavioral intention for mimicking risky health behaviors) of engaging with and being exposed to social media communications regarding problematic drug use was another area of research that has been understudied.

Conclusions: We offer theoretical applications, ethical considerations, and empirical evidence within the scope of social media communication and prescription drug abuse and addiction. Our critical review suggests that social media big data can be a tremendous resource to understand, monitor and intervene on drug abuse and addiction problems.

(*J Med Internet Res* 2017;19(10):e353) doi:[10.2196/jmir.6426](https://doi.org/10.2196/jmir.6426)

KEYWORDS

opioid epidemic; opioid crisis; opioid-related disorders; substance use; substance-related disorders; prescription drug misuse; addiction; Facebook; Twitter; Instagram; big data; ethics

Introduction

User-generated content and user interactions related to drug use (eg, opioid misuse) are prevalent and are rapidly emerging forms of communication across social media platforms [1,2]. Social media big data on this topic offer an avenue for observing and understanding the temporal trends of problematic drug use and relevant risk factors in real time [3], as well as the ability to measure the collective human behavior of an extremely large population of interest [4]. Social media big data involve large, digitized data resources that contain naturally occurring communication phenomena observable through social media platforms and that can be used in computational or scalable solutions to generate data-driven inferences. Not only do some people communicate and share personal experiences, questions, and thoughts about substance use problems on social media, but also persons with addiction problems seek out social support from others with similar addiction problems through networks and communications available on social media [5]. Davey and colleagues argued that online communities and forums are well suited for people to communicate about problematic drug use activities because of their presumed anonymity and relative freedom from geographic constraints and perceived stigma [6]. Shutler and colleagues noted that social media, such as Twitter, can be an observatory platform that can reveal patterns of the current opioid epidemic, as users tweet about illicit, nonmedical use of prescription opioids in part due to users' perceived protection of their real identity [7].

Although well-established resources in the United States such as the Drug Abuse Warning Network and the National Survey on Drug Use and Health offer critical information about substance abuse practices, these traditional platforms are known to lag in time in terms of data availability to the public for possible use in research [3]. On the other hand, online communications regarding drug use problems (eg, opioid use disorders) are surprisingly prevalent on social media [8]. These unsolicited communication datasets provide researchers a novel opportunity to unobtrusively assess and track various health risks, human factors, and emerging trends surrounding drug use [3,9]. Daniulaityte and colleagues stated that online technologies have become one of the leading-edge sources for detecting patterns and trends in illicit drug use [3]. Miller and Sonderlund suggested that online communication technologies can be an effective means for communicating with hard-to-reach populations that have illicit drug use problems [10].

Along with advances in information and communication technologies, seeking health information, disclosing personal health concerns, and exchanging social support are pervasive forms of human communications on social media [2,11-20]. It is critical to develop a conceptual framework to enhance our scientific understanding of these naturally occurring communications on social media related to drug use problems. The use of social media data-driven knowledge may help researchers better identify the utility of these technologies for public health research beyond the knowledge gained from domain experts. Furthermore, computational findings that emerge from self-disclosed social media data may ameliorate concerns about research validity in self-reported data, namely in terms of social desirability, response bias, and memory recall biases.

Due to the analogous and reflective nature of one's social media world to one's own real world, user-generated social media big data are increasingly being embraced and analyzed to observe and predict psychological states and collective human behavior [21,22]. For example, researchers found that social media communications and profiles correlate with real-world reflection of the self [23,24]. Communication patterns and interactions on social media also predict narcissistic personal traits [25], psychological functioning and well-being [26], and personality traits [27,28]. In recent years, studies have demonstrated the utility of social media big data in understanding public health problems, ranging from mental health conditions [29], population-level influenza monitoring [30,31], pronounced use of cannabis concentrates in the marijuana-legalized states in the United States [32], and the prescription stimulant Adderall [2] to perceived risks and sentiment around marijuana use [17].

With a growing line of empirical evidence demonstrating social media's usefulness for observing and predicting health behaviors, social media data on drug use-related communications are being analyzed to address various research inquiries, including temporal trends of problematic substance use [33,34], market changes, social norms and cultural aspects of drug use [35], public perceptions, and relevant psychological factors (eg, sentiments [7]). Systematically assessed results of these social media communication data at scale for drug use problems can further inform key outreach methods, future intervention components, harm-reduction methods, and control and prevention strategies, which, in turn, can be delivered via vital social media channels for public health promotion. To maximize the utility of social media big data in addressing the

urgent public health problems (eg, substance use problems, the opioid crisis) in the United States, we present a conceptual framework designed to guide investigation on problematic drug use–related communications that are observable via social media and to advance their potential impact for public health outreach efforts.

A Multidimensional Framework to Analyze Social Media Communications for Problematic Use of Prescription Drugs

Our conceptualization classifies the communication on drug use problems that is evident on social media into 4 key dimensions: (1) user characteristics, (2) communication characteristics, (3) mechanisms and predictors of problematic substance use–related communications on social media, and (4) psychological and behavioral consequences of these social media communications at an individual and a societal level.

First, understanding user characteristics (eg, who are the users that share drug abusive, risky behaviors and addiction-related psychological states on social media?) permits in-depth subgroup or moderator analyses of the drug-related communications and related risks [36]. More specifically, understanding the demographic characteristics of these communications can advance targeted monitoring of drug abuse trends, as well as enhance the development of moderation modeling for a specific subgroup categorization of interest [37,38].

Second, identifying the communication features of target social media big data can offer insights into temporal, linguistic, and psychological patterns associated with self-disclosed social media communications about problematic use of prescription drugs [35]. For example, Paul and Dredze led a promising study to develop data crawling and surveillance systems to capture contextual factors associated with recreational drug uses through data mining of online communication data [35]. They modified and used a factorial latent Dirichlet allocation, a multidimensional text modeling approach, to incorporate prior knowledge about contextual factors such as drug type (eg, cocaine), delivery method (eg, smoking), and outcome aspects (eg, effects, health). Their approach has demonstrated successful application of data informatics to capture and discover an arbitrary number of contextual factors that are clinically important to understand new recreational drugs and trends.

As emphasized in a recent study incorporating machine learning techniques on Twitter feeds for a mental health diagnosis, analyzing linguistic properties of self-disclosed communication on social media regarding substance use problems provides a novel opportunity to identify communication themes and unmet needs among people with substance use problems [39]. Understanding communication characteristics of these unsolicited social media data will facilitate timely treatment initiation and health communication outreach strategies. Furthermore, the communication properties can be examined in conjunction with analyzing the user characteristics of those communications to identify interacting subgroups of users who share specific topics or valence (eg, anger, fear) regarding problematic substance use.

Third, investigating why and in which contexts people use social media to communicate about substance use–related problems reveals the mechanisms of various unsolicited behaviors (eg, self-disclosing personal stories about the nonmedical use of prescription drugs; using social media to receive social support during the addiction recovery process). Identifying this mechanism of communication behavior through analyzing social media big data, along with user interviews and self-reported surveys, can inform a strategic mediating construct for investigators when developing social media-based prevention or intervention programs (eg, [2]).

Fourth, understanding the psychological effects (eg, self-disclosure) and behavioral outcomes (eg, social influence) that such unsolicited uses of social media communications have on the self and others has received increased public attention [40]. However, this area of research has not been fully explored with a focus on the practical and clinical potential of social media technologies to promote health outcomes [41–43]. Social media big data analytics, in conjunction with mixed methodologies involving longitudinal follow-up and cross-sectional surveys or qualitative interviews, may help researchers identify these potential outcomes (eg, the effects of using recovery support groups on Facebook). Examining the outcomes of naturally occurring social media communications can offer intervention models that target critical moments to deliver a just-in-time intervention via social media at scale.

Integrating these 4 themes into a multidimensional framework enables systematic observation of factors and conditions explaining pervasive uses of social media for drug use–related communications. Development of this multiconceptual framework can also help researchers and clinicians explore the predictive and mechanistic values of social media-based communications in delivering state-of-the-art drug abuse recovery support and engagement systems. Furthermore, learning outcomes relevant to this multidimensional framework will offer data-driven strategies for leveraging social media data, features, and platforms for health promotion (eg, campaigns), as well as for understanding the nature of human communications concerning time-sensitive health issues.

In this work, we begin by reviewing the current use of social media for problematic drug use–related communications. Then, we highlight ethical challenges and methodological considerations when using social media big data for research on drug abuse and addiction. Lastly, based on these 4 conceptual dimensions and ethical considerations, we provide a narrative summary of the literature on social media-based drug communications and propose a typology with key initiatives aimed at addressing knowledge gaps in the use of social media for research on problematic and nonmedical use of prescription drugs.

Prevalence of Drug Use–Related Communication on Social Media and Its Social Impact

Substance Use–Promoting Communication and Its Impact

User-generated content that promotes substance use (eg, positively commenting on pictures of illicit drugs [44]) is

prevalent across market-leading social media platforms, such as Instagram and Twitter [45]. Such content can diffuse rapidly and widely through easily accessible network ties within and across media platforms [44,46]. Capurro and colleagues conducted a systematic review of 73 studies that used social networking sites to understand various public health issues, including sexual risks and mental health [47]. The review reported that 86% of the reviewed studies described user-generated content and served as passive observational investigations for surveillance on target health events among hard-to-reach populations. Their review also implied that researchers are increasingly leveraging social media platforms and data within the domain of various public health issues, thus directly benefitting from the prevalence of user-centered data that indicate risky health behaviors and psychological states. This systematic review, however, did not report research that focused on social media big data–based findings for prescription drug addiction.

Moreno and colleagues [48,49] found that 41% of young adult participants had pictures or messages referencing alcohol, tobacco, or other drug use in their publicly accessible social media profiles. When social media users are frequently and repeatedly exposed to or engage in such substance-promoting communications, they may become more accepting of or immune to these risky behaviors. As noted in media and social influence frameworks, drug-use promotional communications on social media that are shared across social network ties can influence the exposed users to normalize the frequency of these behaviors and, as a result, may change their attitudes toward or risk perceptions of these substances (eg, [50,51]). A national survey of US adolescents found that 40% of all teens in a nationally representative sample had seen pictures on social media depicting other teens getting drunk or using tobacco or illicit drugs [46]. According to the National Survey of American Attitudes on Substance Abuse, adolescents who reported seeing pictures of peers using substances on social media were more likely to use substances than were those who had never seen this peer-generated content on social media [46].

Social psychology and media communication theories explain this link between content exposure and an increased willingness to engage in the behavior being promoted. For example, cultivation theory [52,53] posits that frequent media exposure to risky behavior influences the belief that the mediated version of reality is real, leading to overestimation of the frequency and prevalence of those risky behaviors in the real world. This distorted perception of reality leads people to accept risky or detrimental behaviors portrayed in the media, such as substance use, as relatively normal [46,54]. Social learning theory [55] further buttresses the claim that observing risky behaviors via social media can influence people to mimic behaviors or adopt specific values and thoughts. Bandura's social learning theory [56] posits that media communication can considerably promote changes in human beliefs and behaviors by “informing, enabling, motivating, and guiding” the audience (pg 76).

Social media communication platforms allow substance users to connect with a wide array of social networks and readily accessible substance use–related content. A news feed on a social media site can become a platform that constantly provides

both personalized and socially infused content for social modeling and mimicry. These socially mediated learning processes on social media underscore the importance of considering the consequences (eg, mimicry) of routine exposure to content that is positively framed for and, indirectly or directly, promotes problematic drug use.

Substance Recovery Support and Prevention Communication and Its Clinical Implications

While content that promotes substance use is prevalent on social media, use of social media networks and online communities to explore recovery support for drug addiction problems is an increasingly popular form of communication [42]. Social media-based platforms (eg, Facebook Groups and Pages) are generated and led by self-motivated users with a goal of sharing and providing social support for users who have substance use problems or are in addiction recovery processes. Recently published work led by Rubya and Yarosh examined the characteristics of video-mediated, peer-led synchronous online support communities for addiction recovery [42]. Although the platform they examined in their work is not one of the social media platforms in the current market (eg, Facebook, Twitter), it did offer social media components, such as user interaction features, social networking functions, and engagement tools, that are important for peer-based online social support [57]. Through data crawling, self-selected online surveys, and in-depth interviews, Rubya and Yarosh examined the role of video-based online forums for people in addiction recovery and reported that these forums were viewed as critical in helping people with opioid use disorders overcome any perceived barriers (eg, access, anonymity) to seeking recovery support for drug addiction.

Compared with their application in the drug addiction context, the values of peer support and user engagement on social media platforms have been leveraged in health interventions for other health contexts, such as smoking cessation [58] and weight loss among vulnerable populations [59]. For example, Kim and colleagues [57] used various Facebook Group features for a 6-week smoking cessation and reduction intervention. Their findings support a successful rate of smoking reduction predicted by user engagement (eg, the number of comments posted) and the amount of social support received (eg, the number of “likes” participants received). These mechanisms were facilitated within their Facebook Group intervention that was designed to assist regular smokers to quit cigarette smoking. As discussed in 2 recent systematic reviews on social media-based studies within the public health interest, social media technologies and features have not been fully applied or used for long-term, scalable investigations that can generate successful social network-based health diffusion phenomena for health promotions [47,58]. We further argue that, despite the potential value of unique technology features in social media platforms, most of these platforms have not been systematically leveraged to offer evidence-based content or scientifically guided support for people in opioid addiction recovery processes. Accordingly, the public health and clinical benefits of leveraging social media platforms for substance prevention and recovery support remain underexplored.

A national report released by the US Centers for Disease Control and Prevention stated that almost 48,000 deaths in the United States in 2014 were induced by potent substance poisoning, surpassing the number of deaths caused by motor vehicle accidents [60]. The US National Institute on Drug Abuse reported that the total number of substance use–related deaths in the United States increased since 2001, with as much as a 6.2-fold increase in the number of overdose deaths from heroin from 2002 to 2015 [61]. The substance use epidemic in the United States has caused detrimental consequences on both individual and societal levels that call for urgent research on understanding this phenomenon in a scalable, systematic fashion.

The proliferation of substance use prevention and addiction recovery support communications among social media users provides empirical resources to systematically analyze naturalistic, real-time communications data and to gain time-sensitive insights into substance use risk factors, behavioral patterns, and prevention and recovery processes. Applying big data informatics—the process of examining large unstructured data to discover hidden patterns and useful information [62]—to substance use prevention and recovery support-related social media data will help us gain important and novel knowledge regarding the characteristics of substance use–linked users and their health behaviors [63–65]. Given the prevalence of social media use among people with drug use problems, in addition to the potential opportunities for learning about relevant risk behaviors and factors via user-solicited communication data on social media, we posit that this novel approach will be of importance to the development of evidence-based frameworks that promote addiction prevention and recovery support.

With these clinical implications and opportunities for understanding and enhancing a sensitive health problem, the social media data that contain various levels of user information (eg, age, gender) and users' expectations and nuances embedded in social media contexts require cautionary attention to ethical boundaries and practices. In the following section, we discuss key ethical principles, challenges, and considerations applicable in this novel research setting when incorporating user-centered social media communication data for behavioral health research.

Researchers' Responsibilities and Big Data Ethics in Studying Social Media Users With Substance Use Problems

The use of social media big data for informatics can provide population-level inferences, such as pattern or trend recognition [66] and natural language processing [67], for drug use–related behaviors and psychological states. When using publicly available and personally relevant communication data to understand human health behaviors and risks, special attention should be given to research ethics, especially for data concerning sensitive topics like drug addiction. A balance between ethical principles and scientific discoveries gleaned from novel technology tools should be actively sought and put into practice prior to designing and performing data mining and informatics. Mittelstadt and Floridi [68] explored research ethics-related themes that emerged from meta-analyses of big data–based studies. From those themes, we synthesize relevant principles applicable to each stage of research using social media big data

to understand the drug epidemic. Those stages are identified as follows: (1) determining the scope of search for data mining or retrieval, (2) conducting subgroup analysis or follow-up investigation, and (3) disseminating and presenting social media data-driven findings.

Social Media Big Data Mining and Retrieval

Throughout the process of mining and retrieving social media big data, it is important to balance an understanding of data privacy from the user's perspective beyond the needs of researchers who are interested in the data to provide broader public health benefits. Taking a user's perspective (eg, on beliefs and priorities from a user's mental state) can be a useful practice for ethical considerations. Perspective taking is an other-centered social cognitive process that involves simulation of and making inference about the target counterpart's mind and cognitive state [69]. Accurately understanding a user's mind and expectations requires taking their perspective, perceiving the situation through the lens of the users who generated these social media communications [70]. In the context of social media data mining, it is necessary to take the perspectives of users who self-disclose personal content on social media, as users' expectations of the privacy of their social media posts can differ from researchers' own perceptions regarding the privacy of the posts. For example, one possible scenario might be that Facebook users post personal stories about their addiction problems on "closed" Facebook Groups, expecting that the self-disclosed content will be kept within the groups. Researchers may perceive the content as part of the aggregated piece of information that is technically available to others who join the groups. In this specific scenario, there is a perceptual discrepancy between social media users and researchers regarding how much access researchers should have to the social media data for data mining and analysis. Given this perceptual discrepancy, failure to communicate research purposes and data protection plans up front is likely to leave users feeling deceived, uninformed, or manipulated [71]. Prior to data mining and retrieval, we encourage investigators to understand ethical expectations and the notion of moral harm, which is assessed by the risk level versus the values of scientific discovery from the research activities, from the perspectives of users in a context-specific manner. One method of building mutual understanding between users and investigators might be contacting the groups' administrators to gain insight on in-group norms and user expectations. Researchers might ask for an introduction to the group members or coordinate an announcement within the groups to inform members of the group that researchers may use anonymous data (without disclosing personally identifying user information) in an aggregated manner for research purposes (eg, [72]).

To protect users' privacy and understand their expectations on the use of their potentially sensitive social media content, such as drug use–related communications, we also encourage researchers to construct guidelines for case-specific scenarios concerning potential ethical issues. For example, a research program at Harvard University published a guideline booklet that reviews ethical norms, case reports, and concrete guidance to help researchers who use social media sites for patient recruitment and contact [72]. The guideline booklet provides a series of dilemma-type scenarios that can occur between

researchers, participants, and different stakeholders when social media sites are used for patient recruitment. The guideline booklet also provides multiple case analysis results for each scenario. In line with the guideline booklet, we argue that the same ethical and regulatory practices used in traditional recruitment settings (eg, respect for privacy) can be applied to social media contexts. However, different operational implications should be considered, such as prospectively understanding sensitive values of the social media communities and their users, and the impacts of social media-based research methods on public trust [73]. Given the increasing research attention focus on use of social media for observatory resources and for health communication outreach and delivery platforms, there is a need for developing a concrete guideline on ethical research practices for use of naturally occurring social media communication data and social media-based networks and users for observational purposes and health outreach studies.

As demonstrated in Harvard's guideline for social media use for recruitment research, scenario-specific solutions for ethical challenges on social media-based observational studies can be implemented through "cognitive rehearsals" (or scenario planning) prospectively, rather than retrospectively. Cognitive rehearsal is a cognitive behavioral technique that allows individuals to develop an effective set of responses to a critical event. A cognitive rehearsal approach [74,75] is based on the assumption that a person's reaction to a critical event can be learned and enhanced through reading instructions and building a knowledge base of responses specific to the event. Through cognitive rehearsal techniques, individuals can subsequently adapt learned skills from scripted responses when they face the previously rehearsed event. Cognitive rehearsal approaches have been used as an intervention component to improve responses to impulse control issues and lateral violence [74]. These cognitive rehearsal techniques can be applied during challenging events specific to ethical dilemmas concerning big data use. By "rehearsing" situations and events in advance and generating instructive guidelines, investigators can be better prepared for such situations if they arise, thus establishing a healthy academic culture and mutual trust between the researcher and the community of interest. This scenario-specific, empathetic approach to examine social media big data can be applied to various data mining and retrieval activities, including developing protocols for data anonymization, implementing confidentiality and data protection plans applicable to data mining, and mobilizing plans for strategic communications with social media users for the informed consent process.

Subgroup Analysis and Follow-Up Investigation

Researchers' responsibilities extend to maintaining data confidentiality and protection, and ensuring the transparency of research activities. These activities include not only data mining and analysis, but also subgroup analysis or follow-up investigation (eg, distributing surveys among subgroups, or invitations for interventions). Investigators may need to conduct follow-up surveys or interventions to address questions raised by big data-driven findings. Electronic data are track-(back)-able and technological advances can allow deidentified social media data to be reidentified. Therefore, failures in confidentiality or lack of transparency during the follow-up processes can cause

great concern among users if their data were aggregated from social media without their prior consent. The ethical implications and social consequences of contacting individuals who were attributes of the big data-driven findings should be recognized by researchers and policy makers in relation to data protection and human autonomy issues in big data research, even as the attributes of individuals were imperceptible to the aggregate level of big data-driven knowledge [68,76]. Before pursuing follow-up investigation, researchers should consider the rights of subjects and weigh context-specific risk levels in comparison with the value of scientific discovery from the research activities. Additional caution should be used with respect to not only the confidentiality and privacy of reidentified subjects and data security and protection, but also to the compliance between research activities conducted for follow-up purposes and the policies and terms of use of relevant websites [72].

Dissemination of Social Media Big Data-Driven Findings

Emerging technologies and big data have ethical implications beyond those identified in accomplishing research aims. Researchers are also responsible for ethically disseminating the findings extracted from social media big data [77]. Even if the anonymization of individual-level data is successful, findings that describe participants by specific geographical characteristics, socioeconomic status, health condition, risky behavior, or a combination of these characteristics can cause discrimination and stigmatization of those groups, which in turn can raise group-level harm and risk [78]. The consequences of ignoring group-level harm (eg, increased stigmatization of patients in addiction recovery as a result of big data-driven findings reported in an academic article) are nontrivial and can affect broad members of the identified groups or community [79], including those who have opted out of the study. Group-level effects of reidentifiable anonymized data require further research exploration in terms of their social implications for the groups' users (eg, group-level stigma and group-level privacy) and potential unintended discrimination against subgroups with particular demographic characteristics or health-related problems [80].

Ethical principles, including the fundamental rights of "autonomy, protection, safety, maximization of benefits and minimization of harm, and respect for beneficence" ([81], pg 4), are not difficult concepts to understand. However, applying these principles to a sensitive social media big data context (eg, substance use) can raise ethical challenges. Ethics protocols and guidelines for social media big data-driven health research are evolving relatively slowly, compared with the pace of research outcomes using social media big data. These challenges require special sensitivity to the dynamics of social network-based communities and can only be addressed by carefully engaging in iterative ethical decision-making processes, both prospectively in designing studies and retrospectively by learning from ethical practices conducted in the relevant literature.

Methods

We developed a typology based on a critical review process [82]. The strength of a critical review is not only that the review

process evaluates the previous body of work, but also that it results in models of thought that can offer a new phase of conceptual framework development [82]. The typology is focused on delivering a set of relevant previous studies in terms of 4 conceptual attributes: (1) the characteristics of users who engage in substance use–related communications for promotional or preventive and control purposes; (2) the nature of substance use–related communications, such as their valence, expressed sentiment properties, and patterns of interactions; (3) the social and psychological predictors and mechanisms of those social media communication behaviors; and (4) the effects of problematic substance use–related social media communications on users. This knowledge development method is designed to elaborate on and identify the utility of social media as a means of communication delivery and observational platforms for users with substance use problems. Developing a multidimensional framework can further help theorize and synthesize underlying factors and conditions that influence social media use behaviors among online communities for problematic substance use–related reasons [43].

In our critical review, 1 researcher (SJK) independently searched for relevant literature within the PubMed database and the entire e-collection theme of “infodemiology and infoveillance” of JMIR Publications which includes the *Journal of Medical Internet Research* and articles in other JMIR journals indexed with this topic (<http://www.jmir.org/themes/69>). The search keywords were “prescription drug abuse,” “nonmedical drug use,” “social media,” “social network sites,” “big data,” “data mining,” “social media data,” “informatics,” “machine learning,” “Twitter,” “Instagram,” and “Facebook.” The same researcher (SJK) then evaluated the titles and abstracts of published studies based on their topical relevance (ie, problematic or nonmedical use of prescription drugs) and included studies that used social media communication data and computational analytic methods (eg, data mining, social network analysis, and supervised or unsupervised natural language processing).

Given the scope of the review—that is, focusing on how social media big data can be used to understand communication and behavioral patterns of nonmedical or problematic use of prescription drugs—in our typology, we did not include studies that reported data from nonsocial media platforms (eg, online forums), problematic use of other substances (eg, alcohol, cigarettes, and cannabis), or noncomputational analytic methods (eg, survey only) (eg, [1,3,5,6,35]). Note that demonstrating the systematic nature of the article search and conducting quality assessment are not components of critical review, and, thus, a Consolidated Standards of Reporting Trials (CONSORT) diagram is not required. After screening a large number of searched articles within the PubMed database and the entire e-collection theme of “infodemiology and infoveillance” within the *Journal of Medical Internet Research*, we reviewed findings from 8 highly relevant studies that met all the inclusion criteria. We report empirical evidence from the studies in relation to the 4 research inquiries, the methodological and computational approaches, and ethical practices and considerations discussed in each study (Multimedia Appendix 1 [2,4,7,33,36,83-85]).

Results

Multimedia Appendix 1 lists studies that met eligibility criteria and were closely related to the topic of interest, along with the 4 conceptual dimensions, as well as methodological and ethical domains.

User Characteristics

The characteristics of users who shared substance use–related communications on social media were reported by general group characteristics, such as college students, youth and adolescents, Twitter users, and Instagram users. Different types and levels of substance uses tended to be associated with different demographic characteristics, including sex, age, and socioeconomic status [38]. The demographic and social-psychological characteristics of users can be classification markers for certain types and patterns of substance use–related communications on social media (eg, polydrug use tweets with positive action verbs). Understanding social media communication data with specific user characteristics can inform the division of subgroups when targeting just-in-time interventions, addiction recovery support systems, or antidrug recovery campaigns through social media. However, relatively less research attention was given to analyzing or incorporating user characteristics along with the analyzed communication characteristics.

Communication Characteristics

All the reviewed studies examined various communication characteristics, ranging from geographic and temporal trends associated with nonmedical use of prescription stimulants to sentimental (eg, positive vs negative connotation, emotions), contextual (abusive vs therapeutic), and thematic (eg, feeling high) aspects of social media communications regarding problematic use of prescription drugs (eg, [33]). In a prominent work, Hanson and colleagues analyzed alternative motives and potential side effects of drug intake for nonmedical purposes (eg, as a study aid) through tweets matching [33]. Hanson and colleagues also examined how social network factors explained the nonmedical use of prescription drugs and relevant risk behavior [2]. The empirical evidence in the reviewed articles demonstrated that aggregated, time-stamped social media big data can reveal linguistic characteristics, interaction activities (eg, posting text or image content portraying substance use), needs of and thoughts on drug intake, social relations, and risk behaviors concerning problematic prescription drug and polydrug use.

Mechanisms and Predictors

There was limited work on directly examining moderating or mechanistic factors of drug abuse–related social media communications. The predictors and mechanisms of social media communications for nonmedical use of prescription drugs (eg, attitudes, risk perceptions) were not directly investigated in the studies we reviewed, and thus remain largely underexplored. Applying use and gratification theory of media can shed light on what motivates individuals to share or engage with such content on social media and what kinds of media use gratification (eg, entertainment, sharing problems for moral

support) these media activities provide for people with substance use problems [86,87]. Understanding these mechanisms of social media use for drug-related communications (eg, drug use-related activities and seeking recovery support) can produce clinical insights into how these features can be used in a clinically meaningful manner to fulfill the realistic needs of people with substance use problems. For example, people with substance use problems disclose and share information, and interact with others for substance promotional and prevention- and recovery-related purposes. These self-initiated communication activities might be associated with psychological deficits or the need for social support among substance users, or might coincide with the finding that substance users find self-disclosure activities therapeutically rewarding [57,88]. Mechanistic investigation is likely to help researchers and clinicians identify key factors when considering the designs and development of an intervention to effectively treat substance use problems.

Outcomes

The psychological and behavioral consequences (eg, increased behavioral intention for mimicking risky health behaviors) of engaging with and being exposed to social media communications regarding problematic drug use is another area of research that has been understudied. This dimension was not directly examined or empirically identified in the reviewed studies. However, the importance of understanding the varied aspects of these outcomes was discussed in some studies [7,33]. For example, Shutler and colleagues discussed the potential presumed effects of normalizing illicit drug use behavior on social media [7]. Hanson and colleagues discussed how the prevalence of tweets about nonmedical use of Adderall (eg, as a study aid) may produce a misperception that risky drug use behavior is acceptable among peer groups, and that, in turn, may lead to socially normalizing abusive drug behavior and increasing the levels of abuse [33].

Understanding short-term and long-term effects of media exposure and engagement for problematic drug use-related communications may require population-level-based surveys or longitudinal follow-up investigation in addition to social media big data analytics. By examining the psychological and behavioral effects of using social media for drug promotional or prevention- and control-related purposes, investigators can be well positioned to improve their ability to develop theoretical and methodological models when harnessing social media platforms for health promotion targeting public health problems, such as nonmedical and problematic use of prescription drugs. Investigating the outcome dimension will also help researchers understand the clinical implications and the utility of social media as behavioral intervention platforms.

Methodological Domain

For the data mining process, some studies developed and tested their own social media data monitoring and crawling systems. For example, Cameron and colleagues developed the Prescription Drug Abuse Online Surveillance and Epidemiology (PREDOSE) infrastructure to extract and analyze entities and sentiments of unstructured social media text data regarding prescription drug abuse [83,89]. Data collection periods varies

across studies, ranging from 2 weeks to more than 1 year (eg, [7,36]).

With the exception of 2 studies conducted by Correia and colleagues [4], who used Instagram data, and Cameron and colleagues [83], who did not disclose the name of the analyzed social media platform, all other reviewed studies analyzed Twitter data. Also, with the exception of the Hanson et al study [2] that explored social networks, most of the reviewed studies examined the linguistic properties of Twitter communication data. In doing so, they applied different machine learning (supervised or unsupervised learning) models to a random subset of filtered Twitter text data to identify common latent themes, patterns, and sentiments associated with nonmedical prescription drug or polydrug use (eg, [84]). A study led by Hanson and colleagues examined social networks of nonmedical or abusive use of prescription drugs and polydrug use among college students by selecting 25 subsets of tweet networks that comprised 2227 unique Twitter users. They explored social circles and interaction patterns within each network [2]. They used mixed methods involving human coding, qualitative content analysis, and manual annotation tasks, along with filtered keyword searches, as part of an iterative process to precisely understand a large volume of Twitter content promoting nonmedical use of prescription drugs, such as opioid analgesic drugs [7,33,36,84,85].

Ethical Domain

Among the reviewed studies, 4 reported that an institutional review board (IRB) approved their study [2,33,83]. Two studies reported that the IRB review was waived or was not applicable [7,36]. The status of IRB review and approval was not explicitly reported or discussed in some studies [4,84,85]. The reviewed studies used either Twitter text data or Instagram data, and these datasets were considered publicly open sources, although the topic of interest was problematic use of prescription drugs, which can be personal and risk sensitive. Some researchers reported ethical practices they applied in compliance with their IRB guidelines. For example, Cameron and colleagues [83] did not disclose the name of the social media platform they analyzed, and Kalyanam and colleagues [84] discussed the data anonymization process (eg, removing user name and profile information before analysis). In most studies, potential ethical issues and practices were not discussed in detail. This might be, in part, because the social media data in their studies was considered publicly open or because discussing ethical aspects was not directly within the scope of their study.

Discussion

The detrimental consequences of substance use highlight the urgent need for research to understand the drug epidemic in a scalable, systematic manner. Performing big data analytics on social media content allows researchers to generate data-informed insights into the phenomena of interest, such as the promotional communication of problematic substance use shared on social media platforms. The use of social media data to monitor and observe problematic use of prescription drugs, such as nonmedical use of analgesic opioid drugs, is as yet an unexplored area of biomedical research. In our reviewed studies

reported in a typology, the communication properties were used to identify high-risk events [90], indicating the potential utility of social media data as a resource in scaling up surveillance systems for substance use problems. More specifically, social media communication data aggregated by drug use–related search keywords can indicate the level and stage of drug dependence, the actions of patients engaging in addiction recovery support groups, former users with or without relapse episodes, or current users with or without dependence. Given the large scale of social media communications posted by people who have engaged or are engaging in nonmedical use of prescription drugs such as opioids, harnessing social media platforms and data will provide insight into important novel discoveries of collective public health risk behavior.

In this paper, we propose a multilevel framework and ethical considerations that are applicable to social media communication data to understand problematic drug use phenomena. Based on the 4 dimensions in the framework, along with methodological and ethical domains, we conducted a critical, narrative review of empirical findings that were based on social media communication data involving problematic use of prescription drugs. The 4 primary conceptual dimensions are (1) understanding characteristics of users who share their data (eg, texts, pictures) about nonmedical substance use on social media; (2) the communication characteristics of such self-disclosed data; (3) predictors and underlying mechanisms of social media communications on problematic use of prescription drugs; and (4) the psychological and behavioral consequences that social media use for problematic drug use–related communications may have for users themselves (eg, active users) or others (eg, observers, lurkers). The state of social media uses among people with substance use problems is receiving increased attention. We designed an evidence-based, multiconceptual framework in our typology to inform potential future research directions, which may also offer insights into public health outreach strategies, as well as the development of social media-based substance use prevention and recovery intervention programs.

Advances in communication technology and informatics offer novel opportunities for understanding substance use problems through naturally occurring, self-disclosed communications on social media. This research requires multidisciplinary collaborative efforts between data scientists, social scientists, and clinicians to systematically structure and identify the ongoing substance abuse problems observable through problematic drug use–related communications on social media. Various social media features and characteristics (eg, easy access, perceived anonymity), automated analytic approaches at scale, and the prevalence of sharing and engaging user activities for such communications underscore the benefits of

harnessing social media platforms and data to study drug use trends, patterns, and the underlying psychology and subsequent outcomes. Social media big data on the nationwide public health problem of nonmedical use of prescription drugs in the United States can also have a practical impact at the individual level (eg, seeking social support for addiction recovery support), as well as at the societal level (eg, public health campaign efforts on this topic). Although a different and extended set of ethical challenges exists in the realm of social media big data research, we expect that principle-guided, multidisciplinary, and iterative processes will soon start to converge for this topic. We also raise the importance of the precision and sensitivity of social media big data that can be prone to type I error (ie, falsely identifying and overgenerating inferences from data [3,91]). For future research, mixed methods incorporating survey research and recruitment strategies for longitudinal follow-up investigation can be used to improve the validity of social media data-driven findings.

We acknowledged a lack of theoretical frameworks that are applicable to social media big data for substance use monitoring and observational systems [91]. To bridge this gap in the research, we proposed a typology of the substance use epidemic that was traceable and observed through social media data. This knowledge synthesis is designed to analyze the state of the research on this topic and to guide future research directions. In this typology, we focused on previous work that used automated data analytics such as computational linguistic analysis and social network analysis, rather than using research outcomes purely based on manual coding of content analysis or noncomputational methods. The reviewed articles incorporated findings from supervised or unsupervised machine learning and various computational approaches. The reviewed findings demonstrated that surveillance systems incorporating social media data can produce comparable and valid findings in an epidemiological and scalable manner, in comparison with conventional survey and manual coding content analysis methods (eg, [33,84]).

In this paper, we aimed to offer key conceptual applications, ethical considerations, social media data-based empirical evidence, and a typology framework within the scope of dominant public health issues centered on nonmedical and problematic prescription drug use. As proposed in the novel typology, integrating the 4 conceptual dimensions and multidisciplinary research efforts may advance our knowledge on this nationwide crisis of prescription drug use problems in the United States. Future research may use the proposed conceptual framework and the perspectives delivered in this paper as a leverage in advancing scientific scholarship on this important topic.

Acknowledgments

SJK thanks Lamar D Moss for his insightful edits and our reviewers for suggestions for improving this article.

Authors' Contributions

SJK conceived of this project, conceptualized the study, undertook the literature review and critical review processes, and drafted the first and the last versions of the manuscript. LAM, JTH, and AKD provided recommendations and edits. All authors read drafts and approved the final manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A typology of social media big data analysis for prescription drug abuse and addiction research.

[[PDF File \(Adobe PDF File\), 28KB - jmir_v19i10e353_app1.pdf](#)]

References

1. Morgan EM, Snelson C, Elison-Bowers P. Image and video disclosure of substance use on social media websites. *Comput Hum Behav* 2010 Nov;26(6):1405-1411. [doi: [10.1016/j.chb.2010.04.017](#)]
2. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. *J Med Internet Res* 2013;15(9):e189 [FREE Full text] [doi: [10.2196/jmir.2741](#)] [Medline: [24014109](#)]
3. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. "I just wanted to tell you that loperamide WILL WORK": a web-based study of extra-medical use of loperamide. *Drug Alcohol Depend* 2013 Jun 1;130(1-3):241-244 [FREE Full text] [doi: [10.1016/j.drugalcdep.2012.11.003](#)] [Medline: [23201175](#)]
4. Correia RB, Li L, Rocha LM. Monitoring potential drug interactions and reactions via network analysis of Instagram user timelines. *Pac Symp Biocomput* 2016;21:492-503 [FREE Full text] [Medline: [26776212](#)]
5. Anderson LS, Bell HG, Gilbert M, Davidson JE, Winter C, Barratt MJ, et al. Using social listening data to monitor misuse and nonmedical use of bupropion: a content analysis. *JMIR Public Health Surveill* 2017 Dec 01;3(1):e6 [FREE Full text] [doi: [10.2196/publichealth.6174](#)] [Medline: [28148472](#)]
6. Davey Z, Schifano F, Corazza O, Deluca P, Psychonaut Web Mapping Group. e-Psychonauts: conducting research in online drug forum communities. *J Ment Health* 2012 Aug;21(4):386-394. [doi: [10.3109/09638237.2012.682265](#)] [Medline: [22823094](#)]
7. Shutler L, Nelson LS, Portelli I, Blachford C, Perrone J. Drug use in the twittersphere: a qualitative contextual analysis of tweets about prescription drugs. *J Addict Dis* 2015;34(4):303-310. [doi: [10.1080/10550887.2015.1074505](#)] [Medline: [26364675](#)]
8. Nielsen S, Barratt MJ. Prescription drug misuse: is technology friend or foe? *Drug Alcohol Rev* 2009 Jan;28(1):81-86. [doi: [10.1111/j.1465-3362.2008.00004.x](#)] [Medline: [19320680](#)]
9. Daniulaityte R, Carlson R, Brigham G, Cameron D, Sheth A. "Sub is a weird drug." a web-based study of lay attitudes about use of buprenorphine to self-treat opioid withdrawal symptoms. *Am J Addict* 2015 Aug;24(5):403-409. [doi: [10.1111/ajad.12213](#)] [Medline: [26009867](#)]
10. Miller PG, Sønderlund AL. Using the internet to research hidden populations of illicit drug users: a review. *Addiction* 2010 Sep;105(9):1557-1567. [doi: [10.1111/j.1360-0443.2010.02992.x](#)] [Medline: [20626378](#)]
11. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunaryan K, Sheth A. "When 'bad' is 'good'": identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health Surveill* 2016 Oct 24;2(2):e162 [FREE Full text] [doi: [10.2196/publichealth.6327](#)] [Medline: [27777215](#)]
12. Davis MA, Anthony DL, Pauls SD. Seeking and receiving social support on Facebook for surgery. *Soc Sci Med* 2015 Apr;131:40-47 [FREE Full text] [doi: [10.1016/j.socscimed.2015.02.038](#)] [Medline: [25753284](#)]
13. Gittelman S, Lange V, Gotway CCA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. *J Med Internet Res* 2015;17(4):e98 [FREE Full text] [doi: [10.2196/jmir.3970](#)] [Medline: [25895907](#)]
14. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2015 Jan;22(1):243-256. [doi: [10.1136/amiajnl-2014-002841](#)] [Medline: [25005606](#)]
15. Loss J, Lindacher V, Curbach J. Online social networking sites-a novel setting for health promotion? *Health Place* 2014 Mar;26:161-170. [doi: [10.1016/j.healthplace.2013.12.012](#)] [Medline: [24457613](#)]
16. Gebremeskel RH, Sessoms K, Krehnbrink M, Haney CJ, Coyne-Beasley T. Social media use and adolescent risk taking behavior. *J Adolesc Health* 2014 Feb;54(2):S46-S47. [doi: [10.1016/j.jadohealth.2013.10.106](#)]
17. Cavazos-Reh P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. *J Med Internet Res* 2014;16(6):e157 [FREE Full text] [doi: [10.2196/jmir.3247](#)] [Medline: [24974893](#)]

18. Koskan A, Klasko L, Davis SN, Gwede CK, Wells KJ, Kumar A, et al. Use and taxonomy of social media in cancer-related research: a systematic review. *Am J Public Health* 2014 Jul;104(7):e20-e37 [[FREE Full text](#)] [doi: [10.2105/AJPH.2014.301980](https://doi.org/10.2105/AJPH.2014.301980)] [Medline: [24832403](https://pubmed.ncbi.nlm.nih.gov/24832403/)]
19. Choudhury M, Morris M, White R. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 26-May 1, 2014; Toronto, ON, Canada p. 1365-1376.
20. Kass-Hout TA, Alhinnawi H. Social media in public health. *Br Med Bull* 2013;108:5-24. [doi: [10.1093/bmb/ldt028](https://doi.org/10.1093/bmb/ldt028)] [Medline: [24103335](https://pubmed.ncbi.nlm.nih.gov/24103335/)]
21. Centola D. An experimental study of homophily in the adoption of health behavior. *Science* 2011 Dec 2;334(6060):1269-1272 [[FREE Full text](#)] [doi: [10.1126/science.1207055](https://doi.org/10.1126/science.1207055)] [Medline: [22144624](https://pubmed.ncbi.nlm.nih.gov/22144624/)]
22. Centola D. The spread of behavior in an online social network experiment. *Science* 2010 Sep 3;329(5996):1194-1197 [[FREE Full text](#)] [doi: [10.1126/science.1185231](https://doi.org/10.1126/science.1185231)] [Medline: [20813952](https://pubmed.ncbi.nlm.nih.gov/20813952/)]
23. Gosling SD, Augustine AA, Vazire S, Holtzman N, Gaddis S. Manifestations of personality in online social networks: self-reported Facebook-related behaviors and observable profile information. *Cyberpsychol Behav Soc Netw* 2011 Sep;14(9):483-488 [[FREE Full text](#)] [doi: [10.1089/cyber.2010.0087](https://doi.org/10.1089/cyber.2010.0087)] [Medline: [21254929](https://pubmed.ncbi.nlm.nih.gov/21254929/)]
24. Burke M, Marlow C, Lento T. Social network activity and social well-being. 2010 Apr 10 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 10-15, 2010; Atlanta, GA, USA.
25. Lee J, Sung Y. Hide-and-seek: narcissism and “selfie”-related behavior. *Cyberpsychol Behav Soc Netw* 2016 May;19(5):347-351. [doi: [10.1089/cyber.2015.0486](https://doi.org/10.1089/cyber.2015.0486)] [Medline: [27028460](https://pubmed.ncbi.nlm.nih.gov/27028460/)]
26. Sampasa-Kanyinga H, Lewis RF. Frequent use of social networking sites is associated with poor psychological functioning among children and adolescents. *Cyberpsychol Behav Soc Netw* 2015 Jul;18(7):380-385. [doi: [10.1089/cyber.2015.0055](https://doi.org/10.1089/cyber.2015.0055)] [Medline: [26167836](https://pubmed.ncbi.nlm.nih.gov/26167836/)]
27. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, et al. Automatic personality assessment through social media language. *J Pers Soc Psychol* 2015 Jun;108(6):934-952. [doi: [10.1037/pspp0000020](https://doi.org/10.1037/pspp0000020)] [Medline: [25365036](https://pubmed.ncbi.nlm.nih.gov/25365036/)]
28. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *Proc Natl Acad Sci U S A* 2015 Jan 27;112(4):1036-1040 [[FREE Full text](#)] [doi: [10.1073/pnas.1418680112](https://doi.org/10.1073/pnas.1418680112)] [Medline: [25583507](https://pubmed.ncbi.nlm.nih.gov/25583507/)]
29. Martin EA, Bailey DH, Cicero DC, Kerns JG. Social networking profile correlates of schizotypy. *Psychiatry Res* 2012 Dec 30;200(2-3):641-646 [[FREE Full text](#)] [doi: [10.1016/j.psychres.2012.06.031](https://doi.org/10.1016/j.psychres.2012.06.031)] [Medline: [22796101](https://pubmed.ncbi.nlm.nih.gov/22796101/)]
30. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013;8(12):e83672 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
31. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in Web and social media. *Int J Environ Res Public Health* 2010 Feb;7(2):596-615 [[FREE Full text](#)] [doi: [10.3390/ijerph7020596](https://doi.org/10.3390/ijerph7020596)] [Medline: [20616993](https://pubmed.ncbi.nlm.nih.gov/20616993/)]
32. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. “Time for dabs”: analyzing Twitter data on marijuana concentrates across the U.S. *Drug Alcohol Depend* 2015 Oct 1;155:307-311. [doi: [10.1016/j.drugalcdep.2015.07.1199](https://doi.org/10.1016/j.drugalcdep.2015.07.1199)] [Medline: [26338481](https://pubmed.ncbi.nlm.nih.gov/26338481/)]
33. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *J Med Internet Res* 2013;15(4):e62 [[FREE Full text](#)] [doi: [10.2196/jmir.2503](https://doi.org/10.2196/jmir.2503)] [Medline: [23594933](https://pubmed.ncbi.nlm.nih.gov/23594933/)]
34. West JH, Hall PC, Hanson CL, Prier K, Giraud-Carrier C, Neeley ES, et al. Temporal variability of problem drinking on Twitter. *Open J Prev Med* 2012;02(01):43-48. [doi: [10.4236/ojpm.2012.21007](https://doi.org/10.4236/ojpm.2012.21007)]
35. Paul M, Dredze M. Experimenting with drugs (and topic models): multi-dimensional exploration of recreational drug discussions. 2012 Presented at: AAAI Fall Symposium: Information Retrieval and Knowledge Discovery in Biomedical Text; Nov 2-4, 2012; Arlington, VA, USA.
36. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug Saf* 2016 Mar;39(3):231-240 [[FREE Full text](#)] [doi: [10.1007/s40264-015-0379-4](https://doi.org/10.1007/s40264-015-0379-4)] [Medline: [26748505](https://pubmed.ncbi.nlm.nih.gov/26748505/)]
37. Kim SJ, Marsch LA, Guarino H, Acosta MC, Aponte-Melendez Y. Predictors of outcome from computer-based treatment for substance use disorders: results from a randomized clinical trial. *Drug Alcohol Depend* 2015 Dec 01;157:174-178 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2015.09.019](https://doi.org/10.1016/j.drugalcdep.2015.09.019)] [Medline: [26433562](https://pubmed.ncbi.nlm.nih.gov/26433562/)]
38. Yakushev A, Mityagin S. Social networks mining for analysis and modeling drugs usage. *Procedia Comput Sci* 2014;29:2462-2471. [doi: [10.1016/j.procs.2014.05.230](https://doi.org/10.1016/j.procs.2014.05.230)]
39. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J Med Internet Res* 2017 Aug 14;19(8):e289 [[FREE Full text](#)] [doi: [10.2196/jmir.7956](https://doi.org/10.2196/jmir.7956)] [Medline: [28807891](https://pubmed.ncbi.nlm.nih.gov/28807891/)]

40. Smith M. Science. Washington, DC: American Association for the Advancement of Science; 2016 Jul 13. Can social media help prevent opioid abuse? URL: <http://www.sciencemag.org/news/2016/07/can-social-media-help-prevent-opioid-abuse> [accessed 2017-10-09] [WebCite Cache ID 6u5aA64ou]
41. Daker-White G, Rogers A. What is the potential for social networks and support to enhance future telehealth interventions for people with a diagnosis of schizophrenia: a critical interpretive synthesis. *BMC Psychiatry* 2013;13:279 [FREE Full text] [doi: [10.1186/1471-244X-13-279](https://doi.org/10.1186/1471-244X-13-279)] [Medline: [24180273](https://pubmed.ncbi.nlm.nih.gov/24180273/)]
42. Rubya S, Yarosh S. Video-mediated peer support in an online community for recovery from substance use disorders. 2017 Mar 01 Presented at: 2017 ACM Conference on Computer Supported Cooperative Work Social Computing; Feb 25-Mar 1, 2017; Portland, OR, USA p. 1454-1469.
43. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. *Patient Educ Couns* 2013 Dec;93(3):472-479. [doi: [10.1016/j.pec.2013.07.009](https://doi.org/10.1016/j.pec.2013.07.009)] [Medline: [23928354](https://pubmed.ncbi.nlm.nih.gov/23928354/)]
44. Babb F. venturebeat.com. San Francisco, CA: VentureBeat; 2014 Sep 19. Instagram has a drug problem URL: <http://venturebeat.com/2014/09/19/instagram-has-a-drug-problem> [accessed 2014-10-15] [WebCite Cache ID 6rZEFiVG]
45. van Hoof JJ, Bekkers J, van Vuuren M. Son, you're smoking on Facebook! College students' disclosures on social networking sites as indicators of real-life risk behaviors. *Comput Hum Behav* 2014 May;34:249-257. [doi: [10.1016/j.chb.2014.02.008](https://doi.org/10.1016/j.chb.2014.02.008)]
46. The National Center on Addiction and Substance Abuse at Columbia University. National Survey of American Attitudes on Substance Abuse XVI: Teens And Parents. New York, NY: CASAColumbia; 2011 Aug. URL: <https://www.centeronaddiction.org/addiction-research/reports/national-survey-american-attitudes-substance-abuse-teens-parents-2011> [accessed 2017-10-09] [WebCite Cache ID 6u5aHdmok]
47. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. *J Med Internet Res* 2014;16(3):e79 [FREE Full text] [doi: [10.2196/jmir.2679](https://doi.org/10.2196/jmir.2679)] [Medline: [24642014](https://pubmed.ncbi.nlm.nih.gov/24642014/)]
48. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents: prevalence and associations. *Arch Pediatr Adolesc Med* 2009 Jan;163(1):27-34. [doi: [10.1001/archpediatrics.2008.528](https://doi.org/10.1001/archpediatrics.2008.528)] [Medline: [19124700](https://pubmed.ncbi.nlm.nih.gov/19124700/)]
49. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. *Arch Pediatr Adolesc Med* 2012 Feb;166(2):157-163 [FREE Full text] [doi: [10.1001/archpediatrics.2011.180](https://doi.org/10.1001/archpediatrics.2011.180)] [Medline: [21969360](https://pubmed.ncbi.nlm.nih.gov/21969360/)]
50. Gunther AC, Storey JD. The influence of presumed influence. *J Commun* 2003 Jun;53(2):199-215. [doi: [10.1111/j.1460-2466.2003.tb02586.x](https://doi.org/10.1111/j.1460-2466.2003.tb02586.x)]
51. Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004 Apr;31(2):143-164. [doi: [10.1177/1090198104263660](https://doi.org/10.1177/1090198104263660)] [Medline: [15090118](https://pubmed.ncbi.nlm.nih.gov/15090118/)]
52. Gerbner G, Gross L, Morgan M, Signorielli N. Living with television: the dynamics of the cultivation. In: Bryant J, Zillman D, editors. *Perspectives on Media Effects*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.
53. Bilandzic H. The complicated relationship between media and morality. *J Media Psychol* 2011 Jan;23(1):46-51. [doi: [10.1027/1864-1105/a000032](https://doi.org/10.1027/1864-1105/a000032)]
54. Ajzen I, Fishbein M. Attitude-behavior relations: a theoretical analysis and review of empirical research. *Psychol Bull* 1977;84(5):888-918. [doi: [10.1037/0033-2909.84.5.888](https://doi.org/10.1037/0033-2909.84.5.888)]
55. Bentley RA, O'Brien MJ, Brock WA. Mapping collective behavior in the big-data era. *Behav Brain Sci* 2014 Mar 31;37(01):63-76. [doi: [10.1017/s0140525x13000289](https://doi.org/10.1017/s0140525x13000289)]
56. Bandura A. Social cognitive theory for personal and social change by enabling media. In: Singhal A, Cody MJ, Rogers EM, Sabido M, editors. *Entertainment-Education and Social Change: History, Research, and Practice*. Mahwah, NJ: Lawrence Erlbaum; 2004:75-96.
57. Kim SJ, Marsch LA, Brunette MF, Dallery J. Harnessing Facebook for smoking reduction and cessation interventions: Facebook user engagement and social support predict smoking reduction. *J Med Internet Res* 2017 May 23;19(5):e168 [FREE Full text] [doi: [10.2196/jmir.6681](https://doi.org/10.2196/jmir.6681)] [Medline: [28536096](https://pubmed.ncbi.nlm.nih.gov/28536096/)]
58. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. *Addict Behav* 2017 Oct;73:81-93. [doi: [10.1016/j.addbeh.2017.05.002](https://doi.org/10.1016/j.addbeh.2017.05.002)] [Medline: [28499259](https://pubmed.ncbi.nlm.nih.gov/28499259/)]
59. Patrick K, Marshall SJ, Davila EP, Kolodziejczyk JK, Fowler JH, Calfas KJ, et al. Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). *Contemp Clin Trials* 2014 Jan;37(1):10-18 [FREE Full text] [doi: [10.1016/j.cct.2013.11.001](https://doi.org/10.1016/j.cct.2013.11.001)] [Medline: [24215774](https://pubmed.ncbi.nlm.nih.gov/24215774/)]
60. Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. Drug overdose deaths hit record numbers in 2014. Atlanta, GA: CDC; 2015. URL: <https://www.cdc.gov/media/releases/2015/p1218-drug-overdose.html> [accessed 2017-08-28] [WebCite Cache ID 6t3w2sf97]
61. National Institutes of Health, National Institute on Drug Abuse. Overdose death rates. Bethesda, MD: NIDA; 2017. URL: <https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates> [accessed 2017-08-29] [WebCite Cache ID 6t4ETk7DK]

62. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Prev Med* 2014 Jun;63:112-115 [FREE Full text] [doi: [10.1016/j.ypmed.2014.01.024](https://doi.org/10.1016/j.ypmed.2014.01.024)] [Medline: [24513169](https://pubmed.ncbi.nlm.nih.gov/24513169/)]
63. Bazarova NN, Choi YH. Self-disclosure in social media: extending the functional approach to disclosure motivations and characteristics on social network sites. *J Commun* 2014 Jun 19;64(4):635-657. [doi: [10.1111/jcom.12106](https://doi.org/10.1111/jcom.12106)]
64. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. *Psychol Assess* 2009 Dec;21(4):486-497 [FREE Full text] [doi: [10.1037/a0017074](https://doi.org/10.1037/a0017074)] [Medline: [19947783](https://pubmed.ncbi.nlm.nih.gov/19947783/)]
65. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1-32. [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
66. Rawassizadeh R, Momeni E, Dobbins C, Gharibshah J, Pazzani M. Scalable daily human behavioral pattern mining from multivariate temporal data. *IEEE Trans Knowl Data Eng* 2016 Nov 1;28(11):3098-3112. [doi: [10.1109/TKDE.2016.2592527](https://doi.org/10.1109/TKDE.2016.2592527)]
67. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. *J Med Internet Res* 2017 Feb 28;19(2):e48 [FREE Full text] [doi: [10.2196/jmir.6895](https://doi.org/10.2196/jmir.6895)] [Medline: [28246066](https://pubmed.ncbi.nlm.nih.gov/28246066/)]
68. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 2016 Apr;22(2):303-341. [doi: [10.1007/s11948-015-9652-2](https://doi.org/10.1007/s11948-015-9652-2)] [Medline: [26002496](https://pubmed.ncbi.nlm.nih.gov/26002496/)]
69. Lucas BJ, Galinsky AD, Murnighan KJ. An intention-based account of perspective-taking: why perspective-taking can both decrease and increase moral condemnation. *Pers Soc Psychol Bull* 2016 Sep 20;42(11):1480-9. [doi: [10.1177/0146167216664057](https://doi.org/10.1177/0146167216664057)] [Medline: [27655749](https://pubmed.ncbi.nlm.nih.gov/27655749/)]
70. Eyal T, Epley N. How to seem telepathic: enabling mind reading by matching construal. *Psychol Sci* 2010 May;21(5):700-705. [doi: [10.1177/0956797610367754](https://doi.org/10.1177/0956797610367754)] [Medline: [20483849](https://pubmed.ncbi.nlm.nih.gov/20483849/)]
71. Shah DV, Cappella JN, Neuman WR. Big data, digital media, and computational social science: possibilities and perils. *Ann Am Acad Polit Soc Sci* 2015 Apr 09;659(1):6-13. [doi: [10.1177/0002716215572084](https://doi.org/10.1177/0002716215572084)]
72. Harvard Catalyst Regulatory Foundations Ethics, & Law Program. The use of social media in recruitment to research: a guide for investigators and IRBs. Boston, MA: Harvard Catalyst URL: https://catalyst.harvard.edu/pdf/regulatory/Social_Media_Guidance.pdf [accessed 2017-10-09] [WebCite Cache ID [6u5mOq0nD](https://www.webcitation.org/6u5mOq0nD)]
73. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci U S A* 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: [10.1073/pnas.1320040111](https://doi.org/10.1073/pnas.1320040111)] [Medline: [24889601](https://pubmed.ncbi.nlm.nih.gov/24889601/)]
74. Griffin M. Teaching cognitive rehearsal as a shield for lateral violence: an intervention for newly licensed nurses. *J Contin Educ Nurs* 2004;35(6):257-263. [Medline: [15584678](https://pubmed.ncbi.nlm.nih.gov/15584678/)]
75. Koh WMS. Management of work place bullying in hospital: a review of the use of cognitive rehearsal as an alternative management strategy. *Int J Nurs Sci* 2016 Jun;3(2):213-222. [doi: [10.1016/j.ijnss.2016.04.010](https://doi.org/10.1016/j.ijnss.2016.04.010)]
76. Schroeder R. Big data and the brave new world of social media research. *Big Data Soc* 2014 Dec 17;1(2). [doi: [10.1177/2053951714563194](https://doi.org/10.1177/2053951714563194)]
77. Hancock JT. Introduction to Ethics of Digital Research: The Oxford Handbook of Networked Communication. Oxford, UK: Oxford University Press; 2017.
78. Docherty A. Big data--ethical perspectives. *Anaesthesia* 2014 Apr;69(4):390-391 [FREE Full text] [doi: [10.1111/anae.12656](https://doi.org/10.1111/anae.12656)] [Medline: [24641650](https://pubmed.ncbi.nlm.nih.gov/24641650/)]
79. Fairfield J, Shtein H. Big data, big problems: emerging issues in the ethics of data science and journalism. *J Mass Media Ethics* 2014 Jan 16;29(1):38-51. [doi: [10.1080/08900523.2014.863126](https://doi.org/10.1080/08900523.2014.863126)]
80. Mittelstadt B, Fairweather B, Shaw M, McBride N. The ethical implications of personal health monitoring. *Int J Technoethics* 2014;5(2):37-60. [doi: [10.4018/ijt.2014070104](https://doi.org/10.4018/ijt.2014070104)]
81. Markham A, Buchanan E. Ethical decision-making and Internet research: recommendations from the AoIR Ethics Working Committee (version 2.0). Chicago, IL: Association of Internet Researchers; 2012. URL: <http://aoir.org/reports/ethics2.pdf> [accessed 2017-10-09] [WebCite Cache ID [6u5mwuUEq](https://www.webcitation.org/6u5mwuUEq)]
82. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009 Jun;26(2):91-108 [FREE Full text] [doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)] [Medline: [19490148](https://pubmed.ncbi.nlm.nih.gov/19490148/)]
83. Cameron D, Sheth AP, Jaykumar N, Thirunarayan K, Anand G, Smith GA. A hybrid approach to finding relevant social media content for complex domain specific information needs. *Web Semant* 2014 Dec;29:39-52 [FREE Full text] [doi: [10.1016/j.websem.2014.11.002](https://doi.org/10.1016/j.websem.2014.11.002)] [Medline: [25814917](https://pubmed.ncbi.nlm.nih.gov/25814917/)]
84. Kalyanam J, Katsuki T, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. *Addict Behav* 2017 Feb;65:289-295. [doi: [10.1016/j.addbeh.2016.08.019](https://doi.org/10.1016/j.addbeh.2016.08.019)] [Medline: [27568339](https://pubmed.ncbi.nlm.nih.gov/27568339/)]
85. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data. *J Med Internet Res* 2015 Dec 16;17(12):e280 [FREE Full text] [doi: [10.2196/jmir.5144](https://doi.org/10.2196/jmir.5144)] [Medline: [26677966](https://pubmed.ncbi.nlm.nih.gov/26677966/)]
86. Raacke J, Bonds-Raacke J. MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychol Behav* 2008 Apr;11(2):169-174. [doi: [10.1089/cpb.2007.0056](https://doi.org/10.1089/cpb.2007.0056)] [Medline: [18422409](https://pubmed.ncbi.nlm.nih.gov/18422409/)]
87. Quan-Haase A, Young AL. Uses and gratifications of social media: a comparison of Facebook and instant messaging. *Bull Sci Technol Soc* 2010 Sep 14;30(5):350-361. [doi: [10.1177/0270467610380009](https://doi.org/10.1177/0270467610380009)]

88. Tamir DI, Mitchell JP. Disclosing information about the self is intrinsically rewarding. *Proc Natl Acad Sci U S A* 2012 May 22;109(21):8038-8043 [[FREE Full text](#)] [doi: [10.1073/pnas.1202129109](https://doi.org/10.1073/pnas.1202129109)] [Medline: [22566617](https://pubmed.ncbi.nlm.nih.gov/22566617/)]
89. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. *J Biomed Inform* 2013 Dec;46(6):985-997 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2013.07.007](https://doi.org/10.1016/j.jbi.2013.07.007)] [Medline: [23892295](https://pubmed.ncbi.nlm.nih.gov/23892295/)]
90. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol* 2016 Jun;9:77-82 [[FREE Full text](#)] [doi: [10.1016/j.copsyc.2016.01.004](https://doi.org/10.1016/j.copsyc.2016.01.004)] [Medline: [27042689](https://pubmed.ncbi.nlm.nih.gov/27042689/)]
91. Mounteney J, Fry C, McKeganey N, Haugland S. Challenges of reliability and validity in the identification and monitoring of emerging drug trends. *Subst Use Misuse* 2010;45(1-2):266-287. [doi: [10.3109/10826080903368598](https://doi.org/10.3109/10826080903368598)] [Medline: [20025453](https://pubmed.ncbi.nlm.nih.gov/20025453/)]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

IRB: institutional review board

PREDOSE: Prescription Drug Abuse Online Surveillance and Epidemiology

Edited by G Eysenbach; submitted 29.07.16; peer-reviewed by D Broniatowski, L Thornton, R Daniulaityte, M Conway, J Mitchell; comments to author 02.12.16; revised version received 01.09.17; accepted 20.09.17; published 31.10.17

Please cite as:

Kim SJ, Marsch LA, Hancock JT, Das AK

Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data

J Med Internet Res 2017;19(10):e353

URL: <http://www.jmir.org/2017/10/e353/>

doi: [10.2196/jmir.6426](https://doi.org/10.2196/jmir.6426)

PMID: [29089287](https://pubmed.ncbi.nlm.nih.gov/29089287/)

©Sunny Jung Kim, Lisa A Marsch, Jeffrey T Hancock, Amarendra K Das. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 31.10.2017. 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 <http://www.jmir.org/>, as well as this copyright and license information must be included.