Invited Commentary

Invited Commentary: Recruiting for Epidemiologic Studies Using Social Media

Jenifer E. Allsworth*

* Correspondence to Dr. Jenifer E. Allsworth, Department of Biomedical and Health Informatics, School of Medicine, University of Missouri–Kansas City, 1000 E. 24th Street, Room 5E-103, Kansas City, MO 64108 (e-mail: allsworthj@umkc.edu).

Initially submitted November 18, 2014; accepted for publication December 2, 2014.

Social media–based recruitment for epidemiologic studies has the potential to expand the demographic and geographic reach of investigators and identify potential participants more cost-effectively than traditional approaches. In fact, social media are particularly appealing for their ability to engage traditionally “hard-to-reach” populations, including young adults and low-income populations. Despite their great promise as a tool for epidemiologists, social media–based recruitment approaches do not currently compare favorably with gold-standard probability-based sampling approaches. Sparse data on the demographic characteristics of social media users, patterns of social media use, and appropriate sampling frames limit our ability to implement probability-based sampling strategies. In a well-conducted study, Harris et al. (Am J Epidemiol. 2015;181(10):737–746) examined the cost-effectiveness of social media–based recruitment (advertisements and promotion) in the Contraceptive Use, Pregnancy Intention, and Decisions (CUPID) Study, a cohort study of 3,799 young adult Australian women, and the approximate representativeness of the CUPID cohort. Implications for social media–based recruitment strategies for cohort assembly, data accuracy, implementation, and human subjects concerns are discussed.

cohort studies; contraception; recruitment; sampling strategies; social media

Abbreviation: CUPID, Contraceptive Use, Pregnancy Intention, and Decisions.

In this issue of the Journal, Harris et al. (1) have published results from an important and well-conducted study of online recruitment strategies that is relevant to both epidemiologic and clinical studies. Traditional methods of cohort assembly, such as random digit dialing based on landline telephone service, often do not result in efficient identification of representative samples and may be especially challenging for reaching young, disadvantaged, or minority populations (2). Recent data from the United States have documented that while 2 in every 5 homes had only wireless telephone service, more than half of young adults and persons living in poverty resided in wireless-only households (3). Notably, persons in wireless-only households had higher prevalences of health-related risk behaviors (smoking, binge drinking), a higher likelihood of being uninsured, and a lower likelihood of having a usual place to go for medical care or of having received an influenza vaccine in the past year. While efforts to expand random digit dialing methods to incorporate wireless customers have been investigated, bias is of particular concern for young or low-income adults (4).

Faced with these realities, Harris et al. embarked on an initiative to recruit a representative cohort of young adult women for a prospective study of pregnancy intention and contraceptive use. In their study, the Contraceptive Use, Pregnancy Intention, and Decisions (CUPID) Study, recruitment efforts began with the use of a mailed invitation sent to a stratified random sample of women in the Medicare Australia database (1). The investigators distributed 900 mailings with the assumption of an 18% response rate and a target of 150 respondents (2). This approach, however, yielded only a 5.4% completion rate (consent and survey), with costs of A$101 per participant enrolled (2). In response, the investigators sought to incorporate both online and offline modalities into their recruitment strategy, with the objective of assembling a cost-effective, representative cohort of Australian women between the ages of 18 and 23 years. Researchers have a growing presence online and in social media venues. Social media are particularly appealing for their ability to engage traditionally “hard-to-reach” populations, including young adults and racial, ethnic, and sexual
minorities. As the CUPID Study demonstrates, social media are a useful platform for rapidly reaching and enrolling large numbers of persons who are not reachable via traditional techniques. Social media–based recruitment has the potential to expand the geographic reach of investigators and identify potential participants more cheaply than traditional approaches.

The CUPID researchers employed diverse promotional strategies in their recruitment efforts, including direct advertisements (Facebook ads (Facebook Inc., Menlo Park, California), free classified ads, and posters), unpaid promotional/supportive activities (Facebook, online forums, and Twitter (Twitter Inc., San Francisco, California) posts; radio and print coverage), and convenience (face-to-face events, mailings, and e-mails) and snowball (participant referrals) sampling techniques. The CUPID investigators spent approximately 70% of their recruitment expenditures on direct advertisements, with the remainder spent on convenience sampling approaches; no direct costs were incurred for promotional/supportive activities or snowball sampling. Unfortunately, in their paper, Harris et al. do not present numbers of participants for individual recruitment modalities, limiting our understanding of their impact on enrollment. For most of the 13-month enrollment period, multiple, concurrent recruitment modalities were in active use. What is clear from their Figure 1, however, is that demographically targeted Facebook sidebar ads (October 2012 through June 2013) did not generate adequate or consistent rates of enrollment (1). Sidebar ads were not visible to users of mobile devices, and while it is not possible to estimate person-hours of social media use through mobile devices, the Pew Research Center has found that 67% of young adults (ages 18–29 years) access social media platforms via mobile devices (5). The availability of newsfeed advertisements (June 2013 through September 2013) did dramatically increase recruitment, but the magnitude of this surge is unknown given that the use of Gumtree classified advertisements (eBay International AG, San Jose, California) took place during a similar time period (May 2013 through September 2013). The timing and impact of other online recruitment activities is unclear. The CUPID investigators’ efforts resulted in a nonprobability sample with demographic characteristics and a regional distribution roughly similar to those of their target population. Notably, however, study participants were more educated than the general population—a challenge also experienced by researchers incorporating mobile devices into random digit dialing strategies.

The CUPID Study experience raises a number of important issues that should be considered by epidemiologists. First, investigators should be encouraged to track and report the amount of personnel effort expended on direct individual contacts and supportive activities, incorporate the costs of these activities into estimates of per-participant expenditures, and estimate the impact on recruitment. The CUPID investigators sought to assemble a cohort of young adult women with few exclusion criteria, which could be accomplished with relatively minor personnel effort; however, for harder-to-reach or stigmatized populations, the personnel effort needed to support promotion of the study and to build rapport with the target population may be considerable. For example, in a recent study of Spanish-speaking Latino gay male couples, Martinez et al. (6) used social media platforms and smartphone applications to enroll 14 couples. Half of the couples were enrolled via social media sites following 3,670 messages and 75 participant couple screenings. The remaining couples enrolled were identified through extensive outreach efforts that included contact with more than 50 community-based organizations.

The recruitment impact of unpaid supportive promotion and education on rates of screening and enrollment should be evaluated whenever possible. Both CUPID and the Martinez study employed sustained supportive measures that included extensive outreach to the community of interest, engagement of relevant community organizations, informational posts on Facebook, Twitter, and other platforms, and study websites (1, 6). The individual impacts of these efforts on recruitment rates and participant sociodemographic characteristics in the CUPID Study were not described.

Second, social media–based recruitment requires prospective monitoring of advertisement success rates and changing advertising policies. CUPID investigators monitored the “click-through rates” of individual advertisement content and imagery as well as placement of advertisements (sidebar vs. newsfeed) and adjusted them to maximize outreach to the population of interest. Researchers must also monitor social media platform policies and examine their impact(s) on recruitment. For example, during the CUPID enrollment period, Facebook changed its advertising policy to permit the placement of advertisements within the user’s newsfeed—a change that dramatically expanded the reach of the advertisements for mobile phone users. However, not all changes will be favorable to a research study’s interests. More recent changes to Facebook’s sponsored posts and brand-related posting policies—implemented after the CUPID enrollment period—could decrease the reach and increase the costs of nonadvertisement postings on the study’s Facebook page (7).

Finally, social media–based recruitment raises concerns about data accuracy and ethical handling of information. Researchers engaged in social media recruitment must be wary of duplicate participants (within and across social media platforms) and inaccurate reporting in an environment that prizes anonymity. Social media engagement increases informal contacts between researchers and participants. Traditional recruitment techniques are rigorously overseen by institutional review boards that approve advertisements, telephone scripts, and mailings, but real-time social media engagement with potential participants through promotional posts, replies, and direct messaging is not. Participants may willingly disclose sensitive information that is accessible to study personnel outside of research activities, yet, to date, we lack clear ethical or regulatory guidelines for the appropriate handling of this information as research data. Informed consent procedures should clearly delineate the rules for handling social media–derived information, its use in the conduct of research, and efforts to protect participant privacy and confidentiality.

When beginning assembly of an epidemiologic cohort using social media–based recruitment methods, collaboration with the appropriate methodologist is essential. Surveys of Internet and social media use have shown that significant demographic shifts are common among social media users (8), and depending on the population of interest, complex strategies that incorporate recruitment of participants across
multiple social media platforms may be needed. Survey statisticians may contribute much-needed expertise to assess coverage, sampling, and differential use (frequency and proficiency) across a single or multiple social media platforms when the generation of a statistically representative sample is desired. Currently, a paucity of information on social media user demographics and reliable estimates of sampling frames limits researchers to nonprobability sampling and the construction of approximately representative cohorts (9). However, the pursuit of a representative sample is quite often neither required nor desired for valid scientific inference in epidemiologic studies (10, 11). Scientific inference may be enhanced by deliberate efforts to assemble nonrepresentative samples—for example, cohort studies that limit confounding factors or oversample subpopulations of interest may provide more meaningful inferences than studies of statistically representative cohorts. Social media–based recruitment strategies are well suited to the generation of nonrepresentative cohorts; they can be employed as readily as traditional recruitment strategies, are often more cost-effective, and, for hard-to-reach populations, may have improved reach.

Challenges in participant recruitment can result in significant delays in the conduct of research, increased costs, and increased failure to obtain necessary sample sizes. Efforts to explore and understand the strengths and weaknesses of alternate approaches for cohort identification, such as those described in the article by Harris et al. (1), are urgently needed.

ACKNOWLEDGMENTS

Author affiliation: Department of Biomedical and Health Informatics, School of Medicine, University of Missouri–Kansas City, Kansas City, Missouri (Jenifer E. Allsworth). Conflict of interest: none declared.

REFERENCES