SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 2115-2126
Open Access. Online & Print

DOI: 10.5281/zenodo.11425255

RIVER SAMPLES FOR DEVELOPING ONLINE SAMPLES: A SPIRIT OF RANDOMNESS TECHNIQUE

Mohamed.M.Ibrahim[1*], Fadwa.H.Mohamed[2] and Abdel Hamid,Talaat Asaad[3]

[1]Prof. of Marketing, Faculty of Management & Buisness Intellegence, Modern University for Technolog & Information, Cairo, Egypt. Mohamed_mahmoud2k@yahoo.com, https://orcid.org/0000 -0002-7933-2678
[2]Lecturer, Faculty of Management Business Intelligence, Modern university for Technology& Information, Cairo, Egypt Fadwahassan9710@gmail.com, https://orcid.org/ 0009-0008- 0222-950X
[3] Prof. of Marketing, Faculty of Commerce, Mansoura University, Mansoura, Egypt,tasaad44@gmail.com,https://orcid.org/0000-0001-9006-0477

Received: 11/11/2025
Accepted: 18/12/2025

Corresponding Author: Mohamed.M.Ibrahim (Mohamed_mahmoud2k@yahoo.com)

ABSTRACT

The purpose of this paper is to propose an alternative methodological approach to establish a connection with reality to address one of the most common issues that researchers encounter when choosing a probabilistic sample using online sample development techniques. To achieve this, one sample type used in online surveys the river sample is modified and its surrounding ambiguity is eliminated, according to the identification of the main causes of survey error, which include sampling error as well as nonresponse, coverage, measurement, and representativeness. These sources of error defy the fundamental presumptions of probability sampling, So, by addressing these errors, we can provide a safe and accurate way to generalize the findings by infusing a spirit of randomness, especially since probability sampling is a technique rather than the ideology for randomization.

KEYWORDS: River Sample, Sources Of Errors, Propensity Score.

1. INTRODUCTION

Entering the twenty-first century, the internet is profoundly influencing the survey research industry, much like it is transforming various other aspects of business (Walsh, 1992). In 1996, the first papers about Web surveys were published.

Since then, interest in using the Internet in general and the World Wide Web in particular as a tool for gathering survey data has virtually exploded (Couper &Miller 2008). The rapid growth of web-based surveys has led some to predict that internet surveys may soon replace traditional survey data collection methods (Couper et al 2015). In line with the importance of data collection, which must be responsive to people's changing lifestyles, choices, and attitudes (Goot 2021; Kennedy et al. 2018; Wang et al. 2015). However, Gosling et al. (2004) disagree that Internet-based methods should always replace conventional methods. Instead, there is room for both, and researchers should choose the one that best fits their objectives (Callegaro et al. 2015).

For many years, the standard survey method has been telephone surveys with random-digit dialing. However, there are a number of reasons for this, such as its high cost and the fact that it no longer accurately represents the research population (Baker et al. 2010). In addition to the low response rate—which could be caused by the survey respondent's inconvenient contact time—people who exclusively use mobile phones might not be contacted, so the hunt for more efficient techniques has begun (Keeter et al 2017).

Web-based techniques provide an alternative. Respondents to web-based surveys access a website to submit their answers (Ebert et al 2018). The primary benefits of web-based surveys are their low cost, which comes from not requiring employees to administer the questionnaire or record the results, and their great flexibility, which comes from allowing respondents to respond from anywhere at any time (Wright et al 2005). The past few years have seen a sharp rise in their usage.

Callegaro et al (2015) define web survey as a survey mode using computerized self-administered questionnaires, stored on a specific computer connected to the Internet (i.e. server), which respondents' access via a web browser.

While the terms "Internet surveys" and "Online surveys" are often used interchangeably. Callegaro et al. (2015) point out that the Internet surveys is broader than web surveys and the term "online surveys" generally refers to a process that is more comprehensive than an Internet survey because it deals with computerized survey questionnaires, where data exchange between computers is facilitated by an electronic network (i.e. between a researcher's server and the respondent's device).

Additionally, besides the Internet, other telecommunication networks like fixed telephone networks can also be utilized for conducting online surveys (tele panel surveys, for example) (Regmi et al 2016). The other factor has to do with the computerized questionnaires that are automatically sent out, allowing respondents to send their answers to researchers by simply pressing a button (Vehovar & Manfreda 2017).

Groves et al. (2009) identify various formats of computerized self-administered questionnaires, including disk-by-mail (DBM), telephone self-interviewing, electronic mail surveys (EMSs), email surveys, web surveys across multiple devices, and mobile survey applications. The potential of web surveys for survey researchers is already substantial and is expected to grow even further in the future (Evans & Mathur 2018).

Web surveys will be evaluated for four main sources of errors: sampling, coverage, non-response, and measurement, beside the concerns of representativeness of sample surveys, where the survey industry faces the challenge of conducting research in the presence of these errors that are examples of practical issues that violate the pure assumptions of probability sampling(Brick 2011).

In this context, we note the statement made by Gosling et al. (2004) that they do not think that Internet-based techniques should always take the place of traditional techniques. Rather, there is a room for both, and researchers ought to select the one that best aligns with their goals. There will be suggestions for resolving these issues, considering the error properties and quality data indicators of different Web-based data collection methods.

2. LITERATURE REVIEW

Contemporary research on online sampling highlights both the expanded accessibility enabled by internet-based data collection and the methodological complexities associated with representativeness and inferential rigor. Scholars generally agree that while online sampling introduces efficiency and wider demographic reach, the absence of fully controlled selection mechanisms challenges the foundational assumptions of probability-based inference (Kalton, 2023; Rahman et al., 2022). Accordingly, methodological debates have increasingly focused on understanding how different online recruitment strategies align with or diverge from classical sampling principles.

Four primary online sampling structures are widely recognized in the literature: emailing-list recruitment, online access panels, mixed-mode probability panels, and river sampling (Lehdonvirta et al., 2021; Burns & Veeck, 2020; Callegaro et al., 2015). Email-based recruitment relies on predefined sampling frames but remains sensitive to coverage and frame error due to list quality and restricted access (Couper et al., 2015). Online panels, while scalable and demographically rich, frequently depend on nonprobability volunteer recruitment, introducing self-selection bias and uncertainty in inclusion probability, despite operational attempts to approximate probabilistic representation through demographic matching and controlled quotas.

Mixed-mode probability panels integrate traditional probability sampling procedures—such as address-based sampling and random-digit dialing—with online survey administration, producing more reliable inference by expanding coverage and supporting structured recruitment pathways (Dillman et al., 2014; Couper et al., 2017). However, these approaches require substantial cost and logistical coordination, limiting broad adoption relative to nonprobability alternatives.

River sampling has emerged as a dynamic real-time recruitment framework that intercepts online users while they navigate websites, advertisements, or social platforms (Poynter, 2010; McPhee et al., 2022). Unlike panels that reuse the same pool of respondents, river sampling relies on continuously shifting streams of visitors and thus claims closer alignment with random selection principles. Proponents argue that when recruitment invitations are uniformly displayed and widely distributed, the technique can approximate probabilistic behavior more closely than conventional opt-in samples (Burns & Veeck, 2020). However, critics emphasize that the absence of a fully defined sampling frame and unknown inclusion probabilities restrict the degree to which generalizable inference can legitimately be claimed (Baker et al., 2013). This methodological tension underscores the need for conceptual clarification and empirical validation—motivating the present study's focus on river sampling as a probability-like framework rather than a fully probabilistic one.

The Reasons Underlying Using Online Surveys:

Online surveys are attractive for a variety of reasons, according to the literature review:

(1) they collect data quickly; (Baker et al 2010; Garton et al., 1997) (2) they are expected to be cost reduction than most other methods; (Regmi et al2016; Llieva et al., 2002) (3) their extensive profiling leads to efficient sampling; (Chang & Krosnick 2009) (4) there's no need to code. As a result, statistical analysis can be done on the results almost instantly (Matthias & Couper 2017) (5) the ability to collect data from hard-to-reach samples (Magnani et al 2005) (6) possibility of using the improved graphic and animation capabilities of the web (Callegaro et al. 2015) (7) eliminating interviewer error and bias; (Bound & Mathiowetz 2001) (8) give researchers much more control over data quality; (Jaeger & Cardello 2022) (9) no intermediaries (Couper 2015) (10) communication is asynchronous, in other words dealing with messages is convenient for the user (Fox et al., 2001; Nie & Erbring2000).

2.1. Linked Issues Related To Online Surveys

The majority of researchers understand that the primary causes of survey error are sampling, nonresponse, coverage, and measurement error (Groves 1989). These factors also serve as the basis for many data quality metrics, so each one must be taken into account in relation to costs. It is imperative to address the implications of these multiple sources of error for Web surveys in addition to representativeness issues in order to be confident in generalizing the results (Couper 2000).

Then, McPhee et al (2022) indicate that work toward reducing the negative effects of these errors on the quality of Web surveys while concurrently improving their levels of objectivity and accuracy. In the other words the main challenge is to address and overcome the issues that arise when conducting online surveys.

The first challenge: is sampling error, which happens when a sample is a subset of the population and is not a true representation of it. It is any error in a survey caused by a sample (Wyner, 2007). Sampling bias occurs when the sample is not representative of the population (Baker et al 2013). An over- or under-representation of a particular group may result from the sample selection process favoring or excluding specific categories of people. Finding respondents on the frame is necessary because sampling error happens when a sample is selected from the frame population. According to Couper (2000), there are two significant issues with Web surveys. The first problem is that not everyone in the target population is represented within the frame population, meaning some individuals may be excluded from participation. The second problem involves the challenges of constructing an accurate sampling frame for Web surveys, which complicates the process of selecting a probability sample from the larger population.

According to Baker et al (2013) sampling error can be mitigated by distributing surveys across various online platforms, allowing access to a broader and more diverse group of respondents while still targeting specific demographics. Social media also reflects the demographics and characteristics of target respondents, along with screening questions that they need to respond to in a certain way to enter the survey. In contrast to the standard random-digit-dial phone survey, where, diversity is an important feature of web surveys (Couper and Miller, 2008). Moreover, researchers on the internet can take action to remove repeat responders Gosling et al. (2004).

The second challenge: Nonresponse refers to the failure to gather responses from eligible participants included in a sample (AAPOR, 2009). Using non-probability samples according to Baker et al (2013), researchers should avoid the term "response rate" and instead use a different term. ISO 20252: 2008 defines the term "participation rate" as "the number of respondents who have provided a usable response divided by the total number of initial personal invitations requesting participation" (ISO, 2008).

When some of the sample's respondents are unwilling or unable to finish the survey, nonresponse error takes place. It is dependent upon the nonresponse rate as well as the variations in the variables of interest between respondents and nonrespondents (Groves and Couper 1998).

Lower response rates could be attributed to the need for increased response-stimulating efforts in electronic Web surveys (Dillman 2000), as well as technical difficulties interacting with an Internet survey, which may discourage some from completing the survey.

Also, Callegaro et al. (2015) Point out there are general characteristics effect on nonresponse, where the researcher's control is either indirect (questionnaire design) or relatively weak (sponsor, salience).

Therefore, direct nonresponse interventions pertaining to incentives, the method of contact, and invitation format must be continued by researchers.

In this direction Wu et al (2022) indicates that online surveys aimed at the general public or large groups need to find survey respondents in some way, like by wide publicizing the survey. Additionally, Gosling et al. (2004) affirmed that Internet methods provide means for motivating participants (e.g., feedback) so, the participants will be volunteers who are motivated and interested in accessing and completing the survey, which will enhance the accuracy level.

According to Rivers (2013), it is worth noting that the response rate alone cannot reveal whether and how the sample is biassed, because it is difficult to predict whether respondents It is necessary to understand the differences in variables that may be related to survey responses (Groves & Couper, 1998), rather than random issues (Lavrakas 2008; Groves, 1989).

To improve the response rate of web surveys, Dillman et al. (2009) proposed using Internet panels, which consist of thousands of people recruited through banner ads, pop-up ads, or e-mail advertisements, for probability sampling.

Meanwhile, according to Sekaran & Bougie (2016) nonresponse rates from electronic surveys might not be any lower than those from more conventional techniques, like mail surveys, for instance.

Future growth in electronic questionnaire administration is anticipated given the rise in computer literacy (Sekaran &Bougie 2016).

The third challenge: Coverage error is a function of the mismatch between the target population (The respondents from which the researcher draws inferences) and the frame population (The set of people for whom a listing can be done before choosing a sample) (Couper et al 2015; Groves 1989; Tsao1983). Notably, Burns & Veeck (2020) indicate that the sample frame shape includes some of the area outside the population boundary, but not the entire population shape. Stated differently, there is not always a perfect match between the sample frame and the population.

According to Alvarez & VanBeselaere (2005), non-coverage refers to the situation where a considerable part of the eligible units is excluded from the sample.

The most significant concern is that there isn't a comprehensive list of email addresses that can be used as a sampling frame for online surveys of the general population as a whole. Put another way, don't usually give out participant email addresses. Even if there were, it wouldn't include a sizable portion of the population. In other words, the coverage problems in such a design are those related to the sampling frame (e.g., RDD) (Schonlau and Couper 2017).

According to Callegaro et al. (2015), web surveys are not list-based, but rather probability web intercept surveys in which randomly selected visitors to a specific website are invited to take a web survey, typically via a pop-up window.

Although many researchers, Sekaran and Bougie (2016) think there is a lack of an adequate sampling frame for online surveys, many online survey invitations are published as links on websites or social media, resulting in reducing coverage error.

Some argue, Burns & Veeck (2020) that a river sample is an objective approach to overcoming this challenge. The sample frame of a river sample is the stream of visitors to the site issuing the invitation.

In other words, the frame consists of site visitors, addressing coverage issues, while cookies are used to prevent sending multiple invitations to the same individual (Couper 2000). As a result, it can be classified as a probability sample.

Additionally, randomization in online recruitment, achieved through pop-up invitations, allows for recruitment to be spread over various days or hours, making it comparable to probability web intercept surveys. (Berzofsky,2019, Wu et al, 2022 & Biemer, 2010). However, these approaches also result in improved statistical inference (Steinmetz, Bianchi, Tijdens and Biffignandi, 2014).

On other hand, Ferri García et al. (2020) found that combining propensity score adjustment with calibration significantly enhances coverage bias removal compared to calibration alone, though using population totals from a reference survey does not notably improve estimate accuracy and may slightly increase variance. Additionally, Gosling et al. (2004) noted that while Internet-based findings generally align with traditional research methods, more data is necessary for confirmation. Propensity score adjustment, initially introduced by Rosenbaum and Rubin (1983) as a post-hoc technique in observational studies, is designed to balance covariates between comparison groups and mitigate the confounding effects of selection mechanisms, effectively addressing bias in both probability and non-probability samples (Baker et al. 2013).

The fourth challenge: Measurement error can be defined as the difference between the respondents' answers and their true values on the measure (Callegaro et al. 2015). Four common sources are identified as the origin of these errors: the questionnaire, the respondent, the interviewer (if any), and the method of data collection. Both probability and non-probability samples are covered by this (Baker et al 2013).

In self-administered surveys, the survey instrument or respondent could be the source of measurement errors, as there is no interviewer to act as an intermediary (Couper et al 2015). For the sake of this, authors offer to use conversational chatbot surveys instead of a passive form of questionnaires, where Kim, S., Lee, J., & Gweon, G. (2019) revealed that using conversational chatbot surveys improves data quality and could partially fulfill a human interviewer's role by using effective communication strategies (called virtual interviewer).

In addition, Web surveys may benefit greatly from design due to the abundance of tools available to designers (color, sound, images, animation, etc.) that are not available in traditional design methods, as well as the potential for respondent misinterpretation. It is crucial to remember that the kind of Web survey being conducted and the target population may have an impact on the design (Andrews et al 2007).

Panel conditioning (or time-in-sample bias) is another source of measurement error that is specific to panel or longitudinal surveys (which are frequently used in Web surveys). Panel conditioning is achieved through ongoing participation in a panel (Kalton and Citro 1993). Given their experience with the survey over time, their responses may begin to differ from those of people taking the survey for the first time. This defect is avoided by recruited Internet panels via banner ads and other open solicitations with a random spirit, such as river samples. The fifth challenge: Concerns of representativeness of sample surveys. A subset of data that accurately reflects the features of the broader population from which it was taken is referred to as a representative sample (Fan 2011). Assuming that the sample is representative of the population as a whole, it permits drawing conclusions about the population from the sample.

Kumar et al. (2018) argue that online research can be representative, challenging the notion that it isn't due to limited internet usage. They compare it to traditional methods like random-digit dialing for telephone samples and mall studies, which are considered representative of their respective demographics.

The authors suggest that, similarly, online research samples can be representative of their target populations, depending on the methodology used for sample generation.

In some cases, involving volunteer samples, weighting methods like propensity score adjustment are used to correct for representational biases (Lee, 2006). Its implementation is driven by equality (Lenau et al., 2021). In this context, Hair et al. (2021) demonstrate that respondent weighting can improve both the sample's representativeness and generalizability.

Furthermore, evidence indicates that online and traditional research methods often yield comparable results in both quantitative and qualitative studies. According to Kumar et al. (2018), the web, like any central location, simply serves as a gathering place; it can be used to research subjects that would otherwise be too costly to interview, such as employing a number of different cities to address the issue of how well the shopping center sample is representative.

In addition, an opt-in web-based sampling technique, similar to river sampling, is used. According to Bakr et al. (2013), more formal designs in sample selection have been used to improve representativeness.

3. PROPENSITY SCORE ADJUSTMENT AS SOLUTION FOR SOURCES OF ERROR

Biases in online samples can be reduced by adjusting the samples to reflect online sample characteristics, where reassigning sample weights addresses imbalances in the sample with respect to population targets, including age, gender, and race, etc. (Roshwalb et al 2016), also Fahmi et al (2024) proposed that the composite weighting method offers a practical, robust solution for the efficiency of internet surveys.. Since in the absence of a traditional frame, weighting procedures are used to construct estimates that assign greater weight to respondents from low response rate groups and less weight to those from high response rate groups, thereby closing the gap between the sample and the respondents (Baker et al 2010).

Moreover, Gosling et al. (2004) found that Internet users do not significantly differ from non-users in terms of adjustment and depression. The statistical solution to the selection problem was introduced by Caliendo and Kopeinig (2008). It has been primarily used for weighting adjustment in online panels, telephone, mail, and in-person surveys (Lee and Valliant 2009; Smith 2000; Taylor 2000) in survey statistics. To account for nonresponse, this popular quasi-randomization technique is widely employed (McPhee et al 2022). Moreover, to improve efficacy or address any possible biases caused by nonresponse and coverage errors (Baker et al., 2013), as well as to modify the volunteer sample's distribution to a comparable survey using probability techniques (Lee, 2006).

In this direction, Hair et.al (2021) stated that propensity scoring can be used to adjust the results to look more like those produced by a representative sample, but the procedure's accuracy must be evaluated. To account for sampling inadequacies, propensity scoring weights responses from underrepresented sample members more heavily.

For example, if respondents aged 65 and up are only half as likely to be in an Internet sample as they are in the general population, each senior will be counted twice.

Vehovar et al. (1999) suggested that propensity weighting could improve estimates when facing challenges such as nonresponse, non-coverage, and selection bias, which represent departures from ideal probability sampling. Building on this, Steinmetz et al. (2014) proposed propensity score weighting as a method to address these issues. Such challenges, however, are not exclusive to online surveys.

In the interim, the authors suggest that if the demographic characteristics of respondents on social media platforms are not disclosed, the relative distribution of demographic factors at the research population level (offline) is determined and compared to the active users of the research population on social media platforms (online). The proportional distribution of the research sample is then calculated.

3.1. Procedures For Implementing River Sample To Reveal The Spirit Of Randomness

Taking into account the existing research on the application of river sampling techniques, and in order to bring them closer to the spirit of randomness, other than the statistical calibration that was previously discussed in detail, the researchers propose the following set of procedures to serve this vision:

1. Selecting the websites / platforms based on the statistics (site's viewers and their response patterns, in other words check out everything from demographics and location to behavior and interests by under Audience) available on the website's / platform's visits in order to determine the most traffic rivers of subjects, according to Bock &Van den Poel (2010).
2. Sending out invitations over a long period of time, following up on them, and considering flexible response times, in other meaning usual duration, according to Dillman et al. (2014)
3. Stable and transparent recruitment mechanism. That means the procedures governing banner placement and invitation timing should be clearly defined, consistently reproducible, and insulated from commercial interests or algorithmic optimization that could introduce bias (Groves et al. 2014).
4. Instead of focusing on unusual or exceptional physical traits, the selection criteria for sampling sites should be representative and consistent, based on regular intervals, according to Groves et al. (2009)
5. Determine the data required for a specific entity, also known as relevance, according to Burns and Veeck (2020).
6. Understanding the target audience for each website and how visitors behave when they

visit is critical for effective recruitment, according to (Callegaro et al. 2015).

7. Post banners/ads/links or any other invitation method on multiple sites/ platforms/apps related to the research project and target audience according, to (Kaushik 2009) If banners or invitations are delivered randomly (say, randomized when and to whom they appear), this adds randomness in selection.
8. Using survey's screener section first, and if they meet the requirements there, they are then routed to the questionnaire section, according to Saris & Gallhofer (2014).
9. Using propensity score weighting to solve issues like nonresponse, non-coverage, and selection bias, which are deviations from ideal probability sampling (Hair et.al 2021; Steinmetz et al. 2014; Bethlehem2010 ; Vehovar et al. 1999)

4. DISCUSSION

River sampling incorporates several design features—such as wide intercept recruitment and the distribution of invitations over time—to enhance randomness and mitigate selection bias. Nonetheless, important methodological constraints remain inherent to this approach. Most notably, the absence of a clearly defined sampling frame makes it impossible to calculate precise inclusion probabilities for individual participants. Accordingly, although river sampling offers clear advantages over conventional opt-in samples, it still falls short of the inferential robustness associated with classical probability sampling.

Moreover, the fluid and self-selective character of digital environments can generate unobserved biases linked to respondents' motivations, patterns of platform use, and unequal access to online infrastructures. These sources of distortion restrict the scope of valid generalization and require that study results be interpreted with caution, rather than presented as fully representative of a target population. In this sense, findings derived from river sampling should be understood as yielding structured and informative evidence that can approximate probabilistic reasoning, while not supporting unqualified population-level claims.

Conceptualizing river sampling as a probability-like method captures this dual reality: it recognizes both the methodological strengths of the design and the limits of its inferential power. This framing underscores the need for explicit acknowledgment of underlying assumptions and for realistic expectations about what kinds of empirical conclusions can legitimately be drawn.

From this perspective, river sampling occupies an intermediate position between traditional probability sampling and unrestricted nonprobability approaches. Although it uses organized recruitment mechanisms intended to mimic random selection—such as broad recruitment streams, temporal dispersion, and multiple digital entry points—unknown inclusion probabilities, imperfect coverage, and self-selection processes continue to prevent full alignment with probability theory. For this reason, the present study employs the term probability-like to denote a methodological stance that resides between these two poles.

This label reflects both the promise and the constraints of the approach. When river sampling is combined with post-survey adjustment techniques—such as calibration weighting or propensity score–based corrections—it can yield more stable and broadly applicable estimates than standard volunteer opt-in samples. However, the validity of these gains depends on clearly stated assumptions and should be accompanied by cautious interpretation rather than claims of complete representativeness. Thus, probability-like sampling is best understood as a framework that supports structured, qualified inference, while openly acknowledging the partial and approximate nature of its probabilistic foundations.

5. CONCLUSIONS/RECOMMENDATIONS

The previous review made an effort to connect with reality in order to address one of the most significant problems that researchers face when choosing a probabilistic sample through online sample development methods. These methods offer advantages in terms of cost and technology (Bellhouse 2000; Frankel and Frankel 1987), as well as the flexibility to adapt to environmental pandemics that offline traditional sample methods are unable to provide, wherein social research benefits greatly from the infrastructures of such online surveys (Cornesse et al 2022). Moreover, there are several technical issues with traditional surveys that prompt practitioners to pursue alternative tactics, such as online surveys (Watters et al 2023). These factors significantly contributed to the widespread use of online surveys. Therefore, in order to ensure the accuracy of the conclusions and allow the results to be generalized, it was necessary to look for a probabilistic method for online samples that is flexible, quick to spread, and has a suitable level of accuracy; and the most important, it embodies the spirit of randomness.

The concepts of 'spirit of randomness' and 'river sampling' refer to the practice of recruiting respondents from the general stream of internet traffic, without pre-screening or restricting participants to a predefined panel. This approach aims to enhance randomness and representativeness by capturing a broad and diverse range of individuals as they naturally encounter the survey invitation.

In this context Couper (2000) stated that it is extremely difficult to conduct accurate and unbiased traditional opinion polls, and that this scenario does not exist in reality. Unlike traditional polls, online polls are incapable of bias, and any respondent who is able to participate has the opportunity, often these surveys have no access restrictions. The only bias in online surveys is demographics is that the respondents have access to the Internet's World Wide Web. According to Kocar and Baffour (2023), just because a sample's selection probabilities are unknown doesn't mean they can't be estimated or adjusted even for in a nonprobability sample.

The authors concur with Couper 's (2000) findings in this context, as research indicates that online surveys receive higher response rates than traditional opinion polls. Additionally, obtaining a sample frame is a major challenge because of issues related to confidentiality, law, industry practices, or even competition.

Furthermore, as alternative data sources such as web surveys and big data become more accessible and digitized (Lenau et al., 2021), the authors believe that new opportunities for quasi-probability or pseudo-probability sampling structure (AAPOR 2013), have been opened up.

Since all of the earlier problems have a detrimental effect on the results' accuracy and bias, randomness in the strictest statistical sense is practically nonexistent. The closest approach to achieving the advantages of probability samples may be to maintain a spirit of randomness, as in intercept or river samples. However, this approach requires further work addressing the issues that limit the probability element of these samples (Burns & Veeck 2020; Poynter 2010; Couper 2000).

In fact, the use of online samples has increased dramatically over the past decade, owing to the significant impact of internet technology on data collection options for market research (Bellhouse 2000; Frankel and Frankel 1987). While this method does not involve panel development, it has gained popularity for creating online samples, drawing from deeper and more sustainable sources like expanded river sampling and social networks (Watters et al. 2023).

There is no doubt that additional experiments and scientific research will be required for river recruitment to succeed in replacing it. However, it is unavoidable that this will be the way forward and the eventual winner (Olivier 2011). In this context Kish (1965) showed that for randomization, probability sampling is a technique rather than a dogma, especially when dealing with large numbers.Lastly it is incorrect to write off all Web surveys due to the irrational claims made by a small number of people. In a similar vein, it is unrealistic to think that as the method advances, no significant embarrassing moments will arise (Couper et al 2015; Gosling et al 2004).

Moreover, Fahimi et al. (2015) argue that greater tolerance for such alternatives is unavoidable, especially as the main pillars of traditional methods methodology begin to crumble.

The authors attempted to address issues concerning the sources of error, in addition to the use of statistical adjustments such as weighting methods, specifically propensity score adjustment, which result in more accurate estimates and reliable statistical inferences, as well as suggesting a set of procedures to address the assumption of randomness. Aside from being excited about designing and creating online surveys and their benefits.

Given this, there is a pressing need to develop more reliable sampling techniques using a multi-sourcing or dynamic sourcing model, as online sample providers increasingly rely on various sources beyond their proprietary panels, including social networking sites and general survey invitations placed on multiple websites across the Internet. To improve efficiency or coverage, units are chosen from two or more frames in multiple frame sampling (Brick 2011), which is what river samples provide.

This article adds value to the existing literature on approaches of the online survey by addressing one of the most significant issues that researchers face when dealing with probability samples via the Internet: assessing the river sample method in light of sources of error to improve causal inference, results generalization, and bias mitigation in online panels. As a result, it provides useful guidance to survey practitioners using higher-quality probability online panels, as well as survey researchers interested in implementing other types of online surveys.

5.1. Implications For Scientific Culture And Society:

The methodological developments presented in this study do more than refine statistical practice; they also intersect with broader questions about the nature and health of scientific culture. As empirical research becomes increasingly dependent on digitally mediated modes of data collection, the quality and credibility of online sampling procedures gain critical importance for both scholarly rigor and public confidence in research findings. Improving the transparency and inferential robustness of river sampling—especially through structured quasi-random recruitment and carefully implemented post-survey adjustments—helps to uphold core scientific values such as replicability, fairness in representation, and accountability to empirical evidence.

The enhancement of nonprobability online sampling methods thus has implications that extend into wider debates about the democratization of knowledge production. When online recruitment is designed to be more inclusive and systematically regulated, it creates opportunities for individuals and groups who are often missing from conventional survey frameworks to participate in research. This, in turn, supports a more balanced distribution of epistemic authority and mitigates forms of exclusion that frequently affect marginalized or digitally peripheral populations.

At the same time, strengthening online sampling practices contributes to maintaining public trust in scientific evidence, particularly in an era marked by misinformation, doubts about data-based conclusions, and skepticism toward research institutions. Showing that rigorous, carefully qualified inference can be achieved even within nonprobability designs—by making assumptions explicit, acknowledging limitations, and applying robust corrective procedures—reinforces the role of science as a credible and socially responsive institution.

From this perspective, the introduction of probability-like properties into river sampling should be viewed not solely as a technical refinement, but as a deliberate contribution to nurturing a scientific culture that is simultaneously methodologically innovative, ethically grounded, and socially engaged.

REFERENCES

Alvarez, R. M., & VanBeselaere, C. (2005). Web-based surveys. Encyclopedia of social measurement, 3, 955-962.

Andrews, D., Nonnecke, B., & Preece, J. (2007). Conducting research on the internet: Online survey design, development and implementation guidelines.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., ... & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. Journal of Survey Statistics and Methodology, 1(2), 90-143.

Ball, H. L. (2019). Conducting online surveys. Journal of human lactation, 35(3), 413-417.

Bellhouse, D.R. (2000). Survey sampling theory over the Twentieth Century and its relation to computing technology. Survey Methodology 26: 11 – 20.

Berzofsky, M. E., Langton, L., Krebs, C., Lindquist, C., & Planty, M. (2019). Methods for improving representativeness in a web survey on sexual assault among college students. Journal of interpersonal violence, 34(23-24), 4838-4859.

Bethlehem, J. (2010). Selection bias in web surveys. International statistical review, 78(2), 161-188.

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. Public opinion quarterly, 74(5), 817-848.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In Handbook of econometrics (Vol. 5, pp. 3705-3843). Elsevier.

Brick, J. Michael. 2011. "The Future of Survey Sampling." Public Opinion Quarterly 75(5):872–88.

Burns, A. C., & Veeck, A. (2020). Marketing research. Pearson.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). Web survey methodology. Sage.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. Public Opinion Quarterly, 73(4), 641-678.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. Public Opinion Quarterly, 73(4), 641-678.

Cornesse, C., Felderer, B., Fikel, M., Krieger, U., & Blom, A. G. (2022). Recruiting a probability-based online panel via postal mail: experimental evidence. Social Science Computer Review, 40(5), 1259-1284.

Couper MP, Kelley J, Axinn W, Guyer H, Wagner J, West BT. (2015). Internet and smartphone coverage in a national health survey: implications for alternative modes. Presented at Fed. Comm. Stat. Methodol. Res. Conf., Dec. 1–3, Washington, DC

Couper, M. P. (2000). Web surveys: A review of issues and approaches. The public opinion quarterly, 64(4), 464-494.

Couper, M. P. (2017). New developments in survey data collection. Annual review of sociology, 43(1), 121-145.

Couper, M. P., & Miller, P. V. (2008). Web survey methods: Introduction. Public opinion quarterly, 72(5), 831-835.

De Bock, K., & Van den Poel, D. (2010). Predicting website audience demographics for web advertising targeting using multi-website clickstream data. Fundamental Informaticae, 98(1), 49-70.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method. John Wiley & Sons.

Ebert, J. F., Huibers, L., Christensen, B., & Christensen, M. B. (2018). or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. Journal of medical Internet research, 20(1), e24.

Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. Internet research, 28(4), 854-887.

Fahimi, M., Barlas, F. M., Thomas, R. K., & Buttermore, N. (2015). Scientific surveys based on incomplete sampling frames and high rates of nonresponse. Survey Practice, 8(6).

Fahimi, M., Hair, E. C., Do, E. K., Kreslake, J. M., Yan, X., Chan, E., ... & Osborn, L. (2024). Improving the Efficiency of Inferences From Hybrid Samples for Effective Health Surveillance Surveys: Comprehensive Review of Quantitative Methods. JMIR Public Health and Surveillance, 10, e48186.

Fan, D. P. (2011). Representative Responses from Non-Representative Survey Samples. Midwest Association of Public Opinion Research, Chicago.

Ferri García, R., & Rueda García, M. D. M. (2020). Efficiency of propensity score adjustment and calibration on the estimation from non-probabilistic online surveys.

Fox, S., Rainie, L., Larsen, E., Horrigan, J., Lenhart, A., Spooner, T., & Carter, C. (2001). Wired seniors: the pew Internet and American life project. Washington, DC: Pew Internet and American Life Project. Retrieved May, 4, 2005.

Frankel, Martin R., and Frankel, Lester R. 1987. "Fifty Years of Survey Sampling in the United States." Public Opinion Quarterly 51 Part 2:S127–38

Garton, L., Hay Thornthwaite, C., & Wellman, B. (1997). Studying online social networks. Journal of computer-mediated communication, 3(1), JCMC313.

GfK Knowledge Networks. (2008). Knowledge Panel: A probability-based online panel for survey research. Retrieved from Knowledge Networks internal resources.

Goot, M. (2021). How good are the polls? Australian election predictions, 1993–2019. Australian Journal of Political Science, 56(1), 35-55.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. American psychologist, 59(2), 93.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). Survey methodology. John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R.(2009). Survey Methodology. John Wiley & Sons.

Groves, Robert .1989. Survey Errors and Survey Costs. New York: Wiley.

Groves, Robert.1998.''SurveyError Models and Cognitive Theories of Response Behavior.'' In Cognition and Survey Research, eds. Monroe Sirken, Douglas Hermann, Susan Schechter, Norbert Schwarz, Judith Tanur, and Roger Tourangeau. New York: Wiley, 235–52.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). SAGE Publications.

Jaeger, S. R., & Cardello, A. V. (2022). Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research. Food Quality and Preference, 102, 104676.

Kalton, G. (2023). Probability vs. nonprobability sampling: from the birth of survey sampling to the present day. Statistics in Transition. New Series, 24(3), 1-22.

Kaushik, A. (2009). Web analytics 2.0: The art of online accountability and science of customer centricity. John

Wiley & Sons.
Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). What low response rates mean for telephone surveys. Pew Research Center, 15(1), 1-39.
Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., & Asare-Marfo, D. (2021). Strategies for detecting insincere respondents in online polling. Public Opinion Quarterly, 85(4), 1050-1075.
Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-12).
Kish, L. (1965). Sampling organizations and groups of unequal sizes. American sociological review, 564-572.
Kocar, S., & Baffour, B. (2023). Comparing and Improving the Accuracy of Nonprobability Samples: Profiling Australian Surveys. Methods, data, analyses, 17(2), 36.
Kumar, R. (2018). Research methodology: A step-by-step guide for beginners.
Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. Journal of official statistics, 22(2), 329.
Lee, S., & Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. Sociological Methods & Research, 37(3), 319-343.
Lehdonvirta, V., Oksanen, A., Räsänen, P., & Blank, G. (2021). Social media, web, and panel surveys: using non-probability samples in social and policy research. Policy & internet, 13(1), 134-155.
Lenau, S., Marchetti, S., Mnnich, R., Pratesi, M., Salvati, N., Shlomo, N., ... & Zhang, L. C. (2021). Methods for sampling and inference with non-probability samples. Deliverable D11, 8.
Llieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research: Pros and cons. International Journal of Market Research, 44 (3), 361–367.
Magnani, R., Sabin, K., Saidel, T., & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids, 19, S67-S72.
Matthias, R. D., & Couper, M. P. (2017). Statistical analysis in research practice. New York: Springer.
McPhee, C., Barlas, F., Brigham, N., Darling, J., Dutwin, D., Jackson, C., ... & Wronski, L. Data quality metrics for online samples: considerations for study design and analysis. 2022.
Murray-Watters, A., Zins, S., Silber, H., Gummer, T., & Lechner, C. (2023). River Sampling-a Fishing Expedition: A Non-Probability Case Study. Methods, data, analyses: a journal for quantitative methods and survey methodology (mda), 17(1), 3-27.
Nie, N. H., & Erbring, L. (2000). Internet and Society: A Preliminary Report. Palo Alto, CA: Stanford Institute for the Quantitative Study of Society Press.
Olivier, L. (2011). River sampling non-probability sampling in an online environment. [Web blog post]. Center for Information-based Decision Making and Marketing Research. Retrieved from http://lexolivier/blogspot. com/2011/11/river sampling-non-probability-sampling. html.
Poynter, R. (2010). The handbook of online and social media research: Tools and techniques for market researchers. John Wiley & Sons.
Prepared for the AAPOR Executive Council by a Task Force operating under the auspices of the Couper, M. P., Courtright, M., ... & Zahs, D. (2010). Research synthesis: AAPOR report on online panels. Public Opinion Quarterly
Rahman, M. M., Tabash, M. I., Salamzadeh, A., Abduli, S., & Rahaman, M. S. (2022). Sampling techniques (probability) for quantitative social science researchers: a conceptual guideline with examples. Seeu Review, 17(1), 42-51.
Regmi, P. R., Waithaka, E., Paudyal, A., Simkhada, P., & Van Teijlingen, E. (2016). Guide to the design and application of online questionnaire surveys. Nepal journal of epidemiology, 6(4), 640.
Rivers, D. (2007). Sampling for web surveys. In Joint Statistical Meetings (Vol. 4). Alexandria, VA: American Statistical Association.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.
Roshwalb, A., Lewis, Z., & Petrin, R. (2016). The Efficacy of Nonprobability Online Samples. JSM Proceedings, Survey Research Methods Section, 3657-3666.
Sama, R. (2019). Impact of media advertisements on consumer behaviour. Journal of Creative Communications, 14(1), 54-68.
Saris, W. E., & Gallhofer, I. N. (2014). Design, evaluation, and analysis of questionnaires for survey research.

John Wiley & Sons.
Schonlau, M., & Couper, M. P. (2017). Options for conducting web surveys.
Shaver, L. G., Khawer, A., Yi, Y., Aubrey-Bassler, K., Etchegary, H., Roebothan, B., ... & Wang, P. P. (2019). Using Facebook advertising to recruit representative samples: feasibility assessment of a cross-sectional survey. Journal of medical Internet research, 21(8), e14021.
Shi, L., & Roche, S. P. (2024). Sampling from Online Panels. In Oxford Research Encyclopedia of Criminology and Criminal Justice.
Smith, P.J., Rao, J.N.K., Battaglia, M.P., Daniels, D., and Ezzati-Rice, T. (2000). Compensating for Nonresponse Bias in the National Immunization Survey Using Response Propensities. Proceedings of the American Statistical Association, Section on Survey Research Methods, 641–646.
Steinmetz, S., Bianchi, A., Tijdens, K., & Biffignandi, S. (2014). Improving web survey quality. Online Panel Research: Data Quality Perspective, A, 273-298.
Sudman, S. (1976). Applied sampling. Academic Press.
Taylor, H. (2000). Does Internet Research Work? Comparing Online Survey Result with Telephone Survey. International Journal of Market Research, 42, 58–63.
Vehovar, V. and Manfreda, K.L. (1999). Web Surveys: Can the Weighting Solve the Problem? Proceedings of the American Statistical Association, Section on Survey Research Methods, 962–967.
Vehovar, V., & Manfreda, K. L. (2017). Overview: online surveys. The SAGE handbook of online Research methods
Walsh, J. P., Kiesler, S., Sproull, L. S., & Hesse, B. W. (1992). Self-selected and randomly selected respondents in a computer network survey. The public opinion quarterly, 56(2), 241-244.
Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. International Journal of Forecasting, 31(3), 980-991.
Wejnert, C., & Heckathorn, D. D. (2008). Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research. Sociological Methods & Research, 37(1), 105-134.
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(3), JCMC1034.
Wu, M. J., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. Computers in Human Behavior Reports, 7, 100206.
Wyner, G. A. (2007). Survey Errors. Marketing Research, 19(1).
Yun, G. W., & Trumbo, C. W. (2000). Comparative response to a survey executed by post, e-mail, & web form. Journal of computer-mediated communication, 6(1), JCMC613.