Harris, J, Germain, J, Maxwell, C and Mackay, S

The Ethical Implications of Collecting Data from Online Health Communities

http://researchonline.ljmu.ac.uk/id/eprint/12197/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Harris, J, Germain, J, Maxwell, C and Mackay, S (2020) The Ethical Implications of Collecting Data from Online Health Communities. SAGE Research Methods Cases: Medicine and Health. ISSN 9781529716085 (ISBN)

The Ethical Implications of Collecting Data From Online Health Communities

Jane Harris

Liverpool John Moores University, UK

Jennifer Germain

Liverpool John Moores University, UK

Clare Maxwell

Liverpool John Moores University, UK

Sean Mackay

Liverpool John Moores University, UK

Discipline

Public Health [D26]

Sub-discipline

Health Services Research [SD-PH-3]

Academic Level

Postgraduate

Contributor Biographies

Jane Harris is a Research Fellow at Edge Hill University. Her PhD examined the role that professional YouTubers play in young people's health behavior in the United Kingdom.

Jennifer Germain is a researcher based at the Public Health Institute, Liverpool John Moores University. Her PhD research explored the use of unlicensed weight loss drugs in women.

Clare Maxwell is a registered Midwife and Senior Lecturer in Midwifery at the School of Nursing and Allied Health, Liverpool John Moores University. Her PhD explored mothers' experiences of bottle refusal by their breastfed baby.

Sean Mackay is an Associate Dean and the Program Leader for Primary Care program at the School of Nursing and Allied Health, Liverpool John Moores University. His PhD research focused on online peer support for fathers.

Published Articles

Germain, J., Harris, J., Mackay, S., & Maxwell, C. (2018). Why should we use online research methods? four doctoral health student perspectives. *Qualitative Health Research, 28*, 1650–1657. doi:10.1177/1049732317721698

Abstract

Online communities can provide researchers with a raft of knowledge about a range of populations and groups. While ethical considerations in online research are complicated and nuanced, they are often underreported in existing research. Issues concerning the use of verbatim quotes from online communities, seeking consent, and protecting from harm and distinctions between private and public spaces in particular have generated much debate in recent years. We advocate for greater thought to be given to the ethical implications of online research and reflect on what should be considered public information, the protection of anonymity, and how to protect online users from harm. This case study provides an insight into the practicalities of conducting online research using examples from research conducted by four doctoral students.

Learning Outcomes

By the end of this case, students should be able to

- Identify reasons why online health communities may be appropriate to address their research objectives
- Understand the importance of perceptions of public and private space, seeking consent and verbatim quotes to ensuring ethically sound online research
- Appraise the potential barriers to accessing online communities and the challenges of validating demographic data online
- Demonstrate their ability to enter into meaningful dialogue with ethics committees about the issues pertaining to online research

Case Study

Project Overview and Context

Over the past decade, there has been growing interest in online communities and environments for public health research. The Internet has been described as the “laboratory for the social sciences” and it can be argued that people’s online habitats have become just as important as other environments where human interaction occurs (Eynon, Fry, & Schroeder, 2008; Hallett & Barber, 2014). Studies that adapt traditional quantitative and qualitative methods to online groups, such as online interviewing and online questionnaires, are well defined in the published literature (Ayling & Mewse, 2009; McDermott & Roen, 2012; Seko, Kidd, Wiljer, & McKenzie, 2015). However, methods which involve engaging with online health communities to collect data are less well described. The ethical implications of these methods are the focus of this case study.

This case study draws on experiences from four doctoral projects. The primary case study is PhD research which analyzed 10 online forums to develop a greater understanding of women's use of unlicensed weight loss (UWL) drugs. UWL drugs are drugs which have previously been licensed for weight loss but subsequently withdrawn from the global market, such as sibutramine or rimonabant (appetite suppressants), and those which are untested, such as 2,4-dinitrophenol (DNP; fat burner). These drugs are just one method used to enhance and improve the body in line with social expectations; expectations that are increasingly prominent for women. Despite this, little is known about the use of these drugs in women. Research which does exist focuses mainly upon male-dominated bodybuilding communities or single-case studies in the case of DNP (Lee et al., 2014; McFee, Caraccio, McGuigan, Reynolds, & Bellanger, 2004; McVeigh, Germain, & Van Hout, 2016; Petróczi et al., 2015) or the efficacy of the drugs prior to their being withdrawn from the market in the case of rimonabant and sibutramine (Arterburn, Crane, & Veenstra, 2004; James et al., 2010; Sam, Salem, & Ghatei, 2011; Scheen, 2011).

While we largely focus on the aforementioned UWL drug study, throughout this case study we also present examples from three other PhD projects. These studies explored the role of professional YouTubers in delivering health messages, mothers who are experiencing bottle refusal by their breastfed baby, and peer support for new fathers. All four studies are summarized in Table 1. As our research progressed, many ethical questions about how to best collect data became apparent. To help with these decisions, we formed an Online Methods Group (OMG), involving each student, where we discussed the ethical and methodological issues we were encountering. This case study presents the outcomes of our discussions to help other students who are considering analyzing online data for health research.

Table 1.

Caption: PhD studies referred to in this case study.

Study Code	**Study Title**	**Types of Discussion**	**Population**	**Platforms Accessed**

UWL drug	An exploration of the female use of UWL drugs	Illicit/sensitive drug use	Females: undefined group	Online forums
YouTubers	Examining the role that professional YouTubers play in young people's health behaviors in the United Kingdom	Non-sensitive health messages	Young people living in the United Kingdom 13–18 years	YouTube
Mothers	UK mother's experiences of bottle refusal by their breastfed baby	Sensitive and non-sensitive advice and support	Mothers who are experiencing/have experienced bottle refusal by their breastfed baby	Online forums, Facebook
Fathers	Online peer support for fathers	Sensitive and non-sensitive advice and support	Fathers of pre-school children	Online forum

UWL: unlicensed weight loss.

Research Design

It can seem very tempting when undertaking health research to look at the large amounts of data available online and consider it a quick and easy way to access people's perceptions, experiences, and opinions of health issues. However, when deciding to collect online data, we found several vital considerations for ensuring good-quality and ethical data collection. The first of these is to consider the population of study and whether online methods are an appropriate way to gather data from them which answer the research objectives.

At the planning stage of the UWL drug study, fitness competitions, modeling companies, gyms, and beauty salons were all considered as possible recruitment locations. However, there was no evidence that UWL drug users actually use or engage with any of these locations. Furthermore, those using UWL drugs are unlikely to see themselves as drug users and are less likely to engage with health professionals about their drug use, making them difficult to recruit via services. What was known is this group is active online: UWL drugs are bought online and their use is discussed on online forums. The objective of the UWL drug study was to understand more about this undefined population and their motivations for use. Online forums provide an anonymous venue for those engaging in illicit activity to seek advice and share experiences that they cannot share offline. Also people are more likely to use online forums to discuss illicit or sensitive behavior due to the potential anonymity they offer (Langer, Elliott, & Beckman, 2005), although this is influenced by perceptions of the visibility of the forum and the risk of social stigma (Barratt, 2011). Comments taken from online forum posts such as *"I haven't told any of my family or friends that I am taking sibutramine"* illustrate that forum users seemingly prefer to discuss their drug use on forums rather than with people they know offline. Therefore online forums present a natural opportunity to find out about UWL drug use.

Risky or illicit behaviors are not the only reason that online recruitment might be the optimal method. In the fathers study, the population under investigation was seldom heard and the reasons for this were nuanced, reflecting societal expectations of masculinity and help-seeking behavior. Online forums provide fathers with the opportunity to seek advice with anonymity in a manner that they may struggle to do offline. In contrast, some populations may be suitable because they are well established and highly visible online as was the case for mothers seeking parenting support. Online pregnancy and parenting forums provide huge potential for research with the two largest UK forums, Mumsnet and Netmums, having more than 19 million unique users and 130 million page views per month ("About Us," 2018; "Netmums in Numbers," 2017). The objective of the mothers study was to explore mothers' experiences of bottle refusal by their breastfed babies. Mothers use online forums and social media to discuss topics that are against official health guidance, possibly believing such platforms to be nonthreatening and less judgmental. These forums gave the researcher access to discussion on their research topic that would not necessarily be openly discussed elsewhere. Finally, some populations are defined by characteristics that exist solely online as was the case with YouTubers and their young audience. The objective of the YouTuber study was to understand how YouTube content influenced young people's health behaviors. Since YouTuber fan communities have developed largely online, their demographic characteristics are undefined, and thus, difficult to recruit offline, particularly if young people do not wish to disclose their online activities to their peers.

The populations discussed here present two extremes: populations that are largely invisible and widely visible social networks. By grouping our study populations in this way, we are not suggesting that online research is only suitable in these extreme cases. Online research can also be appropriate for more general identification of online peer support groups (Suzuki & Calzo, 2004) as well as researching specific health conditions (VanDam, Kanthawala, Pratt, Chai, & Huh, 2017; Zhang et al., 2016). Our intention here is simply to highlight the breadth of potential for online methods and the importance of considering your population before you commence research of this nature. When making the decision to collect data from online forums and social network sites, researchers should ask themselves: is online data collection appropriate for my population and research objective? Will the data I need to answer my research question be available? And how will the information individuals share online differ from what they might share with a researcher or health care professional?

Section Summary

- Online research presents opportunities for health researchers to access populations, opinions, and experiences which can be challenging to capture offline.
- Researchers considering using online communities to collect data must consider how appropriate online methods are for their target population and research objectives.

Research Practicalities

Ethical Considerations in Online Research

Online forums were chosen as a suitable place to research female UWL drug users. However, while data on UWL drug use were plentiful and publicly available online, this does not mean researchers should simply help themselves without first considering the ethical implications of taking these data. This is particularly significant if data collection is going to occur without the consent of the forum users who have shared their experiences or the forum moderators who manage the online groups. Scholars have given increasing attention to the ethics of online research over the past decade; however, the existing guidance is limited and on occasion contradictory which can be challenging for those wishing to undertake online research.

Crossing Boundaries? The Public/Private Debate

The British Psychological Society (BPS; 2013) general ethics code describes a public space as situations where individuals *"would expect to be observed by strangers."* Offline, this distinction is quite clear-cut, but what constitutes a public space is more difficult to define online. Internet communication is often conducted in both public (e.g., open discussion forum) and private (e.g., home) spaces simultaneously. The concept of "public/private" in online research is not a binary one and must consider users' expectations and be guided by both consensus and contextual integrity (Beninger, 2017; Markham & Buchanan, 2012). Malin Svenningson-Elm (2009) proposes that our notions of online privacy exist on a continuum from public websites (accessible to all without registration) to semi-public (accessible to all but requiring registration), semi-private (formal requirements for membership) and private (closed group). Where there is any level of ambiguity over data being in the public domain, the BPS recommends that researchers consider the extent to which undisclosed observations may have potentially damaging effects for participants.

The UWL drug study involved analyzing posts from public online forums. One important ethical issue was should online users be informed that their data were being used and should consent be sought? The guidelines around this issue are open to interpretation (BPS, 2013). Essentially, if the space is classified as private, then consent should be sought. However, the blurring of boundaries between private and public space online and the lack of clarity around what constitutes harm have led to differences in how online research is carried out. For the UWL drug study, consent was not obtained from the forum members nor were they informed that the study was taking place. The decision not to inform forum members was mostly for practical reasons. Large volumes of data were collected and it would have been extremely demanding to contact all forum members (between 5,000 and 20,000 members per forum) who had contributed to each thread. Furthermore, the posts were accessed retrospectively so may have been several years old and it was possible the contributors were no longer engaging with the site. If consent was gained, only those posts from members who had consented could be used, which would lead to disjointed threads with lost richness and meaning.

The level of intrusion caused by the research was minimal. There is no interaction between the participants and the researcher when accessing publicly available online posts, and so can be considered less intrusive than face-to-face or even survey-based research. However, researchers must consider the potential impact on individuals and communities of taking data without consent and this is revisited in greater detail later in this case. The BPS (2013) guidelines suggest permission should be gained from moderators but this is only necessary if the forum or online source will be named. In the UWL drug study, the forums were not named. Interviews with moderators of the forums were planned for a subsequent stage. However, when moderators were invited to participate in the interview stage, responses ranged from disinterest to negative. This included receiving no response, being told the research was not relevant or of value, being willing only to take part if there was a financial reward and in some cases resulted in the researcher receiving an automatic ban from the forum. Approximately 30 moderators were contacted with only four willing to take part; it was often difficult for the researcher to identify a specific reason for participation reluctance, particularly where automatic bans occurred or there was a lack of moderator engagement as there was no opportunity for further discussion. This underpins the argument that just because these communities exist, we as researchers cannot automatically assume they will want to engage in research, particularly for those groups which are hidden, hard to reach, or engaging in illicit behaviors.

The reluctance of forum moderators to participate in later stages of the PhD caused the researcher some tension in how those spaces continued to be viewed. If forum moderators did not want to be involved in interviews, would they have been willing to allow their posts to be included in research? This was a tension which was not wholly resolved throughout the research process. For those conducting research of this nature, the potential to cause harm by conducting online research in a covert nature has to be weighed against the value of the research findings and the potential benefits added knowledge and evidence can provide to those groups under investigation.

The decision about public and private spaces is one that must be made in the context of each research study. Similar to the UWL drug study, the YouTuber study did not seek consent to analyze the data from YouTuber videos or attached comments. Data were collected retrospectively, and the videos were in the public domain with no login required to view them and had hundreds of thousands of views. Therefore, it would have been extremely difficult and time-consuming to seek consent from each individual commenter. The YouTubers were pseudonymised (e.g., YouTuber1) and any personally identifiable information (such as names, city of residence) was removed. The fathers study used a semi-private online forum which required an account (which could be anonymous) to post on the forum. This particular forum required researchers to seek consent from the forum moderator (through the private messaging function) before the data were collected. Researchers willing to conduct online research should similarly check forum requirements.

Protecting From Harm

To minimize harm to both the online communities and researcher, the researcher must have a clear and detailed knowledge of the online community they are examining. There is much written on the public versus private space of the Internet (Roberts, 2015), and whether the researcher is eavesdropping on a private conversation in a public space. Although online discussion boards that require no registration or passwords to read their content may be viewed as in the public domain (Svenningson-Elm, 2009), the researcher should consider whether the contributors to the board did or could have an expectation of privacy. In the UWL drug study, the decision was made to only use public forums as this suggested that forum posters were aware that their information was public. However, a lack of engagement from moderators in later stages suggests that this was not universally the case.

A similar dilemma was faced during the YouTuber study. YouTube promotes itself under the tagline "Broadcast Yourself," and YouTubers, by the sheer size of their audience, are aware of the public nature of the content they produce. However, young people posting comments on such videos may not be explicitly aware that they too are posting in the public domain (Reilly, 2014). In the fathers study, where moderator consent was sought, there were requests for information and advice to others "*out there*," which seems to suggest that the site users understand that their messages are open to a wide audience, including strangers, which appears to meet the BPS (2013) definition of a public space. The site's guidance for new members reinforced this by advising against the posting of personally identifying information. In both studies, it was felt that taking appropriate steps to pseudonymize data and ensure verbatim quotes could not be traced back to the original site was sufficient to protect participants from harm, and the steps taken to do so are described in the following section. In the mothers study, interviews with mothers highlighted a small number of closed Facebook groups that the mothers used as sources of advice and support for bottle refusal. The fact that they were "closed" indicated the groups wanted some level of privacy concerning their posts and this was confirmed by the group administrator's rejection of the researcher's request to analyze data from these groups. In accordance with the administrator's wishes, the researcher did not include the closed Facebook groups and only collected data from public forums. These examples highlight that researchers must be responsive to the levels of privacy each online community and their users set.

Determining Whether to Quote or Not to Quote

In qualitative research, participants' words are quoted to provide meaning and richness to data presentation. This is usually verbatim, to reduce the risk of misinterpretation by paraphrasing and to establish the veracity of the researcher's interpretation (Guest, MacQueen, & Namey, 2012). However, both the BPS (2013) and the Association of Internet Research (AoIR; Markham & Buchanan, 2012) invite caution when quoting verbatim from online sources. Researchers must ensure that there are no serious risks of harm to the participants (BPS, 2013; Markham & Buchanan, 2012). In the UWL drug study, the forums did not need a password to access them, and so, it may have been possible to locate the original post by searching for the quote in a search engine. Therefore, it was necessary to decide the probability of harm to the poster if their quote were traced to the original post. This will be different for each research study, depending on the level of personal information and content of the post, and requires the researcher to make a pragmatic decision (Sharkey et al., 2011; Trevisan & Reilly, 2014).

Analyzing online forums is very different from analyzing a spoken interview. Tone of voice, and nonverbal cues are missing, making it difficult to ascertain people's intentions, for example, whether they are serious or making a joke (Kozinets, 2015). Using anonymous verbatim quotes keeps the participant's voice intact and retains a balance between meaning and ownership. Paraphrasing quotes adds an additional layer of researcher interpretation, and the omission of the verbatim quote when presenting findings takes away the opportunity for others to disagree with the researcher's interpretation. A further advantage of using verbatim quotes is that the dialogue between participants can be explored more authentically. Language was often key to the meaning of the posts, for example, using colloquial terms to demonstrate camaraderie.

The UWL drug study generally used verbatim quotes attributed to a pseudonym. The researcher felt that more harm would be caused by misinterpreting forum user's intentions than the possibility of the forum user's original post being found. However, the researcher did attempt to "Google-proof" the quotes. This involved searching for the original quote in the Google search engine and making slight modifications (without changing the meaning) to the quote until they are no longer searchable. Steps were taken to assess the level of personal information provided by a forum user; if forum posters had used what appeared to be a photograph of themselves or their real name as part of their user profiles, then they were excluded from the study (BPS, 2013). The mothers and YouTuber studies took similar steps to "Google proof" quotes and remove personally identifiable information due to the public nature of the online communities. In the fathers study, the forum was semi-private and not named in the research, so, with the exception of personally identifiable information, verbatim quotes were reported without modification.

In the previous sections, we have discussed the potential for online researchers to cause harm and the distinction between private and public spaces. We can see these issues and the use of verbatim quotes are linked intrinsically. If an online space can be deemed public, then consent may be considered less important but verbatim quotes can be searched for and the original post and the poster potentially identified. If the space is deemed private, then consent becomes far more important; however, verbatim quotes are less easily found. Researchers wishing to analyze posts from online communities must therefore consider all three of these issues simultaneously to ensure ethically sound decision making on a case-by-case basis.

Seeking Ethical Approval

Ethical approval was gained from a University Research Ethics committee for each of the studies and three of the studies gained approval from the same ethics committee. The committee in question expressed concern over the use of verbatim quotes in the UWL drug study. This prompted the researcher to reflect on the potential identification of forum users and how “Google proofing” quotes could overcome this. The researcher developed these reflections into a response to the committee which was then used as the protocol for the mothers study. In the case of the YouTubers study, the ethics committee chair referred the researcher to two academics on the committee with expertise in online research who assisted the researcher in amending the application and championed for the research at the committee. From our perspective, this began a collaborative culture between the ethics committee and online researchers at our institution with researcher experience helping to inform future ethical practice. The AoIR (Markham & Buchanan, 2012) similarly recognizes that online research involves tensions, which are often best resolved through a case-based approach as they arise.

Section Summary

This section considers three key ethical issues when using online data in research:

- The distinction between public and private space online is unclearly defined and researcher decisions must respect users’ perceptions of privacy.
- Researchers must consider the potential for harm which could be caused by identification of online users, for example, by tracing verbatim quotes back to their source. Steps such as “google-proofing,” removing identifiable information, and not naming source sites can help protect participants from harm.

Method in Action

Once the appropriate sites have been identified and appropriate ethical approvals granted, the researcher can often face a number of challenges when collecting data from these sites. First navigating the chosen forums including understanding the site structure and “netiquette” (good manners online) can be challenging. Second, the researcher must consider validity; how can the researcher guarantee those posting on online communities are who they purport to be?

Navigating Online Communities

While it is generally assumed that these methods are both cheaper and faster than conventional research recruitment (Freeman & Chapman, 2012), collecting and analyzing online data can be a time-consuming task for the researcher due to the volume of data available. Often posts pertaining to the research objective are buried deeply within threads, and forums are often so large that relevant data could easily be lost among thousands of other threads and posts. Robert Kozinets (2015) recommends researchers take time before the start of their study to familiarize themselves with the relevant online communities. Selected sites should have high traffic, large numbers of discreet posters, and high levels of between member interaction to generate descriptively rich data that are relevant and focused toward the research topic (Kozinets, 2015). The UWL drug study had to develop screening criteria to ensure the most relevant data were selected. This included using key search terms (generic and trade drug names as well as more general searches, for example, weight loss pills, diet pills), developing exclusion and inclusion criteria to best identify online forums which were popular and busy communities with a large proportion of female users, and being able to search the forum easily without a time period restriction.

Meaningful interpretation of data taken from online communities also requires the researcher to understand the netiquette and culture of each particular site. Researchers should engage with online communities to learn as much as possible about the sites and their members to add depth to the understanding of the groups they are studying (Kozinets, 2015). Some researchers advocate partnership with moderators and users to ensure trustworthy interpretation (Barratt & Lenton, 2010; Kozinets, 2002).

Online groups begin organically and grow to provide a network of support and information for the site users. On online discussion sites such as forums and Facebook groups, the moderator can play an important role, and factors such as changes in moderators, lack of moderator confidence in dealing with research requests, and paternalistic attitudes can all influence success in accessing and interacting with online groups. Granting access to researchers may be seen as an additional, or secondary role, and decisions may be inconsistent, or continually deferred. If requests by researchers increase, the site owners may give the responsibility of vetting to one of the staff, and they may even develop a policy and protocol for researchers to apply. However, for the majority of small- or medium-sized networks or online support groups, there may not be a member of staff who is experienced or confident in dealing with researcher requests. In addition, these requests may range from experienced researchers wishing to carry out funded studies to undergraduate students looking to research the site. It may seem easier for the site moderators to turn down the requests, especially if they are not clear how the online community would benefit or if they are worried that it might deter site users.

As a gatekeeper to forum users, moderators have a "*responsibility to best represent and protect a group's interests and ethos"* (Rattani & Johns, 2017, p. 27); however, they may make decisions based on the interests of their online community without consulting their members. This is illustrated in the previously discussed example from the mothers study, where the researcher was rejected by the administrator of a closed Facebook group. While the moderator's motivations were undoubtedly to protect the mothers, this attitude served to potentially restrict the voices of mothers who may have wished their views were captured in the research.

In the UWL drug study, there was a sense that the moderators were the experts. Occasionally, the researcher felt moderators had dismissive attitudes toward academic research as well as a feeling that researchers should be grateful to them for taking part *"I will only do it for a fee. My time is worth money."* For these close-knit groups, perhaps due to the illicit nature of the discussion, there was a strong sense of identity and anyone who falls outside of that identity can struggle to access and engage with their members. This was evident in the negative and dismissive responses toward anyone in the group who disagreed with the use of UWL drugs as well as negative responses toward researchers and academics. In particular, there were feelings of suspicion toward researchers on these forums, in some cases posters advised others to not take part in the research as they believed it was government or police surveillance. Some forum members or moderators had been involved in research previously and reported that they had not liked how they were represented. Similarly, in the fathers study, the site being researched expressed caution because a previous researcher, whom they had welcomed and supported, had published information about the site that site owners felt was unfair and inaccurate. Understandably, they were cautious when approached by another researcher. The site owners requested that they were able to review any information about them before it was published. In this instance, the researcher provided the site owners with copies of the written analysis but received no substantial changes.

Capturing Demographic Data Online

Online research lacks face-to-face validation and it is difficult to authenticate that the persons contributing to forums and social media discussions are really who they purport to be. It is easy for participants to present an “untrue” profile of themselves. However, collecting retrospective data from online sources also has advantages. Data are produced independent of the research agenda so are less likely to be influenced by researcher bias, and participants have space to reflect and validate their own data. The online and untraceable nature of online discussions could produce more honest and authentic data than might be discussed, in a semi-structured interview for example, particularly in relation to sensitive issues.

Online tools can be used to assist researchers in identifying the demographic characteristics of their population. In the UWL study, the online tool Alexa.com was used to provide a basic overview of the forum demographics, for example, age and gender. Similarly, the YouTuber study made use of the tool Socialblade.com to gain a better understanding of YouTubers’ viewing figures and audience. These tools can help ascertain which individuals are most likely to be using each particular online community; however, researchers making use of these commercial tools must also be aware of their limitations. Data sources are often unclear and there is no way of guaranteeing the accuracy of these tools.

For the fathers study, the issue of which contributors were fathers was important, because the study was looking to determine the appropriateness of paternal peer-to-peer support. Some usernames appeared to be derivations of users’ own names, or names that presented their relationship, such as HannahsDad. Additional indicators as to whether the site user was a father were looked for in the content of their posts “I am a father of two beautiful girls,” or in the footer of their posts (a short biography or signature, which appears beneath each post the user makes). Through these profiling methods, the researcher could ascertain that the vast majority of the identified contributors were declaring themselves as fathers. As all of this profiling was self-declared, the researcher had no way to confirm or check this, and there were no visual cues from the online contributor. However, this issue is not unique to online research; one can make parallels with respondents in face-to-face interviews or surveys. Although there may be visual cues in face-to-face interviews, such as gender and approximate age, the researcher is usually reliant on the declaration of the participant of their parenting status, for example. As in all research, it is important that the researcher acknowledges these validation issues and considers how this impacts the transferability and dependability of their data (Kozinets, 2015; Shenton, 2004).

In some cases, researchers may consider using member checking to help validate their online data. Member checking involves presenting some or all of the research findings for comment to the communities who have been studied (Lincoln & Guba, 1985), and in online research, this involves taking the research findings back to the online community studied (Kozinets, 2015). Both the UWL drug study and YouTuber study used Skype and instant messenger interviews with forum moderators and YouTubers, respectively, to confirm the researcher's interpretations. Bringing together data from two different methods can help increase the credibility and dependability of the research findings (Shenton, 2004).

Section Summary

- Each online community will have distinct characteristics which are governed by the site structure, unwritten rules of netiquette, culture of participation, and moderators. It is therefore important that researchers take time to familiarize themselves with each online community to ensure meaningful collection and interpretation of data.
- It can be challenging in online environments to verify that users are who they purport to be due to the increased opportunities for anonymity. Researchers should therefore consider online (e.g., online tools, user profiling) and offline (member checks) verification methods to improve their interpretations.

Practical Lessons Learned

To assist other researchers considering using online recruitment, we have consolidated our experiences into some practical guidance on successful engagement and recruitment.

1. **Take time to identify the most appropriate sites to recruit from**
    - Know your populations and have an understanding of the different online communities they use (e.g. online forums, social networking sites and blogs). Awareness of each of these platforms is important as different topics and groups will be found in different places.
    - Think about your populations' motivations for discussing your area of interest online and which types of site are generating the highest volume of descriptively rich data.
2. **Consider the ethical implications of your research**
    - Consider three key ethical issues before you collect data from online communities: Is this community public or private – how do users perceive this space? What potential harm could be caused by using these data? Should you seek

consent from users and gatekeepers and should you use verbatim quotes from this site?

- Allow time (1–3 months) to seek ethical approval and be open to discussing any tensions in your research with your University Ethics Committee.

3. **Know the netiquette of your online community**
    - Ensure that you understand the rules of your online community – both written and unwritten. Follow the guidelines set out by each community, ensure your posts are placed and worded appropriately and do not send too many reminders. We would advise as a general rule that researchers set a minimum of 2 weeks between reminders and send a maximum of three reminders.
    - Understand the priorities and beliefs of the online community you are researching. In some cases, it may be that your research is not considered “right” for the online community in question, and in these cases, it is important to take no for an answer and withdraw.
4. **Engage with your moderators**
    - In the case of forums and Facebook groups, moderators or administrators often act as a gatekeeper and it is important to engage with these key individuals at an early stage.
    - Moderators may share the prevailing beliefs of their groups, have their own priorities or make decisions without consulting their members. Ensure that moderators and members understand the purpose and benefits of the study for their community.
5. **Consider how you can validate demographic characteristics**
    - Online communities offer individuals extended opportunities to post anonymously, use alternative online identities, and falsify or withhold demographic details. It is important to consider what impact this lack of reliable information could have on your research outcomes.
    - Consider options for validating demographic data online. Options include using online profiling to build up a picture of participant characteristics, using online tools to understand the overall demographic makeup of your research site, or using in person member checks with online community members to confirm your interpretations.

Conclusion

In this case study, we have used our experiences as doctoral researchers using online methods for the first time to highlight some of the ethical limitations and difficulties of online research. Our research has led us to reflect on numerous ethical issues including participants' perceptions of what is public, preserving anonymity, and protecting from harm. Online research is valuable not only in gathering data to complement conventional approaches but also in taking new approaches to data that would otherwise be difficult to obtain. However, what our experiences also highlight is that the reason for using online methods should not be simply because they are easy. Methodological innovation should be balanced with ethically sensitive responses (Barbovschi, Green, & Vandoninck, 2013; Bengry-Howell, Wiles, Nind, & Crow, 2011; Nind, Wiles, Bengry-Howell, & Crow, 2012). Although ethical guidance for online research exists (BPS, 2013; Markham & Buchanan, 2012), often only cursory ethical considerations are discussed in published research. Published studies using online methods rarely report ethical procedure in detail. Questioning the ethical ramifications of our work has caused us all to encounter periods of doubt throughout the research process. However, considering these issues has also been a rewarding process as it has given us opportunity to scrutinize our methods and in some cases begin to lay the groundwork for ethical guidelines in our future research.

The "laboratory" as described by Rebecca Eynon and colleagues (2008) is increasing with online research methods producing several good-quality and innovative research studies. As four research students, we are linked by our genuine fascination with online methods. We caution against innovation for innovation's sake, but we are enthused by the many opportunities that the Internet offers health researchers and others in social science.

Classroom Discussion Questions

1. Suppose you want to collect data from an online forum discussing illicit drug use: Do you need to gain consent? If yes, who from? What are the issues of gaining or not gaining consent?
2. Imagine you want to access posts for your research from a closed Facebook group which you are not a member of. What steps would you take before you access the data?
3. Discuss what potential harm you could cause by publishing information gathered from an online forum where members are discussing sensitive health behaviors? How would you maintain anonymity of forum members?

Multiple Choice Quiz Questions

1. Which populations can online research methods be considered suitable for?

    a. Hidden or seldom heard populations
    b. Highly visible online communities
    c. Both A&B

Correct answer: Both A & B

2. Which of the following ideas is not an online research method ethical stance?
    a. Protecting participants from harm
    b. Anything goes
    c. Use of paraphrased quotations to reduce risk of participant identification

Correct answer:B

3. Why might ethical deviations occur in online research?
    a. Because researchers do not follow an ethical code of conduct
    b. Because researchers are inconsiderate and careless
    c. Because online research is diverse and ethical guidelines vary

Correct answer: C

4. Who might be a suitable gatekeeper for online research methods?
    a. The forum moderator or administrator
    b. The researcher
    c. The Ethics Committee which grants approval

Correct answer: A

5. Which of the following is a way that online demographics can be validated?
    a. Member checks
    b. Verbatim Quotes
    c. Using Public Forums

Correct answer: A

6. Which steps are useful in identifying appropriate communities for inclusion in research?
    a. Developing key search terms to search online
    b. Having inclusion and exclusion criteria
    c. Both A&B

Correct answer: Both A & B

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Further Reading

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (Version 2.0)*. Chicago, IL: Association of Internet Researchers. Retrieved from https://aoir.org/reports/ethics2.pdf

British Psychological Society. (2017). *Ethics guidelines for Internet-mediated research* (INF206/04.2017). Leicester, UK: Author. Retrieved from https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research-2017

Kozinets, R. (2015). *Netnography: Redefined* (2nd Ed.). Thousand Oaks, CA: SAGE.

Web Resources

Alexa: https://www.alexa.com/

Social Blade: https://www. .com

References

“About us.” (2018). *Mumsnet*. Retrieved from https://www.mumsnet.com/info/about-us

Arterburn, D. E., Crane, P. K., & Veenstra, D. L. (2004). The efficacy and safety of sibutramine for weight loss: A systematic review. *Archives of Internal Medicine, 164,* 994–1003. doi:10.1001/archinte.164.9.994.

Ayling, R., & Mewse, A. J. (2009). Evaluating Internet interviews with gay men. *Qualitative Health Research, 19,* 566–576.

Barbovschi, M., Green, L., & Vandoninck, S. (2013). *Innovative approaches for investigating how children understand risk in new media. Dealing with methodological and ethical challenges*. Retrieved from http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20III/Reports/InnovativeMethods.pdf

Barratt, M. J. (2011, June 29–July 2). *Discussing illicit drugs in public Internet forums: Visibility, stigma, and pseudonymity.* Paper presented at the 5th International Conference on Communities and Technologies, Brisbane, Queensland, Australia.

Barratt, M. J., & Lenton, S. (2010). Beyond recruitment? Participatory online research with people who use drugs. *International Journal of Internet Research Ethics, 3,* 69–86.

Bengry-Howell, A., Wiles, R., Nind, M., & Crow, G. (2011). *A review of the academic impact of three methodological innovations: Netnography, child-led research and creative research methods*. Retrieved from http://eprints.ncrm.ac.uk/1844/1/Review_of_methodological_innovations.pdf

Beninger, K. (2017). Social media user’s views on the ethics of social media research. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 57–73). Thousand Oaks, CA: SAGE.

British Psychological Society. (2013). *Ethics guidelines for Internet-mediated research*. Leicester, UK: Author.

Eynon, R., Fry, J., & Schroeder, R. (2008). The ethics of Internet research. In N. Fielding (Ed.), *The SAGE handbook of online research methods* (pp. 22–41). London, England: SAGE.

Freeman, B., & Chapman, S. (2012). Measuring interactivity on tobacco control websites. *Journal of Health Communication, 17,* 857–865. doi:10.1080/10810730.2011.650827

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. London, England: SAGE.

Hallett, R. E., & Barber, K. (2014). Ethnographic research in a cyber era. *Journal of Contemporary Ethnography, 43*, 306–330.

James, W. P. T., Caterson, I. D., Coutinho, W., Finer, N., Van Gaal, L. F., Maggioni, A. P., . . . Renz, C. L. (2010). Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects. *New England Journal of Medicine*, *363,* 905–917. doi:10.1056/NEJMoa1003114.

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research, 39*, 61–72.

Kozinets, R. V. (2015). *Netnography: Redefined*. London, England: SAGE.

Langer, R., & Beckman, S. C. (2005). Sensitive research topics: Netnography revisited. *Qualitative Market Research: An International Journal, 8*, 189–203. doi:10.1108/13522750510592454.

Lee, H. C. H., Law, C. Y., Chen, M. L., Lam, Y. H., Chan, A. Y. W., & Mak, T. W. L. (2014). 2,4-dinitrophenol: A threat to Chinese body-conscious groups. *Journal of the Chinese Medical Association, 77*, 443–445. doi:10.1016/j.jcma.2014.05.003.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: SAGE.

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (Version 2.0)*. Chicago, IL: Association of Internet Researchers. Retrieved from https://aoir.org/reports/ethics2.pdf

McDermott, E., & Roen, K. (2012). Youth on the virtual edge: Researching marginalized sexualities and genders online. *Qualitative Health Research, 22,* 560–570. doi:10.1177/1049732311425052

McFee, R. B., Caraccio, T. R., McGuigan, M. A., Reynolds, S. A., & Bellanger, P. (2004). Dying to be thin: A dinitrophenol related fatality. *Veterinary and Human Toxicology, 46*, 251–254.

McVeigh, J., Germain, J., & Van Hout, M. C. (2016). 2,4-Dinitrophenol, the inferno drug: A netnographic study of user experiences in the quest for leanness. *Journal of Substance Use*. 22(2), 131-138. https://doi.org/10.3109/14659891.2016.1149238

"Netmums in numbers." (2017). *Netmums*. Retrieved from https://advertisements.netmums.com/

Nind, M., Wiles, R., Bengry-Howell, A., & Crow, G. (2012). Methodological innovation and research ethics: Forces in tension or forces in harmony? *Qualitative Research, 13,* 650–667.

Petróczi, A., Ocampo, J. A. V., Shah, I., James, R. A., Naughton, D. P., Jenkinson, C., . . . Taylor, G. (2015). Russian roulette with unlicensed fat-burner drug 2,4-dinitrophenol (DNP): Evidence from a multidisciplinary study of the Internet, bodybuilding supplements and DNP users. *Substance Abuse: Treatment, Prevention, and Policy, 10*(1), 39.

Rattani, A., & Johns, A. (2017). Collaborative partnerships and gatekeepers in online research recruitment. *The American Journal of Bioethics, 17,* 27–29.

Reilly, P. (2014). *The "Battle of Stokes Croft" on YouTube: The development of an ethical stance for the study of online comments* (SAGE Research Methods Cases). London, England: SAGE.

Roberts, L. D. (2015). Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology, 12,* 314–325. doi:10.1080/14780887.2015.1008909.

Sam, A. H., Salem, V., & Ghatei, M. A. (2011). Rimonabant: From RIO to Ban. *Journal of Obesity, 2011,* 432607. doi:10.1155/2011/432607.

Scheen, A. J. (2011). Sibutramine on cardiovascular outcome. *Diabetes Care, 34,* S114–S119. doi:10.2337/dc11-s205

Seko, Y., Kidd, S. A., Wiljer, D., & McKenzie, K. J. (2015). On the creative edge: Exploring motivations for creating non-suicidal self-injury content online. *Qualitative Health Research, 25,* 1334–1346.

Sharkey, S., Jones, R., Smithson, J., Hewis, E., Emmens, T., Ford, T., & Owens, C. (2011). Ethical practice in Internet research involving vulnerable people: Lessons from a self-harm discussion forum study (SharpTalk). *Journal of Medical Ethics, 37,* 752–758.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*, 63–75.

Suzuki, L. K. & Calzo, J. P. (2004). The search for peer advice in cyberspace: An examination of online teen bulletin boards about health and sexuality. *Journal of Applied Developmental Psychology, 25*, 685–698. doi:10.1016/j.appdev.2004.09.002

Svenningson-Elm, M. (2009). How do various notions of privacy influence decisions in qualitative Internet research? In A. B. Markham (Ed.), *Internet inquiry: Conversations about method* (pp. 101–113). Thousand Oaks, CA: SAGE.

Trevisan, F. & Reilly, P. (2014). Ethical dilemmas in researching sensitive issues online: Lessons from the study of British disability dissent networks. *Information, Communication & Society, 17*, 1131–1146.

VanDam, C., Kanthawala, S., Pratt, W., Chai, J., & Huh, J. (2017). Detecting clinically related content in online patient posts. *Journal of Biomedical Informatics, 75*, 96–106.

Zhang, S., O’Carroll Bantum, E., Owen, J., Bakken, S., & Elhadad, N. (2016). Online cancer communities as informatics intervention for social support: Conceptualization, characterization, and impact. *Journal of the American Medical Informatics Association, 24*, 451–459.