

# Blogs as Elusive Ethnographic Texts: Methodological and Ethical Challenges in Qualitative Online Research

*International Journal of Qualitative Methods*  
Volume 16: 1–12  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1609406917705796  
journals.sagepub.com/home/ijq  


Liza C. Kurtz<sup>1</sup>, Sarah Trainer<sup>2</sup>, Melissa Beresford<sup>1</sup>, Amber Wutich<sup>1</sup>, and Alexandra Brewis<sup>1</sup>

## Abstract

Burgeoning online environments offer completely new opportunities for ethnographic and other forms of qualitative research. Yet there are no clear standards for how we study online texts from an ethnographic perspective. In this article, we identify barriers to the application of traditional qualitative methods online, using the example of a systematic thematic analysis of weight-loss blogs. These barriers include the influence of the technology structuring online content, the fluid nature of online texts such as blogs, and the highly connected and public nature of online identities, which may span multiple social media platforms. We discuss some potential approaches to addressing these challenges as preliminary steps toward developing a tool kit suited to ethical, high-quality online modes of ethnographic research.

## Keywords

online, ethnography, text, text analysis, qualitative analysis, ethics, blogs, weight loss, obesity

## What is already known?

Online texts are exponentially increasing in real time, providing new ways and spaces to conduct ethnographic research. The relatively small body of ethnographic literature focused on online spaces that have emerged to date has tended to employ highly specific and nonsystematic sampling techniques to identify and study sites and users of interest. Research into blogging, in particular, has focused on nonrandom, small-scale, longitudinal studies of bloggers across time and has not yet reached out to incorporate larger, comparative, cross-sectional analyses.

## What this paper adds?

Online research is an area of increasing interest to qualitative social science researchers, but is still underexplored, especially given its importance in everyday life for billions of people. One possible reason for the underdevelopment of online research to date is an absence of clear methodological protocols upon which to rely when beginning research. Put more simply, ethnographic researchers, with a focus on in-depth knowledge of people and data, often feel overwhelmed by the sheer quantity of online resources and therefore choose to focus on very small online samples. In this article, we expand research into online

qualitative methods by discussing methods for systematically sampling online blogs, with an emphasis on the technological barriers encountered during initial investigations and potential solutions to these barriers.

## Methodological and Ethical Challenges in Blog-Based Text Analysis

The use of ethnographic techniques to explore online environments is a method of increasing interest to social science researchers, largely because of the increasing importance such spaces play in everyday life worldwide (e.g., Barratt & Maddox, 2016; Bonilla & Rosa, 2015; Bortree, 2005; Gehl, 2016; Graffigna & Bosio, 2006; Hookway, 2008; Horst & Miller, 2013; Huffaker & Calvert, 2005; Karlsson, 2007; Kaun, 2010; Lopez, 2009; McCullagh, 2008; Olive, 2013; Pink et al.,

<sup>1</sup> School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

<sup>2</sup> The Obesity Solutions Initiative, Arizona State University, Tempe, AZ, USA

## Corresponding Author:

Liza C. Kurtz, School of Human Evolution and Social Change, Arizona State University, PO Box 872402, Tempe, AZ 85287, USA.

Email: eckurtz@asu.edu



2016; Pitts, 2004; Postill & Pink, 2012; Qian & Scott, 2007; Sade-Beck, 2004; Steinmetz, 2012; Wilson, Kenny, & Dickson-Swift, 2015). Online environments are used to produce meaningful and complex interactions of all different kinds including providing places for people to connect with others for support and to share thoughts, ideas, and stories. In doing so, online environments provide new spaces in which people are able to perform diverse physical, emotional, and social identities (Boellstorff, 2012; de Laat, 2008; Dumova & Fiordo, 2012; Karlsson, 2007; Mautner, 2005; O'Brien & Clark, 2012; Pitts, 2004; Reed, 2005; Siles, 2011; Wilkinson & Thelwall, 2011). Virtual texts, like discussion boards or blogs, proliferate and would appear to be a rich source of material for qualitative analysis. The peculiarities associated with online interactions and presentations of self, however, mean traditional ethnography—with its reliance on in-person participant observation and interviewing—needs to be rethought. At the same time, however, we argue that ethnography, in particular its emphasis on exploring individuals' discourse and actions within the context of their own chosen milieus via detailed observations and nuanced analysis, has huge potential to illuminate online interactions and identity making. Here we focus on ethnographic approaches to online data collection and analysis.

In this article, we address three main research questions. These are (1) What techniques can be used for systematic sampling in qualitative online research? (2) What technological barriers exist to implementing qualitative systematic sampling and how can they be overcome? and (3) What ethical dilemmas arise in qualitative online research and what strategies can researchers use to deal with them?

To explore these questions, we offer a case study of our own ethnographic study of U.S.-based online weight-loss weblogs ("blogs"). We also outline potential guidelines for ethnographic researchers interested in online research. This is particularly important given that ethnographic researchers, who focus on in-depth analysis of people and culture, often feel overwhelmed by the sheer quantity of online resources, spaces, and users and therefore don't often pursue collection and analysis of larger cross-sectional samples of online texts. We pay particular attention to the problems that arise when designing systematic sampling methods in an online environment formed in part by the technological constraints and economic incentives of blog hosting services and search engines. We also reflect on the ethical implications of deploying qualitative data analysis using online sources that are not intended for either private, diary-like usage nor directed at a specific public audience (McCullagh, 2008; Siles, 2011; Wilson et al., 2015), and we discuss potential ethical problems stemming from researchers' identification of online identities across multiple online platforms including social media.

We suggest that an ethnographic approach to web-based research is generally successful in documenting and analyzing data extracted from blogs, while taking account of the different types of cultural contexts in which such data are constructed, but that an ethnographic approach on its own does not account for many of the other underlying factors structuring data that

are specific to online environments. In the case of our analysis of blogs oriented around weight and weight loss, these factors shaped both the ways blogs were selected for inclusion, and the content of such blogs, which can be more or less targeted at meeting criteria to increase visibility on the web based on the bloggers' familiarity with search engine optimization (SEO) and other technologies. We suggest the need for greater research attention to the trade-offs between privacy and identity in online environments and for greater interdisciplinary collaboration between qualitative researchers and computer scientists. In doing this, our goal is to contribute to what is still a relatively small body of online ethnographic research, outlining practical suggestions for social scientists who are interested in engaging with the burgeoning texts generated by virtual fora.

### Online Research: Three Prior Approaches

Online spaces have already proven to be rich sources of research material for social scientists (e.g., Boellstorff, 2012; Bortree, 2005; Dickens, Thomas, King, Lewis, & Holland, 2011; Hookway, 2008; Huffaker & Calvert, 2005; Karlsson, 2007; Kaun, 2010; Lopez, 2009; McCullagh, 2008; Olive, 2013; Pitts, 2004; Sade-Beck, 2004; Steinmetz, 2012; Wilson et al., 2015). With the possible exception of virtual world-based ethnographies, however, systematic research methods flexible enough to account for researchers' varied agendas, classic ethnographic emphases on understanding individuals in context, and the fluid nature of online content and identities have not yet been fully developed.

We have found it helpful to conceptualize research conducted in online environments into three separate categories: (1) the "big data approach," (2) the "quantitative approach," and (3) the "qualitative approach" (cf. Bernard, Wutich, & Ryan, 2016). Although these categories are a far from complete representation of the literature, they represent a useful tool for considering the methodological and disciplinary implications of online research focused on social patterns as it currently stands. Given the nature of our own research discussed here, we will pay particular attention to the third approach, but the other two remain vitally important for understanding the contours of current web-based research.

The big data approach is typified by initiatives pursued by corporations, government agencies, and universities, all of which employ quantitative methods to mine large data sets to look for patterns or to predict trends (Boyd & Crawford, 2012). Often, these research projects utilize metadata *about* users compiled by hosting services or search engines such as Twitter or Google. Well-known examples include *Google Flu*, which estimated prevalence and incidence for influenza worldwide based on when and where users were querying particular search terms, and Facebook's controversial experiments with user data. Big data analysis has been hailed as a major breakthrough in understanding complex behavioral patterns and is expected to have a significant applied effect, particularly for health care and consumer concerns (Chen, Mao, Zhang, & Leung, 2014;

Hay, George, Moyes, & Brownstein, 2013; Manyika et al., 2011; Murdoch & Detsky, 2013), but its current celebrity status has also been challenged by questions about its lack of granularity and context and its positivist epistemologies (Boyd & Crawford, 2012; Kitchin, 2014).

To date, the qualitative approach to online research has most often been used by anthropologists and other qualitatively oriented social scientists in the context of ethnographic explorations of geographically bounded communities that have an online component (e.g., Coleman, 2010; Malaby, 2009; Miller & Slater, 2000; Ryman, Burrell, Hardham, Richardson, & Ross, 2009; Williams & Jacobs, 2004). The frequent use of geographically bounded communities (like school groups) as online study subjects may reflect qualitative researchers' desire for greater context in which to situate online interactions, but it is not the only strategy for exploring these digital spaces. The qualitative approach has also been deployed to ethnographically explore immersive and self-contained virtual worlds such as *World of Warcraft* and *Second Life*, where the ethnographer becomes participant and the lens for analysis (e.g., Boellstorff, 2012; Boellstorff, 2015; Chen, 2009; Golub, 2010; Kozinets, 2010; Nardi, 2010; Taylor, 2006). Many of the researchers engaged in qualitative online research, particularly anthropologists, work alone or in small teams and focus on bounded communities with clearly defined memberships. Systematic methodological approaches, with their emphasis on reducing bias in data collection and analyses, have not been substantively addressed in much of this research.

A more recent trend in online qualitative research is an increasing interest in users' behavior on social network sites/social media platforms (e.g., Lijadi & van Schalkwyk, 2015; Murthy, 2008; Snelson, 2016; Tufekci, 2008). Boyd and Ellison (2007) define social network sites as

web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and transverse their list of connections . . . within the system. (p. 211)

In part because of this semipublic and highly interconnected nature, research investigating social media has added interesting questions about privacy and authenticity online to the literature on online qualitative research (de Laat, 2008; Livingstone, 2008; Trepte & Reinecke, 2011; Tufekci, 2008). Ethnographic research focused on social media platforms, and social networking sites in particular, has also primarily employed highly specific and nonsystematic sampling techniques to identify and study sites and users of interest.

### *Blogs as Ethnographic Texts*

Despite the increasing interest in online ethnographies, blogs remain fundamentally understudied as ethnographic texts in the qualitative approach to online research. Here we define an ethnographic text as a piece of writing that explores cultural phenomena from the point of view of community "insiders."

A blog is typically defined as a website with a series of frequently updated, reverse chronologically ordered posts containing original content written by one or more authors (bloggers), where each post includes opportunities for comments and links intended to promote a virtual community (Blogpulse, 2015; Blood, 2000; Garden, 2011; Hine, 2008; Hookway, 2008). Blogs have been spaces for both personal use and community building since their original inception in the earliest days of the Internet as compilations of links to other sites of interest (Herring, Scheidt, Bonus, & Wright, 2004). In contrast to these first collections of ephemera, however, modern blogs generally serve as personalized narratives, performative spaces, and self-reflective commentaries for both the blog writers themselves and the readers with whom they establish relationships (Karlsson, 2007; Kononov, Scotch, Post, & Brandt, 2010; de Laat, 2008; Monaghan, 2005; Pitts, 2004; Reed, 2005; Serfaty, 2004).

Social scientists have long studied the effects that different kinds of physical spaces have on the particular types of performativity individuals choose to exhibit and the implications for identity projects (e.g., Bourdieu, 1986, 1984; Butler, 1997, 1993, 1990). In the context of virtual spaces, it has been argued (e.g., Nardi, Schiano, Gumbrecht, & Swartz, 2004) that blogs are imbued with such a strong sense of the author's personality and attitudes that they constitute online diaries, but blogs are also written for an audience and are thus designed to elicit feedback and social networking (de Laat, 2008; Herring et al., 2004; Karlsson, 2007; Pitts, 2004; Rausch, 2006; Reed, 2005; Wilson et al., 2015). This mix of heavily individualized function transposed upon the inherently public space of published online content makes blogs surprisingly difficult to compare to other ethnographic texts such as field notes, interview transcripts, and traditional diaries.

Blogs, situated at the intersection of public and private, can therefore be considered a type of social media, although the practice of blogging predates many other forms of social media associated with particular corporations (e.g., Facebook, Twitter, etc.; Wilson et al., 2015). Furthermore, unlike more bounded social networking sites where all users share a single service provider's infrastructure, online hosting for blogs is not provided by a single organization. Instead, many possible blogging platforms (WordPress, Medium, Blogger, etc.) are available to bloggers. Blogging is also not restricted to a single type of content: Blogs can be

a means for presenting introspective thinking, a record of daily events, a tool for political mobilization, a journalistic project, an open-ended literary experiment, a constant exhibition of images and videos and, in many cases, a combination of all of the above. (Siles, 2011, p. 738)

It is unknown how many blogs exist worldwide, but one popular blogging host supports more than 75.8 million individual blogs as of 2015 (WordPress, 2015). All of these factors mean that obtaining a systematic and representative sample of blogs, a sampling approach that is appropriate for classical content

analysis and some other approaches to qualitative analysis (Bernard et al. 2016), has remained exceedingly difficult.

Prior qualitative research that has analyzed blogs—which are neither physically bounded nor considered “virtual worlds”—has tended to (1) employ small sample sizes, (2) use other forms of data collection to supplement the data extracted from blogs, (3) focus on longitudinal analysis of posts by the same blogger at different points in time, and (4) adopt convenience, snowball, or other nonrepresentative sampling methods (e.g., Bortree, 2005; Davis, 2010; Dickens et al., 2011; Hookway, 2008; Huffaker & Calvert, 2005; Karlsson, 2007; Lopez, 2009; Miura & Yamashita, 2007; Olive, 2013; Pitts, 2004; Qian & Scott, 2007; Sanderson, 2008; Wilson et al., 2015). While each of these trends represents a legitimate choice for approaching research questions in online environments, we argue that they may also reflect mitigation strategies for navigating the difficulties of conducting such research, although they are seldom explicitly framed as such.

### *Blogging as a Weight-Loss Narrative*

Our interest in blogging as a space for public and private identity creation made blogs oriented around weight and weight-loss concerns an appealing source of data. Health, weight, weight loss, and body projects are the common topics of conversation across multiple arenas in the United States today and are simultaneously intensely personal but also open to public debate and comment (Becker, 1995; Boero, 2012, 2010; Bordo, 1993; Brewis, 2014, 2011; Campos, 2004; Casper & Moore, 2009; Granberg, 2006; Greenhalgh & Carney, 2014; Nichter, 2001; Puhl & Heuer, 2010, 2009). It is within this larger discourse on weight and health that the so-called weight-loss blogosphere has developed. Although this term has been used by a few researchers thus far (e.g., Lynch, 2010; Rausch, 2006), it remains vague. Leggatt-Cook and Chamberlain (2011), for instance, in their study of female bloggers, based their definition and selection criteria on whether or not a blogger explicitly stated that she was blogging to support her weight-loss efforts, but they also point out the diversity of approaches and attitudes contained within these parameters. Other researchers (e.g., Boepple & Thompson, 2016) have focused specifically on the more controversial phenomenon of “fitspiration” and “thinspiration” websites (websites that promote fitness and showcase fit bodies and websites that promote and showcase thinness and thin bodies, respectively), but the websites in question are not exclusively blogs, nor representative of the wide array of attitudes toward weight and weight loss present in online environments. These prior attempts illustrate the heterogeneity of data available for analysis: Clearly, online environments offer rich opportunities for qualitative researchers interested in weight loss, but gathering and analyzing this data in a systematic and representative way remain a challenge for researchers.

Leggatt-Cook and Chamberlain (2011) note that weight-loss bloggers are diverse in terms of their chosen method (diet, exercise, surgery, etc.) and express a range of motivations for

choosing to blog about weight loss but most express interest in creating a community that will not only help them in their attempts to lose weight but also support them during the periods of doubt and failure and provide affirmation if and when they successfully lose weight. Thus, public documentation of private endeavors to lose weight is both a strategy to be held accountable but also to garner less critical support of the self-being put on display. As Leggatt-Cook and Chamberlain (2011) also discuss, in a larger environment in which fat bodies are routinely stigmatized, the fact that the blogger has more control over how his or her self is constructed for the online audience—via witty narratives and self-deprecating humor, for instance—is also important. We found the diversity of approaches toward weight loss, as well as the ability of bloggers to manipulate their presentation of self to readers, produced complicated and sometimes conflicting methodological and ethical challenges. The latter issue is certainly something that other researchers interested in social media users more broadly must contend with.

### *The Case of the Weight-Focused Blogger Study*

As previously discussed, social scientists have approached online research in general with a wide array of qualitative methods, but fewer studies have exclusively used blogs in their analysis. The aim of the project was to collect narratives from individually authored, public access blogs that discussed body and weight loss and were produced by individuals of any gender living in the United States, highlighting the potential ways in which blogging about weight loss projects influences health. The challenge was to ensure that we obtained a representative and systematic sample of these U.S.-based weight-loss blogs without having a clearly defined sampling frame, while also staying true to basic ethnographic principles that emphasize and contextualize individual lived experiences and perspectives. Here we present our process for a systematic sampling method for user-generated online texts that are outside the boundaries of a single website (e.g., a message board) or content host (e.g., Facebook).

We began by establishing preliminary inclusion and exclusion criteria for the sample. In order to limit possible variation in cultural norms and messaging, we sampled only blogs written by persons who explicitly disclosed in their narratives or “About Me” pages that they were residing in the United States. Even if bloggers disclosed little else about their life outside virtual reality, every writer we sampled did make some reference at some point to their physical location within United States. Bloggers in the United States are generally exposed to a broadly obesogenic but heavily fat-stigmatizing environment (Boero, 2012, 2010; Brewis, 2014, 2011; Campos, 2004; McCullough & Hardin, 2013; Puhl & Heuer, 2010, 2009). Multiauthored blogs were likewise excluded, as one of our research questions about weight loss explored changes in attitude toward weight loss in a single individual over time (see Trainer, Brewis, Wutich, Kurtz, & Niesluchowski, 2016). Additionally, we sampled only bloggers who were directly pursuing weight loss attempts, discussing

**Table 1.** Cross-Sectional Data Collection Using Weight-Oriented Blog Posts.

Stage	Methodologies	Sampling Strategy	Search Engines Used	N (Total = 234)
Stage 1: Scoping	General search terms, based on existing literature	Nonsystematic, random sample	Google	N/A
Stage 2: Seeding	Blogs from Stage 1 were analyzed for frequent words or phrases, and these words and phrases were used as key words to generate a new sample	Systematic, random sample	Google, Bing, Yahoo	112
Stage 3: Expanding	Purposive sampling was used to increase the diversity of blogs	Purposive sampling	Google, Bing, Yahoo	86
Stage 4: Verifying	The final sample was verified and enlarged by using Stage 2 and Stage 3 key words in a search engine that does not optimize users' results	Systematic, random sample	DuckDuckGo	36

weight loss or fat stigma, or otherwise engaging in an online space defined by the topic of weight. This was further complicated by our choice to use individual blog posts, rather than the blog itself, as the unit of analysis. Thus, the sample included both blogs that self-identified as weight-loss blogs and blogs that were a mix of topics.

Data collection took place in four stages (see Table 1). In Stage 1, a nonsystematic random sample of blogs was selected using Google for the purpose of generating pilot research. Highly general search terms including weight-loss blogs and “weight discrimination” were used to generate a broad selection of results;<sup>1</sup> these terms were selected based on prior research indicating they would yield a wealth of results relating to weight and weight loss (see Bair, Kelly, Serdar, & Mazzeo, 2012; Ballantine & Stephenson, 2011; Boepple & Thompson, 2016; Das & Faxvaag, 2014; Dickens et al., 2011; Harding & Kirby, 2009; Hwang et al., 2010; Leggatt-Cook & Chamberlain, 2011; Manikonda, Pon-Barry, Kambhampati, Hekler, & McDonald, 2014; Pitts, 2004; Rausch, 2006; Saperstein, Atkinson, & Gold, 2007; Tiggemann & Miller, 2010; Walstrom, 2000). The results provided by these terms were not included in the final sample, as the search terms that produced them were not systematically selected (i.e., we selected them ourselves, based on prior research) but were instead analyzed to refine search terms and phrases in the second phase of data collection.

In Stage 2, 12 new search terms based on frequently repeated and/or highly salient words or phrases found in Stage 1 results were used to query the three most popular search engines in the United States: Google, Bing, and Yahoo. These terms also diversified the sample in terms of blogger self-identified gender because of the nature of many of the salient phrases (e.g., “fat guy on a diet” vs. “fat girl on a diet”). We systematically tested the effects of adjacency and proximity operators like quotation marks with each search engine using each search term or phrase. As all three search engines present ranked results based on a website’s relevance to the search terms used, only the first 50 returns meeting the previously mentioned inclusion and exclusion criteria on each search engine were considered, based on the assumption that these results would be most salient. Duplicate results within or across search engines were included only once. For queries that returned more than one blog post from the same blogger, we included whichever blog post was returned first by

search engine. The sample was then screened to ensure blogs met previously listed inclusion criteria. The 112 blog posts were found eligible.

Stage 3 of data collection used purposive sampling to expand the diversity of blogs represented, particularly in terms of blogger sexual orientation, geographic location within the United States, religious affiliation, and race—all demographic factors that previous research on fat and obesity in the United States have posited as influencing attitudes toward fatness and weight loss (e.g., Boero, 2012, 2010; Brewis, 2014, 2011; Granberg, Simons, & Simons, 2009; Greenhalgh & Carney, 2014; McCullough & Hardin, 2013; Puhl & Heuer, 2010, 2009). Frequently occurring phrases from the 112 blogs selected in Stage 2 were used to seed this new round of queries. Otherwise, identical inclusion criteria to those used in Stage 2 were used to screen results in Stage 3. Research centered on weight-loss attitudes and practices in offline environments indicates that although weight loss is a nearly universal concern in the United States, the ways this concern is expressed and acted upon differ markedly, which is why we adopted purposive sampling (Becker, 1995; Boero, 2012, 2010; Bordo, 1993; Brewis, 2014, 2011; Campos, 2004; Casper & Moore, 2009; Granberg, 2006; Greenhalgh & Carney, 2014; McCullough & Hardin 2013; Nichter, 2001; Puhl & Heuer, 2010, 2009). Ultimately, an additional 86 blog posts were added in this stage.

In Stage 4, all previously selected search terms and phrases were queried using DuckDuckGo, a search engine that does not optimize results based on the collection of users’ demographic information or prior search history. This final set of queries helped to validate the representativeness of our sampling results by replicating and expanding the sampling frame extracted from Google, Bing, and Yahoo. This final stage yielded 36 additional blog posts that had not appeared in previous search engine queries.

By the end of data collection, therefore, we had collected 234 blog posts from 234 different bloggers. For further contextualization of blog posts and to ensure the diversity of our sample, we also collected demographic and background information on the bloggers themselves using their About Me pages and/or personal information contained within the blog narratives. We then compiled the information on age, gender, education level, socioeconomic status, weight-loss history, health

concerns, and attitudes toward blogging. The constructed nature of a blog did not allow us to verify background information without a serious breach of bloggers' privacy (just as a background check to "verify" an interviewer's chosen self-presentation could easily violate privacy in an in-person interview-based study). Although previous studies indicate that image management strategies are common on blogs (Bortree, 2005) and bloggers may reduce the amount of personal information shared if privacy is a concern (Krasnova, Günther, Spiekermann, & Koroleva, 2009), a high frequency of deception does not appear to be common. Age was difficult to establish, for example, but all but three bloggers disclosed their gender status.

Data analysis of the 234 blog posts followed qualitative thematic coding methods as described in Bernard, Wutich, and Ryan (2016). Thematic coding is a technique, often used in ethnographic research, designed to draw out salient themes, in this case related to the issues of weight and weight loss, and allows comparisons both across blogs and within the same blog through time. We developed our codes inductively through an iterative process that drew out repeated patterns present in the data (Bernard et al., 2016). Our coding schema addressed the following areas: overall writing style and tone of each blog post, expressed attitudes toward weight and weight loss within the post, and bloggers' portrayals of their own approaches to weight and weight loss.

Using the same sample of 234 bloggers, we then looked longitudinally at their posts across a 10-year period, 2005–2015 (the end date of our data collection). Given the variation in postfrequency across many of these bloggers, as well as the sheer volume of narrative material some of the bloggers generated during the period, we established specific time parameters on our search, examining blog entries that had been written in January and June only, following Rausch's (2006) suggestion that these months witness particular cycles of preoccupation with weight gain and loss in the U.S. context (the former because of New Year's resolutions postholiday feasting and the latter because of the incoming swimsuit season). For authors who did not begin blogging until after 2005, the earliest available post that fits the selection criteria was used. In the longitudinal sample, we switched from a consideration of the blog post itself (the one collected via search engine) as the unit of analysis to a consideration of the blogger as the unit of analysis and we explored shifts in attitudes toward weight and weight loss over time, as well as actual reported changes (or not) in blogger weights (for a more detailed discussion of the methods and results of our longitudinal thematic coding analysis, see Trainer et al., 2016).

Producing a rigorous method for sampling and analyzing textual data collected from blogs proved a complex endeavor. At the project's inception, we anticipated that the opportunity to use machine-driven sampling methods like search engines would accelerate and simplify data collection, while blog narratives would be well suited for text analysis because of their self-contained nature. These assumptions were repeatedly challenged during the course of our study because (a) key

characteristics of search engine technology were structurally contradictory to the principles of systematic sampling and (b) bloggers did not always construct a single identity contained on a single blog but often built multiple, shifting personas across multiple, shifting types of social media.

### *Sampling Challenges in Blog-Based Narrative Data Analysis*

Although scientists are increasingly utilizing online environments as sites for qualitative research, approaches to sampling in these environments have largely reproduced the strategies used for research in offline contexts, while acknowledging the difficulties of sampling in sometimes highly fragmented digital spaces (Davis, 2010; Stern, 2004). Sampling in off-line spaces generally uses a list from which some elements are chosen based on certain criteria (randomness, representation of specific types of cases, etc.; Bernard, 2012). Numerous sampling and recruitment errors and biases have been previously discussed in the literature, particularly with regard to surveys (Dillman, 1991; Groves et al., 2009; Groves & Lyberg, 2010; Wright, 2005) and qualitative methods (Bernard, 2012; Guest, Bunce, & Johnson, 2006; Marshall, 1996). Sampling and recruitment errors can affect the generalizability of results in representative samples and prevent researchers from reaching thematic saturation in ethnographic research using purposive or convenience samples (Bernard, 2012; Groves, et al., 2009). In considering the possibilities for similar errors in online environments, considerable attention has been devoted to sampling methodologies in online survey and focus group research (Bethlehem, 2010; Boydell, Fergie, McDaid, & Hilton, 2014; Couper, 2000; Dillman & Bowker, 2001; Fan & Yan, 2010; Sinclair, O'Toole, Malawaraarachchi, & Leder, 2012). The recruitment and sampling errors potentially present in online *ethnographic* research, however, as well as strategies for managing and mitigating them, have not yet been fully documented. Toward that end, we present here some considerations, based on our own case study of weight-loss blogs.

We attempted to answer our first research question (what techniques can be used for systematic sampling in qualitative online research?) through the adaptation of nondigital sampling methods to create a novel method for sampling online texts through the use of "seed" search terms, as described in Stage 1 of our sample. We then utilized an iterative sampling strategy (Stages 2–4) in an attempt to ensure a systematic and representative sample. We encountered several technological impediments to these goals. The sampling strategy outlined in our study is predicated upon the use of search engines (Google, Bing, Yahoo, and DuckDuckGo) to identify and extract potentially eligible blog posts. Employing a search engine such as Google to access this type of online content is intuitive and natural to anyone who has spent a significant amount of time on the Internet, whether in an academic or lay context. Several aspects of search engine-based sampling frames may be problematic from a methodological standpoint, however, and it is worth discussing each of these in detail.

The first issue stems from the fact that search engine query results are presented to the user hierarchically based on certain criteria. These criteria can include attempts by search engines to determine a result's salience to a user's query. Google's *PageRank algorithm*, which prioritizes results that are frequently linked to by other Internet content on the assumption that more frequent linkages indicate higher salience, is an example of criteria-based ranking by a search engine. Using DuckDuckGo in conjunction with the other search engines addressed some of these concerns.

Another issue arises from the fact that search engines further enhance their results by considering a user's previous search history and which results received a click-through in order to personalize the results received by a user. To track this information, many websites use individualized text files called browser "cookies" that are placed on personal computers when the website is visited. Each unique file contains information that notifies the website of the user's previous visits. Functionally, cookies act as a website's "memories" of users. In the case of search engines, cookies help track queries from specific computers, and this information is then used to predict which results are most desirable to that user in the future (Google, 2015). Search engines that do not employ cookies or other tracking mechanisms are available—again, our use of DuckDuckGo to check sample results in this case study was important. Here we wish to draw attention to the fact that cookies and related tracking technologies are a potential source of error when constructing samples in online environments. They are also largely invisible to the user. A researcher, therefore, must have prior familiarity with the underlying technologies and strategies used to produce search engine results before this source of error can be accounted for.

Rankings can also be manipulated, however, by blogger or other content producer-driven strategies, which are collectively often referred to as SEO. SEO is now an industry of its own, with consultants frequently hired by online businesses to increase their rankings in search engine results in order to receive greater visibility or to increase income from advertising hosted on their webpages (Furnell & Evans, 2007; van Couvering, 2004, 2007). "Cloaking," or designing a separate page optimized for discovery by search engine indexing algorithms, is an example of SEO (Malaga, 2008). Cloaked pages are inaccessible using standard web browsing methods and will never be seen by web users; their only function is to increase the ranking of the website in search query results by meeting search engine criteria. Factors such as ranking and SEO strongly affect search engine results but are not discussed in search engine results' pages or otherwise made obvious to users. Because these techniques are also invisible to the user, we cannot even detail the ways in which this may have affected our sample, but we certainly noticed that three or four bloggers appeared in every single one of our search phrase queries.

A final problematic aspect of search engines that we wish to bring up in this article is that search terms commensurate with search engine algorithms are difficult to generate empirically and systematically. Wilkinson and Thelwall (2011) point out

that the construction of relevant key words is surprisingly difficult, due to a range of factors including polysemy, synonymy, and outright spam (messages sent via automated messaging systems). As previously discussed, we used an unsystematic and informal list of key words (e.g., weight-loss blogs and weight discrimination) as seed terms to locate and analyze a first round of blog posts. The content of these blog posts was then used to create a combination of natural language and key word search phrases (e.g., "my struggles with weight loss," fat girl on a diet, and "dieting body acceptance"). These phrases were based on empirical data (blog posts) rather than heuristics (what the researchers thought would yield good results), but they may also be out of sync with search algorithms, which attempt to interpret natural language in a machine-compatible way (Google, 2015). Since search engines make certain assumptions about topic words and allows particular operators, the phrase my struggles with weight loss yields different results from the collection of key words "weight-loss struggles." Furthermore, over the course of our research, we found major disparities in bloggers' comfort and expertise with online platforms and technologies. Bloggers with greater Internet savvy are more likely to be oriented toward key word rather than natural language terms in order to facilitate high rankings on search engines (Furnell & Evans, 2007), while bloggers with less familiarity with key word optimization may be listed lower in the query results or even omitted altogether. Using empirical data to produce search terms, as we have done here, pushes forward online sampling methods but produces contradictions between systematic term generation and the underlying technological and economic structures shaping online content.

Ultimately, researchers working in online environments must, at minimum, document errors and biases in standard methodological techniques, but they can only do so when they are aware of them. Prior research indicates that search engine returns are biased in favor of commercial sites, popular sites, and U.S.-based sites (Croteau & Hoynes, 2006; van Couvering, 2004). By focusing on U.S.-based sites only, we eliminated one source of bias; by focusing only on private-authored blogs, we eliminated another. We noticed a trend in our results, however, whereby more widely read blogs, blogs written by individuals with more technological savvy, and blogs written by individuals who had successfully attracted commercial advertising were more likely to appear multiple times across many of our different search terms. These trends are possible consequences of the influence of search engine rankings, SEO, cookies, and the varying levels of Internet expertise across bloggers. The development of systematic methods documenting and correcting for these possible sources of error would be an important next step in the development of online research methods.

### *Ethical Challenges in Blog-Based Narrative Data Analysis*

As previously discussed, the ability of blog authors and other online content producers to manipulate their presentation of self, as well the unclear boundaries between blogs as spaces for private reflection versus as content for public consumption,

have ethical implications of interest for social scientists pursuing research in online environments. In an effort to respect at least the most immediately apparent of these boundaries, we chose to sample only blogs that were public access rather than part of a restricted access forum. Public access blogs are written for an unrestricted audience and are thus officially in the public domain (Sixsmith & Murray, 2001). Public-access blogs, however, may still contain information that would be deemed highly personal in other contexts (McCullagh, 2008; Siles, 2011), and academics hotly debate the implications of using personal information gathered online. Some feel that bloggers and their material must remain anonymous in academic work (even though many bloggers and commentators use fictionalized web names to begin with), while others argue that it is an issue of document copyright and intellectual property, and thus, bloggers and their material must be cited properly (Pitts, 2004; Wilkinson & Thelwall, 2011). We adhere to the latter perspective here—with important exceptions. We follow Snodgrass (2015) in his suggestion that researchers adhere to “local understandings of what constitutes public as opposed to private online exchange” (p. 473). In our context, the bloggers we studied overwhelmingly treated text reproduction and publication as a copyright issue, that is, their blogs indicated they wish to be cited correctly, and we prioritized their explicit wish for correct citation over the possible additional privacy of remaining anonymous. An important subset of blogs did contain warning messages that the blogger did *not* want his or her content reproduced elsewhere, and in these cases, we respected these stated wishes. Nevertheless, this did not appear to be the standard blogger approach to blog content.

Additionally, while it may be tempting for researchers to anonymize authors when sharing results based on the highly personal stories laid out in blogs, attempts to protect bloggers through anonymization also carry an ethical risk. Prior research suggests that blogger disclosures are intentional and people are aware of the trade-offs between privacy versus developing recognizable online identities and followers through repeated sharing of personal experiences (McCullagh, 2008). In some cases, therefore, anonymization may actually erase the bloggers’ own judgments about what is too risky or intimate to share and works against individuals’ attempts to construct online identity and community. Ultimately, we deferred to the blog authors themselves, insofar as possible, by using the stated requests on their blogs for quotations to be correctly cited (or, in some cases, their requests for their content not to be quoted) as a guide for their level of inclusion in our reporting of results. In cases where researchers are pursuing a more embedded ethnography that involves contact with participants engaged in digital content creation, it may be appropriate to ask informants directly what level of privacy they prefer.

We found that the public–private duality became even more complex when considered across multiple contexts. Producers of online ethnographic texts—in our case, bloggers—easily move across content platforms, and, perhaps even more importantly, researchers can *follow* their movements. If a blogger, for example, was consistent with their username or other

identifying information, we could in theory track them across social media sites, personal websites, blogs, directories, domain lookups, and other sources, and through time either concurrently or retroactively. In a project ostensibly about weight-loss *blogging*, we struggled with the interconnectedness of online content mediums: Was it ethically and methodologically acceptable to use a blogger’s Twitter account to confirm demographic data or characterize weight-loss journeys? What if the Twitter feed was linked on the blog? What if it wasn’t, but the usernames and the avatar photo were the same? Ultimately, in our own research, we adhered to the original parameters of our institutional review board (IRB) proposal and focused solely on blogs, but we admit the boundaries we had delineated sometimes felt highly artificial. One more involved approach might be to recruit key informants from within the sample who are comfortable sharing their linked identities and additional details of their online experience with researchers, then use this supplementary data to further validate and contextualize results. This separate layer of recruitment would allow informants to make informed decisions about the use of their content beyond an initial sampling of the public blog text. Furthermore, both the means of digital surveillance and its venues can (and did, in our case) change with incredible speed. Ethical safeguards such as universities’ IRBs require the delineation of research boundaries in advance but how can principal investigators satisfy ethical compliance guidelines when multiple digital platforms might be created, adopted, cross-linked with users’ other digital presences in additional online spaces, and then—not infrequently—abandoned by large numbers of users as they migrate to new venue, all within the span of a study? In the case of the weight-loss blogging study, we restricted ourselves only to the traditional form of a weblog—a website with a series of frequently updated, reverse chronologically ordered posts—in order to most clearly state the boundaries of our research to our IRB. This helped place reasonable limits on our sampling frame and ensured that we did not violate privacy by pursuing bloggers across multiple content platforms, but we doubtless missed many informative and enlightening weight-loss experiences that were chronicled by users of more recent platforms such as Twitter and Instagram. We suggest that researchers may need to consider a closer and more ongoing dialogue with IRB, to allow for greater flexibility in research methods and sites in response to the dynamic nature of online spaces.

Even the very concept of “online” as a separate space from in-person interactions was challenged during the course of our research. Certain bloggers, for instance, not only attracted followers but also online trolls (defined variously by Urban Dictionary as “One who purposely and deliberately starts an argument in a manner which attacks others on a forum without in any way listening to the arguments proposed by his or her peers” and “Being a prick on the Internet because you can”). Some of these trolls not only made negative comments in the online comments section but also made threats against the bloggers that affected them offline as well (e.g., disrupting speaking events and public appearances.) In these instances,

therefore, What is the role of the researcher? Does one passively observe (“lurk”)? Does one weigh in? Does one publish controversial blog content that seems certain to attract more negative attention toward a particular blogger—even if that blogger allows the reproduction of the text of her blog? Online identities can merge with off-line experiences to create ethically complex hybrid situations, and it is quite probable that future researchers will need to consider similar interstitial spaces, as the use of highly charged insults and harassment, including rape and death threats, is well-documented online (Consalvo, 2012; Jane, 2015; Mantilla, 2013).

## Conclusion

In our attempts to understand what techniques are available to perform systematic sampling for online qualitative research and what technological barriers and ethical dilemmas researchers might encounter, we articulate a number of key challenges for researchers. During our qualitative study of weight-loss bloggers, we found that the translation of research methods between standard and online texts is not a simple one. Where online text analysis solves some problems faced in traditional qualitative research (like easy access to that text), it also creates other, different challenges. Technological barriers, like search engine bias, could be managed but not eliminated and will likely require collaborations with computer scientists to deal with effectively. Ethical considerations, like the complicated ethics of working with bloggers and others who have developed unique and potentially highly identifiable identities across multiple online platforms, are yet to be fully appreciated and may require ongoing dialogue between researchers, ethical reviewer boards, and key ethnographic informants.

A goal here is to open a wider discussion about how we can harvest blog and other online texts in ways that are systematic and replicable, while also respecting and protecting those who produce them. We suggest that this is a major emerging set of issues that will require broad consideration among qualitative researchers, complicated by rapidly changing technologies and a multitude of complex and sometimes conflicting strategies for creating and navigating identity in online spaces. Despite these methodological and ethical challenges, the use of blogs as ethnographic texts—and ethnography in online spaces more generally—is filled with truly promising new types of possibilities for studying text and those who produce it.

## Acknowledgments

This study was supported in part by the Virginia G. Piper Charitable Trust through an award to the Mayo Clinic/Arizona State University Obesity Solutions Initiative. We also wish to thank Monet Neisluchowski and Deborah Williams for assistance with conceiving and executing the blog project.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Note

1. For a complete list of search terms, refer to Trainer, Brewis, Wutich, Kurtz, and Niesluchowski (2016).

## References

- Bair, C. E., Kelly, N. R., Serdar, K. L., & Mazzeo, S. E. (2012). Does the Internet function like magazines? An exploration of image-focused media, eating pathology, and body dissatisfaction. *Eating Behaviors, 13*, 398–401.
- Ballantine, P. W., & Stephenson, R. J. (2011). Help me, I'm fat! Social support in online weight loss networks. *Journal of Consumer Behaviour, 10*, 332–337.
- Barratt, M. J., & Maddox, A. (2016). Active engagement with stigmatised communities through digital ethnography. *Qualitative Research, 16*, 701–719.
- Becker, A. E. (1995). *Body, self, and society: The view from Fiji*. Philadelphia: University of Pennsylvania Press.
- Bernard, H. R. (2012). *Social research methods: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review, 78*, 161–188.
- Blood, R. (2000). Weblogs: A history and perspective [blog post]. Rebecca's Pocket. Retrieved from [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html)
- Boellstorff, T. (2012). *Ethnography and virtual worlds: A handbook of method*. Princeton, NJ: Princeton University Press.
- Boellstorff, T. (2015). Coming of age in Second Life: An anthropologist explores the virtually human. Princeton, NJ: Princeton University Press.
- Boepple, L., & Thompson, J. K. (2016). A content analytic comparison of fitspiration and thinspiration websites. *International Journal of Eating Disorders, 49*, 98–101.
- Boero, N. (2010). Fat kids, working moms, and the epidemic of obesity: Race, class, and mother-blame. In E. Rothblum & S. Solovay (Eds.), *The fat studies reader* (pp. 113–119). New York: New York University Press.
- Boero, N. (2012). Killer fat: Media, medicine, and morals in the American “obesity epidemic.” Newark, NJ: Rutgers University Press.
- Bonilla, Y., & Rosa, J. (2015). # Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist, 42*, 4–17.
- Bordo, S. (1993). Feminism, Foucault, and the politics of the body. In C. Ramazanoglu (Ed.), *Up against Foucault: Exploring some of the tensions between Foucault and feminism* (pp. 179–202). New York: Routledge.
- Bortree, D. S. (2005). Presentation of self on the web: An ethnographic study of teenage girls' weblogs. *Education, Communication & Information, 5*, 25–39.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice Cambridge, MA: Harvard University Press.

- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Santa Barbara, CA: Greenwood.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*, 662–679.
- Boyd, D., & Ellison, N. (2007). Social network sites: Definitions, history, and scholarship. *Journal of Computer-Mediated Communication, 13*, 210–230.
- Boydell, N., Fergie, G., McDaid, L., & Hilton, S. (2014). Avoiding pitfalls and realizing opportunities: Reflecting on issues of sampling and recruitment for online focus groups. *International Journal of Qualitative Methods, 13*, 206–223.
- Brewis, A. A. (2011). *Obesity: Cultural and biocultural perspectives*. Newark, NJ: Rutgers University Press.
- Brewis, A. A. (2014). Stigma and the perpetuation of obesity. *Social Science & Medicine, 118*, 152–158.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York, NY: Routledge.
- Butler, J. (1993). *Bodies that matter: On the discursive limits of sex*. New York, NY: Routledge.
- Butler, J. (1997). *Excitable speech: A politics of the performative*. New York, NY: Routledge.
- Campos, P. (2004). *The obesity myth: Why America's obsession with weight is hazardous to your health*. New York, NY: Penguin Group.
- Casper, M. J., & Moore, L. J. (2009). *Missing bodies: The politics of visibility*. New York, NY: New York University Press.
- Chen, M. G. (2009). Communication, coordination, and camaraderie in World of Warcraft. *Games and Culture, 4*, 47–73.
- Chen, M., Mao, S., Zhang, Y., & Leung, V. C. (2014). *Big data: Related technologies, challenges and future prospects*. New York, NY: Springer.
- Coleman, E. G. (2010). Ethnographic approaches to digital media. *Annual Review of Anthropology, 39*, 487–505.
- Consalvo, M. (2012). Confronting toxic gamer culture: A challenge for feminist game studies scholars [online article]. *Ada: A Journal of Gender, New Media, and Technology*, (1). Retrieved from [http://adanewmedia.org/2012/11/issue1-consalvo/?utm\\_source=rss&utm\\_medium=rss&utm\\_campaign=issue1-consalvo](http://adanewmedia.org/2012/11/issue1-consalvo/?utm_source=rss&utm_medium=rss&utm_campaign=issue1-consalvo)
- Couper, M. P. (2000). Review: Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494.
- Croteau, D., & Hoynes, W. (2006). *The business of media: Corporate media and the public interest*. New York, NY: Pine Forge Press.
- Das, A., & Faxvaag, A. (2014). What influences patient participation in an online forum for weight loss surgery? A qualitative case study. *Interactive Journal of Medical Research, 3*, e4.
- Davis, K. (2010). Coming of age online: The developmental underpinnings of girls' blogs. *Journal of Adolescent Research, 25*, 145–171.
- De Laat, P. B. (2008). Online diaries: Reflections on trust, privacy, and exhibitionism. *Ethics and Information Technology, 10*, 57–69.
- Dickens, M., Thomas, S. L., King, B., Lewis, S., & Holland, K. (2011). The role of the fatosphere in fat adults' responses to obesity stigma: A model of empowerment without a focus on weight loss. *Qualitative Health Research, 21*, 1679–1691.
- Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225–249.
- Dillman, D. A., & Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. In B. Batanic, U. D. Reips, M. Bosnjak, & A. Werner (Eds.), *Online social sciences* (pp. 53–71). Seattle, WA: Hogrefe & Huber.
- Dumova, T., & Fiordo, R. (2012). Blogging in the global society: Cultural, political and geographical aspects. In T. Dumova & R. Fiordo (Eds.), *Blogging in the global society: Cultural, political and geographical aspects* (pp. vii–xiv). Hershey, PA: IGI Global.
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*, 132–139.
- Furnell & Evans, M. P. (2007). Analysing google rankings through search engine optimization data. *Internet Research, 17*, 21–37.
- Garden, M. (2011). Defining blog: A fool's errand or a necessary undertaking? *Journalism, 13*, 483–499.
- Gehl, R. W. (2016). Power/freedom on the dark web: A digital ethnography of the dark web social network. *New Media & Society, 18*, 1219–1235.
- Golub, A. (2010). Being in the world (of Warcraft): Raiding, realism, and knowledge production in a massively multiplayer online game. *Anthropological Quarterly, 83*, 17–45.
- Google. (2015). How Google uses cookies—privacy & terms. Retrieved from <https://www.google.com/policies/technologies/cookies/>
- Graffigna, G., & Bosio, A. C. (2006). The influence of setting on findings produced in qualitative health research: A comparison between face-to-face and online discussion groups about HIV/AIDS. *International Journal of Qualitative Methods, 5*, 55–76.
- Granberg, E. (2006). “Is that all there is?” Possible selves, self-change, and weight loss. *Social Psychology Quarterly, 69*, 109–126.
- Granberg, E. M., Simons, L. G., & Simons, R. L. (2009). Body size and social self-image among adolescent African American girls the moderating influence of family racial socialization. *Youth & Society, 41*, 256–277.
- Greenhalgh, S., & Carney, M. (2014). Bad biocitizens? Latinos and the US “obesity epidemic.” *Human Organization, 73*, 267–276.
- Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: Wiley.
- Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*, 849–879.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*, 59–82.
- Harding, K., & Kirby, M. (2009). *Lessons from the fat-o-sphere: Quit dieting and declare a truce with your body*. New York, NY: Penguin Group.
- Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine, 10*, e1001413.
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (pp. 1–11). Honolulu, HI: IEEE.

- Hine, C. (2008). Virtual ethnography: Modes, varieties, affordances. In N. G. Fielding, M. L. Raymond, & G. Blank (Eds.), *The Sage handbook of online research methods* (pp. 257–270). Thousand Oaks, CA: Sage.
- Hookway, N. (2008). “Entering the blogosphere”: Some strategies for using blogs in social research. *Qualitative Research*, 8, 91–113.
- Horst, H., & Miller, D. (2013). *Digital anthropology*. London, England: A & C Black.
- Huffaker, D. A., & Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 10. doi:10.1111/j.1083-6101.2005.tb00238.x
- Hwang, K. O., Ottenbacher, A. J., Green, A. P., Cannon-Diehl, M. R., Richardson, O., Bernstam, E. V., & Thomas, E. J. (2010). Social support in an Internet weight loss community. *International Journal of Medical Informatics*, 79, 5–13.
- Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, 17, 65–87.
- Karlsson, L. (2007). Desperately seeking sameness: The processes and pleasures of identification in women’s diary blog reading. *Feminist Media Studies*, 7, 137–153.
- Kaun, A. (2010). Open-ended online diaries: Capturing life as it is narrated. *International Journal of Qualitative Methods*, 9, 133–148.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1–12.
- Konovalov, S., Scotch, M., Post, L., & Brandt, C. (2010). Biomedical informatics techniques for processing and analyzing web blogs of military service members. *Journal of Medical Internet research*, 12, e45.
- Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. Thousand Oaks, CA: Sage.
- Krasnova, H., Günther, O., Spiekermann, S., & Koroleva, K. (2009). Privacy concerns and identity in online social networks. *Identity in the Information Society*, 2, 39–63.
- Leggatt-Cook, C., & Chamberlain, K. (2011). Blogging for weight loss: Personal accountability, writing selves, and the weight-loss blogosphere. *Sociology of Health and Illness*, 34, 963–977.
- Lijadi, A. A., & van Schalkwyk, G. J. (2015). Online Facebook focus group research of hard-to-reach participants. *International Journal of Qualitative Methods*, 14, 1–9.
- Livingstone, S. (2008). Taking risky opportunities in youthful content creation: Teenagers’ use of social networking sites for intimacy, privacy and self-expression. *New Media & Society*, 10, 393–411.
- Lopez, L. K. (2009). The radical act of “mommy blogging”: Redefining motherhood through the blogosphere. *New Media & Society*, 11, 729–747.
- Lynch, M. (2010). Healthy habits or damaging diets: An exploratory study of a food blogging community. *Ecology of Food and Nutrition*, 49, 316–35.
- Malaby, T. (2009). *Making virtual worlds: Linden lab and second life*. Ithaca, NY: Cornell University Press.
- Malaga, R. A. (2008). Worst practices in search engine optimization. *Communications of the Association for Computing Machinery*, 51, 147–150.
- Manikonda, L., Pon-Barry, H., Kambhampati, S., Hekler, E., & McDonald, D. W. (2014). Discourse analysis of user forums in an online weight loss application. In A. Oh, B. Van Durme, D. Yarowsky, O. Tsur, & S. Volkova (Eds.), *Proceedings of the joint workshop on social dynamics and personal attributes in social media* (pp. 28–32). Baltimore, MD: Association for Computational Linguistics.
- Mantilla, K. (2013). Gendertrolling: Misogyny adapts to new media. *Feminist Studies*, 39, 563–570.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey Global Institute.
- Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, 13, 522–526.
- Mautner, G. (2005). Time to get wired: Using web-based corpora in critical discourse analysis. *Discourse & Society*, 16, 809–828.
- McCullagh, K. (2008). Blogging: Self-presentation and privacy. *Information & Communications Technology Law*, 17, 3–23.
- McCullough, M. B., & Hardin, J. A. (Eds.). (2013). *Reconstructing obesity: The meaning of measures and the measure of meanings*. Oxford, NY: Berghahn Books.
- Miller, D., & Slater, D. (2000). *The Internet: An ethnographic approach*. New York, NY: Berg.
- Miura, A., & Yamashita, K. (2007). Psychological and social influences on blog writing: An online survey of blog authors in Japan. *Journal of Computer-Mediated Communication*, 12, 1452–1471.
- Monaghan, L. F. (2005). Big handsome men, bears, and others: Virtual constructions of “fat male embodiment.” *Body & Society*, 11, 81–111.
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Journal of the American Medical Association*, 309, 1351–1352.
- Murthy, D. (2008). Digital ethnography: An examination of the use of new technologies for social research. *Sociology*, 42, 837–855.
- Nardi, B. (2010). *My life as a night elf priest: An anthropological account of World of Warcraft*. Ann Arbor: University of Michigan Press.
- Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, 47, 41–46.
- Nichter, M. (2001). *Fat talk: What girls and their parents say about dieting*. Cambridge, MA: Harvard University Press.
- O’Brien, M. R., & Clark, D. (2012). Unsolicited written narratives as a methodological genre in terminal illness: Challenges and limitations. *Qualitative Health Research*, 22, 274–284.
- Olive, R. (2013). “Making friends with the neighbours”: Blogging as a research method. *International Journal of Cultural Studies*, 16, 71–84.
- Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2016). *Digital ethnography: Principles and practice*. Thousand Oaks, CA: Sage.
- Pitts, V. (2004). Illness and internet empowerment: Writing and reading breast cancer in cyberspace. *Health*, 8, 33–59.
- Postill, J., & Pink, S. (2012). Social media ethnography: The digital researcher in a messy web. *Media International Australia*, 145, 123–134.
- Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, 17, 941–964.

- Puhl, R. M., & Heuer, C. A. (2010). Obesity stigma: Important considerations for public health. *American Journal of Public Health, 100*, 1019–1028.
- Qian, H., & Scott, C. R. (2007). Anonymity and self-disclosure on weblogs. *Journal of Computer-Mediated Communication, 12*, 1428–1451.
- Rausch, P. (2006). *Cyberdieting: Blogs as adjuncts to women's weight loss efforts* (Unpublished master's thesis). Gainesville, FL: University of Florida.
- Reed, A. (2005). "My blog is me": Texts and persons in UK online journal culture (and anthropology). *Ethnos, 70*, 220–242.
- Ryman, S. E., Burrell, L., Hardham, G., Richardson, B., & Ross, J. (2009). Creating and sustaining online learning communities: Designing for transformative learning. *International Journal of Pedagogies and Learning, 5*, 32–45.
- Sade-Beck, L. (2004). Internet ethnography: Online and offline. *International Journal of Qualitative Methods, 3*, 45–51.
- Sanderson, J. (2008). The blog is serving its purpose: Self-presentation strategies on 38pitches.com. *Journal of Computer-Mediated Communication, 13*, 912–936.
- Saperstein, S. L., Atkinson, N. L., & Gold, R. S. (2007). The impact of internet use for weight loss. *Obesity Reviews, 8*, 459–465.
- Serfaty, V. (2004). Online diaries: Towards a structural approach. *Journal of American Studies, 38*, 457–471.
- Siles, I. (2011). From online filter to web format: Articulating materiality and meaning in the early history of blogs. *Social Studies of Science, 41*, 737–758.
- Sinclair, M., O'Toole, J., Malawaraarachchi, M., & Leder, K. (2012). Comparison of response rates and cost-effectiveness for a community-based survey: Postal, internet and telephone modes with generic or personalized recruitment approaches. *BMC Medical Research Methodology, 12*, 132.
- Sixsmith, J., & Murray, C. D. (2001). Ethical issues in the documentary data analysis of Internet posts and archives. *Qualitative Health Research, 11*, 423–432.
- Snelson, C. L. (2016). Qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods, 15*, 1609406915624574.
- Snodgrass, J. (2015). Ethnography of online cultures. In H. R. Bernard & C. C. Gravlee (Eds.), *Handbook of methods in cultural anthropology* (2nd ed.) (pp. 465–495). Lanham, MD: Rowan & Littlefield.
- BlogPulse. (2015). Statistics. BlogPulse Homepage. Retrieved from <http://blogpulse.co/>
- Statistics. (2015). WordPress. Retrieved from <https://en.support.wordpress.com/stats/>
- Steinmetz, K. F. (2012). Message received: Virtual ethnography in online message boards. *International Journal of Qualitative Methods, 11*, 26–39.
- Stern, S. R. (2004). Expressions of identity online: Prominent features and gender differences in adolescents' World Wide Web home pages. *Journal of Broadcasting & Electronic Media, 48*, 218–243.
- Taylor, T. L. (2006). Does WoW change everything? How a PvP server, multinational player base, and surveillance mod scene caused me pause. *Games and Culture, 1*, 318–337.
- Tiggemann, M., & Miller, J. (2010). The Internet and adolescent girls' weight satisfaction and drive for thinness. *Sex Roles, 63*, 79–90.
- Trainer, S., Brewis, A., Wutich, A., Kurtz, L., & Niesluchowski, M. (2016). The fat self in virtual communities: Success and failure in weight-loss blogging. *Current Anthropology, 57*, 523–528.
- Trepte, S., & Reinecke, L. (2011). *Privacy online: Perspectives on privacy and self-disclosure in the social web*. Berlin, CT: Springer Science & Business Media.
- Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society, 28*, 20–36.
- Van Couvering, E. (2004). *New media? The political economy of Internet search engines*. Paper presented at the Annual Conference of the International Association of Media & Communications Researchers, Porto Alegre, Brazil.
- Van Couvering, E. (2007). Is relevance relevant? Market, science, and war: Discourses of search engine quality. *Journal of Computer-Mediated Communication, 12*, 866–887.
- Walstrom, M. (2000). "You know, who's the thinnest?" Combating surveillance and creating safety in coping with eating disorders online. *Cyberpsychology and Behavior, 3*, 761–783.
- Wilkinson, D., & Thelwall, M. (2011). Researching personal information on the public web: Methods and ethics. *Social Science Computer Review, 29*, 387–401.
- Williams, J. B., & Jacobs, J. S. (2004). Exploring the use of blogs as learning spaces in the higher education sector. *Australasian Journal of Educational Technology, 20*, 232–247.
- Wilson, E., Kenny, A., & Dickson-Swift, V. (2015). Using blogs as a qualitative health research tool: A scoping review. *International Journal of Qualitative Methods, 14*, 1609406915618049.
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*. doi:10.1111/j.1083-6101.2005.tb00249.x