P&I
POLICY & INTERNET

Social Media, Web, and Panel Surveys: Using Non-Probability Samples in Social and Policy Research

Vili Lehdonvirta, Atte Oksanen, Pekka Räsänen, and Grant Blank

The use of online surveys has grown rapidly in social science and policy research, surpassing more established methods. We argue that a better understanding is needed, especially of the strengths and weaknesses of non-probability online surveys, which can be conducted relatively quickly and cheaply. We describe two common approaches to non-probability online surveys—river and panel sampling—and theorize their inherent selection biases: namely, topical self-selection and economic self-selection. We conduct an empirical comparison of two river samples (Facebook and web-based sample) and one panel sample (from a major survey research company) with benchmark data grounded in a comprehensive population registry. The river samples diverge from the benchmark on demographic variables and yield much higher frequencies on non-demographic variables, even after demographic adjustments; we attribute this to topical self-selection. The panel sample is closer to the benchmark. When examining the characteristics of a non-demographic subpopulation, we detect no differences between the river and panel samples. We conclude that non-probability online surveys do not replace probability surveys, but augment the researcher's toolkit with new digital practices, such as exploratory studies of small and emerging non-demographic subpopulations.

KEY WORDS: online survey, probability survey, non-probability survey, panel survey, methodology, selection bias

网络调查的使用已在社会科学和政策研究中快速增长，使用次数比更早确立的方法还多。我们主张，应更好地了解网络调查的优缺点，尤其是非概率网络调查，后者能以相对快速和耗资较少的形式进行。我们描述了两种常用的非概率网络调查方法—随机抽样和面板抽样—并将其内在的选择偏差进行理论化，即主题性自我选择和经济性自我选择。我们用一个全面人口登记处的基准数据，对两个随机样本(脸书和基于*Web*样本)和一个面板样本(选自一个大型调查研究公司)进行了一项实证比较。随机样本偏离人口变量基准，其出现非人口变量的频率高出许多，即使对人口数据进行调整后也是如此；我们将这一现象归因于主题性自我选择。面板抽样则更接近基准数据。当检验不基于人口数据的亚群时，我们发现随机样本与面板样本之间不存在差异。我们的结论认为，非概率网络调查无法代替概率调查，但能以新的数字实践扩大研究人员的工具组，例如探究小型或新兴的不基于人口数据的亚群。

关键词： 网络调查，概率抽样调查，非概率抽样调查，面板抽样调查，方法论，选择偏差

doi: 10.1002/poi3.238

El uso de encuestas en línea ha crecido rápidamente en las ciencias sociales y la investigación de políticas, superando los métodos más establecidos. Argumentamos que se necesita una mejor comprensión, especialmente de las fortalezas y debilidades de las encuestas en línea no probabilísticas, que se pueden realizar de manera relativamente rápida y económica. Describimos dos enfoques comunes para las encuestas en línea no probabilísticas (muestreo de ríos y paneles) y teorizamos sus sesgos de selección inherentes: a saber, la autoselección tópica y la autoselección económica. Realizamos una comparación empírica de dos muestras de río (Facebook y muestra basada en la web) y una muestra de panel (de una importante empresa de investigación de encuestas) con datos de referencia basados en un registro de población integral. Las muestras de ríos divergen del punto de referencia en variables demográficas y producen frecuencias mucho más altas en variables no demográficas, incluso después de ajustes demográficos; atribuimos esto a la autoselección tópica. La muestra del panel está más cerca del punto de referencia. Al examinar las características de una subpoblación no demográfica, no detectamos diferencias entre las muestras de río y panel. Concluimos que las encuestas en línea sin probabilidad no reemplazan las encuestas con probabilidad, sino que aumentan el conjunto de herramientas del investigador con nuevas prácticas digitales, como estudios exploratorios de subpoblaciones no demográficas pequeñas y emergentes.

PALABRAS CLAVE: encuesta en línea, encuesta de probabilidad, encuesta de no probabilidad encuesta de panel, metodología, sesgo de selección

Introduction

Survey researchers face growing difficulties with traditional recruitment methods. Costs have been increasing for all types of surveys (Willems, van Ossenbruggen, & Vonk, 2006), response rates have been falling (Bethlehem, 2016), particularly for random digit dialing (RDD) phone surveys (Curtin, Presser, & Singer, 2005), and the rise of mobile phones has created problems for phone survey coverage (Son, Khattak, & Kim, 2013). Because of these difficulties, phone-survey-based research published in social science journals has steadily declined over the past 10 years (Figure 1). At the same time, the use of online surveys has exploded. Social scientists and policy researchers have enthusiastically embraced online surveys for their low costs and fast turnaround times (Fang, Wen, & Prybutok, 2013; Yun & Trumbo, 2000).

However, there is a growing gap between the rapid adoption of online surveys in social research and our methodological understanding of their usefulness (Bosnjak, Das, & Lynn, 2016). Many online surveys, including many published in high-quality, peer-reviewed journals (e.g., Griswold & Wright, 2004; Martin, 2009; O'Brien, 2017; Sagar, Jones, Symons, Tyrie, & Roberts, 2016), are non-probability surveys, where the probability of a given population member ending up in the sample is unknown. Moreover, there are many different ways of conducting non-probability online surveys, with different methodological implications. The purpose of this article is to compare the main approaches to non-probability online survey research today, and to analyze their strengths and weaknesses in studying the incidence of a phenomenon in a national population and the characteristics of non-demographic subpopulations. We focus on fundamental issues of representativeness, self-selection, and cost-efficiency. By "online survey" we mean a

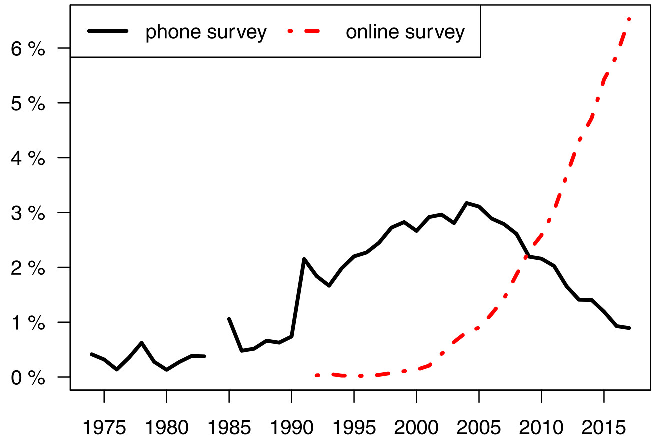

Figure 1. Phone and Online Survey Publications as Percentage of All Survey Publications in the Social Sciences Citation Index (SSCI). *Note:* All survey publications defined as search results returned by SSCI Topic Search for "survey"; phone survey publications defined as results for "*phone survey" OR "random digit dial*" (note wildcards); online survey publications defined as results for "online survey" OR "web survey" OR "internet survey." These are not exhaustive terms, and there is error caused by e.g. the presence of reviews and methodological articles, but arguably the results capture the main contours of the phenomenon.

survey where the respondents are recruited by an online method (e.g., via a website advertisement or a mass email) and the survey is completed online. We conduct an empirical comparison of three online survey samples and benchmark data from a rigorous probability survey conducted by a national statistical agency. In particular, we compare river samples with other non-probability and probability samples, which is rare in previous literature, and conclude with a discussion of emerging research practices enabled by non-probability online surveys.

Background

In social science and policy research some of the most common uses of surveys are estimating population means, such as the prevalence or incidence of a phenomenon in a national population, and the characteristics of the affected population. The ideal sample for generalizing to a population is fully probabilistic, that is, one where each person in the population has a known, non-zero chance of being selected (Groves, 1989). Since the 1970s, survey researchers have used RDD to create probability samples (Keeter, Kennedy, Dimock, Best, & Craighill, 2006). Until the rise of mobile telephony, landline phones provided excellent coverage of geographically defined populations in rich countries. Differences in landline coverage between sociodemographic groups (e.g., wealthy and poor households) were approximately known, so survey results could be weighted to correct for these known inequalities in the probability of selection. In countries with comprehensive population registries, an alternative to RDD is to select participants randomly from the population registry (Statistics Finland, 2016). In practice, the samples are never fully representative due to non-response bias: not everyone selected for inclusion agrees to participate, and refusal and non-completion are not random, but systematically linked to respondents' attributes (Groves, 1989; Lavrakas, 2008). The best phone and mail surveys repeatedly contact people in the sample until they respond, maximizing the response rate and thus minimizing the possibility of non-response bias.

Online surveys first emerged in late 1980s (Walsh, Kiesler, Sproull, & Hesse, 1992). Some online survey panels are based on probabilistic methods, where respondents are initially recruited via telephone RDD or other traditional means, and subsequently access the survey via the Internet (Bosnjak et al., 2016; Callegaro et al., 2014a; Yeager et al., 2011). To eliminate coverage bias, those without Internet access are given a free computer and Internet connection or asked to return responses via other means. However, probability-based online panels remain rare, existing only in a handful of countries (Bosnjak et al., 2016). They are also relatively expensive to set up and maintain. For these reasons, the majority of Internet survey research today is in practice conducted with inexpensive and widely accessible non-probability "convenience" sampling methods. There are also digital means of obtaining probability samples from specific populations, such as when a list of the email addresses of the employees of a firm is used to obtain a random sample of the employees of the firm (Couper, 2000), but these are not discussed in this article.

Types of Non-probability Online Surveys: River and Panel Sampling

The original and simplest non-probability approach to recruiting respondents online is "river" sampling, also known as intercept sampling or real-time sampling (Olivier, 2011; Walsh et al., 1992; Yun & Trumbo, 2000). River sampling means recruiting respondents by inviting them to follow a link to a survey placed on a web page, email, or somewhere else where it is likely to be noticed by members of the target population. The name refers to the idea of researchers dipping into the traffic flow of a website, catching some of the users floating by.

A basic problem with river sampling is the coverage bias. Coverage bias occurs because not every subpopulation is represented proportionately or indeed at all in digital media (Räsänen, 2006). This is partly a problem of unequal access to the Internet—the digital divide—but further problems are caused by so-called second-level digital divides (Hargittai, 2002), manifesting as disparities in usage style and frequency (Van Dijk, 2005). Different subpopulations access very different websites and services, and at very different frequencies. Without knowing the demographic distribution of the users of a service and how frequently they access the service, a researcher cannot establish a probability model linking the river sample to a national population. A river sample is not formally generalizable even to the user population of the service in question, only to the population of "active" users defined as users accessing the service within the sampling period. In the worst case, river samples can be vulnerable to attacks by organized groups who intentionally bias the survey.

Besides river sampling, a second approach to online survey research today is the use of commercial non-probability online panel providers. There are many such providers today, ranging from new startup companies to established media and research outfits such as Ipsos Mori, Qualtrics, and Survey Sampling International (SSI) (Callegaro et al., 2014a). In contrast to probability-based online panels, typical recruitment strategies of non-probability online panels include placing ads on websites or social media, and distributing invitations to newsgroups and mailing lists (for more examples, see Callegaro, Villar, Yeager, & Krosnick, 2014b). Interested users opt in to become panel members, and users from multiple sources are often blended into a single panel (Lorch, Cavallaro, & van Ossenbruggen, 2014). A key difference to river sampling is that the panel providers undertake to manage the demographic compositions of their respondent pools, preventing organized attacks and trying to correct the biases stemming from digital medias' uneven coverage so that the panels would be demographically similar to national populations. Online panel surveys are widely used by social scientists as well as policy consultants and think tanks.

However, previous work suggests that panel surveys still systematically over-represent some groups and under-represent others (Willems et al., 2006). For instance, U.S. online panel surveys over-represent white, better educated, active Internet users (Chang & Krosnick, 2009; Dever, Rafferty, & Valliant, 2008; Malhotra & Krosnick, 2007). One explanation is found in second-level digital divides: even if a panel's registered membership is demographically similar to the population, subpopulations within the panel may still differ in how frequently they go online and complete surveys. Another possible explanation is that panel providers are not as scientifically rigorous about their panels' demographic composition as their marketing materials suggest (Willems et al., 2006). In practice, researchers often find it necessary to impose quotas on respondents from different sociodemographic groups to obtain samples whose demographic composition mirrors the population (Callegaro & DiSogra, 2008). Panel samples undoubtedly offer better coverage than river samples, but the difference may be more a matter of degree than fundamental quality.

Crowdsourcing providers such as Amazon Mechanical Turk and Prolific are similar to panel providers, in that they provide access to opt-in respondents. Their main difference for the purposes of survey research aimed at estimating population means is that crowdsourcing providers make no pretense of trying to curate panels that in some ways mirror the general population (Weinberg, Freese, & McElhattan, 2014). Their value proposition is essentially providing easy access to large convenience samples.

Types of Self-selection Biases in Online Surveys: Topical and Economic

Online survey researchers must not only reach the desired kinds of users (coverage), but also convince them to respond. A possibly more consequential difference between river sampling and panel sampling is found in the methods through which they coax users to respond, and the different self-selection biases that these methods give rise to. Self-selection refers to respondents selecting themselves for participation in a survey; self-selection bias arises when the propensity to self-select differs systematically between subpopulations (Bethlehem, 2010). It is a reverse of the problem of non-response bias in probability surveys, where participants selected by the researchers de-select themselves in uneven ways. Though self-selection bias has been extensively studied and theorized (Groves, 1989; Lavrakas, 2008), we argue that the specific forms of self-selection inherent in different forms of non-probability online surveys warrant further conceptualization.

In the early years of the Internet, researchers could sometimes obtain large sample sizes simply by posting their survey link to a mailing list or discussion group (Walsh et al., 1992). Today, online surveys are no longer novel and potential respondents are rarely attracted to yet another survey, so this is no longer possible. Surveys have to compete for attention with a bewildering variety of digital content, including commercial content designed to capture users' attention. Researchers have responded in different ways. One common method is to collaborate with a media outlet, publishing a link to the survey as part of a news article or other piece of content related to the survey topic. The survey piggybacks on the attention given to the news article, reaching a larger audience. Another method is to place paid advertisements. The "click-through rate" (CTR) or percentage of users who follow the link depends on how effectively the ad's text and imagery capture the users' attention; in the language of the online advertising industry, the ad has to be "relevant" to the audience. The two main modes of river sampling today—advertising and piggybacking on media content—are thus fundamentally similar, in that they recruit respondents by attracting their attention with content.

A key weakness of these methods is that not everyone is equally attracted. Specifically, participants self-select on the basis of how salient they find the ad or media content (Chang & Krosnick, 2009). If the salience is systematically related to the attributes under study, then the sample is likely to be biased (Bethlehem, 2010). For example, Couper, Singer, Conrad, and R. Groves (2008) found that volunteer respondents to a privacy survey were more interested in privacy than the general population. This form of selection can be referred to as *topical self-selection:* the advertised topic of the study ends up determining who responds to it. A related weakness is *priming* or pretest sensitization: respondents exposed to a piece of content on a topic just before taking a survey may be inclined to reinterpret their situations and experiences through frames put forward in that content. For instance, if a respondent reads an online newspaper article about other peoples' victimization experiences just before responding to a survey question on whether they have such victimization experiences themselves, then they might be more inclined to recall or reinterpret experiences from their past in that light. Subjective experiences are more susceptible to such priming than objective attributes such as age. Because of topical self-selection and priming, river samples are thus likely to generate higher incidences of the attributes under study than probability samples.

In a panel sample or crowdsourced sample, the method of attracting responses is different: the primary enticement to participate is typically not attention, but compensation. "Professional" respondents enrolled by a panel provider can participate in multiple surveys and are compensated for each survey with money, discount vouchers, prize draws, or similar (Hillygus, Jackson, & McKenzie, 2014). Research on online panel respondents' motivations for participation shows that enjoyment and altruism are also important motivations, but studies consistently find that those participating in more surveys are more likely to be doing so for monetary reasons (Hillygus et al., 2014; Sparrow, 2007). In contrast to river sampling, prospective participants typically receive very little prior information about the topic of the survey; a vague one-line description or less is usual. Panel sampling is thus likely to suffer less from topical self-selection than river sampling.

However, due to the compensation, panel sampling may be more susceptible to another form of selection, which can be termed *economic self-selection:* the prospective participant weighs the proposed compensation (monetary or other) against their opportunity cost and decides on that basis whether to participate. This biases the sample towards people with a low opportunity cost, such as the unemployed, retired, and students. Willems et al. (2006) found that the most active online panel respondents in the Netherlands were less satisfied with their incomes and more often unfit for work than the average respondent. Weinberg et al. (2014) found that U.S. respondents from Amazon Mechanical Turk were younger and poorer than average U.S. adults. By biasing the socioeconomic composition of the sample, economic self-selection may indirectly influence the non-demographic attributes of interest as well.

Using Online Surveys to Study National Populations?

Despite these inherent biases, non-probability online surveys are now frequently used to make claims about the general population in social science and policy research (e.g., Martin, 2009; O'Brien, 2017; Petzold, 2017). Is this justified? In the empirical part of this article we ask:

RQ1. How accurately do non-probability river and online panel samples reflect the general population?

Few studies have directly compared non-probability online samples with benchmark data (for an overview, see Callegaro et al., 2014b). Chang and Krosnick (2009) compared a compensation-based non-probability online panel sample with an RDD sample and a probability Internet sample. They found that incidences of non-demographic variables (concerning political participation) were about 10 percent higher in the non-probability sample, even after weighting. They attributed this to panel respondents having self-selected into the survey on the basis of a one-sentence description of the survey (which we conceptualized as topical self-selection). Distributions of demographic variables in the unweighted panel sample deviated on average 9.3 percentage points from U.S. Current Population Survey (CPS) data, compared to 4.5 and 4.3 points for the probability samples. In part, this could be attributable to economic self-selection. But weighting with Internet access propensity reduced the river sample's average deviation to only 3.0 points, which can be seen as a very good result. However, the data used in the study was collected in 2000, when the Internet's attention economy was very different from what it has subsequently become.

Three more recent studies have been conducted in the Netherlands. Scherpenzeel and Bethlehem (2010) compare opt-in non-probability online panel surveys and a probability Internet survey, selected using a true random sample from the population register, against benchmark data. They show that the results from the probability Internet survey are much closer to benchmark data than are the results of the non-probability panels. Bethlehem (2015) compared three surveys: two self-selected surveys and one that was partially randomly selected from a population register and partially self-selected. Means in the two self-selected surveys are wildly different from the partial random sample, by as much as 30 percentage points. Brüggen, Van Den Brakel, and Krosnick (2016) compare the results of 18 opt-in online panels with the results from two benchmark surveys based on random samples of the Dutch population. They found that the non-probability panels produced deviations from the reference surveys ranging from 3.74 to 9.75 percentage points. Furthermore, the results from the non-probability panels could not be adjusted. Post-stratification weighting did not correct these deviations, and in many cases actually made them larger. They conclude that their "results fundamentally challenge the quality of nonprobability panels and suggest sticking to the foundations of survey sampling" (Brüggen et al., 2016, p. 22).

Yeager et al. (2011) compared two probability samples (an RDD phone sample and a probability Internet sample) with seven non-probability online samples (six panel samples and one river sample), using 2004–2009 U.S. government registry and survey data as benchmarks. The probability samples were more accurate. Non-probability online samples deviated on average 5.2 percentage points from the benchmark on secondary demographics and non-demographics (ranging from passport ownership to smoking). Post-stratification weights did not significantly improve the accuracy, and in some cases made the non-probability samples *less* accurate than the unweighted data. The accuracy of the online surveys remained about the same from 2004 to 2009.

The accuracy of Yeager et al. (2011) non-probability samples is actually quite reasonable for many purposes, perhaps as a result of their sophisticated recruitment methods. Their river sample was recruited with pop-up invitations that "appeared on the computer screens of users of a popular Internet service provider" (Yeager, 2011, p. 713). This probably resulted in very good coverage of Internet users compared to typical river sampling efforts. But very few social researchers have access to such coverage, especially today, when many people access Internet services primarily through mobile apps (Lugtig & Toepoel, 2016). With samples that are more recent and more typical of everyday social and policy research, we would expect river samples to be significantly more biased.

Using Online Surveys to Study Subpopulations?

The comparison studies reviewed above are positioned in the field of public opinion studies, where generalizing to national populations is paramount. In other social and policy research it is often important to reach non-demographic subpopulations. These groups are not defined by standard demographic variables like age, gender, and ethnicity, but by circumstance or behavior. Examples include gig workers (Lehdonvirta, 2018), e-cigarette users (Dawkins, Turner, Roberts, & Soar, 2013), or cyberharassment victims (Dreßing, Bailer, Anders, Wagner, & Gallas, 2014; Näsi et al., 2014). They are often low-incidence subgroups and they can be considered a subcategory of what Callegaro et al. (2014a) refer to as "specialty panels." National probability samples have significant drawbacks in studying such subpopulations. If the subpopulation's size is small, then a very large overall *n* is needed to capture a sufficient number of subpopulation members. This problem is compounded by the fact that for novel social phenomena such as e-cigarette use or cyberharassment, researchers must often collect their own data, as government data are not available. In some cases, the population of interest crosses national boundaries, as with online gig workers. Against this background, it is unsurprising that many articles on hard-to-reach non-demographic subpopulations published in leading journals are now based entirely on non-probability online surveys (e.g. Dawkins et al., 2013; Näsi et al., 2014; Sagar et al., 2016).

As discussed above, there are many theoretical problems with generalizing from non-probability online samples. Articles published in top journals often avoid this by simply examining the characteristics of the sample without making any formal inferences to the population of interest. It is implicitly assumed, but not necessarily explicitly stated, that the sample can give researchers some hints as to the characteristics of the population, making the exercise worthwhile. However, this assumption remains largely untested. We therefore ask:

RQ2. How accurately do non-probability river and online panel samples reflect a non-demographic subpopulation?

In particular, we are interested in providing empirically backed advice for social and policy researchers who need to choose which mode of online survey sampling—river or panel—to use for studying a non-demographic subpopulation. The preceding discussion on topical self-selection suggests that river sampling methods could yield higher proportions of respondents who belong to the subpopulation of interest, reducing the total *n* needed and thus the cost. But they could also have different characteristics from subpopulation members reached through online panel sampling, due to priming and other effects. We will examine these issues with an empirical study.

Data and Methods

We address the research questions by comparing responses from three online surveys conducted in Finland in 2013. We thus respond to Brüggen et al.'s (2016, p. 23) call to replicate their work in Nordic countries, which have very high rates of Internet use and also reliable government registers that can provide high-quality benchmark data. We moreover use an array of different online sampling methods, representing methods frequently used in small and medium-sized social and policy research projects. We (i) use a commercial online panel, (ii) recruit respondents through a link in a news article published on the website of Finland's public broadcaster, and (iii) attract respondents via advertisements posted on social media. The responses are all elicited using an identical online survey instrument. The samples are compared with benchmark data from a rigorous probability survey conducted simultaneously by Statistic Finland, weighted to match population registry data. This design addresses some of the challenges of comparing survey outcomes across different methods (DiSogra & Callegaro, 2016).

The survey instrument included standard sociodemographic questions as well as questions on various online and offline behaviors and experiences, including questions related to "cyberhate" (i.e., hate expressed via online media). The survey was conducted as part of a research project focusing on cyberhate, but it was marketed as a general survey concerning Internet use. The instrument was accessible to both desktop and mobile Internet users. A disadvantage of the survey for the purposes of methodological comparison is that it targeted only 16–30 years old, so coverage biases arising from digital divides between young and old people cannot be addressed in this study. But a notable advantage is that two of the instrument's non-demographic items were also incorporated into the well-resourced national statistical survey that was executed simultaneously, providing excellent demographic and non-demographic benchmark data for examining self-selection biases. In the following sections, we introduce the samples, measures, and analysis techniques in more detail.

Samples

The study has four samples using different recruitment methods, summarized in Table 1. The *YLE sample* consists of respondents who found the survey through a link attached to an online news article published by Finland's public broadcaster YLE in May 2013. The article was about cyberhate and the link to the survey was active for two weeks. Another sample labeled as the *Facebook sample* consists of respondents recruited using three Facebook advertising campaigns targeted at Finnish users aged 15–30. The campaigns were launched between April and May 2013. Four images and four texts were used in 15 combinations to entice users to click or tap on the advertisement, detailed in Supporting Information Appendix 1. A prize draw (movie tickets) was mentioned in all ads. The campaigns received a

Table 1. Samples and Their Recruitment Methods, Field Periods, and Responses

	YLE	Facebook	SSI	Statistics Finland
Type	River	River	Online panel	Probability
Sampling method	Link in news article	Targeted ads with prize draw	Non-interlocking quotas	Population registry
Interview mode	Online survey	Online survey	Online survey	CATI
Start date	May 2013	April 2013	May 2013	April 2013
End date	May 2013	May 2013	June 2013	May 2013
N	493	1,337	555	2,945
n (16–30 years old)	232	1,089	544	640

total of 6,074 clicks from unique users, of which 1,337 proceeded to complete the survey. From both the YLE and Facebook samples, respondents aged 16–30 were selected for analysis in this study.

The *SSI sample* consists of respondents recruited from May to June 2013 from a Finnish non-probability online panel administered by SSI. SSI does not provide details on how the panel members were recruited nor what the response rate was, but in general the company blends users from different sources, and attempts to maintain consistency of the multi-sourced sample by pre-screening respondents with reference to standard sociodemographic and other factors (for details, see Lorch et al., 2014). SSI sent email invitations to our survey to the panel members, and prospective respondents were asked to complete additional screening questions on age group, gender, and residential area; non-interlocking quotas were applied to obtain respondents that mirror the Finnish population aged 15–30 in terms of these factors ($n = 555$). From these data, 544 respondents aged 16–30 were selected for this study. The SSI sample thus consists of "semi-professional" multiple-survey takers, whereas the YLE and Facebook samples represent river samples with "naïve" respondents. The YLE sample is purely attention-based, whereas the Facebook sample attracted participants with a combination of attention-seeking and a minor economic compensation (prize draw).

The benchmark data comes from Statistics Finland's *Use of Information and Communications Technology by Individuals Survey*, fielded to 16-74-year-olds in April–May 2013. The sample was drawn from the national population registry using simple random sampling. Informants were contacted by telephone, and responses were obtained over the phone using computer-assisted telephone interviewing (CATI). The survey yielded a response rate of 54 percent. Non-response bias was addressed using Statistics Finland's post-stratification weight variable, which weights by respondent age, gender, and residential area to match the actual population distribution as revealed by the population registry (for details, see Statistics Finland, 2016). From this data, we selected respondents aged 16–30 as our benchmark data ($n = 640$).

Measures

Non-Demographic Variables. Two non-demographic variables are available in both the online surveys and the benchmark data. One, exposure to cyberhate, was elicited with, "In the past three months, have you seen hateful or degrading writing or speech online, which inappropriately attacked certain groups of people or individuals?" (yes/no). Another is cyberharassment victimization: "In your own opinion, have you been a target of harassment online, for example when people have spread private or groundless information about you or shared pictures of you without your permission?" (yes/no). The two form an interesting pair, as both are highly policy-relevant, but the former can be expected to have a much higher incidence in the general population than the latter.

Demographic Variables. Our demographic and background variables consist of gender, age, self-reported Facebook use (yes/no), and whether the respondent is a full-time student, reflecting an aspect of their economic status. We treat age as a continuous variable, others as categorical.

Measures of Subpopulation Characteristics. We compare the cyberharassment victim subpopulations across the three online samples using attitudinal and behavioral characteristics that often associate with cyberharassment victimization. These items are not available in the benchmark data. The items are detailed in Supporting Information Appendix 2.

Statistical Techniques

Our aim in the explanatory analysis is first to compare the raw frequencies of online hate exposure and cyberharassment victimization in the three online samples against the benchmark data, and then to evaluate the effect of using post-stratification weighting based on national demographic statistics (Statistics Finland, 2016) to see whether this can be used to produce frequencies that are closer to the benchmark data (RQ1). After that, we compare victim subpopulations in the three online survey samples in terms of several attitudinal and behavioral characteristics, to see how similar they are (RQ2). In the ordinary least squares regression models, we report the adjusted mean-values, or predictive marginal coefficients, of the attitudinal and behavioral scales. We report 95% confidence intervals based on a 10,000 replication bootstrap.

Results

Online Surveys Versus the General Population

Table 2 shows descriptive statistics of demographic variables in the four samples. The two river samples (YLE and Facebook) are very different from the benchmark, and also from each other. In contrast, the SSI panel sample composition is very close to the benchmark, which is partly unsurprising as quotas were imposed on gender, age group, and residential area. However, the SSI data is also very close to the benchmark on the percentage of students. This is notable, because being a student implies a different economic status from being in the labor market,

Table 2. Descriptive Statistics of Demographic Variables

	Male %	Mean Age (SD)	Student %	*n*
Statistics Finland	52.5	23.19 (4.12)	46.5	640
SSI	50.0	22.7 (4.11)	45.8	544
Facebook	38.6	19.51 (3.87)	74.8	1,089
YLE	25.0	23.69 (4.19)	60.3	232

so economic self-selection effects might have been expected to bias this variable, given that panel participation is compensated.

Table 3 shows the percentage distributions of the non-demographic variables. In the two river samples, the distributions are very different from the benchmark data, reporting cyberhate exposure approximately 1.5 times as frequently and cyberharassment victimization approximately six times as frequently as benchmark respondents. The online panel survey (SSI) is closer to the benchmark. On cyberhate exposure, the panel is only 1.2 percentage points away from the benchmark. On cyberharassment victimization, the panel produces a higher incidence than the benchmark, but by a smaller margin than the river samples. The panel survey thus performs better across the board than the river samples. One possible explanation is that incidence in the general population varies across sociodemographic groups; since the panel survey used quota sampling, it captured respondents whose sociodemographic composition is much closer to the general population, and thus yielded incidences closer to the benchmark. If this explanation is correct, then we should be able to obtain better accuracy from the river samples by weighting them to match the population on sociodemographic background.

The second column of Table 3 shows the distributions of the non-demographic variables when post-stratification weights are applied to match the samples with the national population on gender, age, and student status. This weighting causes cyberharassment incidence in the YLE sample to fall by 9.5 percentage points, but otherwise the weighted distributions are surprisingly similar to the unweighted. The river samples still yield incidences that are far greater than the benchmark. This indicates that the large non-demographic differences between the samples cannot be explained simply by differences in respondents' demographic backgrounds. The people who volunteered for the river samples are different from the general population in ways that go beyond demographic differences.

A more likely explanation for the divergent results is self-selection. It seems likely that the YLE sample is biased by topical self-selection: people who paid attention to an article on cyberhate often did so because they had relevant personal experience, resulting in the very high incidences reported. In the Facebook sample,

Table 3. Distributions of Non-Demographic Variables (95% Confidence Intervals)

Exposure (yes)	Unweighted %	Weighted %
Statistics Finland		
Cyberhate	48.7 (44.3–52.2)	48.7 (44.8–52.6)
Cyberharrasment	5.3 (3.3–7.2)	5.3 (3.6–7.1)
SSI		
Cyberhate	47.6 (43.4–51.8)	46.4 (42.2–50.6)
Cyberharrasment	19.5 (16.1–22.9)	19.7 (16.3–23.1)
Facebook		
Cyberhate	68.1 (65.4–70.9)	67.7 (64.9–70.5)
Cyberharrasment	35.9 (32.8–39.1)	35.7 (32.6–38.8)
YLE		
Cyberhate	89.3 (85.2–93.4)	87.6 (83.3–91.9)
Cyberharrasment	30.1 (23.8–36.4)	20.6 (15.1–26.2)

the potential self-selection mechanism is less obvious. Four different ad texts were used, none of which directly referenced cyberhate or cyberharassment (Table 4). But the copy text that attracted most responses by far can be seen as containing indirect references: "Are you worried? Have you come across strange things on the net?" This could have disproportionately attracted respondents exposed to cyberhate or victimized by cyberharassment. Details on the ad campaigns are described in Supporting Information Appendix 1.

Another possible explanation for the very high incidences reported by the river samples is priming. In the case of the YLE sample, people who read the article on cyberhate might have become more attuned to interpreting their past experiences in that light. In the case of the FB sample, any priming effects would have been more abstract and indirect. This possibility is examined in the next section. On RQ1, we can conclude that river samples were not a viable way to gauge the incidence of the phenomenon in the general population, even when the responses were weighted to match the national population on gender, age, and student status. Non-probability panel sampling offered a considerably more accurate reflection of the general population, though it performed unevenly across the two non-demographic variables.

Online Surveys Versus a Non-Demographic Subpopulation

Even if online surveys tend not to accurately reflect the general population, could they be used to study non-demographic subpopulations? We saw that the river samples in particular attracted respondents that were disproportionately likely to report belonging to a subpopulation of affected individuals. This suggests that river samples could reach a sizeable sample of subpopulation members with a relatively small total n, decreasing survey cost. This would still be a non-probability sample, so any inferences drawn from it would be tentative at best. Regardless, river samples are now widely used to shed light on the characteristics of non-demographic subpopulations, so it is worthwhile to examine how they work.

The notion that river samples reach disproportionately large numbers of subpopulation members rests on the assumption that the high incidences they produce result from topical self-selection rather than from priming. If the higher incidences reported by the respondents are not due to self-selection but due to priming, then

Table 4. Facebook Ads and Total Clicks Across Three Campaigns

Ad Title (In Italics) and Text	Total Clicks
"*Are you worried?* Have you come across strange things on the net? Respond to a survey and win movie tickets!"	5,120
"*Feeling confused?* Have you found surprising paths online? Respond to a survey and win movie tickets!"	427
"*Feeling warm?* Have you found new interests on the net? Respond to a survey and win movie tickets!"	398
"*Feeling connected?* Have you found new friends online? Respond to a survey and win movie tickets!"	129

the survey is actually failing to reach the correct subpopulation. This would be a fatal problem even for research that uses river samples to draw tentative conclusions about a subpopulation's characteristics.

To examine this, we compared the characteristics of the cyberharassment victim subpopulations reached through the three online samples (comparable characteristics are not available in the benchmark data). The YLE sample can be assumed to be the most susceptible to priming due to its recruitment method, and the Facebook sample less susceptible. The SSI online panel sample should not be susceptible to cyberharassment-related priming at all, because the survey was marketed as a general survey on Internet use. The subpopulation samples are naturally smaller than our full samples, especially for SSI with its low incidence of cyberharassment victimization. But they are not atypical of subpopulation sample sizes in online survey research published in leading journals (e.g., $n = 209$ current smoker subpopulation sample in Dawkins et al., 2013).

We compared the samples on all attitudinal and behavioral characteristics available across the data sets that have been linked to or could plausibly be linked to cyberharassment victimization (Table 5). For instance, multiple studies have shown the overlap between offline and online victimization (Helweg-Larsen, Schutti, & Larsen, 2012; Keipi, Näsi, Oksanen, & Räsänen, 2017). We would expect "real" cyberharassment victims to differ on these characteristics from respondents who simply interpreted past experiences as cyberharassment due to priming. Yet, no statistically significant differences between the subpopulation samples were detected (two-sided *t* tests of means). Larger sample sizes could have revealed statistically significant differences, but the practical effect sizes would still be limited. For the purposes of online survey research with typical subpopulation sample sizes, there do not seem to be significant differences in characteristics between the river samples and the online panel sample. The river samples thus do not seem to be significantly biased by priming or similar effects in comparison to the online panel sample. They did, however, provide a higher subpopulation sample size in relation to the total number of respondents, making them potentially attractive for survey researchers hoping to reach non-demographic subpopulations at low cost.

Table 5. Characteristics of Cyberharassment Victims (Adjusted Means and Bootstrapped 95% Confidence Intervals [CI])

	Panel Survey (SSI), Mean (95% CI)	River Sample (FB), Mean (95% CI)	River Sample (YLE), Mean (95% CI)
Happiness	6.44 (6.04–6.84)	6.33 (6.08–6.58)	6.50 (5.94–7.06)
Economic satisfaction	4.49 (4.02–4.97)	4.91 (4.63–5.19)	4.71 (4.18–5.24)
Self-esteem	5.65 (5.18–6.12)	5.83 (5.53–6.12)	5.40 (4.78–6.03)
Generalized trust	4.31 (3.91–4.71)	4.26 (4.01–4.50)	4.11 (3.55–4.67)
Meeting friends	4.75 (4.47–5.04)	5.32 (5.15–5.49)	4.87 (4.55–5.19)
Offline harassment	1.16 (0.95–1.37)	1.24 (1.12–1.36)	1.16 (0.89–1.43)

Notes: SSI, $n = 103$; FB, $n = 355$; YLE, $n = 65$. Regression models adjusted for age, gender, Facebook usage, and student status.

Discussion and Conclusions

New information and communication technologies have opened up ways of collecting research data at a fraction of the cost and at many multiples of the speed of conventional social science survey research. According to Mayer-Schönberger and Cukier (2013, p. 30), "highly skilled survey specialists of the past [have] lost their monopoly on [...] empirical data." Survey research used to be the province of skilled specialists who often dedicated their careers to it, but with the rise of inexpensive online panels, tools like SurveyMonkey, and social media like Facebook, any graduate student or think-tank intern can now conduct surveys with little methodological know-how. According to this reading, the rapid growth of online survey research is thus the result of a democratization of a sort.

However, evaluated against the criteria of survey research theory and practice developed throughout the twentieth century, there is no doubt that non-probability online surveys fall short in representativeness. In our study, river samples recruited via Facebook ads and via a public broadcaster differed significantly from benchmark data provided by Statistics Finland. On demographic variables the differences ranged from 13.8 to 28.3 percentage points, and on non-demographic variables 15.3 to 38.9 percentage points, even after weights were applied. Demographic differences were expected, given the stark demographic disparities in how often people access different kinds of online services (Van Dijk, 2005). The non-demographic differences were likewise expected and consistent in direction with what we termed topical self-selection: people who are concerned with the topic of a survey are more likely to take the survey. Topical self-selection is logically associated with river sampling, as river sampling relies on attracting potential respondents' attention.

The non-probability online panel sample (SSI) fared better, deviating from benchmark data only 0.7 to 2.5 percentage points on demographics (partly expected as some demographics were used to construct quotas) and 2.3 to 14.4 points on non-demographics after adjustment. We expected that what we termed economic self-selection could potentially bias the panel sample on relevant demographic variables that were not used in quotas, given that panel respondents are essentially paid to participate. But the proportion of students in the sample differed from benchmark data by only 0.7 percentage points. This finding stands in slight contrast to earlier research, where economic self-selection was apparent (Weinberg et al., 2014; Willems et al., 2006). More detailed demographic variables might have revealed differences, however, and the unevenness of the non-demographic results nevertheless places doubts on this sample as a basis for generalizations to a national population.

Our river samples, in particular, were even more divergent from the general population than previous comparison studies have found (Chang & Krosnick, 2009; Yeager et al., 2011). We attribute this to the relative sophistication of the online sampling methods used in previous studies; for instance, the river sampling method in Yeager et al. (2011) was possible only with assistance from a popular Internet services provider, which is not available to every survey researcher. Our methods are more likely to reflect what the majority of non-probability online survey research today is like, with small budgets, fast turnaround times, and lack of privileged access (e.g., Dawkins et al., 2013; Martin, 2009; O'Brien, 2017). As suggested by Brüggen et al. (2016), our study also addressed questions about benchmark data quality by using a probability sample drawn from a comprehensive population registry and weighted to match the same (Statistics Finland, 2016). Overall, the findings suggest that even if traditional survey methods such as RDD no longer offer the performance they once did, they are still likely to be far superior to non-probability online surveys for estimating population means.

Online Surveys and New Research Practices

Must we conclude, then, that the explosion of online survey research is a misguided attempt to cut costs and satisfy impatient non-experts, and that it will end up retarding social sciences and misinforming policy? Not necessarily. We argue that it is important to consider not only how well non-probability online sampling performs in the role traditionally ascribed to survey research, but also what kinds of new research practices it may be enabling.

In this study, we considered the use of online surveys for exploratory research on emerging social and policy issues. Phenomena such as cyberbullying, online gig work, and e-cigarette use did not exist until recently. National population-level surveys to study such subpopulations can be very costly or result in small sample sizes if the incidence in the general population is low. Innovative use of online surveys can help. For instance, Lepanjuuri, Wishart, and Cornick (2018) used a combination of a national survey to estimate the incidence of gig work in the population, and an online survey to zoom in on gig workers' characteristics. Another problem with some national surveys is long turnaround times, potentially forcing policy makers to take positions on current topics without the benefit of any research evidence (probability online panels are better in this respect). Dawkins et al. (2013) used a river sample to produce a quick exploratory study, probing the characteristics of a new policy-relevant subpopulation. For these purposes, the fact that river samples attract disproportionately large numbers of subpopulation members can be an advantage, because it means that a smaller overall n is required. We compared subpopulations reached through river and panel samples and found that they were not statistically or practically different on relevant attitudinal and behavioral characteristics. This suggests that river samples can reach genuine subpopulation members rather than respondents primed to interpret their experiences according to the frame put forward by the researcher.

Of course, the conclusions that can be drawn from such samples are still limited, given that they are non-probability samples. They can be used to show that certain characteristics or phenomena exist (have non-zero probability), but the parameter estimates or percentages that the samples produce are unlikely to accurately represent the subpopulation. They can be used to make tentative inferences regarding demographic differences in subpopulation characteristics (e.g., female e-cigarette users were significantly more likely than male to have heard about e-cigarettes from a friend, in Dawkins et al., 2013), but the effect sizes will not be accurate, and could result from selection effects. In non-probability online survey research, the statistical focus thus shifts from traditional non-response bias to understanding and mitigating "response bias," such as topical self-selection in river sampling and economic self-selection in panel surveys. The safest use of such samples is in examining members of the sample itself, which can be worthwhile if the sample is large or comprehensive enough to be an interesting population (e.g., Sagar et al., 2016; though especially when studying contentious issues, care is needed to prevent river samples from being intentionally polluted by organized online groups). The rapid growth of online survey research depicted in Figure 1 is thus likely to be attributable to the use of online surveys for new research practices such as this, rather than to their direct substitution for population-level probability surveys.

Moving beyond our empirical study, another new research practice enabled by online surveys is combining self-report survey data with observational digital trace data or "big data" (González-Bailón, 2013). For instance, Lehdonvirta, Ratan, Kennedy, and Williams (2014) combined self-reported measures of gender and age with multiple observational measures of user behavior on two platforms. In practice, digital trace data consists of limited and unsystematic information and its coverage is constrained by the same first- and second-level digital divide issues as river sampling (Blank, 2017a, 2017b). But digital trace data and online survey data in combination may be able to reduce the burden on respondents and address measurement errors inherent to both traditional and online survey research, analogous to how conventional social surveys can sometimes be combined with administrative registry data.

While this study focused on the use of non-probability online surveys for estimating population means and examining non-demographic subpopulations, it is also common to use such surveys for other purposes, such as to conduct experiments. Coppock, Leeper, and Mullinix (2018) found that the population treatment effects estimated from survey experiments conducted with online convenience samples corresponded very closely with those estimated from nationally representative samples, despite significant differences in the samples' demographic compositions. They note that this is the expected result when an experiment's treatment effect heterogeneity is low, making online surveys potentially well suited for conducting such experiments.

Finally, the fact that online surveys reach different demographics than phone and postal surveys could be embraced by future survey researchers. If the twentieth-century ideal of reaching a nationally representative sample of respondents through a single medium of communication is increasingly unviable in the fragmented media environment of the twenty-first century, then new survey research practices based on multi-modal and blended samples could increasingly become the norm (Fang et al., 2013; Lorch et al., 2014). In such practices, online, mobile, and other sources of respondents are seen as complements rather than as substitutes to older channels. For now, though, conventional probability samples are likely to remain the most accurate sources of nationally representative data.

One of the limitations of this study is that we did not examine the potential effects of survey mode. While the effects of survey mode are complex, in general interviewer-based surveys are more susceptible to social desirability bias than are self-administered questionnaires, such as the online surveys analyzed in this study. Thus lower incidences in the Statistics Finland sample might for instance have been caused by respondents hesitating to disclose sensitive cyberharassment experiences to an interviewer. However, we do not think such mode effects are likely to explain the results. The Statistics Finland survey was conducted following best practices (Statistics Finland, 2016), including practices to address interviewer effects. Moreover, the river samples and the panel sample all used the same self-administered mode and yet the panel produced much lower incidences.

Another potential data quality issue that we did not examine was that attention-based and compensation-based surveys might differ in terms of respondent inattentiveness or propensity to satisfice (Chang & Krosnick, 2009; Willems et al., 2006). Respondents motivated by economic returns could be expected to satisfice more than respondents motivated by an intrinsic interest in the survey topic, disadvantaging panel samples compared to river samples. Finally, though our empirical study is situated in a national context of particular interest (Brüggen et al., 2016), generalizations to other countries must naturally be treated with caution.

Besides offering an empirical comparison of survey methods—including river samples, which have rarely been compared with other non-probability and probability modes—our article has contributed to theorizing the types of self-selection biases inherent to different forms of non-probability online surveys, and to identifying emerging research practices enabled by online surveys that augment the social and policy researchers' conventional toolkit.

Vili Lehdonvirta, Oxford Internet Institute, University of Oxford, Oxford, UK [vili.lehdonvirta@oii.ox.ac.uk].
Atte Oksanen, Faculty of Social Sciences, Tampere University, Tampere, Finland.
Pekka Räsänen, Department of Social Research, University of Turku, Turku, Finland.
Grant Blank, Oxford Internet Institute, University of Oxford, Oxford, UK.

References

Bethlehem, J.G. 2010. "Selection Bias in Web Surveys." *International Statistical Review* 78: 161–88.

Bethlehem, J.G. 2015. "Sunday Shopping—The Case of Three Surveys." *Survey Research Methods* 9: 221–30.

Bethlehem, J.G. 2016. "Solving the Nonresponse Problem With Sample Matching?" *Social Science Computer Review* 34 (1): 59–77.

Blank, G. 2017a. "The Digital Divide Among Twitter Users and Its Implications for Social Research." *Social Science Computer Review* 35: 679–97. https://doi.org/10.1177/0894439316671698

Blank, G. 2017b. "Online Research Methods and Social Theory." In *Handbook of Online Research Methods*, 2nd ed., eds.N. Fielding, R.M. Lee, and G. Blank. Thousand Oaks, CA: Sage Publications, 628–41.

Bosnjak, M., M. Das, and P. Lynn. 2016. “Methods for Probability-Based Online and Mixed-Mode Panels: Selected Recent Trends and Future Perspectives.” *Social Science Computer Review* 34 (1): 3–7.

Brüggen, E., J. Van Den Brakel, and J. Krosnick. 2016. “Establishing the Accuracy of Online Panels for Survey Research.” Report 2016|04, Statistics Netherlands, The Hague, The Netherlands.

Callegaro M., R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas eds. 2014a. “Online Panel Research: History, Concepts Application and a Look at the Future.” In *Online Panel Research: A Data Quality Perspective*. New York: Wiley, 1–22.

Callegaro, M., and C. DiSogra. 2008. “Computing Response Metrics for Online Panels.” *Public Opinion Quarterly* 72 (5): 1008–32.

Callegaro, M., A. Villar, D. Yeager, and J.A. Krosnick. 2014b. “A Critical Review of Studies Investigating the Quality of Data Obtained With Online Panels Based on Probability and Nonprobability Samples.” In *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas. New York: Wiley, 1–22.

Chang, L., and J.A. Krosnick. 2009. “National Surveys Via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality.” *Public Opinion Quarterly* 73 (4): 641–78.

Coppock, A., T.J. Leeper, and K.J. Mullinix. 2018. “Generalizability of Heterogeneous Treatment Effect Estimates Across Samples.” *Proceedings of the National Academy of Sciences* 115 (49): 12441–46.

Couper, M.P. 2000. “Web Surveys: A Review of Issues and Approaches.” *Public Opinion Quarterly* 64: 464–94.

Couper, M.P., E. Singer, F. Conrad, and R. R. Groves. 2008. “Risk of Disclosure, Perceptions of Risk, and Concerns About Privacy and Confidentiality as Factors in Survey Participation.” *Journal of Official Statistics* 24: 255–75.

Curtin, R., S. Presser, and E. Singer. 2005. “Changes in Telephone Survey Nonresponse Over the Past Quarter Century.” *Public Opinion Quarterly* 69: 87–98.

Dawkins, L., J. Turner, A. Roberts, and K. Soar. 2013. “‘Vaping’ Profiles and Preferences: An Online Survey of Electronic Cigarette Users: Vaping Profiles and Preferences.” *Addiction* 108 (6): 1115–25.

Dever, J.A., A. Rafferty, and R. Valliant. 2008. “Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?” *Survey Research Methods* 2: 47–62.

DiSogra, C., and M. Callegaro. 2016. “Metrics and Design Tool for Building and Evaluating Probability-Based Online Panels.” *Social Science Computer Review* 34 (1): 26–40.

Dreßing, H., J. Bailer, A. Anders, H. Wagner, and C. Gallas. 2014. “Cyberstalking in a Large Sample of Social Network Users: Prevalence, Characteristics, and Impact Upon Victims.” *Cyberpsychology, Behavior and Social Networking* 17 (2): 61–7.

Fang, J., C. Wen, and V. Prybutok. 2013. “The Equivalence of Internet Versus Paper-Based Surveys in IT/IS Adoption Research in Collectivistic Cultures: The Impact of Satisficing.” *Behaviour & Information Technology* 32 (5): 480–90.

González-Bailón, S. 2013. “Social Science in the Era of Big Data.” *Policy & Internet* 5 (2): 147–60.

Griswold, W., and N. Wright. 2004. “Cowbirds, Locals, and the Dynamic Endurance of Regionalism.” *American Journal of Sociology* 109 (6): 1411–51.

Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York, NY: Wiley.

Hargittai, E. 2002. “Second-Level Digital Divide: Mapping Differences in People's Online Skills.” *First Monday* 7 (4). http://firstmonday.org/article/view/942/864

Helweg-Larsen, K., N. Schutti, and H. Larsen. 2012. “Predictors and Protective Factors for Adolescent Internet Victimization: Results From a 2008 Nationwide Danish Youth Survey.” *Acta Paediatrica* 101: 533–39.

Hillygus, S.D., N. Jackson, and Y. McKenzie. 2014. “Professional Respondents in Non-Probability Online Panels.” In *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas. New York: Wiley, 219–37.

Keeter, S., C. Kennedy, M. Dimock, J. Best, and P. Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates From a National RDD Telephone Survey." *The Public Opinion Quarterly* 70 (5): 759–79.

Keipi, T., M. Näsi, A. Oksanen, and P. Räsänen. 2017. *Online Hate and Harmful Content: Cross-National Perspectives*. Abingdon & New York: Routledge.

Lavrakas, P. 2008. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage.

Lehdonvirta, V. 2018. "Flexibility in the Gig Economy: Managing Time on Three Online Piecework Platforms." *New Technology, Work & Employment* 33 (1): 13–29.

Lehdonvirta, V., R.A. Ratan, T.L. Kennedy, and D. Williams. 2014. "Pink and Blue Pixel$: Gender and Economic Disparity in Two Massive Online Games." *The Information Society* 30 (4): 243–55.

Lepanjuuri, K., R. Wishart, and P. Cornick. 2018. *The Characteristics of Those in the Gig Economy: Final Report*. London: BEIS.

Lorch, J., K. Cavallaro, and R. van Ossenbruggen (2014). Sample Blending: 1+1 > 2. Survey Sampling International. https://www.surveysampling.com/site/assets/files/1584/sample-blending-1-1-2.pdf

Lugtig, P., and V. Toepoel. 2016. "The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey." *Social Science Computer Review* 34 (1): 78–94.

Malhotra, N., and J.A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences About Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys With Nonprobability Samples." *Political Analysis* 15 (3): 286–323.

Martin, K.A. 2009. "Normalizing Heterosexuality: Mothers' Assumptions, Talk, and Strategies with Young Children." *American Sociological Review* 74: 190–207.

Mayer-Schönberger, V., and K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.

Näsi, M., P. Räsänen, A. Oksanen, J. Hawdon, T. Keipi, and E. Holkeri. 2014. "Association Between Online Harassment and Exposure to Harmful Online Content: A Cross-National Comparison Between the United States and Finland." *Computers in Human Behavior* 41: 137–45.

O'Brien, R. 2017. "Redistribution and the New Fiscal Sociology: Race and the Progressivity of State and Local Taxes." *American Journal of Sociology* 122 (4): 1015–49.

Olivier, L. 2011. "River Sampling Non-Probability Sampling in an Online Environment." Center for Information-Based Decision Making and Marketing Research. http://lexolivier.blogspot.com/2011/11/river-sampling-non-probability-sampling.html

Petzold, K. 2017. "Cosmopolitanism Through mobility: Physical-Corporeal or Virtual-Imagined?" *British Journal of Sociology* 68 (2): 167–93.

Räsänen, P. 2006. "Consumption Disparities in Information Society: Comparing the Traditional and Digital Divides in Finland." *International Journal of Sociology and Social Policy* 26 (1/2): 48–62.

Sagar, T., D. Jones, K. Symons, J. Tyrie, and R. Roberts. 2016. "Student Involvement in the UK Sex Industry: Motivations and Experiences." *British Journal of Sociology* 67 (4): 697–718.

Scherpenzeel, A., and J. Bethlehem. 2010. "How Representative are Online-Panels? Problems of Coverage and Selection and Possible Solutions." In *Social Research and the Internet: Advances in Applied Methods and New Research Strategies*, eds. M. Das, P. Ester, and L. Kaczmirek. New York: Routledge Academic Press, 105–32.

Son, S., A. Khattak, and N.-K. Kim. 2013. "Noncoverage Errors in Travel Surveys Due to Mobile Phone-Only Households." *Transportation Research Record* 2354: 29–39.

Sparrow, N. 2007. "Quality Issues in Online Research." *Journal of Advertising Research* 47 (2): 179–82.

Statistics Finland. 2016. "Väestön tieto- ja viestintätekniikan käyttö." http://www.stat.fi/til/sutivi/meta.html

Van Dijk, J. 2005. *The Deepening Divide: Inequality in the Information Society*. London: Sage.

Walsh, J.P., S. Kiesler, L.S. Sproull, and B.W. Hesse. 1992. "Self-Selected and Randomly Selected Respondents in a Computer Network Survey." *Public Opinion Quarterly* 56: 241–244.

Weinberg, J., J. Freese, and D. McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsource-Recruited Sample." *Sociological Science* 1: 292–310.

Willems, P., R. van Ossenbruggen, and T. Vonk. 2006. The Effects of Panel Recruitment and Management on Research Results: A Study Across 19 Online Panels. ESOMAR Panel Research.

Yeager, D.S., J.A. Krosnick, L. Chang, H.S. Javitz, M.S. Levendusky, A. Simpser, and R. Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted With Probability and Non-Probability Samples." *Public Opinion Quarterly* 75 (4): 709–47.

Yun, G.W., and C.W. Trumbo. 2000. "Comparative Response to a Survey Executed by Post, E-mail, & Web Form." *Journal of Computer-Mediated Communication* 6 (1). https://doi.org/10.1111/j.1083-6101.2000.tb00112.x JCMC613

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website.