Title

Social Media as an Emerging Data Resource for Epidemiologic Research: Characteristics of Regular and Nonregular Social Media Users in Nurses Health Study II.

Journal

American Journal of Epidemiology, 189(2)

Authors

Kim, Eric

James, Peter

Zevon, Emily

et al.

Publication Date

2020-02-28

DOI

10.1093/aje/kwz224

Peer reviewed

American Journal of Epidemiology

Vol. 189, No. 2
DOI: 10.1093/aje/kwz224
Advance Access publication:
October 8, 2019

Practice of Epidemiology

Social Media as an Emerging Data Resource for Epidemiologic Research: Characteristics of Regular and Nonregular Social Media Users in Nurses' Health Study II

Eric S. Kim*, **Peter James, Emily S. Zevon, Claudia Trudel-Fitzgerald, Laura D. Kubzansky, and Francine Grodstein**

* Correspondence to Dr. Eric S. Kim, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02215 (e-mail: eskim@hsph.harvard.edu).

Initially submitted December 22, 2018; accepted for publication September 20, 2019.

With advances in natural language processing and machine learning, researchers are leveraging social media as a low-cost, low-burden method for measuring various psychosocial factors. However, it is unclear whether information derived from social media is generalizable to broader populations, especially middle-aged and older adults. Using data on women aged 53–70 years from Nurses' Health Study II (2017–2018; $n = 49{,}045$), we assessed differences in sociodemographic characteristics, health conditions, behaviors, and psychosocial factors between regular and nonregular users of Facebook (Facebook, Inc., Menlo Park, California). We evaluated effect sizes with phi (φ) coefficients (categorical data) or Cohen's d (continuous data) and calculated odds ratios with 95% confidence intervals. While most comparisons between regular and nonregular users achieved statistical significance in this large sample, effect sizes were mostly "very small" (conventionally defined as φ or d <0.01) (e.g., optimism score: $\text{mean}_{\text{regular users}} = 19$ vs. $\text{mean}_{\text{nonregular users}} = 19$ ($d = -0.03$); physical activity: $\text{mean}_{\text{regular users}} = 24$ metabolic equivalent of task (MET)-hours/week vs. $\text{mean}_{\text{nonregular users}} = 24$ MET-hours/week ($d = 0.01$)). Some factors had slightly larger differences for regular users versus nonregular users (e.g., depression: 28% vs. 23% ($\varphi = 0.05$); odds ratio = 1.27 (95% confidence interval: 1.22, 1.33); obesity: 34% vs. 26% ($\varphi = 0.07$); odds ratio = 1.42 (95% confidence interval: 1.36, 1.48)). Results suggest that regular Facebook users were similar to nonregular users across sociodemographic and psychosocial factors, with modestly worse health regarding obesity and depressive symptoms. In future research, investigators should evaluate other demographic groups.

epidemiologic methods; machine learning; natural language processing; psychology; public health; social media

Abbreviations: CI, confidence interval; NHS II, Nurses' Health Study II; OR, odds ratio.

Substantial evidence demonstrates that the social milieu in which we live (e.g., neighborhoods, workplaces, schools, faith-based organizations, and family/home environments) impacts our health; however, mechanisms underlying these effects are not fully understood (1). One proposed pathway through which social conditions may operate is psychological factors (e.g., depression) and the biological and behavioral responses they evoke (2, 3). However, most epidemiologic studies are unable to routinely capture data on a wide array of psychosocial factors, because regularly administering these assessments at scale causes substantial participant burden and is financially prohibitive for many researchers. Social media data may provide an innovative, low-cost, low-respondent-burden solution for regularly capturing psychosocial information.

Each day, billions of people utilize social media, creating a vast data resource of unrealized public health potential. According to a 2018 Pew Research Poll (4), the most commonly used social media platforms among US adults are YouTube (73%; YouTube, San Bruno, California), Facebook (68%; Facebook, Inc., Menlo Park, California), and Instagram (35%; Facebook, Inc.). Facebook has emerged as a social media tool of particular interest to the health research community because of its large base of daily users of all

ages (5). As of early 2019, 76%–84% of persons under age 50 years, 68% of adults aged 50–64 years, and 38% of adults aged ≥65 years use Facebook, with usage in all age groups generally increasing in comparison with previous years, except for teenagers (4–6).

With recent advances in natural language processing and machine learning, researchers have begun leveraging digital footprints in social media and applying novel algorithms that assess multiple psychosocial factors. After an algorithm for predicting a psychological characteristic is trained and validated, applying the model to millions of social media posts may take only hours (7). Initial studies suggest that characteristics such as personality (7–9), psychological well-being (e.g., life satisfaction) (10), and psychological distress (e.g., depression) (11) could be inferred using social media digital footprints.

However, before merging social media data and health research data (assuming explicit participant consent), better characterization of regular social media users is important. To our knowledge, little work has addressed whether substantial differences exist between regular users of social media and nonregular users, especially among middle-aged and older adults, who comprise the populations of interest for much health research. Therefore, we used data from Nurses' Health Study II (NHS II), an ongoing cohort study of women aged 53–70 years, to compare numerous characteristics of regular users versus nonregular users of Facebook, which was, by far, the most common social media platform used by cohort participants.

METHODS

Study population

NHS II is an ongoing cohort study of 116,430 female nurses aged 25–42 years at cohort initiation in 1989. Participants complete biennial online (approximately 50%) or paper questionnaires to update information on lifestyle, psychosocial factors, health behaviors, and chronic health conditions (12). The overall follow-up rate is approximately 90% (12). Participants were queried about regular Facebook usage in the 2017–2018 online questionnaire, and data from 49,045 women have been fully processed to date. This study was approved by the institutional review board at Brigham and Women's Hospital (Boston, Massachusetts).

Measures

Facebook use. Participants' use of Facebook was assessed in 2017 by asking, "Do you regularly post updates or information on social media (rather than just viewing or 'liking' posts)?". Among people who answered yes, a follow-up question asked which of the following sites participants used: 1) Facebook, 2) Instagram, 3) Twitter (Twitter, San Francisco, California), or 4) other. In this study, we categorized women according to whether they reported regular use of Facebook or not.

Sociodemographic factors. On the questionnaires, participants reported the following information in 2017 (unless the information was not available at that assessment, in which case we noted otherwise): age (years; continuous), race (white, black, or other; assessed in 1989), marital status (married, divorced/separated/single, or widowed; assessed in 2013), and occupational status (disabled/retired/other, part-time nurse, nonnurse, or homemaker; assessed in 2013). The mailing addresses of all NHS II participants have been geocoded, and we considered the following census-tract-level factors: urbanicity of residence (urban (urban area with ≥50,000 people), suburban (an urban cluster of 10,000–49,999 people), or small town/rural (an urban cluster of <10,000 people)), population density (number of people per square mile; continuous), median annual household income (dollars; continuous), and median home value (dollars; continuous). All census variables were based on data from the 2000 US Census.

Psychosocial factors. Women were categorized as depressed if they reported a Center for Epidemiologic Studies Depression Scale–Revised symptom score greater than 10 (13) (assessed in 2013), physician-diagnosed depression in the past 2 years (yes/no; assessed in 2017), or use of antidepressant medication in the past 2 years (yes/no; assessed in 2017). Women were categorized as having elevated anxiety if they scored ≥5 on the Generalized Anxiety Disorder 7 Scale (assessed in 2013) (14). Optimism was assessed with the 6-item Life Orientation Test–Revised in 2017 (continuous) (15). Purpose in life was captured with the 3-item subscale of the Ryff Scales of Psychological Well-Being in 2017 (continuous) (16). Social integration was assessed with the Berkman-Syme Social Network Index (assessed in 2013)—a composite measure of social relations: marital status, number of close friends and close relatives, and frequency of attendance at religious and nonreligious/community activities (categorical) (17).

Health conditions. On each biennial questionnaire from 1989 to 2017, participants were asked about physician diagnoses (yes/no) of heart disease, cancer (other than nonmelanoma skin cancer), type 2 diabetes, high cholesterol, and hypertension. We also calculated body mass index (weight (kg)/height $(m)^2$) from self-reported weight and height assessed in 2017 (continuous) (18).

Health behaviors. Health behavior data from the questionnaires included information on smoking status (never smoker, former smoker, or current smoker; assessed in 2017), physical examination in the last 2 years, as a proxy for health-conscious behaviors (yes/no; assessed in 2017), diet (assessed with the Willett food frequency questionnaire (19) in 2011 and quantified using Alternate Healthy Eating Index 2010 (score range, 0–100)), physical activity (assessed across 6 types of exercise in 2009 (20) and summarized as continuous weekly energy expenditure in metabolic equivalent of task–hours per week), and combined alcohol intake from wine, beer, and liquor (g/day, continuous; assessed in 2011).

Statistical analysis

To compare regular Facebook users with nonregular users, we first performed χ^2 tests for categorical data and t tests for continuous data, which indicated whether there were

Table 1. Sociodemographic Characteristics of Participants in Nurses' Health Study II, by Category of Facebook Use, United States, 2017–2018[a]

Sociodemographic Factor	Nonregular Facebook Users, % ($n = 35,263$)	Regular Facebook Users, % ($n = 13,782$)	Effect Size (d or φ)[b]	Regular vs. Nonregular Facebook Use			
				Odds Ratio	95% CI	Mean Difference	95% CI
Mean age, years[c]	59 (5)	58 (5)	0.07			0.3	0.3, 0.4
Race			0.03				
White	93	95		1.23	1.13, 1.34		
Black	1	1		0.66	0.54, 0.82		
Other	5			0.85	0.77, 0.93		
Marital status			0.03				
Married	76	75		0.96	0.92, 1.00		
Divorced, separated, or single	19	21		1.00	0.95, 1.05		
Widowed	5	4		1.13	1.02, 1.24		
Occupational status			0.02				
Disabled/retired/other	27	27		0.98	0.94, 1.02		
Full- or part-time nurse	63	62		1.08	1.00, 1.17		
Nonnurse	6	7		0.84	0.76, 0.94		
Homemaker	4	4		1.02	0.97, 1.06		
Urbanicity of residence			0.007				
Urban	84	83		0.96	0.91, 1.01		
Suburban	9	9		1.02	0.96, 1.09		
Small town/rural	7	7		1.06	0.98, 1.14		
Census tract characteristics							
Mean population density, no. of people/square mile[c]	3,129 (9,694)	2,778 (7,625)	−0.04			−340	−518, −163
Median annual household income, thousands of dollars[d]	67 (25)	65 (24)	−0.06			−1.6	−2.1, −1.1
Median home value, thousands of dollars[d]	177 (129)	169 (117)	−0.07			−8.1	−10.4, −5.7

Abbreviation: CI, confidence interval.

[a] Values for categorical variables may not sum to 100% because of rounding or missing data.

[b] Effect sizes less than 0.01 and less than 0.2 are conventionally defined as very small and small effects, respectively. Values are Cohen's d (continuous data) or φ coefficients (categorical data).

[c] Values are expressed as mean (standard deviation).

[d] Values are expressed as median (standard deviation).

statistical differences between the 2 groups. If data were missing, participants were excluded from univariate analyses; however, missingness was minimal (<10% for most variables). With the large sample size and the relevance of differentiating clinical significance from statistical significance, we also calculated effect sizes using the phi coefficient (φ) for categorical data and Cohen's d for continuous data, to capture the magnitude of differences, independently of sample size. Effect sizes less than 0.01 and less than 0.2 for both the φ coefficient and Cohen's d are conventionally defined as very small and small effects, respectively (21, 22). We also compared regular users with nonregular users by calculating odds ratios and 95% confidence intervals for categorical variables and mean differences for continuous variables. Analyses were conducted in SAS, version 9.3 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Participants who completed the 2017–2018 online questionnaire and those who completed the paper-and-pencil questionnaire were similar in age (mean age: 59 years vs. 58 years). Their socioeconomic status, as determined by spouse's education, was reasonably comparable (graduate degree: 29% vs. 23%), as was their marital status (married: 76% vs. 72%). Among women who provided data on social media use on the 2017–2018 online questionnaire, 29% reported regularly posting updates or information on social

Table 2. Psychosocial, Health Condition, and Health Behavior Characteristics of Participants in Nurses' Health Study II, by Category of Facebook Use, United States, 2017–2018[a]

Characteristic	Nonregular Facebook Users, % ($n = 35,263$)	Regular Facebook Users, % ($n = 13,782$)	Effect Size (d or φ)[b]	Regular Versus Nonregular Facebook Use			
				Odds Ratio	95% CI	Mean Difference	95% CI
Psychosocial factors							
Depression (CESD-R score >10)[c]	23	28	0.05	1.27	1.22, 1.33		
Anxiety (GAD-7 Scale score ≥5)	23	25	0.02	1.10	1.05, 1.15		
Optimism score (LOT-R)[d]	19 (5)	19 (5)	−0.03			−0.2	−0.3, −0.1
Ryff Purpose in Life score[d]	13 (2)	13 (2)	−0.04			−0.1	−0.1, −0.1
Berkman-Syme Social Integration Index			0.02				
Socially isolated	12	11		0.88	0.82, 0.94		
Moderately isolated	25	25		0.98	0.94, 1.03		
Moderately integrated	32	33		1.04	0.99, 1.08		
Socially integrated	30	31		1.02	0.98, 1.07		
History of health conditions							
Heart disease	5	6	0.02	1.19	1.09, 1.29		
Cancer	15	14	−0.002	0.99	0.93, 1.04		
Type 2 diabetes	10	11	0.02	1.16	1.09, 1.23		
High cholesterol	59	62	0.03	1.15	1.11, 1.20		
Hypertension	41	46	0.04	1.21	1.17, 1.26		
Obesity (BMI ≥30)[e]	26	34	0.07	1.42	1.36, 1.48		
Health behaviors							
Smoking status			0.04				
Never smoker	66	62		0.83	0.80, 0.87		
Past smoker	30	34		1.23	1.18, 1.28		
Current smoker	4	4		0.96	0.87, 1.05		
Recent physical examination	86	87	0.01	1.08	1.02, 1.15		
Diet score (AHEI-2010)[d]	66 (13)	65 (13)	0.006			0.1	−0.2, 0.4
Total physical activity, MET-hours/week[d]	24 (29)	24 (29)	0.01			0.5	−0.1, 1.0
Alcohol consumption, g/day[d]	7 (11)	7 (11)	−0.05			−0.5	−0.8, −0.3

Abbreviations: AHEI, Alternate Healthy Eating Index; BMI, body mass index; CESD-R, Center for Epidemiologic Studies Depression Scale–Revised; CI, confidence interval; GAD, Generalized Anxiety Disorder; LOT-R, Life Orientation Test–Revised; MET, metabolic equivalent of task.

[a] Values for categorical variables may not sum to 100% because of rounding or missing data.

[b] Effect sizes of <0.01 and <0.2 are conventionally defined as very small and small effects, respectively. Values are Cohen's d (continuous data) or φ coefficients (categorical data).

[c] Depression status determined by: CESD-R ≥10, doctor diagnosed depression in the last 2 years, or antidepressant use in the last 2 years.

[d] Values are expressed as mean (standard deviation).

[e] Weight (kg)/height (m)^2.

media. Virtually all of these regular users (28%) used Facebook; 5% used Instagram, and 2% used Twitter. Our data on Facebook usage converged with Pew data, which estimate that approximately 22%–36% of middle-aged and older women use Facebook frequently (6).

Overall, across all factors, regular Facebook users were fairly similar to nonregular users. While most differences were statistically significant (Tables 1 and 2), effect sizes fell within the definitions of the "very small" (<0.01) to "small" (<0.2) range (21, 22); that is, Cohen's d and φ coefficients ranged from −0.07 to 0.07. Specifically, regular users were comparable to nonregular users with regard to sociodemographic characteristics (Table 1). For example, the mean ages of regular and nonregular users were 58 years and 59 years, respectively ($d = 0.07$; mean difference = 0.3 years, 95% confidence interval (CI): 0.3, 0.4), median annual household incomes were $65,440 and $67,003, respectively ($d = -0.06$; mean difference = $1,615, 95%

CI: –2,097, 1,132), and both groups were similarly likely to have remained in the nursing profession (62% and 63%, respectively; $\varphi = 0.02$; odds ratio (OR) = 1.08, 95% CI: 1.00, 1.17).

Table 2 further shows characteristics of regular users versus nonregular users across psychosocial factors, chronic health conditions, and health behaviors. For optimism, regular and nonregular users were similar, with very small effect sizes between optimism scores ($\text{mean}_{\text{regular users}} = 19$ and $\text{mean}_{\text{nonregular users}} = 19$ ($d = -0.03$); mean difference = −0.2, 95% CI: –0.3, −0.1). Differences in the number of women who were depressed among regular and nonregular users were somewhat stronger (28% with depression vs. 23% ($\varphi = 0.05$); OR = 1.27, 95% CI: 1.22, 1.33). The prevalences of chronic health conditions among regular and nonregular users were largely similar. For example, heart disease was prevalent in 6% of regular users and 5% of nonregular users ($\varphi = 0.02$; OR = 1.19, 95% CI: 1.09, 1.29); although regular users had a somewhat higher prevalence of obesity, the magnitude of the difference was very small (34% of regular users vs. 26% of nonregular users ($\varphi = 0.07$); OR = 1.42, 95% CI: 1.36, 1.48). Finally, the prevalence of past smoking was 34% among regular users versus 30% among nonregular users ($\varphi = 0.04$; OR = 1.23, 95% CI = 1.18, 1.28).

DISCUSSION

Regular and nonregular Facebook users generally displayed few differences across numerous sociodemographic, health, and psychosocial factors, with modestly worse health for a few factors, mainly obesity and depression. While this is important initial evidence indicating that Facebook could yield valuable data for epidemiologic research, additional studies should evaluate whether these similarities are consistent in men and in other age and racial/ethnic groups.

Social media potentially provide a new format for rapidly obtaining information on large populations and may yield research measures that are less reliant on self-reporting or conscious self-insight (23, 24). If findings derived from frequent social media users are reasonably generalizable and do not reflect systematic differences between those who do and do not use the platforms regularly, investing in algorithms that harness such data to assess psychosocial factors may prove worthwhile. These methods could provide an innovative, low-cost, low-respondent-burden approach for capturing data on various psychosocial factors in large epidemiologic cohorts at greater frequency than is currently feasible.

One caveat in using social media data for research is the potential for misuse of private information. Such data collection differs from traditional approaches in myriad important ways, and because of the rapidly evolving nature of social media, policies regarding privacy and ethical use of data have not been universally defined. To realize the potential of social media data for public health, future research will need to carefully consider ethical challenges. Such efforts are under way, and several investigators have proposed best practices to assist researchers in navigating this new terrain (25–27).

Our study had several limitations. NHS II includes middle-aged, educated, and largely white women. Thus, our findings may not apply to other sociodemographic groups, and further research in other populations is needed. Additionally, we focused on self-reporting of "regular" Facebook usage; subsequent work with finer granularity will be able to more finely dissect our findings. Preliminary findings suggest that the use of social media itself could alter psychosocial characteristics, notably depressive symptoms and decreased life satisfaction (28–30), but the findings remain mixed (31–33). Thus, further research is needed to clarify such potential effects and to better understand how these issues would affect research using social media. Moreover, investigators should continue to carefully evaluate the potential tradeoffs between use of more traditional methods of data collection (e.g., self-report surveys) and the use of emerging methods (e.g., passive monitoring through social media).

Some significant strengths of our research include the use of data from a large and richly characterized cohort and the use of validated, high-quality measures of sociodemographic, psychosocial, behavioral, and physical health.

In conclusion, with recent advances in natural language processing and machine learning, Facebook and other social media may open new avenues for assessing psychosocial factors on an unprecedented scale. This could allow the evaluation of novel hypotheses linking psychosocial factors with health. The initial findings here—that sociodemographic characteristics, health characteristics, and psychosocial factors do not appear meaningfully different in middle-aged and older NHS II women according to regular versus nonregular Facebook use—indicate that social media may be able to provide a relevant data resource for health research. Important next steps include further studies in other cohorts, careful consideration of privacy and ethical concerns, and, eventually, assessment of whether social-media-derived measures can predict or inform future health in meaningful ways.

ACKNOWLEDGMENTS

Author affiliations: Department of Social and Behavioral Sciences, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts (Eric S. Kim, Emily S. Zevon, Claudia Trudel-Fitzgerald, Laura D. Kubzansky); Lee Kum Sheung Center for Health and Happiness, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts (Eric S. Kim, Claudia Trudel-Fitzgerald, Laura D. Kubzansky); Human Flourishing Program, Institute for Quantitative Social Science, Harvard University, Cambridge, Massachusetts (Eric S. Kim); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Peter James); Department of Epidemiology, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts (Francine Grodstein); and Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital

and Harvard Medical School, Boston, Massachusetts (Francine Grodstein).

E.S.K. and P.J. are joint co–first authors of this paper.

This work was supported by National Institutes of Health grants R01AG053273, R00CA201542, K99AG055696, UM1CA176726, and R01CA67262. C.T.F. received a postdoctoral fellowship from the Fonds de Recherche du Québec–Santé.

We thank the participants and staff of Nurses' Health Study II for their valuable contributions.

E.S.K. has worked as a consultant with AARP (Washington, DC) and UnitedHealth Group (Minnetonka, Minnesota).

REFERENCES

1. Berkman LF, Kawachi I, Glymour MM. *Social Epidemiology*. 2nd ed. New York, NY: Oxford University Press; 2014.
2. Kivimäki M, Batty DG, Kawachi I, et al., eds. *The Routledge International Handbook of Psychosocial Epidemiology*. 1st ed. London, United Kingdom: Routledge; 2017.
3. Kubzansky L, Huffman JC, Boehm JK, et al. Positive psychological well-being and cardiovascular health promotion. *J Am Coll Cardiol*. 2018;72(12):1382–1396.
4. Facebook, Inc. Our mission. https://newsroom.fb.com/company-info/. Accessed February 6, 2018.
5. Smith A, Anderson M. Social media use in 2018. Washington, DC: Pew Research Center; 2018. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/. Accessed May 9, 2018.
6. Anderson M, Perrin A. Tech adoption climbs among older adults. Washington, DC: Pew Research Center; 2017. http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/. Accessed February 6, 2018.
7. Park G, Schwartz HA, Eichstaedt JC, et al. Automatic personality assessment through social media language. *J Pers Soc Psychol*. 2015;108(6):934–952.
8. Azucar D, Marengo D, Settanni M. Predicting the big 5 personality traits from digital footprints on social media: a meta-analysis. *Personal Individ Differ*. 2018;124:150–159.
9. Park G, Schwartz HA, Sap M, et al. Living in the past, present, and future: measuring temporal orientation with language. *J Pers*. 2017;85(2):270–280.
10. Schwartz HA, Sap M, Kern ML, et al. Predicting individual well-being through the language of social media. *Pac Symp Biocomput*. 2016;21:516–527.
11. Guntuku SC, Yaden DB, Kern ML, et al. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci*. 2017;18:43–49.
12. Bao Y, Bertoia ML, Lenart EB, et al. Origin, methods, and evolution of the three Nurses' Health Studies. *Am J Public Health*. 2016;106(9):1573–1581.
13. Andresen EM, Malmgren JA, Carter WB, et al. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994;10(2):77–84.
14. Spitzer RL, Kroenke K, Williams JBW, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–1097.
15. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J Pers Soc Psychol*. 1994;67(6):1063–1078.
16. Ryff CD, Keyes CLM. The structure of psychological well-being revisited. *J Pers Soc Psychol*. 1995;69(4):719–727.
17. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol*. 1979;109(2):186–204.
18. Rimm EB, Stampfer MJ, Colditz GA, et al. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology*. 1990;1(6):466–473.
19. Yuan C, Spiegelman D, Rimm EB, et al. Validity of a dietary questionnaire assessed by comparison with multiple weighed dietary records or 24-hour recalls. *Am J Epidemiol*. 2017;185(7):570–584.
20. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol*. 1994;23(5):991–999.
21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
22. Sawilowsky SS. New effect size rules of thumb. *J Mod Appl Stat Methods*. 2009;8(2):597–599.
23. Sinnenberg L, Buttenheim AM, Padrez K, et al. Twitter as a tool for health research: a systematic review. *Am J Public Health*. 2016;107(1):e1–e8.
24. Sinnenberg L, DiSilvestro CL, Mancheno C, et al. Twitter as a potential data source for cardiovascular disease research. *JAMA Cardiol*. 2016;1(9):1032–1036.
25. Kosinski M, Matz SC, Gosling SD, et al. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *Am Psychol*. 2015;70(6):543–556.
26. Hunter RF, Gough A, O'Kane N, et al. Ethical issues in social media research for public health. *Am J Public Health*. 2018;108(3):343–348.
27. Vayena E, Salathé M, Madoff LC, et al. Ethical challenges of big data in public health. *PLoS Comput Biol*. 2015;11(2):e1003904.
28. Shakya HB, Christakis NA. Association of Facebook use with compromised well-being: a longitudinal study. *Am J Epidemiol*. 2017;185(3):203–211.
29. Tromholt M. The Facebook experiment: quitting Facebook leads to higher levels of well-being. *Cyberpsychol Behav Soc Netw*. 2016;19(11):661–666.
30. Hunt MG, Marx R, Lipson C, et al. No more FOMO: limiting social media decreases loneliness and depression. *J Soc Clin Psychol*. 2018;37(10):751–768.
31. Orben A, Dienlin T, Przybylski AK. Social media's enduring effect on adolescent life satisfaction. *Proc Natl Acad Sci*. 2019;116(21):10226–10228.
32. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. *Nat Hum Behav*. 2019;3(2):173–182.
33. Orben A, Przybylski AK. Screens, teens, and psychological well-being: evidence from three time-use-diary studies. *Psychol Sci*. 2019;30(5):682–696.