Open Access   Original Article

Tracing participants for longitudinal environmental health research using social networking sites: a pilot study

Rogih Riad Andrawes[1] , Susan Jamuria Yousufzai[1] , Susan Sahar Mattin[1] , Susan Elliott[2] , Caroline Barakat[1]*

[1]Faculty of Health Sciences, Ontario Tech University, Oshawa, ON L1G 0C5, Canada
[2]Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

*Correspondence: Caroline Barakat, Faculty of Health Sciences, Ontario Tech University, 2000 Simcoe St N, Oshawa, ON L1G 0C5, Canada. caroline.barakat@ontariotechu.ca

Abstract

Aim: Longitudinal cohort study designs are considered the gold standard for investigating associations between environmental exposures and human health yet they are characterized by limitations including participant attrition, and the resource implications associated with cohort recruitment and follow-up. Attrition compromises the integrity of research by threatening both the internal and external validity of empirical results, weakening the accuracy of statistical inferences and the generalizability of findings. This pilot study aimed to trace participants from a historical cohort study, the Hamilton Child Cohort Study (HCC) (n = 3,202), (1976–1986, 2003–2008) which was originally designed to examine the relative contribution of indoor and outdoor exposure to air pollution on respiratory health.

Methods: Original participants were traced through social networking sites (SNS) by leveraging personal identifying data (name, age, sex, educational affiliation, and geographical locations) from the HCC entered into SNS search engines.

Results: Of the original cohort (n = 3,166), 21% (n = 665) were identified as having social media presence (SMP) on a single social media platform, with 15% (n = 479) found on Facebook, 6% (n = 185) on LinkedIn, < 1% (n = 9) on Instagram, and n = 1 participant on Twitter. However, 68% (n = 2,168) of the cohort were associated with multiple SNS with the same features (matching names, ages, and locations), making conclusive identification challenging. The remaining 11% (n = 334) of the cohort had no SMP (NSMP). Statistical differences in sample characteristics of each cohort were analyzed using the Pearson chi-square test. Significant differences between the SMP and NSMP cohorts were found in relation to sex (p < 0.001), and childhood neighborhood of residence (p < 0.05).

Conclusions: This study underscores social media's potential for tracing participants in longitudinal studies while advising a multi-faceted approach to overcome inherent limitations and biases. A full-scale study is necessary to determine whether utilizing SNS to trace participants for longitudinal research is an effective tool for re-engaging research participants lost to attrition.

Keywords

Introduction

Participant recruitment and retention for longitudinal studies remain a major hurdle in cohort research and must be addressed to increase data collection efficiency and scientific robustness [1, 2]. Although longitudinal cohort research designs offer unique strengths mainly due to their capacity for assessing temporal changes and causal effects, participant retention and engagement over time is a challenge that intensifies with increasing length of follow-up. Indeed, prospective longitudinal cohort studies are time-consuming and require substantial financial investments to initiate and collect data over time [3]. For instance, one of the longest longitudinal studies, the Nurses' Health Study established in 1976, led to an expenditure of over one hundred million US dollars [4]. Furthermore, attrition is commonly reported in prospective studies where participants are hard to trace for follow-up research [5, 6]. Attrition compromises the integrity of research by threatening both the internal and external validity of empirical results, weakening the accuracy of statistical inferences and the generalizability of findings [7–13].

A possible option that has the potential to facilitate longitudinal research is to trace original participants from historical cohorts, and ultimately conduct follow-up research. The term "tracing" or "tracking" is used to outline the process of locating, targeting, and re-engaging with participants during specific periods to measure changes and trends in variables of interest [14, 15]. Traditionally, common methods such as advertisement posters, business cards, mailshots, newspapers, and radio advertisements have been used to recruit participants, but tend to be time-consuming, costly, and produce low response rates [16]. The past couple of decades have seen remarkable advancements in information systems including new platforms that have contributed to an evolution in participant recruitment and tracing methods. For example, while web search phone directories were used in research to trace cohort participants [11], such methods have now become less useful or obsolete. Recently, social networking sites (SNS) have become an avenue for participant recruitment [5, 17, 18] and may be promising tools for participant tracing and retention. Accordingly, with the increasing use of SNS such as Facebook, Twitter (X), LinkedIn, Instagram, etc., there is an opportunity to leverage these platforms for recruitment and cohort study follow-up, potentially improving efficiency and facilitating the collection of more accurate and timely data. Indeed, with over 4.5 billion global users of SNS [19], researchers can now reach a large and diverse pool of potential participants in a timely manner.

Given the size and reach of SNS, it represents a potential avenue to mitigate the challenges related to participant tracing and attrition in longitudinal cohort studies, including offsetting financial burdens and increasing efficiencies. The use of Facebook in particular has been implemented by researchers to re-engage participants in follow up studies, leading to improved response rates and reduced attrition [5, 6, 18, 20]. For example, a study by Bolanos et al. [20] aimed to maximize follow-up response rates in a longitudinal study of adults using methamphetamine and successfully located 8.7% of participants through Facebook, with 2% completing the follow-up interview. In another study, the use of Facebook led to a 16% reduction in loss to follow-up in a longitudinal research project with two follow-up points at seven and ten years [5]. Moreover, using SNS may also maximize the recruitment and retention of certain cohorts that engage in social media daily, such as adolescents [18]. For example, a study found that targeting adolescent girls through Facebook minimized loss of follow-up, enabling them to find 45% of the lost cohort ($n = 175$) [18]. SNS searching may be particularly useful for addressing stigmatized health-related concerns and in situations where in-person methods may be restricted, such as during the COVID-19 pandemic [21]. In addition to the use of SNS, studies suggest using a multi-pronged strategy to increase participant recruitment rates by incorporating branding, outreach/networking, on-site presence, and incentives [22, 23].

While previous literature has demonstrated the potential benefits of using SNS to re-engage participants in longitudinal research [5, 22–24], the effectiveness of SNS for participant tracing and re-engagement in more recent years remains underexplored [25]. To our knowledge, no research has explored the effectiveness of using SNS to trace participants after decades of no contact. Considering the exponential growth and changing landscape of SNS, it is essential to re-evaluate the effectiveness of these various platforms in not only recruiting and retaining participants in longitudinal research but also tracing original participants from historical cohorts.

Materials and methods

Ethics consent to conduct this study and secondary analysis of the Hamilton Child Cohort Study (HCC) dataset, was received from Ontario Tech University (REB File #17547). Original participants of the HCC consented to be contacted for follow up research. Participants were traced using publicly available information on SNS.

Use of SNS to trace participants

We piloted the use of SNS including Facebook, Instagram, LinkedIn, and X to trace original participants from an earlier large-scale cohort study that examined the relative contribution of childhood exposure to indoor and outdoor air pollution to respiratory health outcomes, the HCC (n = 3,202, 1976–1986), please refer to Pengelly et al. [26] for a detailed description, as well as a related cohort study, the Reconstructed HCC (RHCC) (n = 397, 2003–2008) [27, 28]. From the total sample of the original HCC (n = 3,202), 36 participants who chose not to be contacted for follow-up were excluded from the tracing process. For the current study, 3,166 participants were traced through SNS over the span of 11 weeks from October 12th to January 6th, 2022. Personal identifying data of participants was leveraged from the initial study completed in 1986, including their first and last name, age, sex, educational affiliation, and geographic locations. These children participated in an earlier study based in Hamilton, Ontario between 1976 and 1986 that examined the link between exposure to air pollution and respiratory health. At the start of the study, participants were between the ages of 6 and 8 years and resided in four distinct neighborhoods characterized by gradients in residential mobility and air pollution levels [26, 28]. The neighborhoods included the east lower (EL), west lower (WL), east upper (EU), west upper (WU), and industrial core of Hamilton. Participants were also previously contacted in 1986 and traced again from 2005–2007 using two web-based directories: Yellowpages.ca and Canada411.ca [28].

For the current pilot study, social media accounts with similar names [Cohort Study on Health and Air Quality (C-SH*AIR*)], were created on four different SNS, including Facebook, LinkedIn, and X, and Instagram. Each account included details about the research team and about the original study. To identify participants, the search engine of each SNS was used to search for individuals from the original cohort by inputting their first and last name. Facebook and LinkedIn were chosen as the first method of tracing participants, followed by Instagram and X, if participants were not identified using the former SNS. Each profile corresponding to that name was reviewed. The biographies of the accounts that matched the names were examined to see if potential similarities were found when cross-referencing the information to the data collected from the original study. An excel datasheet was used to keep note of recruitment methods, the number of participants traced to each social media account, current demographic information of participants, and relevant communication with participants. Individuals linked to one social media account based on their identifying information were labelled in this study as having social media presence (SMP). Those that we did not find with a social media account were labelled as having no SMP (NSMP).

Updated demographic characteristics retrieved from the social media profiles of potential participants included their current location, educational affiliations (secondary and post-secondary schools), and marital status. The updated database, including the SMP and NSMP cohort's demographic details retrieved from SNS, was merged with the original study database with the participant's corresponding ID numbers and names. Demographic characteristics and health-related information from childhood of each cohort

were included from the original study database that took place in 1976–1986 by researchers from McMaster University.

Analysis

Results of the online search strategy using SNS, in relation to proportions of participants identified with SMP and those with NSMP were analyzed. Characteristics of the SMP and NSMP cohorts were analyzed in relation to childhood demographic characteristics, including biological sex (male or female), family income during childhood, and neighborhood of residence. Health-related information from childhood included smoking during childhood, asthma diagnosis, hospitalization, chest illness, and persistent cough. Statistical differences in participant characteristics among the SMP and NSMP cohorts were compared using the Pearson chi-square statistic. Descriptive statistics were used to summarize the current demographic characteristics of the SMP cohort, including their location of residence, educational affiliations, and marital status. All analysis was conducted using IBM SPSS Statistics (Version 28) with significance level set at p = 0.05.

Results

Using Facebook, a total of 479 individuals were identified who were considered to be participants of the HCC study. Using LinkedIn, 185 participants were identified. Instagram was used to identify nine potential participants. Only one participant was identified on X. Of the 3,166 participants traced, 21% were linked to one social media account and thus identified as having SMP, while 10% were not found or traced to any social media accounts (NSMP). Overall, 68% of participants of the original cohort were not able to be conclusively identified for reasons such as being linked to multiple accounts with the same name, accounts with not enough information, or if ages did not match the estimated cohort age. Of the people identified to have SMP, 15% (n = 479) and 6% (n = 185) of the sample were found through Facebook and LinkedIn, respectively (Figure 1).
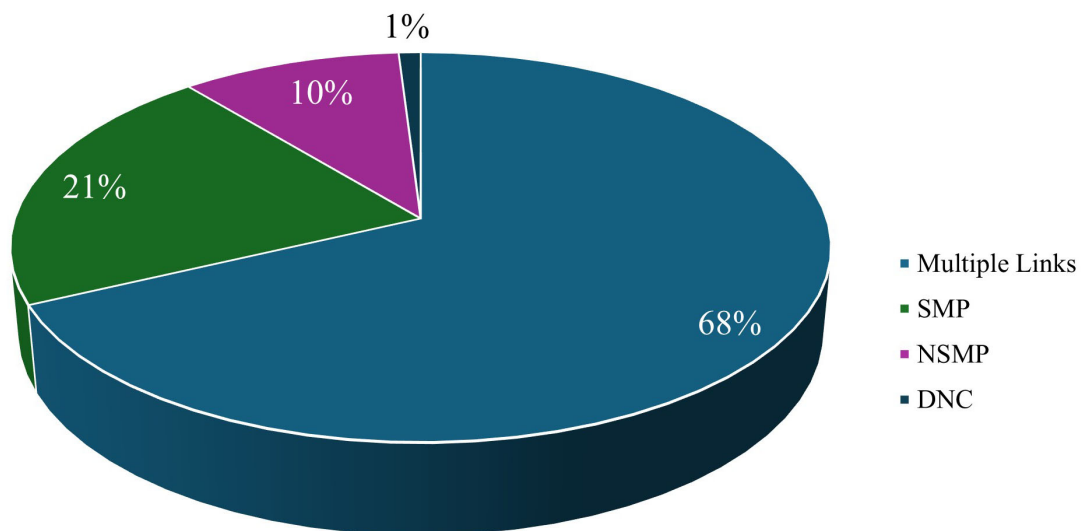

Figure 1. Findings of the total Hamilton Child Cohort Study (HCC) sample that were traced through social networking sites (n = 3,166). DNC: participants who requested in 2006 to not be contacted again. SMP: social media presence; NSMP: no SMP

Demographic and health-related characteristics of the SMP and NSMP cohort are summarized in Table 1. There was a significant difference in relation to the proportion of males (59%) compared to females (41%) traced through social media (χ^2 = 12.83, p < 0.001). In the current study, more individuals of the SMP cohort resided in the WU neighbourhood during childhood (31%), whereas the highest proportion of the NSMP cohort resided in the EL neighbourhood (34%) (χ^2 = 8.23, p = 0.042). A higher proportion of participants whose family income in childhood was above $20,000 were located through social media, compared to their counterparts with NSMP (56% and 53%, respectively). In relation to health

characteristics, a higher proportion of individuals with NSMP reported ever having asthma (9%), compared to their counterparts traced to a social media account (7%). In addition, about 7% of the NSMP cohort were found to report chest illness during childhood, compared to 6% of the SMP cohort. A higher proportion of individuals in the NSMP cohort were hospitalized at least once compared to those of the SMP cohort (6.9% vs 5.6%, respectively). Compared to their counterparts, a higher proportion of individuals of the NSMP cohort were reported to have smoked during childhood (24% vs 22%). These differences were not statistically significant.

Table 1. Demographic and health-related characteristic comparisons between the SMP and NSMP cohorts

Characteristics	Social media presence (n = 665)		No social media presence (n = 334)	
	n	%	n	%
Sex**				
Female	271	40.8	176	52.7
Male	394	59.2	158	47.3
Neighborhood of residence in childhood*				
West upper	207	31.6	78	23.5
West lower	168	25.6	85	25.6
East upper	99	15.1	57	17.2
East lower	182	27.7	112	33.7
Family income in childhood				
Over $20,000	362	56.1	167	52.7
$10,000–20,000	192	29.8	105	33.1
Under $10,000	91	14.1	45	14.2
Persistent cough in childhood				
No	537	96.2	261	95.3
Yes	21	3.8	13	4.7
Chest illness in childhood				
No	623	94.1	309	93.1
Yes	39	5.9	23	6.9
Hospitalization before the age of two				
Unknown/NA	628	94.4	311	93.1
Hospitalized at least once or more	37	5.6	23	6.9
Asthma ever				
No	619	93.1	305	91.3
Yes	46	6.9	29	8.7
Smoked during childhood				
No	481	77.7	236	76.1
Yes	138	22.3	74	23.9
Current residence				
Hamilton	409	61.5	-	-
Halton region cities	20	3.0	-	-
Toronto	15	2.3	-	-
Other	76	11.4	-	-
Out of province	10	1.5	-	-
Missing	135	20.3	-	-

Significance tests indicate differences between SMP cohort and NSMP cohort. *N* may not add up due to missing data. NA: not applicable. *: $p < 0.05$; **: $p < 0.001$; -: no data

Current demographic characteristics of the SMP cohort

The study was able to deduce the current location of 80% of the SMP cohort (Table 1). Of those individuals, 409 (61.5%) reported that they still live in Hamilton, followed by 20 individuals currently living in other Halton Region cities (e.g., Burlington, Oakville), and the rest dispersed across provinces in Canada, or

residing in different countries. In relation to educational status, 64% of individuals had data available on the secondary school (*n* = 159) or post-secondary institution (*n* = 266) attended. The proportion of participants of the SMP cohort that had data available on the high school they attended, and who still reside in Hamilton was 61%. The most attended high schools included Hill Park Secondary School (14%), Scott Park Secondary (13%), and Westmount Secondary School (9%). Furthermore, of the SMP cohort that had available data on their post-secondary institution, 70% still resided in Hamilton (Figure 2). The most attended post-secondary institution was Mohawk College (54%), followed by McMaster University (24%), Sheridan College (3%), Humber College (2%), and others (18%). This information suggests that a majority of the original cohort stayed in Hamilton for post-secondary education and indicates that many stayed in this area throughout their lifespan. In relation to marital status, 141 individuals had their marital status available on their social media page, and 60% reported being married, 11% were in a relationship, and 22% were single. The remainder indicated that they were either divorced or in a complicated relationship.
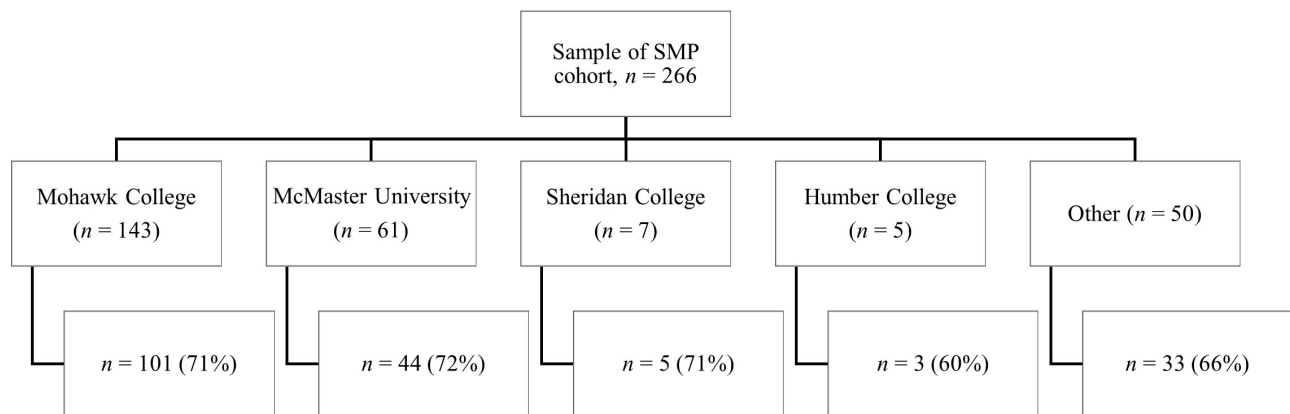

Figure 2. Flow diagram of the most prevalent post-secondary schools that participants of the SMP cohort attended and the proportion of individuals that are still located in Hamilton, Ontario. SMP: social media presence

Discussion

We traced participants from a previous cohort study, the HCC study, which was conducted between 1978 and 1986. We utilized only publicly available information through social media platforms to determine the effectiveness of using SNS to trace and recruit a reconstructed cohort for follow up research. Through SNS such as Facebook and LinkedIn, we were able to trace 21% of participants to one social media account. Characteristics between participants identified and linked to one social media account and determined as having SMP and those with NSMP differed in terms of sex and neighborhood of residence. In addition, the NSMP cohort had higher rates of childhood illness.

The initial study involved children as participants and aimed to evaluate the long-term effects of childhood exposure to air pollution. As the years have passed and the children from the original study have transitioned into adulthood, locating them for follow-up research has become increasingly challenging due to various factors such as changes in contact information, relocation, and potential name changes. Even for participants that were traced in 2006, the likelihood of maintaining accurate contact information decreased over time. Participants may have changed their names, no longer have contact with parents, passed away, or moved away, further complicating the tracing process. To address these challenges, our study leveraged the power of social media as a tool for locating participants. By utilizing SNS, we were able to conduct searches by names, which facilitated the process of finding participants who may have been otherwise difficult to locate. However, tracing participants with common names through social media sources proved to be a challenge due to the higher frequency of possible matches.

In a previous related study [28] that analyzed predictors of locating participants of the original cohort study, males were more likely to be located when compared to females (59% vs 41%). This is consistent with the current research as more males were found through social media when compared to females.

Various other studies also convey the same findings, proposing sex or gender as a factor that influences the success rate of tracking cohorts [29]. This may be due to the fact that most women after marriage are more likely to adopt their partner's name, making it more difficult for researchers to track them after years have passed with no contact. Moreover, we found that the SMP cohort was more likely linked to the WU region, which indicated that they came from a relatively higher socioeconomic background in childhood. In addition, childhood health history of the SMP cohort was relatively better in comparison to the NSMP cohort. A previous study of the HCC by Barakat-Haddad et al. [27], found that those residing in the EL and WL were more likely to have a family income below the income cut off, higher levels in utilizing gas cooking, parental smoking, and sharing a room with at least two family members. Thus, these demographic characteristics linked to health-related outcomes seem to influence tracing success. Indeed, Barakat-Haddad et al. [28] previously found that participants who were above the low-income cut off in childhood and those who were not exposed to smoking were more likely to be traced after 20 years of no contact.

Other factors that may impact the success rate of tracing participants include the periods between follow up. For instance, findings of a study that focused on the effectiveness of cancer screening trials employed two and five years follow up periods throughout the study and found better results in the shorter periods when contrasted to the longer follow up [30]. When a two year follow up was employed, 2,372 participants out of the original 154,901 were traced. However, after a five year follow up period, this decreased to 909 individuals being located [30]. Another study that implemented a two year follow up survey located 86% of a cohort of 2,631 participants [31]. In this study, between 17 years and 37 years had passed with no contact with participants, which hindered the success rate of tracing cohorts [14, 15]. Our results suggest that participants who included information on their location, age, academic enrollment, and given name were more likely to be successfully traced on Facebook and LinkedIn. Overall, 21% of the cohort was traced using this method. In comparison, a study conducted in 2019 examining the success of relocating 237 participants to complete a postpartum survey after five years had passed—using Facebook search engines, was able to find 47% of the cohort [7].

The use of social media has proven to be a more efficient method of tracing participants, as compared to traditional methods of data collection [32]. Our study found that a notable number of participants were identifiable through their social media profiles, which allowed for effective and efficient re-engagement with the cohort. This aligns with previous literature that also found the use of recruitment through social media to be more efficient when compared to traditional methods of recruitment [33]. In addition, the use of social media platforms such as LinkedIn and Facebook enabled us to reach participants who had moved away from the study area, which would have been difficult to achieve using traditional methods such as phone calls or mailing letters. Facebook's extensive user base and LinkedIn appeal to professionals and older demographics were also a strength in tracing our cohort. Furthermore, information related to where they attended school may be used to locate and re-engage with participants by contacting that school and organizing alumni events in order to invite participants back as an incentive for future follow-up research. In addition, marital status can be used to reach participants/partners. The cost-effectiveness of using social media is also notable, as it eliminates the need for more expensive traditional methods. This not only saves time and resources but also reduces the potential for errors that may occur during the manual collection of data. Furthermore, the use of social media platforms allowed for easy dissemination of study updates and reminders to participants which may contribute to higher participation rates in subsequent follow-up studies.

Our study has demonstrated the efficacy of SNS in tracing research participants while also revealing demographic limitations that could affect the representativeness of certain groups within the general population. Specifically, our research focused on childhood illness, and we found no significant differences in the prevalence of these conditions between participants identified through social media and those who were not. However, given that this was a pilot study and that the NSMP cohort had relatively worse health outcomes based on the retrieved childhood health history in comparison to the SMP cohort, there may be differences during future research as we trace and re-engage with more participants.

Furthermore, unlike previous studies conducted [18] which focused solely on Facebook, we expanded our approach to include multiple platforms. Notably, our study included LinkedIn, which proved particularly effective for tracing participants due to the rich professional and educational information available on profiles. This addition was useful for our research as it provided deeper insights into the backgrounds of participants, which can be valuable in public health studies.

It is important to note the limitations imposed by using social media for tracing participants. For example, the tracing team encountered problems being labeled as spam on Facebook due to a high volume of messages being sent in a short period of time to potential participants. This resulted in the suspension of the account which delayed the tracing process. The tracing team had to request a re-evaluation of the suspension for Facebook to unsuspend the account. Another barrier when tracing participants on Facebook was that the messages were being sent to their spam box instead of message requests. We addressed this barrier by shortening the informative message of the study to the participants. Facebook also consistently asked for the account to be authorized, putting a halt to advertisements being viewed by users and reaching out to potential participants. To avoid barriers in future projects, researchers should consider requesting verification of a Facebook account before starting tracing and include paid advertisements as this increases the viewership of the Facebook page created for the respective research study.

Likewise, LinkedIn posed limitations as well, including the restriction on messaging non-connected users. In order to overcome this limitation, our research team opted to acquire a LinkedIn business premium account. This provided us with access to the InMail feature, enabling us to send messages to potential participants without being connected to them. However, it is worth noting that the cost of such premium subscriptions can be significant.

Another limitation we encountered was that a number of individuals were not active on social media platforms, which made it difficult to track them. Specifically, the lower success rate on Instagram and X can be attributed to multiple potential factors. These platforms predominantly attract younger users, which conflicts with the study demographic as our participants are middle aged (50–65). For example, a study found that around 71% of Instagram users ranged within the emerging adult population (18 to 29) [34]. Individuals ranging within the older age group are often less represented, having limited to no presence on X and Instagram. This generational gap as well as the platforms' designs being less focused on sharing personal identifying details—crucial for the accurate tracing of individuals, are potential reasons and possible limitations of using specific social media sites for tracing the sample demographic of this study.

The potential of biases arising from samples for the follow-up study must be considered to ensure accuracy. Sample bias may occur if individuals being traced in the study are more likely to be selected than participants who are lost to follow up due to differences in certain variables that are being examined [35]. To combat possible bias and loss of accuracy in the results of cohort follow up, studies have suggested an ideal percentage of loss to follow up. According to various studies, 5–60% is the agreed upon ideal loss to follow up that will result in an unbiased estimate of influence [36]. In general, researchers trust that an 80% capture rate is sufficient [37].

Moreover, when using SNS to trace participants, many individuals had common names with multiple links, which made it challenging to confirm if the individual found was the original participant. Another limitation was the risk of selecting the wrong individual due to the lack of demographic data available on social media platforms. This limitation made it challenging to differentiate between individuals with the same name, as social media platforms do not always provide enough information to confirm their identity. These limitations highlight the need for caution when using social media as a tracing tool and the importance of utilizing multiple methods of contact to increase the chances of locating all participants.

As the use of social media emerges, opportunities present to enhance the recruitment process for longitudinal cohort studies. Given the growing trend of paid advertisements, it is imperative to assess their efficacy and weigh their advantages against their drawbacks. Moreover, an integrated approach that merges conventional tracing methods (e.g., referrals, phone calls, flyers/posters, and post mail) with modern digital strategies such as SNS, appears promising and warrants further investigation. For future

projects involving tracing, it is recommended to allocate a dedicated budget for the purposes of connecting with participants on LinkedIn—particularly given that the sales navigator core subscription—which allows for up to 50 InMails per month, comes at a monthly cost of over $85 CAD. In addition, it is recommended that to increase viewership of recruitment material through Facebook advertisements, a one-time fee ranging between $5–10 CAD should be considered, which can potentially reach around 1,000 individuals. Advertisements can also be adjusted to the specific audience of interest such as age, interests, email address, and location. In upcoming research, our intention is to not only recruit but also re-engage participants who may have been lost to follow-up. As we move forward with subsequent phases of our study, leveraging the power of digital platforms in participant recruitment and retention will be of utmost importance.

Overall, this study has demonstrated the potential of using social media to trace participants from a historical cohort study. Despite the challenges faced in relation to significant time gaps between studies, and the limitations inherent to using social media platforms, research and results have shown that this method of participant tracing can be efficient and cost effective. However, while social media is a valuable tool to track participants, it is key to acknowledge its limitations such as account suspensions, spam bots, difficulties differentiating between individuals with similar names, and sampling bias to name a few. It is also important to recognize that due to a range of structural factors, and from an equity perspective, those most marginalized are the most difficult to find. To reduce these challenges and minimize bias, future research should implement a multi-faceted approach to participant re-engagement, using both traditional and modern methods. This study adds to the growing body of literature that supports the use of social media as a participant tracing tool in longitudinal studies. This may help improve processes of retaining participants for longitudinal research and the validity of findings from the long-term effects of various exposures and interventions on human health and wellbeing.

Abbreviations

EL: east lower

HCC: Hamilton Child Cohort Study

NSMP: no social media presence

SMP: social media presence

SNS: social networking sites

WU: west upper

X: Twitter

Declarations

Author contributions

RRA, SJY, and SSM: Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing. SE and CB: Supervision, Conceptualization, Data curation, Resources, Software, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical approval

Ethics consent to conduct this study and secondary analysis of the HCC study dataset, was received from Ontario Tech University (REB File #17547), and complies with the Declaration of Helsinki.

Consent to participate

Informed consent to participate in the study was obtained from all participants.

Consent to publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

Copyright

References

1. Desai M. Recruitment and retention of participants in clinical studies: Critical issues and challenges. Perspect Clin Res. 2020;11:51–3.
2. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemp Clin Trials Commun. 2018;11:156–64.
3. Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to Establish and Follow up a Large Prospective Cohort Study in the 21st Century - Lessons from UK COSMOS. PLoS One. 2015;10: e0131521.
4. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. Am J Nurs. 1978;78: 1039–40.
5. Mychasiuk R, Benzies K. Facebook: an effective tool for participant retention in longitudinal research. Child Care Health Dev. 2012;38:753–6.
6. Masson H, Balfe M, Hackett S, Phillips J. Lost without a Trace? Social Networking and Social Research with a Hardto-Reach Population. Br J Soc Work. 2011;43:24–40.
7. Stephenson NL, Hetherington E, Dodd S, Mathews A, Tough S. Mitigation of Participant Loss to Follow-Up Using Facebook: All Our Families Longitudinal Pregnancy Cohort. J Med Internet Res. 2019;21: e10441.
8. Price SM, Park CH, Brenner RA, Lo A, Adams S, Baetz RA, et al. Participant retention in a longitudinal study: do motivations and experiences matter? Surv Pract. 2016;9.
9. Davis E, Demby H, Jenner LW, Gregory A, Broussard M. Adapting an evidence-based model to retain adolescent study participants in longitudinal research. Eval Program Plann. 2016;54:102–11.
10. Gustavson K, Soest Tv, Karevold E, Røysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health. 2012;12:918.
11. Kleschinsky JH, Bosworth LB, Nelson SE, Walsh EK, Shaffer HJ. Persistence pays off: follow-up methods for difficult-to-track longitudinal samples. J Stud Alcohol Drugs. 2009;70:751–61.
12. Young AF, Powers JR, Bell SL. Attrition in longitudinal studies: who do you lose? Aust N Z J Public Health. 2006;30:353–61.
13. Ribisl KM, Walton MA, Mowbray CT, Luke DA, Davidson WS, Bootsmiller BJ. Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. Eval Program Plann. 1996;19:1–25.

14. Teague S, Youssef GJ, Macdonald JA, Sciberras E, Shatte A, Fuller-Tyszkiewicz M, et al. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Med Res Methodol. 2018;18:151.

15. Ermakova T, Fabian B, Bender B, Klimek K. Web Tracking – A Literature Review on the State of Research. AIS eLibrary. 2018.

16. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. Adv Psychiatr Treat. 2003;9:229–38.

17. Barakat C, Dash S, Klicnik I, Mohammed S, Yousufzai S. A cross-sectional study to examine constraints to sport participation among ethnically diverse female adolescents from Ontario, Canada. Int J Sport, Exercise Health Res. 2021;5:68–75.

18. Jones L, Saksvig BI, Grieser M, Young DR. Recruiting adolescent girls into a follow-up study: benefits of using a social networking website. Contemp Clin Trials. 2012;33:268–72.

19. Dixon SJ. Number of social media users worldwide from 2017 to 2028 [Internet]. [cited 2023 Dec 12]. Available from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/#:~:text=Social%20media%20usage%20is%20one,almost%20six%20billion%20in%202027

20. Bolanos F, Herbeck D, Christou D, Lovinger K, Pham A, Raihan A, et al. Using facebook to maximize follow-up response rates in a longitudinal study of adults who use methamphetamine. Subst Abuse. 2012;6:1–11.

21. Walsh LE, Carter-Bawa L. Using social media to recruit individuals for health-related research: Feasibility and lessons learned. J Health Psychol. 2023;28:599–606.

22. Hurwitz LB, Schmitt KL, Olsen MK. Facilitating Development Research: Suggestions for Recruiting and Re-Recruiting Children and Families. Front Psychol. 2017;8:1525.

23. Michalek AK, Kan D, Prochaska J. Engaging veterans with substance abuse disorders into a research trial: success with study branding, networking, and presence. Transl Behav Med. 2015;5:167–76.

24. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs. 2010;19:227–33.

25. Pekarsky C, Skiffington J, Leijser LM, Slater D, Metcalfe A. Social Media Recruitment Strategies to Recruit Pregnant Women Into a Longitudinal Observational Cohort Study: Usability Study. J Med Internet Res. 2022;24:e40298.

26. Pengelly LD, Kerigan AT, Goldsmith CH. Sensitivity of children with asthma to chronic low-level sulphur dioxide exposure. In: Brasser LJ, Mulder WC, editors. Man and his Ecosystem. Proceedings to the 8th World Clean Air Congress; 1989 Sep 11-15; The Hague, The Netherlands. Amsterdam: Elsevier; 1989. pp. 7–12.

27. Barakat-Haddad C, Elliott SJ, Pengelly D. Does Chronic Exposure to Air Pollution in Childhood Impact Long-Term Respiratory Health? Prof Geogr. 2012;64:446–63.

28. Barakat-Haddad CP, Elliott SJ, Eyles J, Pengelly D. Predictors of locating children participants in epidemiological studies 20 years after last contact: Internet resources and longitudinal research. Eur J Epidemiol. 2009;24:397–405.

29. Hampson SE, Edmonds GW, Barckley M, Goldberg LR, Dubanoski JP, Hillier TA. A Big Five approach to self-regulation: personality traits and health trajectories in the Hawaii longitudinal study of personality and health. Psychol Health Med. 2016;21:152–62.

30. Marcus PM, Childs J, Gahagan B, Gren LH. Tracking and tracing of participants in two large cancer screening trials. Contemp Clin Trials. 2012;33:601–5.

31. Warner ET, Glasgow RE, Emmons KM, Bennett GG, Askew S, Rosner B, et al. Recruitment and retention of participants in a pragmatic randomized intervention trial at three community health clinics: results and lessons learned. BMC Public Health. 2013;13:192.

32. Darko EM, Kleib M, Olson J. Social Media Use for Research Participant Recruitment: Integrative Literature Review. J Med Internet Res. 2022;24:e38015.

33. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh. 2013;45:256–64.

34. Brooke Auxier MA. Social Media Use in 2021 [Internet]. Pew Research Center; 2024 [cited 2023 Sep 15]. Available from: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

35. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010;126:619–25.

36. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19:751–60.

37. Fischer EH, Dornelas EA, Goethe JW. Characteristics of people lost to attrition in psychiatric follow-up studies. J Nerv Ment Dis. 2001;189:49–55.