A retrospective analysis of NIH-funded digital health research using social media platforms

Camille Nebeker[1] (iD), Sarah E. Dunseath[2] and Rubi Linares-Orozco[3]

Abstract

Objective: Social network platforms are increasingly used in digital health research. Our study aimed to 1. qualify and quantify the use of social media platforms in health research supported by the National Institutes of Health (NIH) and document changes occurring between 2011 and 2017 and 2. examine whether institutions hosting these studies provided public-facing guidelines on how to conduct ethical social media health research.

Methods: The NIH RePORTER (Research Portfolio Online Reporting Tools) database was searched to identify research utilizing Instagram, Pinterest, Facebook, or Twitter. Studies included used social media for observational research, recruitment, intervention delivery or to assess social media as an effective research tool. Abstracts were qualitatively analyzed to describe the population and health topic by year. Websites of organizations receiving funding for this research were searched to identify whether guidance or policy existed.

Results: Studies ($n = 105$) were organized by population targeted and health focus. Main "Health" themes were labeled: 1. substance use, 2. disease/diagnosis, 3. psychiatry/mental health, and 4. weight and physical activity. The populations most involved included adolescents and young adults, and men who have sex with men. The number of research studies using social media increased approximately 590% between 2011 and 2017. Studies were linked to 56 organizations of which 21% ($n = 12$) provided some accessible guidance with 79% ($n = 44$) offering no guidance specific to social media health research.

Conclusions: Social media research is conducted with vulnerable populations that are traditionally difficult to reach. There is a compelling need for resources designed to support ethical and responsible social media-enabled research to enable this research to be carried out safely.

Introduction

The practice of health research has changed dramatically over the past decade, largely due to technological advancements. Scientists can now access individual-level participant data from mobile apps, social networks and a variety of pervasive sensing technologies, like wearable fitness tracking devices. Specifically, social network platforms like Facebook, Twitter and Instagram are increasingly being leveraged by researchers for study recruitment as well as surveillance and intervention delivery.[1,2] These platforms provide data

[1]Department of Family Medicine and Public Health, School of Medicine, University of California, San Diego, USA
[2]School of Medicine, University of California, San Diego, USA
[3]Compliance Advisory Services, University of California, San Diego, USA

Corresponding author:
Camille Nebeker, Department of Family Medicine and Public Health, UC San Diego, 9500 Gilman Drive, MC 0811, La Jolla, CA, 92093, USA.
Email: nebeker@eng.ucsd.edu
Twitter: @cnebeker

sources that serve as attractive alternatives to existing methods for detecting, for example, early signs of pending natural disasters as well as infectious disease outbreaks.[3,4] Publicly available tweets have also been used for early earthquake detection, which can facilitate rapid notification of at-risk populations.[5] As with many big data sources, these data can be used passively (users are not aware of how their data are used) and actively (users are aware they are contributing), which raises interesting ethical issues around informed consent, privacy and data management (e.g., collection, storage and sharing) practices.

Over the past decade, consumer use of social media platforms has increased dramatically: in 2005 5% of the US population reported having at least one social media account, whereas in 2017 it rose to 70% with the majority of users accessing these platforms daily.[6] As more people use social media to access news, health information, and their personal and professional contacts, and as the social media user-base continues to expand to reflect broader demographics, there are increasingly novel research opportunities. To support these opportunities, the National Institutes of Health (NIH) has increased its support for research that leverages social network platforms. While potentially transformational, these new and emerging technologies are challenging the existing regulatory infrastructures originally developed to protect people who agree to participate in health research.[7,8]

The regulatory and ethical landscape

The ethical challenges, in some respects, are no different than those encountered with traditional research, which tend to focus on informed consent, participant privacy, and data management, including practices to ensure confidentiality of information collected. Yet, when using social media as a research tool, there are different challenges that are both new and nuanced. For example, the controversial "Facebook Emotional Contagion" study challenged the norms for informed consent when Facebook researchers manipulated the news feeds of its user base without directly obtaining informed consent from those involved in the experiment.[9] When consumers learned that Facebook was conducting experiments on them, there was public outcry that participation in research without explicit informed consent was unacceptable, which led to dialogue about whether Facebook's terms of service (ToS) could claim to be a proxy for informed consent to participate in research.[10] More recently, Cambridge Analytica accessed information from Facebook user profiles to micro-target political campaign messages.[11] Grindr, a social network for gay, bi and trans men, allowed third parties to access encrypted forms of sensitive data, such as the HIV status of its users.[12] OkCupid, an online dating social network, also experienced problems when researchers accessed its publicly accessible user data and then posted user demographics, personality traits, and relationship preferences without user permission.[13]

These examples demonstrate that the risks of using social media in research may not be immediately apparent to scientists conducting the research. For academic researchers, the Institutional Review Board or IRB is the resource for ethical and regulatory guidance; however, Gelinas et al. found that some IRBs have guidance but most do not, and even when there are guidelines there is a lack of consensus on defining and managing the issues.[7] For example, a recent study reported that Twitter users expect to provide consent if their public tweets are analyzed for research purposes,[14] and another study suggested that consent should be obtained prior to quoting a publicly accessible tweet in an academic manuscript.[15] However, we found no published guidelines that reflect these expectations and potential violations of the ethical principles that guide human research. In spite of these publicized breaches and new ethical challenges, the federal offices in the United States that provide policy and guidance on human research protections have not developed standards specific to research involving social media platforms.

The purpose of this study was to qualify and quantify the use of social media platforms in health research supported by the NIH and to document changes over time by using the NIH RePORTER (Research Portfolio Online Reporting Tools) database. We then examined the extent to which the institutions hosting these studies provided public-facing guidelines or policy to assist investigators using social media strategies in health research and conclude with recommendations and resources.

Methods

The NIH RePORTER database was accessed to 1. quantify how social media was used as a research tool and 2. qualify participant groups involved and health areas targeted. The NIH RePORTER database includes research projects funded by the NIH and its affiliated agencies and institutes. The time frame selected for this search was research funded between 2011 and 2017. The NIH RePORTER query form was used to identify potentially relevant projects using the four key terms of "Facebook," "Instagram," "Twitter," and "Pinterest" in the title or abstract. The keyword that identified each project was recorded. Projects identified in multiple, distinct keyword searches were consolidated to eliminate duplicates. Only projects in which any

of the four major social media platforms played an integral role in methodology were included. Specifically, in order to meet inclusion criteria, one or more of the following criteria must have been met: 1. social media was used to deliver the intervention, 2. social media was used to recruit participants into a study, 3. social media data were analyzed by researchers as a means of surveillance, or 4. social media platforms were evaluated for their feasibility and compared to preexisting methods used in research. NIH-supported social media studies conducted over multiple years were identified to account for duplicates and counted only in the first year of funding (Figure 1).

Projects meeting inclusion criteria ($n = 105$) were organized by population targeted and health focus. Populations were labeled 1. "age" with groupings of adolescents/young adults, children, geriatrics; 2. "race/ethnicity" with groupings of African American, Hispanic, and Asian; 3. "sexual orientation" consisting primarily of men who have sex with men (MSM); 4. "women" grouped as pregnant or postpartum, and 5. "other," which accounted for all populations that were not discussed in at least two studies. For example, one project in the data set studied hospice workers, so this population was placed in the "other" category. Some studies targeted multiple populations, each of which was accounted for in the data set. For example, if a study specifically targeted young adult African American MSM, then three populations would be listed for that project: 1. young adult, 2. African American, and 3. MSM.

Main "health" themes were labeled 1. substance use, 2. disease/diagnosis, 3. psychiatry/mental health, and 4. weight and physical activity. Studies were categorized into their appropriate themes via a keyword search (see Table 1).

The health themes were further organized into subthemes. For example, the subtheme labeled "substance use" included categories labeled alcohol, smoking (including tobacco, e-cigarettes, and marijuana), opioids, and stimulants (including amphetamines and cocaine). Disease/Diagnosis was broken down into "psychiatry and mental health," "cancer," "diabetes," "HIV," "cardiovascular disease," "vaccination," and "other" yielded by a search using the keyword terms "disease" and "diagnosis." The "weight and physical activity" subtheme included categories labeled "physical activity," "weight loss," and "obesity." Some themes overlapped, for example diabetes and weight. Some subthemes within a theme overlapped, such as marijuana and tobacco/smoking within the theme of substance use (see Table 1). A researcher (CN) independently reviewed 10% of these projects to confirm the organization of results. Disputed placement of projects within the themes and subthemes was discussed between two researchers (CN and SD) and 100% agreement was reached. We then identified patterns over time specific to new populations and themes.

Our second research question focused on the extent to which institutions hosting social media-enabled health research provide guidance to investigators. Of the 56 institutions identified using the NIH RePORTER query search, the institutional website was searched using the following keywords: "Institutional Review Board," "IRB," "Human Research Protections Program," and "HRPP." The IRB website was then reviewed for publicly accessible policies and guidance using the following keywords: "internet research" and "social media." Institutions without a public-facing institutional or company website were noted as "missing" and not explored further. Institution webpages that required a username and
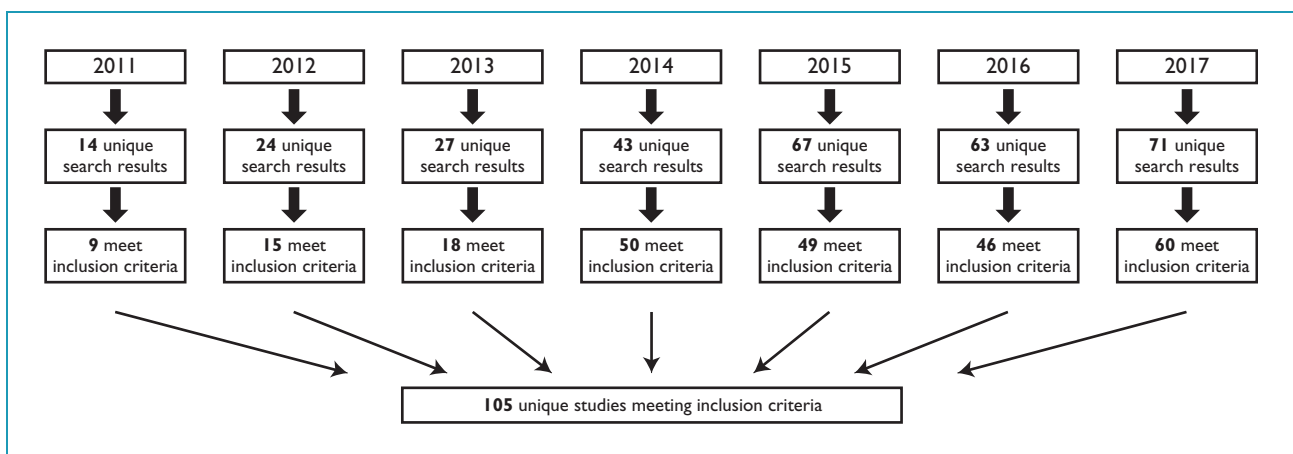

Figure 1. Process of inclusion of NIH-funded social media research.

Table 1. Themes in social media research as identified by keyword search 2011–2017.

Broad theme	Substance use	Disease/diagnosis	Weight & activity	Psychiatry & mental health
Terms	Substance use/abuse Alcohol Smoking Tobacco Marijuana Opioid/opiate Heroin Cocaine Methamphetamine Ecstasy/MDMA e-cigarettes LSD Prescription pills Stimulants	Cancer Oncology Carcinoma Sarcoma Diabetes HIV Cardiovascular disease Hypertension Heart disease Osteoporosis Vaccination/immunization	Physical activity Weight loss Weight gain Obesity Sedentary behavior	Psychiatry Psychology Mental health Depression Post-traumatic stress disorder Anxiety Suicide/suicidal Bipolar disorder

password before allowing access to the website were coded as "inaccessible."

Results

The number of research studies using social media increased approximately 570% from nine studies receiving NIH funding in 2011 to 60 studies in 2017. Of those, 50% ($n = 53$) used Facebook, 26% ($n = 27$) used Twitter, 14% ($n = 15$) used both Facebook and Twitter with the remaining 10% using a combination of Instagram or Pinterest with either Facebook or Twitter. In this period, the number of specifically targeted populations expanded to include broader demographics (see Table 2) and health foci (see Table 3). In 2011, the only specific populations examined were African Americans, Hispanics, adolescents/young adults, and MSM. By 2017, populations had expanded to include infants and children, prenatal and postpartum women, and transgender and lesbian subjects. The most consistently targeted populations were adolescents/young adults who comprised between 33.3 and 56.5% of the total NIH-funded social media research across all 6 years, followed by MSM at 6.7–33.3% across the 6 years (see Table 2).

From 2011 to 2017, social media continued to be used as a means of intervention in health research (33.3–66.7%), as a tool for platform assessment (33.3–45.6%), as a recruitment tool (11.1–30.0%), and as a means of surveillance (36.7–56.7%). Over this period, social media was increasingly used to recruit participants in studies. In 2011, 11.1% of studies used social media for recruitment, and the proportion increased each year until 2017, at which point 30.0% of studies used social media for recruitment (See Table 3).

Primary populations of focus

Adolescent and Young Adults. From 2011–2017 adolescent/young adult social media research increased. In 2011, the only funded project that focused on young adults studied physical activity, weight gain, depression, HIV and substance use. By 2017, young adult research included the diverse domains of mental health including specifically depression, suicidality, anxiety, post-traumatic stress disorder (PTSD) and health-related topics including cancer, diabetes, hypertension, heart disease, vaccination and immunization. In this population, social media was most consistently used as a means of intervention delivery (33.3–100.0%), and was also commonly used for platform assessment (32.4–53.8%), and surveillance (32.0–66.7%). Less commonly, social media was used as a recruitment tool, though the use of social media in this capacity did consistently increase each year over the course of this study (0.0–26.5%) (see Table 4).

Men who have sex with men (MSM). MSM, the second most studied population, had a total of 16 unique studies from 2011–2017. Of these, four also focused specifically on transgender populations and three on lesbian populations. In this time frame, 13 funded studies focused on HIV (81.3%), four focused on substance use (25.0%), three on mental health (18.8%), one on cancer (6.3%), and one on physical activity (6.3%). Social media was most consistently used as intervention delivery in the MSM population (33.3–100.0%). It was also consistently used for platform assessment (33.3–100%) and surveillance (33.3–66.7%). Social media was not used for recruitment of MSM participants until 2015, at which point 33.3% of studies used social media for recruitment purposes. In 2016 and 2017 the percentage of studies using social

Table 2. NIH-funded social media studies targeting specific populations.

Year/key term	African American N (%)	Hispanic N (%)	Infants/ children N (%)	Young adults/ adolescents/teenagers N (%)	MSM, lesbian, transgender, gay N (%)	Pregnant, pregnancy N (%)
2011 n = 9	1 (11.1)	1 (11.1)	0 (0)	3 (33.3)	1 (11.1)	0 (0.0)
2012 n = 15	3 (20.0)	1 (6.7)	0 (0)	7 (46.7)	5 (33.3)	0 (0.0)
2013 n = 18	3 (16.7)	1 (5.6)	1 (5.6)	9 (50.0)	3 (16.7)	1 (5.6)
2014 n = 30	2 (15.4)	1 (3.3)	1 (3.3)	13 (43.3)	2 (6.7)	1 (3.3)
2015 n = 49	8 (16.3)	3 (6.1)	2 (4.1)	25 (51.0)	6 (12.2)	0 (0)
2016 n = 46	7 (15.2)	2 (4.3)	1 (2.2)	26 (56.5)	5 (10.9)	2 (4.3)
2017 n = 60	11 (18.3)	5 (8.3)	1 (1.7)	34 (56.7)	8 (13.3)	3 (5.0)
Total unique studies accounting for duplicates across multiple years N = 105	15 (14.3)	7 (6.7)	2 (1.9)	51 (48.6)	16 (15.2)	4 (3.8)

The following keywords did not yield any projects: e-cigarettes, stimulants, prescription pills, LSD, MDMA, bipolar disorder, oncology, sarcoma, osteoporosis, sedentary behavior, geriatric, Indian, Middle Eastern.

Table 3. Use of social media in NIH-funded research.

Year/use	2011 n = 9 (%)	2012 n = 15 (%)	2013 n = 18 (%)	2014 n = 30 (%)	2015 n = 49 (%)	2016 n = 46 (%)	2017 n = 60 (%)
Intervention	6 (66.7)	8 (53.3)	8 (44.4)	10 (33.3)	18 (36.7)	20 (43.5)	30 (50.0)
Platform assessment	5 (55.6)	6 (40.0)	6 (33.3)	11 (36.7)	18 (36.7)	19 (41.3)	27 (45.0)
Recruitment	1 (11.1)	2 (13.3)	3 (16.7)	6 (20.0)	12 (24.5)	12 (26.1)	18 (30.0)
Surveillance	4 (44.4)	7 (46.7)	9 (50.0)	17 (56.7)	24 (49.0)	23 (50.0)	22 (36.7)

media increased to 60.0% and 50.0% respectively (see Table 5).

Key health focus

Substance use. A total of 50 social media research studies focused on substance use between 2011 and 2017, and the scope of this research broadened between these time points. Social media in substance use research only targeted alcohol or smoking/tobacco use in 2011, but expanded to include cocaine and methamphetamine in 2012, and marijuana, and opioids/ heroin use in 2014. Research focused on alcohol and tobacco/marijuana have continued to be examined via social media research, accounting for 64% and 54% respectively of substance use research from 2011–2017 (Table 6).

Table 4. Use of social media research in adolescents/young adults 2011–2017.

Year/keyword	2011 n = 3 (%)	2012 n = 7 (%)	2013 n = 9 (%)	2014 n = 13 (%)	2015 n = 25 (%)	2016 n = 26 (%)	2017 n = 34 (%)
Substance use							
Alcohol	2 (66.7)	4 (57.1)	4 (44.4)	7 (53.8)	10 (40.0)	12 (46.2)	13 (38.2)
Tobacco/smoking	1 (33.3)	1 (14.3)	1 (11.1)	6 (46.2)	6 (24.0)	7 (26.9)	9 (26.5)
Marijuana	0 (0.0)	0 (0.0)	0 (0.0)	3 (23.1)	4 (16.0)	3 (11.5)	4 (11.8)
Opioids	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Cocaine	0 (0.0)	1 (14.3)	1 (11.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Methamphetamine	0 (0.0)	1 (14.3)	1 (11.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Ecstasy/MDMA	0 (0.0)	0 (0.0)	0 (0.0)	1 (7.7)	1 (4.0)	0 (0.0)	0 (0.0)
Psychology/mental health							
Depression	1 (33.3)	4 (57.1)	4 (44.4)	3 (23.1)	3 (12.0)	3 (11.5)	3 (8.8)
PTSD	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.9)
Anxiety	0 (0.0)	1 (14.3)	1 (11.1)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.9)
Suicide	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.9)
Schizophrenia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disease/ diagnosis							
Cancer	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (7.7)	5 (14.7)
Diabetes	0 (0.0)	0 (0.0)	0 (0.0)	1 (7.7)	1 (4.0)	1 (3.8)	2 (5.7)
HIV	1 (33.3)	3 (42.9)	2 (22.2)	2 (15.4)	7 (28.0)	5 (19.2)	6 (17.6)
Cardiovascular hypertension, heart failure	0 (0.0)	0 (0.0)	0 (0.0)	1 (7.7)	1 (4.0)	1 (3.8)	2 (5.9)
Vaccination/ Immunization	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (4.0)	0 (0.0)	1 (2.9)
Physical activity/weight							
Physical activity	1 (33.3)	1 (14.3)	2 (22.2)	1 (7.7)	1 (4.0)	1 (3.8)	2 (5.9)
Weight loss	0 (0.0)	0 (0.0)	0 (0.0)	1 (7.7)	1 (4.0)	1 (3.8)	1 (2.9)
Weight gain	1 (33.3)	1 (14.3)	2 (22.2)	3 (23.1)	3 (12.0)	2 (7.7)	2 (5.9)
Obesity	0 (0.0)	0 (0.0)	1 (11.1)	3 (23.1)	4 (16.0)	3 (11.5)	2 (5.9)
Use of social media							
Intervention	2 (66.7)	3 (42.9)	4 (44.4)	6 (46.2)	13 (52.0)	16 (61.5)	20 (58.8)
Platform assessment	1 (33.3)	3 (42.9)	3 (33.3)	7 (53.8)	10 (40.0)	11 (42.3)	11(32.4)

Table 4. Continued

Year/keyword	2011 n = 3 (%)	2012 n = 7 (%)	2013 n = 9 (%)	2014 n = 13 (%)	2015 n = 25 (%)	2016 n = 26 (%)	2017 n = 34 (%)
Recruitment	0 (0.0)	1 (14.3)	2 (22.2)	3 (23.1)	6 (24.0)	7 (26.9)	9 (26.5)
Surveillance	2 (66.7)	4 (57.1)	4 (44.4)	5 (38.5)	8 (32.0)	9 (34.6)	16 (47.1)

Table 5. Use of social media research in MSM 2011–2017.

Year/key term	2011 n = 3	2012 n = 5	2013 n = 3	2014 n = 2	2015 n = 6	2016 n = 5	2017 n = 8
Substance use							
Alcohol	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (16.7)	1 (0.20)	1 (12.5)
Cocaine	0 (0.0)	1 (20.0)	1 (33.3)	0 (0.0)	1 (16.7)	1 (0.20)	0 (0.0)
Ecstasy	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (16.7)	1 (0.20)	0 (0.0)
Methamphetamine	0 (0.0)	1 (20.0)	1 (33.3)	0 (0.0)	1 (16.7)	1 (0.20)	0 (0.0)
Smoking/tobacco	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (12.5)
Psychology/mental health							
Mental health	0 (0.0)	1 (20.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (12.5)
PTSD	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Suicide	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disease/diagnosis							
HIV	**3 (100.00)**	**5 (100.0)**	**3 (100.0)**	**2 (100.0)**	**6 (100.0)**	**4 (80.0)**	**5 (62.5)**
Cancer	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (12.5)
Physical activity/weight							
Physical activity	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (12.5)
Use of social media in research							
Intervention	3 (100.0)	3 (60.0)	1 (33.3)	2 (100.0)	3 (50.0)	3 (60.0)	6 (75.0)
Platform assessment	1 (33.3)	2 (40.0)	2 (66.7)	2 (100.0)	2 (33.3)	2 (40.0)	3 (37.5)
Recruitment	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (33.3)	3 (60.0)	4 (50.0)
Surveillance	1 (33.3)	3 (60.0)	2 (66.7)	0 (0.0)	2 (33.3)	0 (0.0)	3 (37.5)

Disease/diagnosis. Research on disease/diagnosis including cancer, diabetes, HIV, cardiovascular disease and vaccinations/immunization made up 41.0% of NIH-funded social media research from 2011–2017. Four of the nine projects in 2011 were related to one of these conditions and, of these, three examined HIV and one examined cancer. In 2017, 26 of the 61 NIH-funded social media research projects examined one of these health conditions, of which nine were related to cancer, seven to diabetes, 10 to HIV, five

Table 6. Number of substance use projects receiving NIH funding from 2011–2017.

Year/key term	2011 n = 5 (%)	2012 n = 9 (%)	2013 n = 11 (%)	2014 n = 16 (%)	2015 n = 21 (%)	2016 n = 24 (%)	2017 n = 25 (%)	Total unique studies across all years accounting for duplicates n = 50 (%)
Alcohol	2 (40.0)	4 (44.4)	4 (36.4)	12 (75.0)	17 (81.0)	18 (75.0)	14 (56.0)	32 (65)
Smoking/tobacco	4 (80.0)	6 (66.7)	8 (72.7)	10 (62.5)	10 (47.6)	11 (45.8)	14 (56.0)	27 (54)
Marijuana	0 (0.0)	0 (0.0)	0 (0.0)	4 (25.0)	5 (23.8)	5 (20.8)	5 (20.0)	9 (18)
Opioid	0 (0.0)	0 (0.0)	0 (0.0)	2 (12.5)	2 (9.5)	2 (8.3)	0 (0.0)	2 (4)
Cocaine	0 (0.0)	1 (11.1)	1 (9.1)	1 (6.3)	2 (9.5)	2 (8.3)	0 (0.0)	3 (6)
Meth-amphetamine	0 (0.0)	1 (11.1)	1 (9.1)	0 (0.0)	1 (4.8)	1 (4.2)	0 (0.0)	2 (4)
Heroin	0 (0.0)	0 (0.0)	0 (0.0)	1 (6.3)	1 (4.8)	1 (4.2)	0 (0.0)	1 (2)
Ecstasy	0 (0.0)	0 (0.0)	0 (0.0)	1 (6.3)	2 (9.5)	1 (4.2)	0 (0.0)	2 (4)

Table 7. Number of disease/diagnosis projects receiving NIH funding from 2011–2017.

Year/key term	2011 n = 4 (%)	2012 n = 7 (%)	2013 n = 5 (%)	2014 n = 10 (%)	2015 n = 22 (%)	2016 n = 17 (%)	2017 n = 26 (%)	Total unique studies across all years accounting for duplicates n = 43 (%)
Cancer	1 (25.0)	1 (14.2)	1 (20.0)	3 (30.0)	3 (13.6)	5 (29.4)	9 (34.6)	12 (27.9)
Diabetes	0 (0.0)	0 (0.0)	0 (0.0)	4 (40.0)	6 (27.3)	4 (23.5)	7 (26.9)	10 (23.3)
HIV	3 (75.0)	6 (85.7)	4 (80.0)	3 (30.0)	12 (54.5)	8 (47.1)	10 (38.5)	21 (48.8)
Cardiovascular disease	0 (0.0)	0 (0.0)	0 (0.0)	1 (10.0)	2 (9.1)	2 (11.8)	5 (19.2)	5 (11.6)
Vaccinations/ immunizations	0 (0.0)	0 (0.0)	0 (0.0)	1 (10.0)	2 (9.1)	1 (5.9)	2 (7.7)	3 (7.0)

to cardiovascular disease, and two to vaccination/immunization (see Table 7). HIV continued to be highly researched using social media tools, comprising 43.5% of the studies related to the above listed diseases/diagnoses and 9.4% of social media research overall in 2017. Of the 12 unique social media-enabled cancer studies performed between 2011 and 2017, four were specific to skin cancer, one to cervical cancer, three to cancer prevention, two to cancer caregivers, and two focused on survivors of any type of cancer.

Psychiatry or mental health. Between 2011 and 2017 there were a total of 28 NIH-funded social media research studies focused on psychiatry/mental health. In 2011, one of the nine was about depression. By 2017, social media research had expanded to mental health topics including PTSD and suicide.

Weight and physical activity. Over the 2011–2017 time frame, 14 unique studies were funded that examined weight and physical activity (see Table 8). In 2011

Table 8. Number of social media studies on physical activity and weight receiving NIH funding 2011–2017.

Year/key term	2011 n = 1 (%)	2012 n = 1 (%)	2013 n = 3 (%)	2014 n = 3 (%)	2015 n = 7 (%)	2016 n = 5 (%)	2017 n = 8 (%)	Total unique studies across all years accounting for duplicates n = 14 (%)
Physical activity	1 (100.0)	1 (100.0)	3 (100.0)	1 (33.3)	4 (57.1)	3 (60.0)	6 (75.0)	10 (71.4)
Obesity	0 (0.0)	0 (0.0)	2 (66.7)	3 (100.0)	5 (71.4)	4 (80.0)	5 (62.5)	10 (71.4)
Weight gain	1 (100.0)	1 (100.0)	1 (33.3)	3 (100.0)	3 (42.9)	3 (60.0)	4 (50.0)	7 (50.0)
Weight loss	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	2 (28.6)	4 (80.0)	4 (50.)	4 (28.6)

and 2012 studies focused on physical activity and weight gain with the first obesity studies occurring in 2013 followed by weight loss in 2014. Of note, the key term "sedentary behavior" did not yield any studies in the years 2011–2017.

Regulatory environment

With the increased use of social media in research involving human participants, we examined the extent to which institutions provided guidance to researchers on how to assess and mitigate research risks. Of the 56 institutions receiving NIH support to carry out this research, two required an institutional affiliation to gain access to the website and nine were private entities with limited or no information about their research and/or regulatory review processes. The IRB or Human Research Protection Program pages were reviewed for the remaining 45 institutions with four institutions providing specific "social media" guidance or related policy with an additional eight institutions having specific "internet research" guidance or related policy. The remaining 33 institutions did not include visible guidelines related to social media use as a research tool for recruitment, surveillance, or intervention delivery.

The few institutions that published social media research guidance for its research community, (e.g., University of Rochester, University of Pennsylvania (UPenn), and City University of New York (CUNY)) generally did not provide detail, but indicated that social media research would be evaluated on a case-by-case basis and was dependent on the platform and proposed research. The responsibility for understanding the technology and related ToS was placed on the investigator. For example, CUNY's guidance document (http://www2.cuny.edu/wp-content/uploads/site s/4/page-assets/research/research-com

pliance/human-research-protection-program-hrpp/hrpp-policies-procedures/Internet_Based_Research.pdf) states,

> principal investigators are responsible for becoming familiar with the terms of service and privacy policy for each internet venue or mobile technology to be used in their respective research prior to the implementation of human subject research activities, and need to provide the IRB with their assessment of how best to safeguard subject privacy and confidentiality based on the tools being used

UPenn states, "the IRB requires that investigators carefully consider a plan for protections that will be utilized for each social media application to minimize privacy, confidentiality and safety risks to subjects" (https://irb.upenn.edu/sites/default/files/IRB%20Social%20Media%20Guidance_2018.8.pdf). So, while these organizations are addressing the use of social media in research, the information is not actually guidance. Placing responsibility on the research to review and understand the ToS agreements is not a solution that is in the service of protecting human subjects. In fact, the ToS agreements often conflict with the federal regulations in that they require disputes to be addressed through mediation and prevent the end user from filing a claim for damages.[8]

Discussion

As social media networks evolve and increasingly become a part of our daily fabric, leveraging these networks for use in the delivery of health research is on the rise. This study demonstrates that social media-enabled research has not only grown in volume, but also in its breadth. While initially considered a useful tool for engaging with adolescent and young adult research

participants, it has also emerged as a means to reach other populations also underrepresented in research including those identifying with minority groups that are often excluded in biomedical research. Moreover, we documented that social media methods are being used to explore complex social issues such as substance abuse and mental health as well as infections and/or chronic health conditions such as HIV, diabetes, cardiovascular disease, and cancer. Social media research is becoming increasingly equipped to understand human behavior as well as to deliver interventions that can have a significant and long-lasting impact in terms of preventing disease development, improving mental health outcomes, and promoting healthy living through physical activity and weight loss.

Given the sensitivity of the subject matter being studied combined with populations considered vulnerable by most accounts, it is surprising that little to no formal guidance is readily available to assist researchers in designing studies for deployment on social media platforms. Presently, there is no guidance from US federal regulatory bodies, and recent changes to the Common Rule do not directly address research using social network platforms.[16] Moreover, in addition to the growth of social media-enabled health research, the broader field of digital health research involving pervasive sensors and artificial intelligence carried out by researchers in both regulated and unregulated sectors, is outpacing what ethics boards are equipped to handle.[8,17] A slowly growing literature is developing on how to evaluate and mitigate potential harms involving social media research.[1,15,18,19] IRB's find it difficult to determine what risks and privacy expectations are unique to social media and what challenges can be addressed by modifications of known and understood risks inherent in research.[20] Benton et al. proposed recommendations for research using social media that include 1. concern of inadvertently compromising user privacy by linking data, 2. alternative processes for obtaining informed consent, 3. additional steps to de-identify social media data before analysis and dissemination, and 4. care when attributing presenting information in public forums.[18] While we recognize the pivotal role an IRB can play as gatekeeper to research, there is an urgent need for the research community as whole to consider societal and ethical consequences when using social media to carry out scientific research, to ensure the respect and trust of the public.[21]

In fact, it is essential that the leaders within the digital health research sector act to take a pivotal role in setting ethical standards, including journal editors, sponsors, government regulatory entities, as well as the technology companies that are increasingly involved in health research.[22] For example, the Federal Trade Commission (FTC) is well positioned to require that US companies honestly disclose their research practices to users, in the same manner as their privacy practices, and the FTC can seek consent orders requiring comprehensive research oversight and compliance programs.[22] Recent advances in privacy law, such as the California Consumer Privacy Act of 2018 that went into effect 1 January 2020 (https://legi nfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_ id=2 01720180AB375) giving consumers control of their data including the right to know what information is collected and why as well as who has access. In fact, consumers will be able to prohibit companies from selling their data and can also require that their personal information be deleted. This act is intended to protect Californians and will apply to companies within and outside the state.

These new privacy protections are similar to those implemented in the European Union (EU) via the General Data Protection Regulation (GDPR) in 2018. The GDPR gives EU citizens control over their personal data, but holds organizations accountable for gathering data via consent and transparency, and under strict conditions. Those who manage (controllers) and those who receive the data (processors) are equally responsible and liable for protecting the data of individuals residing in the EU. The GDPR has increased awareness among both consumers and organizations of the need for explicit consent and transparency of data and data flows, and there are significant financial consequences for non-compliance. Furthermore, grassroots Initiatives taking place that have global impact, like Connected and Open Research Ethics (CORE),[23] are attempting to bridge the gap between policy and practice by conducting research on the ethics of social media-enabled research. Those conducting research on the ethical dimensions of social media research would be able to advance this area more rapidly if study sponsors required those proposing this research to include research questions that could inform ethical practices. Professional societies provide another avenue for setting standards for research professionals and could advocate for developing guidance to support best practices in health research using social media.

Limitations

This study was limited by the four keywords that bounded the scope of social media research discovered via the NIH RePORTER database. In addition, use of NIH RePORTER exclusively as a data source meant only global studies that in some way partnered with a US-based research team were included. The ability to identify institutional guidance specific to social media

use in research was limited by how the web pages were tagged to optimize search engine functionality. As this study was unfunded, a lack of resources prevented our request for policy documents from each institution, which could have provided a more complete picture of available guidance. Aside from these limitations, this is the first study that we know of that has mined the NIH RePORTER data to qualify and quantify the use of social media tools to better understand important health issues affecting people worldwide.

Conclusions

The use of social media platforms in research involving human participants is increasing. This study documented how platforms are used for surveillance, recruitment, and intervention delivery in hard to reach populations to study issues ranging from mental health and substance use to cancer and cardiovascular disease. However, as social media-enabled health research increases, little guidance exists to aid researchers in evaluating the ethical, legal/regulatory, and social implications surrounding social media in human research. In light of the growing number of research studies utilizing social media and recent data breaches surrounding popular social media platforms, establishing ethical standards in social media research is a priority. Working across disciplines and sectors is critical if we are to open a global dialogue toward the development of ethical and responsible social media research and balance participant protections with the advancement of important health discoveries.

ORCID iD: Camille Nebeker ⓘ https://orcid.org/0000-0001-6819-1796

References

1. Arigo D, Pagoto S, Carter-Harris L, et al. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. *Digit Health* 2018; 4: 2055207618771757.
2. Nebeker C, Lagare T, Takemoto M, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. *Transl Behav Med* 2016; 6: 577–586.
3. Chunara R, Andrews JR and Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg* 2012; 86: 39–45.
4. Saltzman M. Social media mining: can we prevent the apocalypse? *J Biosecurity Biosaf Biodefense Law* 2017; 81: 19–37.
5. Sakaki T, Okazaki M and Matsuo Y. Earthquake shakes Twitter users: real-time event detection by social sensor. In: *19th international conference on the World Wide Web* Raleigh, North Carolina, 26–30 April 2010, p. 851–860. New York: ACM.
6. Smith A and Anderson M. Social media use in 2018, http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ (2018, accessed 4 January 2019).
7. Gelinas L, Pierce R, Winkler S, et al. Using social media as a research recruitment tool: ethical issues and recommendations. *Am J Bioeth* 2017; 17: 3–14.
8. Nebeker C, Harlow J, Espinoza Giacinto R, et al. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. *AJOB Empir Bioeth* 2017; 8: 266–276.
9. Kramer A, Guillory JE and Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci* 2014;111:8788–8790.
10. Meyer MN. Misjudgements will drive social trials underground. *Nature* 2014; 511: 265.
11. Granville K. Facebook and Cambridge Analytica: what you need to know as fallout widens. *The New York Times*, 19 March 2018, https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html (accessed 31 December 2018).
12. Neuman S and Domonoske C. Grindr admits it shared HIV status of users: NPR. *The Two Way*, 3 April 2018, https://www.npr.org/sections/thetwo-way/2018/04/03/599069424/grindr-admits-it-shared-hiv-status-of-users (accessed 31 December 2018).
13. Zimmer M. OkCupid study reveals the perils of big-data science. *Wired*, 14 May 2016, https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/ (accessed 31 December 2018).

14. Fiesler C and Proferes N. "Participant" perceptions of Twitter research ethics. *Soc Media Soc* 2018; 4: 2056305118763366.

15. Ayers JW, Caputi TL, Nebeker C, et al. Don't quote me: reverse identification of research participants in social media studies. *npj Digit Med* 2018; 1: 30.

16. Menikoff J, Kaneshiro J and Pritchard I. The Common Rule, updated. *N Engl J Med* 2017;376:613–615.

17. Bloss C, Nebeker C, Bietz M, et al. Reimagining human research protections for 21st century science. *J Med Internet Res* 2016;18:e329.

18. Benton, Adrian GC and Dredze M. Ethical research protcols for social media health research. In: *Proceedings of the first workshop on ethics in natural language processing* Valencia, Spain, April 2017, p. 94–102. Valencia, Spain: Association for Computational Linguistics.

19. Paul MJ and Dredze M. A model for mining public health topics from Twitter. *Health (Irvine Calif)* 2012;11:16, http://www.cs.jhu.edu/~mdredze/publica tions/2011.tech.twitter_health.pdf (accessed 4 January 2019).

20. Buchanan E and Zimmer M. *Internet research ethics, Stanford Encyclopedia of Philosophy*. Redwood City, CA: Stanford University, https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry = ethics-internet-research (2018, accessed 1 January 2019).

21. Crowcroft J, Haddadi H and Henderson T. Responsible research on social networks: dilemmas and solutions, https://haddadi.github.io/papers/digitalethics2017.pdf (2017, accessed 2 January 2019).

22. Grimmelmann J. Law and ethics of experiments on social media users. *Colorado Tech Law J* 2015;219, https://scholarship.law.cornell.edu/facpub/1487/ (accessed 4 January 2019).

23. Torous J and Nebeker C. Navigating ethics in the digital age: introducing connected and open research ethics (CORE), a tool for researchers and institutional review boards. *J Med Internet Res* 2017; 19:e38.