

Common Pitfalls in using Online Platforms for Data collection in COVID times and its implications

Hariom Kumar Solanki¹, P. Giridara Gopal², Rama Shankar Rath³

Correspondence: Dr Rama Shankar Rath, Assistant Professor, All India Institute of Medical Sciences, Gorakhpur, Uttar Pradesh, India

Email: ramashankar.aiims@gmail.com

Received 28 Sept 2020/**Revised** 21 Dec 2020/**Accepted** 21 Dec 2020

Citation: Solanki HK, Gopal PG, Rath RS. Common Pitfalls in using Online Platforms for Data collection in COVID times and its implications. Nepal J Epidemiol. 2020;10(4); 930-932.

DOI: 10.3126/nje.v10i4.31614

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Copyright © 2020 CEA& INEA. Published online by NepJOL-INASP.

www.nepjol.info/index.php/NJE

Sir,

In recent COVID-19 time when the direct physical data collection is quite impossible, we authors came across various online methods of data collection related to various social, behavioral research. We congratulate those researchers who have the burning desire to conduct research and they find various ways to do the same. With this letter we want to highlight few pitfalls related use of online data collection tools. Online forms allow for remote and easy data collection [1]. Availability of smart phone with internet access on the part of study participants as well as the investigators is sufficient for developing and carrying out data collection for studies based on self-filled questionnaires. In Indian context where many of the studies are done by early career researchers are not supported (non-funded), the cost saved on paper and printing can be considerable [2]. This form of data collection also ensures security of data collected and issues associated with handwriting deciphering.

Recently, there has been spurt in studies being conducted using online forms. In the prevailing situation of COVID19 pandemic where measures like social distancing and avoiding unnecessary travel are important to prevent outbreak exacerbation, it seems to be a pragmatic option given the benefits listed above. However, there are certain pitfalls to doing studies in such a way not meeting certain condition.

First issue pertains to exclusion of people based on their inability to afford, or their unwillingness to use electronic gadgets like smartphone and computer and internet [3]. This will to certain extent make the study participants non-representative of study population which the researcher intended to cover.

Second important issue is fulfillment of condition of probability sampling [3]. In the probability sampling every unit has a known probability of getting selected in the study. However, dissemination of online form for data collection from participants is akin to convenient sampling method where

people within the networks of the researcher/s or networks accessible to the researcher/s will be tapped into, especially when there is no clearly defined group of eligible people. Most of the times a request to forward the link to the form is also made by the research team. Some of the participants may circulate it, others may not. Thus, the degree of penetration of the link into the sampling frame may be unknown for many of the studies done through online forms. Both these factors contribute to difficulty in ascertaining the probability of inclusion of sampling units in the study and thus affecting the external validity of the study. In the same way self-selection of individuals is another part which may also lead to selection bias [3,4].

Third issue pertains to the issue of response rate. In absence of knowledge of the number of eligible people reached, the responses rate cannot be ascertained. As explained above it is exceedingly difficult to ascertain the number of eligible people reached in this method especially when the study is not limited to a defined number of institutions with a common communication group [5]. However, issue of higher nonresponse rate in web based and online surveys is not a problem in current scenario of increased use of internet [6].

Similarly, there is an issue related to the responders of such studies. Filling the form multiple times by the same responder/s due to misunderstanding/ mistakes like filling the form again on receiving the link repeatedly with request to fill the form without realizing that the form was meant to be filled only once; may bias the study results [7]. Sometimes some responders may also have strong interest in the academic area study is covering and they might strongly feel that study should favor result in some particular way, this may also motivate some responders to fill the form multiple times or forward the form to individuals in his/ her network who think like them and likely to favor a particular pattern of responses. Both these situations introduce bias into the study which are difficult to identify and control for once the study has been completed. This issue however may be solved in closed group studies, the researchers can limit multiple submission of forms by participants by providing a unique response code to each intended participant in advance which he or she can use only once in the duration of data collection of a particular study. In contrary nonresponse in an internet based or online survey may be due to disease condition studied [8]. Thus, higher and lower response rate in web-based surveys must be considered while interpreting the results.

Another issue related to responders in open study groups arises when the responder is not a part of the intended study population. For example, if a group of researchers want to study an aspect of occupational health in workers of a particular occupational group (say health care workers level of stress in the times of COVID19 pandemic) in a particular administrative

unit/ region/ area. It is possible that healthcare workers will have professional links outside that administrative unit/ area and the link is shared with professional working in other administrative units, where the independent/ exposure variable is different from the unit where the study was intended for. The responses registered from outside the intended study population will thus bias the results depending upon multiple independent variables in unknown ways.

We therefore recommend using online form-based study only in well identified group of people preferably in a closed group through a common and secure thread. We also discourage the practice of asking to forward the online link into the contacts and networks of individual study participants except when snowball sampling is being attempted or list of all eligible participants is available from reliable databases. The number of registered medical practitioner in a country or state for example may be available from relevant medical council. When reporting such a study the total number of people who received an online link -whether directly or indirectly - should be incorporated in the body of the report.

In conclusion, though useful when used thoughtfully and for specific well identified groups of individuals, online form based studies might suffer from certain significant infirmities from epidemiological point of view which may cast a shadow of doubt on the reliability and validity of the findings of such studies. This should be kept in mind by both the researchers as well as readers of the research findings/ articles, especially in the prevalent COVID- 19 pandemic situation when researchers are rushing to complete their studies on one or other aspect of COVID-19 and report their findings as soon as possible for publication in academic journals on the one hand and journals are speeding up the process of editorial and peer review for quick dissemination of research findings.

Dr. Hariom Kumar Solanki, Assistant Professor, Department of Community Medicine, Greater Noida Institute of Medical Sciences, Greater Noida, Uttar Pradesh

Dr. P. Giridara Gopal, Scientist C, Centre for Community Medicine, All India Institute of Medical Sciences, New Delhi

Dr. Rama Shankar Rath, Assistant Professor, All India Institute of Medical Sciences, Gorakhpur, Uttar Pradesh

Dated the 28 September 2020

Acknowledgement

None

Competing interests

There is no conflict of interest for any author of this manuscript.

Authors' affiliations:

¹Department of Community Medicine, Greater Noida Institute of Medical Sciences, Greater Noida, Uttar Pradesh

²Centre for Community Medicine, All India Institute of Medical Sciences, New Delhi

³All India Institute of Medical Sciences, Gorakhpur, Uttar Pradesh

References

1. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology?. *Am J Epidemiol.* 2010;172(11):1292-1298. doi:10.1093/aje/kwq29
<https://doi.org/10.1093/aje/kwq291>
PMid:20880962
2. Wyatt JC. When to use web-based surveys. *J Am Med Inform Assoc.* 2000;7(4):426-429. doi:10.1136/jamia.2000.0070426
<https://doi.org/10.1136/jamia.2000.0070426>
PMid:10887170 PMCID:PMC61446
3. Bethlehem J. Selection bias in web surveys, *Int Stat Rev* 2010; 78(2): 161-178
<https://doi.org/10.1111/j.1751-5823.2010.00112.x>
4. Greenacre ZA, The importance of selection bias in internet surveys, *Open J Stat*, 2016, 6, 397-404
<https://doi.org/10.4236/ojs.2016.63035>
5. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies, *Modern Epidemiology*, Rothman KJ, Greenland S, ed. 3rd ed Philadelphia, PA: Lippincott Williams and Wilkins; 2008:128-147
6. Lefever, S., Dal, M. and Matthíasdóttir, Á. Online data collection in academic research: advantages and limitations. *Br J Educ Technol.* 2007;38:574-82.
<https://doi.org/10.1111/j.1467-8535.2006.00638.x>
7. Steiner N, David L, *Health measurement Scales: A practical guide to their development and use* 5th Edition, Oxford Publication, 2014
8. Janssens AC, Kraft P. Research conducted using data obtained through online communities: ethical implications of methodological limitations. *PLoS Med.* 2012;9(10):e1001328.
<https://doi.org/10.1371/journal.pmed.1001328>
PMid:23109913 PMCID:PMC3479089