

Web surveys in the time of COVID-19

Websurveys nos tempos de COVID-19

Websurveys en tiempos de la COVID-19

Raquel Brandini De Boni ¹

doi: 10.1590/0102-311X00155820

Life and research work in the age of COVID-19 can be compared to the epistolary courting between Florentino Ariza and Fermina Daza, in Gabriel García Márquez' classic novel *Love in the Time of Cholera* ¹. In the midst of the pandemic, we are using correspondence more than ever. By "correspondence" in this case, we specifically mean the use of e-mail and web surveys. Among the various kinds of surveys that can be conducted via the internet, the book *Internet Research Methods* ² provides a didactic review for novices, besides several key insights for more experienced researchers. The current article focuses on web surveys, also known as web-based surveys, e-surveys, and online surveys. The article briefly presents some of the advantages that may have fueled their growing popularity in the current context (besides the possibility of collecting data from a distance, a clear advantage when social distancing is needed), as well as the spinoffs and ethical issues that need to be considered when planning such surveys and interpreting their results.

Web surveys are strategies used to obtain primary data and have been performed since the 1990s. Since then, different fields of knowledge such as the social sciences, political science, behavioral science, and communications science have used these surveys and discussed their possible limitations and the need for methodological improvements ^{3,4,5}. Web surveys can generally be conducted using e-mail lists (members of a profession or university students, for example ^{6,7}), internet user panels ^{8,9}, or publication of the survey's link on websites and social networks ^{10,11}. Considering that each of these approaches has unique characteristics, the focus here will be on web surveys that included volunteers recruited through websites and social networks.

A quick search in PubMed/MEDLINE using the terms "(websurvey' OR 'web-based survey' OR online survey) AND COVID-19", approximately six months after diagnosis of the first case of COVID-19 in China, yielded 146 published studies. This figure does not come close to the total number of web surveys that are being conducted worldwide, but it clearly indicates one of the method's advantage: the speed between planning the survey and publishing the results. Importantly, this does not even include the results of reports, opinion polls, or other instruments, but only articles published in scientific journals, selected via peer review, in the health field, and in a single database. The speed in obtaining and publishing information is particularly important in crisis situations. A minimal understanding of perceptions, beliefs, knowledge, behaviors, and physical and mental health problems can back new studies and interventions when other data are not available. Besides the speed in data collection, large numbers of individuals may be included in the surveys (thousands or tens of thousands

¹ Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

Correspondence

R. B. De Boni
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Av. Brasil 4365, Rio de Janeiro, RJ 21040-360, Brasil.
raqueldeboni@gmail.com



of volunteers can be interviewed within a couple of weeks), not to mention the wide geographic range and the possibility of crossing borders almost instantaneously. Nationwide web surveys and even those encompassing several countries can yield results in a few weeks, especially when the recruitment is performed via social networks (e.g., Facebook, Instagram, Twitter, and WhatsApp).

However, speed, large samples, and wide scope do not guarantee external validity (representativeness or generalization of the results), which can only be achieved with probabilistic sampling methods or censuses. This limitation has been reported since the first web surveys, and various groups and institutions have attempted to develop recruitment, sampling, and analytical methods to overcome it. Some issues that make this task more difficult are: (1) internet coverage (for example, the Brazilian survey tool *TIC Domicílios 2019*¹², reports that 74% of Brazilians use the internet, but that the proportion drops to 57% in lower-income socioeconomic classes D and E); (2) lack of a single registry of internet users; (3) quantification of non-response; (4) selection bias; and (5) the possibility of a single internet user answering the questionnaire several times. To make a long story short, when we post a survey on a website or push a banner with a survey's link in a social network, we have no way of guaranteeing the either exact number of users that were reached or their characteristics. Even using measures of webpage traffic or the social networks' estimated reach, we only have approximations that may not reflect the reality. For example, when a given IP (Internet Protocol) accesses a webpage several times, the same user may be returning to it repeatedly, or another person that uses the same computer may also be accessing the page. Both the number and characteristics of these users define the population from which we obtain a sample, based on which we wish to make inferences. These are also essential variables for calculating the probabilities of inclusion and the proportion and evaluation of non-response (i.e., necessary information to know whether the users that participated are similar to those that did not). Another relevant issue is self-selection: the characteristics of the volunteers participating in the study may be related to the outcomes (overestimating or underestimating prevalence rates and measures of association). When we know the associations in advance, we can at least discuss in which direction we are erring. For example, in a sample consisting mainly of women, we expect the prevalence of depression to be higher than in the general population (since depression is more frequent in women). However, there are variables that we are unable to measure and that can alter the results in ways we are unable to identify. Finally, a participant may answer the questionnaire more than once. In anonymous surveys (when no form of the participant's identification is collected), it is practically impossible to guarantee that a participant answered just once. In the attempt to increase web surveys' transparency, the possibilities for interpretation of the results, and their reproducibility, a set of guidelines was published in 2004 called the *Checklist for Reporting Results of Internet E-Surveys* (CHERRIES)¹³. Although the guidelines are not widely used and the suggestion to avoid duplication (blocking the IP) is not universally accepted², all these considerations are well-described and can assist the scientific community in increasing the studies' quality and interpretation.

Two other advantages of web surveys may also explain their popularity during the COVID-19 pandemic: cost and ease of implementation. As an example of the cost issue, two Australian economists compared the cost of a web survey to that of a mail survey (a strategy with numerous limitations, but low-cost) among individuals that visited the world's largest sand island, Fraser Island (Queensland, Australia). Considering that the authors purchased a software to program the online questionnaire, and not counting the cost of their own labor, each valid questionnaire in the web survey cost USD 1.94, while a valid questionnaire in the mail survey cost USD 8.13¹⁴. As for implementation, with the large number, variety, and flexibility of available software packages and servers for programming and storing online questionnaires, the initial difficulties with implementation were overcome. Even someone with no knowledge of programming language can post a simple questionnaire. This last advantage also eliminates the need to plan an additional data entry strategy (a central issue when planning off-line studies).

Even so, the best and most precisely programmed software and most secure servers cannot guarantee consistent questionnaires and accurate data. The drafting and validation of questionnaires, especially those that assess subjective aspects and are self-completed, is a complex process that usually requires one or more studies these preliminary by psychometric (by the way, these preliminary studies can also be performed online¹⁵). Thus, the use of previously validated questionnaires, when available, is an excellent decision when a streamlined solution is needed. Even so, the online version needs to

be tested before the survey begins (small pilot assessments are usually performed), both to assess the comprehensibility and ease of completion and to evaluate aspects involving the programming and the resulting databases. Despite these, Zhang et al.¹⁶ showed that the same psychological assessment scales whose answers were obtained in face-to-face interviews (using *CASI – computer-assisted self-interviewing*), telephone interviews, and online interviews presented different social desirability biases (according to the sensitivity and acceptability of the study's topic).

Thus, web surveys, like any other research method, present limitations and results must be interpreted taking these into account. Furthermore, as in other methods, web surveys also require clearly defining in advance the overall and specific objectives, the hypotheses to be tested, and the statistical analyses to be performed. Finally, it is necessary to rigorously comply with the ethical aspects involved in studies with human subjects and to extensively discuss and assess the new ethical challenges¹⁷. Two examples are the precautions with anonymity and data security. It is important for researchers, when using commercial software, to determine whether the IP (or other data that may identify participants such as zip code or e-mail) is collected and/or stored together with information furnished by the user and how this information is communicated to participants and protected. The servers must also be reliable and secure in order for the data not to be hacked or modified by third parties. There are also discussions on the format and size of the informed consent form. One of the potential problems is that consent may be interpreted merely as a disclaimer, as we have seen in many websites and apps (because of the fine print, the position on the page, or the length, considering that a short text may not contain all the necessary information, while an overly lengthy text may induce the user not to read it carefully).

None of these topics is new to the field. The scientific literature identifies various suggestions and solutions and the challenges for dealing with each of the aspects cited above (and numerous others not listed but equally relevant). Still, it is not an easy task to guarantee the integrity of science in moments as complex as the current pandemic, meanwhile quickly generating information that can be useful for everyone's health. It is unlikely for a successful survey to be performed by a single researcher, and often more than one research group needs to be involved in the process. Our capacity for connection and teamwork/networking and collaboration is thus essential and will continue to be if we wish to generate creative solutions for the near future. In the specific area of behavior and mental health, some of these solutions will only be possible if we can reconcile our technical capacity with the technology and effective communication strategies¹⁸. Building and reinforcing these bridges for the common good can be a positive legacy from this unprecedented crisis.

Additional information

ORCID: Raquel Brandini De Boni (0000-0002-2455-5997).

References

1. Marquez GG. O amor nos tempos do cólera. Rio de Janeiro: Editora Record; 1986.
2. Hewson C, Vogel C, Laurent D. Internet research methods. London: Sage Publishing; 2016.
3. Kaye BK, Johnson TJ. Research methodology: taming the cyber frontier. Techniques for improving online surveys. *Soc Sci Comput Rev* 1999; 17:323-37.
4. Solomon DJ. Conducting web-based surveys. *Practical Assessment, Research, and Evaluation* 2001; 7:19.
5. Sills S, Song C. Innovations in survey research: an application of web-based surveys. *Soc Sci Comput Rev* 2002; 20:22-30.
6. Joncew CC, Cendon BV, Ameno N. Websurveys como método de pesquisa. *Informação & Informação* 2014; 19:192-218.
7. Gallè F, Sabella EA, Da Molin G, Liguori G, Montagna MT, Orsi GB, et al. A cross-sectional study investigating lifestyle and weight perception of undergraduate students in southern Italy. *BMC Public Health* 2019; 19:1316.
8. Domche GN, Valois P, Canuel M, Talbot D, Tessier M, Aenishaenslin C, et al. Telephone versus web panel National Survey for monitoring adoption of preventive behaviors to climate change in populations: a case study of Lyme disease in Québec, Canada. *BMC Med Res Methodol* 2020; 20:78.
9. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality. *Public Opin Q* 2009; 73:641-78.
10. Marcon G, Monteiro GMC, Ballester P, Cassidy RM, Zimmerman A, Brunoni AR, et al. Who attempts suicide among medical students? *Acta Psychiatr Scand* 2020; 141:254-64.
11. Torres TS, Bastos LS, Kamel L, Bezerra DRB, Fernandes NM, Moreira RI, et al. Do men who have sex with men who report alcohol and illicit drug use before/during sex (chemsex) present moderate/high risk for substance use disorders? *Drug Alcohol Depend* 2020; 209:107908.
12. Núcleo de Informação e Coordenação do Ponto BR. TIC Domicílios 2019. <https://cetic.br/pesquisa/domicilios/indicadores/> (accessed on 28/May/2020).
13. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004; 6:e34.
14. Fleming CM, Bowden M. Web-based surveys as an alternative to traditional mail methods. *J Environ Manage* 2009; 90:284-92.
15. Lee SA, Mathis AA, Jobe MC, Pappalardo EA. Clinically significant fear and anxiety of COVID-19: a psychometric examination of the Coronavirus Anxiety Scale. *Psychiatry Res* 2020; 290:113112.
16. Zhang XC, Kuchinke L, Woud ML, Velten J, Margraf J. Survey method matters: online/offline questionnaires and face-to-face or telephone interviews differ. *Comput Human Behav* 2017; 71:172-80.
17. Salganik MJ. Ethics. In: Salganik MJ, editor. *Bit by bit: social research in the digital age*. Princeton: Princeton University Press; 2018. p. 281-325.
18. Balanzá-Martínez V, Atienza-Carbonell B, Kapczinski F, De Boni RB. Lifestyle behaviours during the COVID-19 – time to connect. *Acta Psychiatr Scand* 2020; 141:399-400.

Submitted on 06/Jun/2020

Approved on 11/Jun/2020