

## **BIAS IN ONLINE HEALTH SURVEYS: IDENTIFYING AND OVERCOMING CHALLENGES**

KAROLINA SZALAŁATA

Jan Mikulicz-Radecki University Hospital in Wrocław, Poland

E-mail address: karolina.szalata1@gmail.com

ORCID: <https://orcid.org/0009-0006-0042-0082>

MONIKA CZEKALSKA

University Clinical Hospital No. 1 of the Medical University of Lodz, Poland

E-mail address: monmalx@gmail.com

ORCID: <https://orcid.org/0009-0004-7091-5369>

LUIZA ŁABUZIŃSKA

University Clinical Hospital No. 2 of the Medical University of Lodz, Poland

E-mail address: luizalabuzinska@gmail.com

ORCID: <https://orcid.org/0009-0004-7404-1662>

MARTYNA BUKOWIEC

Central Clinical Hospital of Medical University of Lodz, Poland

E-mail address: bukowiec.martyna97@gmail.com

ORCID: <https://orcid.org/0009-0007-2901-9230>

WERONIKA WOROSZ

St. Queen Jadwiga Clinical District Hospital No. 2 in Rzeszów, Poland

E-mail address: weronik06@gmail.com

ORCID: <https://orcid.org/0009-0008-5284-1633>

LAURA MAGDALENA SIKORSKA,

Andrzej Frycz Modrzewski University, Kraków, Poland

E-mail address: laura666@onet.pl,

ORCID: <https://orcid.org/0009-0009-5326-779X>

GABRIELA SZPIŁA

4 Military Clinical Hospital with Polyclinic in Wrocław, Poland

E-mail address: gabrielaszpila99@gmail.com

ORCID: <https://orcid.org/0009-0008-5105-1189>

ANGELA CŹWIL

Copernicus Memorial Hospital, Lodz

E-mail address: angela.cwil@gmail.com

ORCID: <https://orcid.org/0009-0006-9680-9606>

## ABSTRACT

**Aim.** Online surveys have become a vital tool in health research due to their cost-effectiveness, speed, and global reach. However, their methodological limitations – particularly various forms of bias – may affect the validity and generalisability of findings. This review aims to provide an overview of the most common types of bias in online health surveys and to discuss strategies for their mitigation.

**Methods.** A narrative literature review was conducted to identify and categorise the primary sources of bias associated with online survey research in health-related contexts. The review focused on three main areas: survey design and presentation bias, cognitive and psychological response biases, and selection bias. Studies addressing these issues were examined to extract examples and recommended mitigation approaches.

**Results.** The review identified several key sources of bias. Poor survey design, excessive length, or misleading visual formatting – was associated with reduced data quality and increased dropout rates. Response biases, such as social desirability, recall errors, or avoidance of sensitive questions, were found to skew data, particularly in self-reported health measures. Selection bias emerged as a major concern, as access to and familiarity with digital technologies significantly influences who participates in online surveys. Individuals with poorer health or limited digital literacy are often underrepresented.

**Conclusions.** Despite their numerous advantages, online health surveys are vulnerable to various forms of bias that can compromise data validity. Researchers must carefully consider survey design, sample recruitment, and respondent characteristics. Applying strategies such as pre-testing, use of clear language, adaptive sampling techniques, and bias correction methods can enhance the quality of online survey data. When ethical standards are upheld and methodological rigour is applied, online surveys can remain a powerful and reliable tool in public health research.

**Keywords:** online surveys, bias, online research, data validity, bias mitigation

## INTRODUCTION

In contemporary research, online surveys have become a crucial tool in social, medical, and epidemiological sciences. Their growing popularity is largely due to their low cost, speed of execution, and the ability to reach a broad audience worldwide (Eysenbach & Wyatt, 2002). In public health research, online surveys allow for the rapid and efficient collection of data, which is particularly important during dynamic epidemiological events. An example of this is the COVID-19 pandemic, where traditional data collection methods were hindered, and online surveys served as a vital alternative for conducting research. Other data collection methods, such as in-person or telephone surveys, require more effort and often yield lower response rates (Geldsetzer, 2020; Hlatshwako et al., 2021). However, it is important to note that differences in the results obtained depending on the mode of administration (i.e., different survey methods) may lead to systematic measurement errors (Singh et al., 2021). Furthermore, studies indicate that participants may be more likely to provide truthful answers in online surveys regarding sensitive topics, as opposed to computer-assisted telephone interviews or interactive voice response systems (Kreuter, 2008).

Despite numerous advantages, online surveys are subject to significant methodological limitations, including systematic biases that can affect the credibility and generalisability of research findings. One of the key issues is selection bias, which arises from uneven access to the internet and varying digital literacy across populations, potentially leading to the overrepresentation of certain social groups. Additionally, the lack of control over the survey environment and the anonymity of respondents may result in lower-quality responses (Wright, 2005).

This article reviews the literature on various types of bias in online health surveys and methods for minimising them. Particular attention is given to issues related to sample representativeness, self-selection effects, and the impact of survey design and external factors on data quality. A better understanding of these issues, along with the application of appropriate correction methods and careful sampling, can significantly enhance research quality and facilitate more accurate inferences (Hlatshwako et al., 2021). When properly applied, bias reduction methods and adherence to ethical standards can make online surveys a highly effective tool in public health and medical research.

## TYPES OF BIAS IN SURVEY RESEARCH

### **The Impact of Survey Design on Response Quality**

The structure and length of a survey can significantly influence data quality and respondent engagement levels. Motivation to participate directly affects the reliability of results, and phenomena such as dishonest survey completion and survey fatigue are substantial sources of bias (Singh et al., 2021). Lengthy and complex forms often lead to decreased attention and what is known as “survey fatigue,” which results in incomplete responses or premature abandonment of the survey. Research suggests that the relevance of the survey is more important than its length in determining follow-up rates (McCambridge et al., 2011).

Issues related to question formulation include the use of unclear language, leading questions, or inconsistencies in questionnaire construction (Choi & Pak, 2005). Ambiguous wording can lead to divergent interpretations, and the use of specialised terminology without prior explanation can result in incorrect answers, particularly when respondents are unfamiliar with the language used. For example, a question about physical activity, such as “Do you regularly engage in moderate physical activity?” – the term “moderate activity” can be interpreted subjectively and vary across respondents.

In online surveys, visual elements such as question formatting and response presentation can exacerbate these effects, leading to biased outcomes. Stylistic changes in a survey may influence how respondents process and respond to questions. For example, respondents may be more likely to choose options listed at the beginning or end of a multiple-choice list. Additionally, grouping answer options together or highlighting specific choices graphically can make those responses appear more important, thus introducing bias in selection (Smyth et al., 2006).

### **Cognitive Factors Affecting Respondents' Answers**

When constructing questionnaires, it is essential to account for errors related to subconscious and psychological reactions that may influence the validity of data. One such error is central tendency bias, where respondents avoid extreme values on scales and favour middle options. Another example is positive skew, where respondents tend to give positive responses to satisfaction-related questions. Additionally, studies have shown a tendency to select the first options in self-administered surveys (primacy bias) or the last ones in telephone or in-person interviews (recency bias) (Choi & Pak, 2005).

Social desirability bias is another significant category of response bias. It refers to providing answers that align with social expectations and avoiding responses considered socially unacceptable. Respondents might downplay their body weight, exaggerate their physical activity levels, or claim better mental health than is actually the case. Respondents may also consciously provide false answers, particularly in health-related surveys where socially undesirable diseases or behaviour (such as venereal diseases or suicide attempts) may be under-reported, and socially undesirable exposures (such as smoking or substance use) may be omitted (Choi & Pak, 2005). However, a study by Kreuter et al. (2008) found that online surveys are less likely to be affected by this bias compared to telephone or in-person interviews, as participants were more likely to provide truthful responses in online surveys, as confirmed by university records. Recall bias, where respondents struggle to accurately remember past events, can also significantly affect data quality. The literature highlights selective memory bias, where difficulty recalling health-related symptoms or events from the past leads to errors in the data (Tourangeau, 2000). Another factor that affects the accuracy of reported data is the telescoping effect, where events are mistakenly recalled as occurring more recently than they actually did, resulting in overestimations of their frequency (Choi & Pak, 2005).

### **Selection Bias**

Selection bias is a significant issue in survey research, especially in online surveys, as respondents may not be representative of the entire population. Participation in a survey requires access to specific distribution channels, which can limit the diversity of the sample. This may result in insufficient representation of those less familiar with technology, older individuals, or patients with serious health conditions. Furthermore, individuals completing online health surveys may be more motivated to respond if they have specific health experiences (Bethlehem, 2010; van Gelder, Bretveld & Roeleveld, 2010).

Differences in data collection methods (e.g., paper surveys vs. online surveys) and modes of access (e.g., via smartphones vs. desktop computers) can also lead to statistically significant differences in responses. Respondents using different devices often differ demographically, which can impact research outcomes (Toepoel & Lugtig, 2019). Studies show that paper survey participants are more likely to report health issues and have more negative attitudes toward environmental stressors than those responding to online surveys (Cantuaria &

Blanes-Vidal, 2019). Furthermore, internet users generally report better health than non-users, suggesting an additional selection factor in health-related studies (Schnell, Noack & Torregroza, 2017).

### **The Specificity of Bias in Online Health Surveys**

Health research conducted online is prone to various types of systematic errors, arising both from the characteristics of the studied population and the medium itself. One of the key challenges is the reliance on subjective health assessments, which may vary depending on factors such as age, education, and comorbidities. Respondents often base their evaluations on personal feelings, which can lead to inaccurate self-reports. For instance, individuals with undiagnosed hypertension may perceive their health as good, despite an actual health risk (Jylhä, 2009).

Another challenge is the lack of representativeness in the online population. Internet users differ demographically from the general population, limiting the generalisability of results. Older adults, rural residents, and individuals with lower education levels are often underrepresented, leading to potential bias in the findings (van Gelder et al., 2010). Self-selection is particularly significant in health research, as the health status of various social groups may differ significantly. Inadequate representation of certain groups can distort the actual picture of the phenomenon being studied. The topic of the survey should be tailored to the online population to minimise selection bias, for example, surveys about technology are more appropriate for internet users than surveys on homelessness (Eysenbach & Wyatt, 2002).

The volunteer effect is another factor that contributes to selection bias. People who are more interested in the survey topic are more likely to participate, creating systematic differences between respondents and the general population. For example, research has shown that women are more likely to engage with health-related topics, which may lead to their overrepresentation in health surveys (Eysenbach & Wyatt, 2002). Consequently, online surveys often yield inflated results concerning positive health habits, such as physical activity or a healthy diet, while those less concerned with their health or with limited access to medical information are underrepresented (Tourangeau & Yan, 2007).

Sensitive topics such as mental health, substance use, or abortion may further influence the quality of data collected, as respondents may avoid providing honest answers due to concerns about privacy (Tourangeau & Yan, 2007). This effect can be exacerbated by unacceptable disease bias and unacceptable exposure bias, leading individuals with certain conditions or those exposed to socially undesirable risk factors to intentionally omit crucial information. The anonymity of online surveys is both an advantage and a potential source of error. On one hand, it may encourage more honest responses, but on the other, the lack of direct oversight can lead to incomplete or unreliable answers due to low respondent motivation. Additionally, concerns about data privacy may cause participants to avoid disclosing sensitive information, which can negatively affect the credibility of the results (Tourangeau & Yan, 2007).

Bias in online health surveys can also be intensified by the low level of public knowledge about mental health disorders. Many people fail to recognise specific symptoms or hold incorrect beliefs about the causes and treatment of these conditions, leading to an underestimation of the prevalence of such disorders in studies (Jorm, 2000). Furthermore, the respondent's mood at the time of completing the survey can influence their responses. Individuals in a negative mood are more likely to report negative health symptoms and lower quality of life, which can lead to misleading conclusions about the true health status of the population (Jylhä, 2009).

In summary, health research conducted online is particularly vulnerable to various forms of bias, which can affect its reliability and the ability to generalise the results. Understanding these mechanisms and implementing strategies to minimise errors are crucial for improving the quality of online survey-based research.

### **Strategies for Minimising Bias**

Bias in online surveys is a significant methodological challenge, impacting the credibility and representativeness of the collected data. To minimise errors related to sample selection bias, ensuring the representativeness of the study group is crucial. Groves and Peytcheva (2008) suggest strategies such as shortening and simplifying the survey form, reducing cognitive load for respondents, and clearly communicating the study's purpose, which enhances participant engagement and reduces the nonresponse rate. Personalising invitations and sending reminders via email or SMS can also increase response rates. However, it is important to note that a high response rate does not necessarily eliminate bias, which is why analysing differences between respondents and non-respondents, as well as using statistical weights to adjust results, is recommended (Groves & Peytcheva, 2008).

To enhance the representativeness of the sample, random sampling from medical registers or databases can be helpful (CDC, 2013). Offline recruitment, such as partnering with local institutions or social organisations, enables the inclusion of underrepresented groups, such as older adults, who are often less represented in online surveys. Additionally, offering financial or material incentives can increase participation from hard-to-reach groups and diversify the respondent structure (Singer & Ye, 2013).

Survey question design is a key element in reducing bias both at the level of the research tool and the responses provided by participants. Willis (2013) emphasises the importance of cognitive interviews as a tool for improving question quality. Questions should be formulated neutrally and precisely, avoiding suggestions or ambiguity. Testing different question formats and tailoring them to the specific characteristics of respondent groups, considering cultural and social factors, promotes better understanding of the survey content. In health research, avoiding complex medical terminology and piloting the survey with smaller groups can help identify potential issues and improve clarity, thus enhancing the reliability of the data. It is also important to avoid leading ques-

tions, which steer respondents toward a particular answer through their construction or phrasing. Testing different ways of asking questions, such as open versus closed-ended questions, and adjusting them to the specifics of different respondent groups can help reduce bias (Choi et al., 2004).

Another crucial aspect is data anonymisation, which is essential for improving the credibility of responses, especially on sensitive topics such as mental health, substance use, or sexual behaviour. Ensuring participants' anonymity can also reduce socially desirable responding and encourage more honest answers (Tourangeau & Yan, 2007). In health-related surveys, this is especially important, as questions often address private and intimate issues, and protecting respondents' data increases the likelihood of obtaining truthful responses.

Combining different data collection methods, such as integrating subjective assessments with objective data from medical records, can enhance the accuracy of results. Integrating online survey data with electronic health records allows for verification of respondents' declarations and improves the credibility of the research (Jensen et al., 2014). In studies on physical activity, using data from fitness apps or health monitoring devices (e.g., smartwatches) can provide objective measurements, complementing respondents' subjective self-reports. Such technologies enable precise data collection on steps taken, heart rate, or calories burned, allowing for more accurate analysis of the studied phenomena. Evenson et al. (2015) suggest that activity tracking devices are a reliable source of data in health studies.

Implementing reporting standards such as CHERRIES (Checklist for Reporting Results of Internet E-Surveys) can help in the design and evaluation of online surveys, ensuring transparency and completeness in the methodology description (Eysenbach, 2004). Establishing uniform guidelines for conducting online research allows for better control over potential sources of bias and improves the comparability of results across different studies.

In conclusion, effectively minimising bias in online research requires a multifaceted approach, including recruitment strategies, carefully designed questions, anonymity, and integration with other data sources. Employing these methods increases the reliability and credibility of the results, enabling more accurate analysis of the phenomenon under investigation.

## CONCLUSIONS

Online surveys have become an essential tool in health-related research due to their cost-effectiveness, speed, and ability to reach diverse populations. However, they are susceptible to various biases that may impact data quality and generalizability.

One of the most significant challenges is selection bias, which results from unequal internet access and differing digital competencies among respondents. This can lead to an overrepresentation of certain demographic groups while excluding others, particularly older individuals or those with lower socioeconomic status. Furthermore, individuals with specific health concerns may be more motivated to participate, potentially skewing the findings.

Survey design plays a crucial role in mitigating bias. Factors such as question wording, structure, and length influence respondent engagement and answer accuracy. Poorly formulated questions can lead to misinterpretations, while lengthy or complex surveys may cause respondent fatigue, reducing data reliability. Additionally, response bias—such as social desirability bias—can lead participants to provide answers that align with perceived social norms rather than their actual experiences.

Cognitive biases also affect data validity. Recall bias, central tendency bias, and primacy-recency effects can alter the way respondents answer questions, particularly in health-related studies where accurate self-reporting is critical. The tendency to underreport stigmatised behaviour or overestimate desirable attributes further complicates data integrity.

Despite these challenges, strategies exist to minimise bias and improve data quality. Ensuring diverse recruitment channels and using mixed-mode survey designs can enhance sample representativeness. Additionally, optimising survey wording, layout, and visual presentation can reduce response distortions and improve participant engagement.

In conclusion, while online surveys present methodological challenges, they remain a valuable tool in health research when biases are properly acknowledged and mitigated. Researchers should employ best practices in survey design and data collection to enhance reliability, validity, and applicability of findings in public health and medical studies.

## REFERENCES

- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78(2), 161–188. doi: 10.1111/j.1751-5823.2010.00112.x.
- Cantuaria, M. L., & Blanes-Vidal, V. (2019). Self-reported data in environmental health studies: Mail vs. web-based surveys. *BMC Medical Research Methodology*, 19, 238. doi:10.1186/s12874-019-0882-x.
- Centers for Disease Control and Prevention. (2013). *The BRFSS data user guide*. Centers for Disease Control and Prevention. Retrieved from [https://www.cdc.gov/brfss/data\\_documentation/index.htm](https://www.cdc.gov/brfss/data_documentation/index.htm)
- Choi, B. C., & Pak, A. W. (2005). A catalog of biases in questionnaires. *Preventing chronic disease*, 2(1), A13.
- Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 159. doi: 10.1186/s12966-015-0314-1.
- Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, 6(3), e34. doi: 10.2196/jmir.6.3.e34.
- Eysenbach, G., & Wyatt, J. (2002). Using the Internet for surveys and health research. *Journal of Medical Internet Research*, 4(2), E13. doi: 10.2196/jmir.4.2.e13.
- Geldsetzer, P. (2020). Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: A cross-sectional survey on COVID-19. *Journal of Medical Internet Research*, 22(4), e18790. doi: 10.2196/18790.
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189. doi: 10.1093/poq/nfn011.
- Hlatshwako, T. G., Shah, S. J., Kosana, P., Adebayo, E., Hendriks, J., Larsson, E. C., Hensel, D. J., Erausquin, J. T., Marks, M., Michielsen, K., Saltis, H., Francis, J. M., Wouters, E., & Tucker, J. D. (2021). Online health survey research during COVID-19. *The Lancet Digital Health*, 3(2), e76–e77. doi: 10.1016/S2589-7500(21)00002-9.

- Jensen, R. E., Snyder, C. F., Abernethy, A. P., Basch, E., Potosky, A. L., Roberts, A. C., Loeffler, D. R., & Reeve, B. B. (2014). Review of electronic patient-reported outcomes systems used in cancer clinical care. *Journal of oncology practice*, 10(4), e215–e222. doi: 10.1200/JOP.2013.001067.
- Jorm, A. F. (2000). Mental health literacy: Public knowledge and beliefs about mental disorders. *The British Journal of Psychiatry*, 177(5), 396–401. doi.org/10.1192/bjp.177.5.396.
- Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3), 307–316. doi: 10.1016/j.socscimed.2009.05.013.
- Kevin, B. W. (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). doi.org/10.1111/j.1083-6101.2005.tb00259.x.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865. doi: 10.1093/poq/nfn063.
- McCambridge, J., Kalaitzaki, E., White, I. R., Khadjesari, Z., Murray, E., Linke, S., Thompson, S. G., Godfrey, C., & Wallace, P. (2011). Impact of length or relevance of questionnaires on attrition in online trials: Randomized controlled trial. *Journal of Medical Internet Research*, 13(4), e96. doi: 10.2196/jmir.1733.
- Singh, S., & Sagar, R. (2021). A critical look at online survey or questionnaire-based research studies during COVID-19. *Asian Journal of Psychiatry*, 65, 102850. doi: 10.1016/j.ajp.2021.102850.
- Singer, E., & Ye, C. (2012). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112–141. doi: 10.1177/0002716212458082.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Effects of using visual design principles to group response options in web surveys. *International Journal of Internet Science*, 1(1), 6–16.
- Schnell, R., Noack, M., & Torregroza, S. (2017). Differences in general health of internet users and non-users and implications for the use of web surveys. *Survey Research Methods*, 11(2), 105–123. doi: 10.18148/srm/2017.v11i2.6803.
- Toepoel, V., & Lugtig, P. (2015). Online surveys are mixed-device surveys: Issues associated with the use of different (mobile) devices in web surveys. 2190-4936. 2015. doi: 10.12758/mda.2015.009.
- Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (29–48). Lawrence Erlbaum Associates Publishers.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883. doi: 10.1037/0033-2909.133.5.859.
- van Gelder, M. M., Bretveld, R. W., & Roeleveld, N. (2010). Web-based questionnaires: The future in epidemiology? *American Journal of Epidemiology*, 172(11), 1292–1298. doi: 10.1093/aje/kwq291.
- Willis, G. B., & Artino, A. R., Jr (2013). What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. *Journal of graduate medical education*, 5(3), 353–356. doi: 10.4300/JGME-D-13-00154.1.