Title: Why should we use online research methods? Four doctoral health student perspectives

Short title: Why should we use online research methods?

Authors and affiliations:

Jennifer Germain¹
Jane Harris¹
Sean Mackay²
Clare Maxwell²

1. Public Health Institute, Liverpool John Moores University
2. School of Nursing and Allied Health, Liverpool John Moores University

Acknowledgments:

Thanks to our respective Directors of studies and Dr Dave Harriss for their useful ethics conversations.
Abstract
The growth of the internet has led to an increase in researchers utilizing online methods. Online communities such as forums, blogs and video platforms are particularly useful for research involving populations that are internet savvy, seldom heard or discussing sensitive or illicit behavior. Drawing upon the experiences of four doctoral health students who are using online methods, this article discusses the value and benefits of conducting online research as well as the limitations and difficulties encountered. Consideration is given to the methodological and ethical implications of online research. Our own research leads us to reflect on: participants’ perceptions of what is public, preserving anonymity and protecting participants from harm.

Introduction
The internet forms part of our everyday lives. Estimates indicate that there are 3.5 billion internet users worldwide and approximately 2.34 billion of these users have a social media account (Statista, 2017a). Given the recent surge in research examining online communities and environments, it is evident that researchers are becoming increasingly aware of the importance of the internet in our day-to-day lives. Eynon, Fry, and Schroeder (2008) describe the internet as the ‘laboratory for the social sciences’ and it can be argued that people’s online habitats are just as important as other environments where human interaction occurs (Hallett & Barber, 2014). Studies which adapt traditional qualitative methods to online groups, such as online interviewing, are well defined in the published literature (Ayling & Mewse, 2009; McDermott & Roen, 2012; Seko, Kidd, Wiljer, & McKenzie, 2015). However, online methods such as advertising your research study online, actively seeking out specific groups or analyzing and engaging with online communities are less well described. These areas of research are the focus of this article. We, the authors of this article, are all PhD students in health carrying out qualitative or mixed methods research, and are making use of online methods within our research. This includes recruiting from online
communities and analyzing data from those groups through online forums, blogs and video platforms. We all come from different research backgrounds and have different research interests and expertise. Our PhD’s explore the role of professional ‘Youtubers’ in delivering health messages, the use of unlicensed weight loss (UWL) drugs in females, mothers who are experiencing bottle refusal by breastfed babies and peer support for new fathers. Whilst our research areas are different we are united in our choice to use online methods in our research. We are also united in the level of consideration we gave to choosing these methods. We are researching groups of people who are: internet savvy and actively engaged online, who are discussing sensitive or illicit behavior where disclosure may be judged in ‘real life’ or, who are hidden or seldom heard outside of the online communities they inhabit. Despite our commitment to using online methods, we have each still encountered periods of doubt when we have questioned the ethical ramifications of using online research as a method. This commentary, drawing upon our own experiences, discusses our use of online methods, the value and benefits of using them and finally questions whether, from an ethical perspective, it is appropriate to use online research methods at all.

**Target Practice - Which populations are online research best for?**

Online methods can be used in research with a whole range of different populations. As four public health research students, we have online methods in common but the populations we are studying are very different. By comparing four different studies, we have found common ground, which has been used to structure this article, however, we have also reflected on the differences pertinent to each study. Some populations are well established and known online: as is the case for mothers seeking parenting and wider support. Some are even defined by common characteristics that exist solely online: such as the communities (or fandoms) which exist around professional YouTubers. For other populations, online research is useful because these groups are hard to reach offline either because they are engaging in risky or illicit activities, such as the case of unlicensed weight-loss (UWL) drug users, or because they are
discussing topics they may not feel comfortable sharing elsewhere, as in the case of new fathers. These four populations, which are the focus of our individual research studies, will be used in the remainder of this section to highlight the benefits of online research methods when recruiting both: established online and hard to reach populations in health research. For ease of reading, our studies will be referred to as Youtubers, UWL drugs, Mothers and Fathers (table 1).

<table>
<thead>
<tr>
<th>Study code</th>
<th>Study title</th>
<th>Type of discussion</th>
<th>Population</th>
<th>Platforms accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWL drug</td>
<td>An exploration of the female use of unlicensed weight loss drugs.</td>
<td>Illicit/sensitive Drug use</td>
<td>Females: Undefined group</td>
<td>Online forums</td>
</tr>
<tr>
<td>YouTubers</td>
<td>Examining the role that professional YouTubers play in young people’s health behaviors and identity in the UK</td>
<td>Non-sensitive Health messages</td>
<td>Young people living in UK 13-18 years</td>
<td>YouTube</td>
</tr>
<tr>
<td>Mothers</td>
<td>UK Mothers’ experiences of bottle refusal by their breastfed baby</td>
<td>Sensitive and non-sensitive advice and support</td>
<td>Mothers who are experiencing/have experienced bottle refusal by their breastfed baby</td>
<td>Online forums, Facebook</td>
</tr>
</tbody>
</table>
Table 1: Summary of the four PhD topics

<table>
<thead>
<tr>
<th>Fathers</th>
<th>Online peer support for fathers</th>
<th>Sensitive and non-sensitive advice and support</th>
<th>Fathers of pre-school children</th>
<th>Online forum</th>
</tr>
</thead>
</table>

The world at our fingertips - The internet as a global network

Perhaps one of the most visible online communities is that which has grown up around parenting particularly in relation to motherhood. Online pregnancy and parenting forums/discussion groups provide huge potential to the online researcher with the two largest forums: Mumsnet and Netmums having over 17 million unique users and 130 million page views per month. For the researcher looking to study parenting from a worldwide perspective, Babycentre describes itself as reaching 1 in 5 mothers globally (Babycentre, n.d.). Recent statistics regarding Facebook depict its user profile as mostly female with the largest age group being 20-39 years – providing an almost ‘age perfect’ sample of women of reproductive age (Statista, 2017b). Based on these statistics, one could argue that discussions surrounding sample bias in relation to using online participants are negated due to the vast numbers of those who consult online sources (Bridges, 2016). In relation to UK mothers, it could be postulated that those who are not actively consulting online communities are becoming the minority. It is evident that mothers use online forums and social media to discuss scenarios and topics that are against official health guidance – possibly believing such platforms to be non-threatening and less judgmental. This gives the researcher obvious benefits of being able to study online content on subjects that would not necessarily be openly discussed. This should be tempered, however, with the amount of forum moderation that takes place. Comments and chat can also often originate from negative personal experiences and the authenticity of participants cannot be regulated.
The mothers study used an online questionnaire as one of the stages of data collection. A URL to an online questionnaire was shared on Facebook and in relevant threads on parenting forums, attracting 841 respondents in two weeks. Recruiting online allows the researcher to reach large numbers of would-be participants in a short period. This should however be balanced with a potential lack of control regarding dispersal which could result in a study going ‘viral’ (Ellis-Barton, 2016). Shortening of URL’s (e.g. using goo.gl or tiny.url) can provide useful analytical data such as where participants accessed the online survey, e.g. Twitter, Facebook, and their broad geographical location. This gives useful data as to how far the questionnaire has travelled online, although it does not take into account snowballing – a recruitment method that was commonly used in the mothers study.

Online discussion groups and forums are among the longest established and researched online social platforms; providing the foundations for well-known online research methods such as netnography (Kozinets, 2002). However, the past decade has seen the development of numerous other social media platforms providing a rich source of data for researchers. YouTube is one such global social network with over 1 billion users uploading hundreds of millions of hours of video daily (YouTube, 2014). Young people are the biggest users of these social media platforms; in a recent survey, just over a quarter of 12-15 year olds said they would turn to YouTube for accurate and true online information (Ofcom, 2014). The way in which young people engage with social media has allowed the rise of the YouTube celebrity or YouTuber. There are approximately 150 YouTubers in the UK with over 1 million subscribers. YouTubers could potentially be an important source of health information for young people. YouTubers present a magnified version of young people’s own networked lives, with the same issues of misinterpretation and surveillance which may be more relatable than the experiences of other adults providing health and wellbeing advice (Beer, 2008; Boyd, 2008; Uhls & Greenfield, 2012).
The YouTube study also used an online survey as one stage of data collection. YouTuber fan communities have developed largely online. Their demographic characteristics are therefore undefined and they are thus difficult to recruit offline. A URL to the online survey was shared on Twitter, Tumblr and Instagram using hashtags related to the top 50 UK YouTubers. YouTubers themselves were also invited to promote the survey through their social media accounts.

While the two populations above differ in terms of their health priorities and behaviors, these two online communities are linked by their visibility and scale; with both providing a potential gateway between the researcher and a network of millions. However, converse to this, some communities are easier to recruit online because of their lack of visibility.

*Exploring the unknown - Invisible populations online*

UWL drug users are one such undefined population. Very little is known about the types of people who use them and the motivations for use. Whilst there is some literature on the use of such drugs, this has predominately been within male dominated bodybuilding settings (Petróczi et al., 2015). Fitness competitions, modelling companies, gyms and beauty salons were considered as possible locations to recruit UWL drug users at the planning stage. However, there was no evidence that UWL drug users actually use or engage with any of these locations. This group is known to be active online: UWL drugs are bought online and their use is discussed on online forums. People are more likely to use online forums to discuss illicit or sensitive behavior due to the potential anonymity they offer (Elliott, Langer, & Beckman, 2005) although this is influenced by perceptions of the visibility of the forum and the risk of social stigma (Barratt, 2011). Comments in the UWL drug study such as “I haven’t told any of my family or friends that I am taking reductil” illustrate that forum users seemingly prefer to discuss their drug use on forums rather than with people in their ‘real life’.
Likewise, for fathers needing support in early and pre-school parenting, the usual channels of advice are largely aimed at mothers, and frequently the support and information is provided from an authoritative source, such as health professionals or official health service websites. Face to face services may be difficult for fathers to access, because they have returned to work, for example, or because they feel that they stand out in the female dominated environments such as health clinics. Online support groups are available for many health conditions, and importantly offer peer support in addition to, or rather than, official information. Whilst there are the very popular parenting websites like Mumsnet, there is no father-specific space on these sites, and men may feel uncomfortable searching and posting on them. One UK website provides peer-to-peer discussion boards, and the fathers study looks at the experiences of fathers expressed on these threads, and the characteristics of the communication in the posts. These two populations are united by the fact that they are hard to reach offline. For UWL drug users the reason for this lack of visibility is clear; online forums provide and anonymous venue for those engaging in illicit activity to seek advice and share experiences that they cannot share offline. In the case of fathers, the reasons are more nuanced, reflecting societal expectations of masculinity and help seeking behavior. However, online forums are appealing to this group because they provide fathers with the opportunity to seek advice with anonymity in a manner that they may struggle to do offline. The populations discussed here present two extremes: widely visible social networks and populations that are largely invisible. By grouping our study populations in this way, we are not suggesting that online research is only suitable in these extreme cases. Our intention here is simply to highlight the breadth of potential for online methods. The next section considers some key ethical and methodological issues in using online methods.

**Ethical considerations in online research**
Online research is complex. Both guidance and examples of methodological frameworks are limited. The purpose of this section is to give an overview of some of the key concerns from the authors’ perspectives. This section aims to explore some of these issues with a particular focus on honesty and consent, quotations and the potential to cause harm.

_Crossing boundaries? - The public/private debate_

The British Psychological Society (BPS) general ethics code describes a public space as situations where individuals “would expect to be observed by strangers”. In the offline world this distinction is quite clear cut but the idea of a public space is more difficult to define online. Firstly, because internet communication is often conducted in public (e.g. open discussion forum) and private (e.g. home) spaces simultaneously and secondly, because it is not always easy to determine which spaces individuals consider to be “public” or “private”. The concept of “public/private” in online research is not a binary one and so must consider users expectations guided by both consensus and contextual integrity (Markham & Buchanan, 2012). Where there is any level of ambiguity over data being in the public domain, the BPS (Hewson et al., 2013) recommend that researchers consider the extent to which undisclosed observation may have potentially damaging effects for participants.

Three of the studies (UWL drugs, mothers and fathers) include analyzing posts from online forums and one study (YouTubers) involves analyzing YouTube videos and comments. One important ethical issue in this debate is the notion of honesty: whether researchers inform online users that they are using their data and also whether consent is sought. The guidelines around this issue are open to interpretation with the BPS discussing both issues of honesty and consent within the context of public versus private spaces (Hewson et al., 2013). Essentially, if the space can be classified as private then consent should be sought. The blurring of boundaries between private and public space online and the lack of clarity in what constitutes harm have led to differences in how online research is carried out. There is no clear guide to
measuring potential harm. Our decisions in how to deal with this matter within our own research have differed. The UWL drugs and fathers studies both involve the analysis of online forums. Whilst neither researcher obtained consent from their participants (the forum members), consent was sought in the father study from the website owners. For these studies, the decision not to inform forum members were mostly for practical reasons; the posts were accessed retrospectively so they may be several years old, and the contributors may no longer engage with the site. The level of intrusion was minimal, accessing online posts can be considered less intrusive than face-to-face or even survey-based research as there is no interaction between the participants and the researcher. The studies involved large volumes of data and therefore it would have been extremely demanding to contact all forum members who had contributed to each thread. Additionally, 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 guidelines (Hewson et al., 2013) suggest permission from moderators is necessary only if the forum or online source will be named. Whilst the fathers study did gain consent from forum leaders, the UWL drug study did not. In the case of the UWL drug study, this was not a deliberate attempt to deceive. The forums were not named and interviews with moderators of the forum were planned for a subsequent stage. However, when moderators were involved in the research, responses ranged from disinterested to negative; in some cases the researcher was automatically banned from the forum. The reluctance of forum moderators to participate in later stages of the PhD caused 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?

The UWL drug and father studies discuss the issues around honesty and consent in online forum analysis. However, these issues are pertinent to all social media platforms. In the YouTube study, consent will not be sought to transcribe videos as they are in the public domain (with no login required to view videos).
and data collection will be largely retrospective. However, like the fathers study, the mothers study will be seeking to gain forum moderator consent to access closed Facebook groups as posts are not usually visible to the general public. Posts could be accessed if the researcher became a member of the closed Facebook groups; however, it was felt that seeking consent would be the preferred approach.

To quote or not to quote – that is the question.

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, because these forums do not need a password to access them, it is possible to locate the original post by putting the quote into a search engine. Therefore, it is necessary to decide the probability of harm to the poster if their quote was traced to the original post. This depends on the level of personal information and the content of the post.

The UWL drug and fathers studies used verbatim quotes. In both cases, it was 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. Analyzing online forums is very different to analyzing a spoken interview. Tone of voice, and nonverbal cues are missing, making it difficult to ascertain peoples intentions, for example whether they are serious or joking. Using anonymous verbatim quotes keeps the participants 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. Part of the mothers study is to focus on online advice giving and support between mothers whose breastfed baby
refuses a bottle. To use paraphrased quotes would potentially limit the richness of these conversations and miss the subtle nuances that occur during their online chat.

In these studies, it was decided to quote directly, since the language was the key to the meaning of the posts, for example using colloquial terms to demonstrate camaraderie. One potential solution to make quotes more “Google-proof” was to cut the length of the verbatim quotes used. However, there was a desire to use the most evocative quotes to illustrate the results, and in some cases, a two-word phrase allowed Google to identify the source. Since the forum is open and not password protected, the quotes were presented verbatim, attributed to a pseudonym. Other options are to assess the level of personal information provided by a forum user. In the UWL study, if forum posters had used a photograph of themselves or their real name with their user profiles then they were excluded from the study.

How do we protect from harm?

In the previous sections, we have already touched upon the potential we, as online researchers, have to cause harm and the distinction between private and public spaces. These issues are linked intrinsically. If the online space can be deemed public, then consent may be considered less important. Verbatim quotes can be searched for and the original post and the poster identified. If the space is deemed private, then consent becomes far more important, however, verbatim quotes are less easily found.

To minimize harm, 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; Stevens, Donnell, & Williams, 2015), and whether the researcher is eavesdropping on a private conversation in a public space. Whilst discussion boards that require no registration or passwords to read their content may be viewed as in the public domain, the researcher should consider whether the contributors to the board did or could have an expectation of privacy. For the forum used in the fathers study, 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 definition of a public space (Hewson et al., 2013). The site’s guidance for new members reinforces this by advising against posting of personally identifying information. People can also register to post messages, or they can post anonymously. In the mothers’ study, follow up interviews highlighted a small number of closed Facebook groups that they 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. Taking this into consideration, it was decided to have an open dialogue with forum moderators regarding how best to capture data from the groups without crossing possible boundaries set by them.

In the UWL drug study, the decision was made to only use open forums, this suggested that forums 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 is faced by the Youtube 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).

**Whose digital footprints are these? - 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. This could be true of all research where participants may feel pressure to respond in a certain way. Collecting retrospective data from online sources has advantages because data are produced independent of the research agenda so are less likely to be influenced by researcher bias. Additionally, participants have space to reflect and validate their own
data. It can provide an outlet for a more honest discussion of sensitive issues thus providing data that is potentially more authentic due to the very nature of it being online and untraceable.

Three of the studies (UWL, YouTubers and Mothers) combined online data collection with face to face member checks and triangulated the findings. In contrast data for the fathers study was collected solely online therefore the issue of which contributors were fathers was important, since the study was looking at help seeking from a masculine perspective, and it was seeking to determine the appropriateness of peer-to-peer support for fathers. Some usernames appeared to be derivations of their own names, or names that presented their relationship, such as HannahsDad. Additional indicators to whether the site user was a father was looked for in the content of their posts “I am a father of two beautiful girls”, or in the footer of their posts (which can be set on registration). There were also anonymous posts, and these became a separate group. All of this profiling was therefore self-declared; the researcher has no way to confirm or check this, and there were no visual cues from the online contributor. However, one can make parallels with respondents in face-to-face interviews or surveys. Whilst 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. While the fathers study was qualitative, it was possible to demonstrate that the vast majority of the individually identified contributors were declaring themselves as fathers.

Unlike the fathers study, the UWL drug study did not examine the profile of forum users. Steps were taken to include the most relevant forums and posts in the study. Online tools, such as Alexa.com do exist which provide a basic overview of the demographics, e.g. age, gender. These tools can help ascertain which individuals are most likely to be using a particular forum, however there is no way of guaranteeing the accuracy of these tools and often data sources are unclear. For the UWL drug study, the researcher was
only interested in the views of female drug users so employed these tools to seek forums with a high proportion of female users.

**Your request has been approved: Seeking ethical approval**

Ethical approval was sought for each study from Liverpool John Moores University Research Ethics Committee and The University of Lancaster Faculty of Health and Medicine Research Ethics Committee. Three of the studies sought 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 give greater consideration to the potential for the identification of forum users through Google searching and “Google-proofed” quotes in order to overcome this. The subsequent response to the committee was then used as the protocol for the later two studies (YouTubers and Mothers). 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 Association of Internet Research (Markham & Buchanan, 2012) similarly recognizes that online research involves tensions which are often best resolved through a case based approach as they arise.

**Conclusion**

In this commentary, we have used our experiences as doctoral researchers using online methods for the first time to highlight some of the limitations and difficulties of online research. We are all supervised by experienced and innovative qualitative researchers but it has been our fresh perspective as novice researchers that has led us to ask the most interesting questions and challenge the conventions of online research. Our research has led us to reflect on numerous issues including participants’ perceptions of what is public, preserving anonymity and ultimately protecting from harm.

From our experiences, we believe online research methods provide a useful approach. Online research is valuable not only 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; Nind, Wiles, Bengry-Howell, & Crow, 2012). While ethical guidance for online research exists, (Hewson et al., 2013; Markham & Buchanan, 2012) often only cursory ethical considerations are discussed in published research. Published studies using online methods very 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 Eynon et al. (2008) is increasing with online research methods producing a number of 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.
Reference


