Guidance for research on the COVID-19 disease in times of pandemic

Oya KALAYCIOGLU

Affiliations:
1 PhD, Assistant Professor, Department of Biostatistics and Medical Informatics, Bolu Abant Izzet Baysal University, Bolu, Turkey.

Corresponding author:
Oya Kalaycioglu, Department of Biostatistics and Medical Informatics, Bolu Abant Izzet Baysal University, Golkoy 14280, Bolu, Turkey. E-mail: oyakalaycioglu@ibu.edu.tr

Abstract

This study is aimed at examining the manner of use of internet and phone surveys and making suggestions related to sample selection strategies during the COVID-19 outbreak. As the influence of the COVID-19 pandemic spreads worldwide, many researchers are engaged in telephone and internet surveys to assess the socioeconomic impacts of the outbreak of the disease and the emerging needs of societies during this period. However, the data collection methods requiring access to technology and the ability to use it may lead to under-representativeness of socioeconomically deprived groups. In emergency situations, although this can be a legitimate limitation, it should be reported by the researchers and optimal effort should be made to increase representativeness. In order to ensure that the population is adequately represented with a sample, the choice of survey mode and sampling strategy should be determined based on the research objectives, socio-demographic characteristics of the target population and the available resources. However, most survey research being conducted during the outbreak of COVID-19 disease lack proper justification of the sampling strategy used and neglect to address the limitations due to under-represented communities. This may consequently result in misleading and overgeneralized findings. Therefore, guidance to researchers concerning appropriate survey methods and sample selection strategies that could be used in surveys during the COVID-19 pandemic are needed. The comparison of advantages and disadvantages of different survey strategies presented in this study are expected to contribute to the methodological literature in data collection.
Riassunto


TAKE-HOME MESSAGE

*With phone and internet surveys during the outbreak of COVID-19, those who are most likely to be affected by the pandemic may not be reached. In emergency situations, the under representation of marginalised groups can be a legitimate limitation, but should be reported by the researchers.*

Competing interests - none declared.

Copyright © 2020 Oya Kalaycioglu Edizioni FS Publishers

This is an open access article distributed under the Creative Commons Attribution (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. See http://www.creativecommons.org/licenses/by/4.0/.


DOI 10.19204/2020/gdnc3
INTRODUCTION

Surveys provide valuable information on emergency management efforts in disaster environments by exploring the state of society in terms of people’s health and their economic and social needs. Pandemics are disasters in the context of public health that may significantly increase morbidity and mortality on a wide geographical area while leading to negative social and economic consequences as well [1]. The COVID-19 pandemic has spread rapidly since the beginning of the epidemic in China in December 2019 and there is an urgent need to conduct well designed and effectively implemented surveys to determine how societies are affected by this pandemic, their priorities, needs, the level of compliance with restrictions imposed and socioeconomic vulnerabilities. Therefore, in addition to collecting data regarding the disease incidence and health outcomes, there is an emerging need to collect data on socioeconomic measures as well [2]. Surveys are thus launched to reach wider groups of people and to identify quickly the overall trends and impacts of COVID-19. For the outcomes of these surveys to be comprehensive enough and close to reality, however, the survey mode and participants must be selected carefully by considering both the objective of the survey and the available means [3].

The survey research is mostly carried out through three different modes, namely face-to-face interviews, phone survey and internet survey. The decision on choosing the optimal survey mode while minimising the overall total survey error requires consideration of several issues such as representativeness, sampling, non-response rate and measurement [4]. While there are many studies comparing the advantages and disadvantages of survey modes [5–8], there is no literature on the implementation of these modes during a pandemic. Of all survey modes, the gold standard remains, in terms of representing all sections of society, the household surveys that use probability sampling methods and face-to-face interviews [9]. However, data collection with face-to-face interviews and field research becomes nearly impossible during COVID-19 due to the dynamic nature of pandemics, their global scale, uncertainties regarding the normalization period and various restrictions on the course of daily life. Furthermore, as face-to-face interviews involve the necessity of travelling and having close contact with participants, the use of phone or internet surveys are suggested during the COVID-19 pandemic for remote data collection [10].

Besides the survey modes used, selecting the survey participants with either probability or non-probability-based sampling methods has its effects on the representativeness of the survey. In probability sampling, each participant to be selected from the population has a known and non-zero probability of being selected, whereas this probability is unknown in non-probability sampling [11]. While in probability sampling the researcher identifies those to be invited to take part in the survey, that is beyond the researcher’s control in non-probability based sampling. Probability-based sampling is recommended since it enhances the representative capacity of the sample [12]. However, this method usually requires a sample framework list and getting such a list may not always be possible in emergency situations, may be costly or take too much time. Thus, non-probability methods can be preferred under these constraints [11]. Scientists, as well as researchers from different institutions and agencies, are presently intensively engaged in collecting data through phone and internet surveys to assess the impact of the outbreak of COVID-19 disease [13–17]. Furthermore, many national and international organizations have launched internet surveys to make a swift assessment of economic and social effects of the COVID-19 pandemic (Eurofund, Erasmus Student Network, Harvard Humanitarian Initiative, UNESCO). The use of phone and internet surveys for collecting data during a pandemic is not new. During the outbreak of Ebola disease in 2014 [18–20] and influenza pandemic in 2009–2010 [21–24], phone and internet surveys were employed to assess
how the disease affected the behaviours and perceptions of the public, and to guide policy makers and public health professionals on focusing particular issues.

In order to ensure that these surveys reflect existing realities and cover all segments of society, the objective of the survey and available means must be taken into account. The aim of this paper is to provide a guidance for the optimal survey mode and sampling strategies for research geared to identifying the social and economic needs of people during the COVID-19 outbreak and to examine the effects of the pandemic on society. In this context, the manner of use of the internet and phone surveys are examined and suggestions are made related to sample selection during the COVID-19 outbreak. Furthermore, the pros and cons of internet and phone surveys are examined and compared to each other with reference to the nature of the pandemic period.

DISCUSSION

Phone and Internet surveys during COVID-19 pandemic

Selecting the participants in phone surveys

Widespread ownership and use of mobile phones all around the world has motivated the use of phone surveys as an alternative to face-to-face interviews [25]. With specific reference to the COVID-19 pandemic, the use of phone surveys are recommended by the World Bank in April 2020, in order to monitor the burden of disease, manage the allocation of economic resources, identify psycho-social vulnerabilities and evaluate public health policies [26].

There are various probability-based sampling methods that can be employed in phone surveys during a pandemic. In cases were researchers had earlier developed a sampling framework representing society that contained up-to-date phone numbers, selecting the sample from this framework on the basis of probability sampling during the COVID-19 pandemic will improve the sample's representative capacity [26]. In doing so, new findings as to common and varying effects of the outbreak can be observed and measuring new effects stemming from the COVID-19 would be possible. Surveys of this kind, which include repeated measurements from the same participant (which are often referred as longitudinal surveys), also give the opportunity of examining changes taking place in the course of time and making causal inferences [27]. The approach also provides opportunities to minimize non-response error on the basis of past interaction between researchers and participants [26]. Nevertheless, it is important to note that this sample framework is not too old to preserve the validity of contact details. In case the researcher has no prior sampling framework, he/she can secure, after taking necessary permissions and abiding by relevant code of ethics, a list of active phone numbers can be obtained from a telecom company or firm. The researchers can then call randomly selected numbers from this list and conduct phone conversations in the form of interview. However, obtaining these lists may be a time consuming endeavour and often entails some costly procedures. Furthermore, lists may not be representative for the whole population. For instance, if data is obtained from only one telecom company, while there are more lists in a country, it would not represent subscribers to other companies.

Another method of sampling used in phone surveys is to make calls by randomly picking numbers, which is called random-digit dial (RDD). These calls are typically made by official and private institutions and are based on interactive voice response (IVR) or Short Messaging Service (SMS) system. This sample selection method in phone surveys has a cost advantage as there is (i) no need for an interviewer with IVR or SMS surveys, and (ii) obtaining the sample framework is cost-free. Thus, due to low cost, the sample size may be as large as desired. However, randomly determined numbers may not match an active phone number or these numbers may belong to firms instead of real persons.
Selecting the participants in internet surveys

First emerging in the 1990s, internet surveys became preferred to phone surveys for their lower cost and quicker responses. During the outbreak of COVID-19 disease, given the cost limitations and urgent need for national and international surveys, internet surveys stand out as the frequently preferred method [13]. Particularly in developing countries, however, internet surveys may fall short of covering all segments of society [5] and lower participation rates [28]. Nevertheless, given that people mostly stay at home as a result of lockdowns during the pandemic there may be higher participation to such surveys when compared to normal times.

While conducting internet surveys, probability-based samples can be selected by using one of two different strategies. The first is to obtain a sample framework with a list of e-mails, to then randomly select individuals from this list and send them an invitation via e-mail to take part in the survey. These lists are usually obtained from large-scale institutions like universities, large firms, and governmental agencies once having obtained necessary ethical permission. It is possible for sample frameworks obtained in this way to be homogeneous with respect to a specific characteristic. As the second way, a representative sample framework which was determined before the outbreak with available e-mail addresses may be used to invite the participants by e-mail.

Instead of ensuring the participation of only a selected target population, the internet survey may also appear on a website accessible to all internet users. Information regarding the survey can be announced via advertisements and referrals from different websites or in social media or printed and visual media. There is no limitation to participation to such surveys and selection of participants is therefore beyond the control of the researcher. The researcher may however have some level of control over participants by limiting IP addresses or adding some filter questions asked prior to the main survey. This method is economical and yields quicker outcomes since it does not require any sampling framework nor any effort from the researcher to identify the sample. Neither is there the problem of cost increase stemming from sample size since the survey is open to all.

Comparative advantages and disadvantages of phone and internet surveys

There are many factors affecting the quality of outcomes in phone and internet surveys. While identifying the optimal survey method, researchers should consider whether there is any sampling framework containing contact information, the target population to be represented, rates of participation and response envisaged, question types, and time and budget constraints. As a guide for researchers during the COVID-19 pandemic, Table 1 summarizes the pros and cons of various survey methods and sample selection strategies.

One distinguishing feature in survey methods is the participation of the interviewer in the data collection process and the level of their interaction with the respondent. While phone surveys are managed by interviewers who read out the questions to participants and record their responses (on paper or computer), in internet surveys or interactive voice response surveys no such interaction with the participants is possible. The interactive nature of interviews reduces measurement errors stemming from item non-response and misunderstanding on the part of the respondent [5]. In the COVID-19 pandemic, which has brought circumstances which are unprecedented to the large majority of people, it is important to improve the intelligibility of survey questions. Also, in interview surveys there may be open-ended questions from which deeper information can be obtained about the participants. The interviewer’s direct communication with the participant enables the supply of detailed information about the survey which can also play a motivating role in participation. Thus, the rates of participation to phone surveys are higher compared to internet surveys [7]. In case the
researcher identifies an existing representative sample and includes in the survey those with whom she/he had established contact earlier for a different survey, the tie between the researcher and participant is expected to increase the rate of participation to the phone survey [26]. Besides the advantages mentioned above, interview type surveys have some disadvantages. The first is related to the higher cost of such surveys, resulting from the need to employ interviewers and data entry staff, as well as phone call charges. Also, since data is not automatically recorded in a database it is often time consuming to record their responses (on paper or computer), with the problem compounded by possible errors that can happen during data entry. The sample size in phone surveys may be smaller relative to internet surveys and interactive voice response surveys due to cost and time constraints.

If the survey contains sensitive questions such as those regarding substance addiction, psychological and physical abuse, sexual orientation, political orientation and income, the rate of responding to such surveys is lower in live phone surveys conducted by interviewers than in internet surveys [30]. Comparing the internet and interactive voice response based methods it was found that there are higher rates of response to sensitive questions in internet surveys [31]. The benefits of internet surveys become clearer when considering the cost, quick data access and the rate of response to sensitive questions.

Both phone and internet surveys, however, may not be satisfactorily representative of the society concerned. As far as geographical representation is concerned, internet surveys usually underrepresent rural areas since internet access is limited in such areas. In phone surveys there may also be problems of representation in the case where phone numbers are obtained from a company that does not cover or has only limited coverage of rural areas. Education and income levels affect mobile phone ownership and internet use to different extents. Low education and income levels both lower the rate of internet use, but the most important factor lowering the rate of mobile phone use is low income [36]. The effect of education and income levels on phone and internet use is also interrelated with age. As levels of education and income rise there is significant increase in phone and internet use in the elderly [37].

**Implications for research**

In general, during the COVID-19 outbreak surveys managed by interviewers that use a representative sample framework obtained from earlier surveys must be preferred whenever possible to increase coverage and participation rates. However, situations confronted during the pandemic require urgent intervention, and as the internet surveys offer quick outcomes with lower cost relative to phone surveys, they have become more attractive in this regard. As opposed to its significant practical advantages, there is a possibility that those who preferred to take part in the internet survey may not be representative of general population and socio-demographic sub-groups [29]. It is important to note that digital literacy is needed to participate in random-digit-dial or internet based surveys which is less common among rural people, women, elderly persons and groups at low education and income levels, who are more likely to experience social, economic and psychological consequences of COVID-19. Meanwhile, in spite of their disadvantages in terms of representation, internet surveys have the possibility of reaching large number of people in a short period of time. Enlarging the sample size, by keeping the survey available for longer on the website, may add value to its statistical power, but it must not be forgotten that power does not guarantee unbiasedness of the statistical analysis. The validity of the research findings depends more on the representative capacity of the sample than its size. If there are systematic differences with respect to socio-demographic characteristics between those who volunteered to participate and those who abstained to participate to the survey, low rate of participation may
Table 1. Overview of advantages and disadvantages of survey modes and strategies to select participants.

<table>
<thead>
<tr>
<th>Strategy for selecting the participants</th>
<th>Phone Survey</th>
<th>Internet Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting the sample from a representative sampling framework obtained prior to COVID-19</td>
<td>Selecting the sample from a list of active phone numbers that can be obtained from a telecom company</td>
<td>Making calls with IVR or send SMS message to randomly dialled numbers (Random-digit-dial)</td>
</tr>
<tr>
<td>Explanation of questions</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Participation</td>
<td>High</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Item nonresponse</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Time required for data collection</td>
<td>Long</td>
<td>Long</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Response to sensitive questions</td>
<td>Lowest</td>
<td>Low</td>
</tr>
<tr>
<td>Open-ended or complex questions</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data entry bias</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample size</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Coverage bias in terms of:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>Gender</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>Age</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>Income</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Education</td>
<td>Small</td>
<td>Small</td>
</tr>
</tbody>
</table>

| | | | | | |

* a Based on the live conversation between the interviewer and the respondent  b Provided that the list contains up-to-date phone numbers  c If the selected telecom company’s coverage is poor in rural areas. IVR: Interactive voice response. Bolded cells indicate the optimal method for the selected criteria.

lead to *non-response bias* in survey outcomes. Researchers are therefore advised to provide reminders and explanations to encourage participation. With respect to comparative rates of participation, it was found that there is higher participation in phone surveys than in internet surveys [39, 40]. Therefore, it is recommended to employ mixed-mode initiatives involving both e-mail and mobile phone SMS messaging for reducing the participation related disadvantages of internet surveys [41].

**CONCLUSION**

Information gathered so far concerning COVID-19 suggests that collecting data on socio-demographic and socio-economic factors are decisive with respect to the effects of the outbreak [2]. However, an essential point to be considered when collecting data remotely is the presence of limitations that preclude access to population groups that are most deeply affected by adverse circumstances caused by the outbreak. These groups include those who cannot be communicated with due to their stay in intensive care units, persons with mental or physical disabilities with limited use of internet and phone, elderly persons, care home residents, homeless, refugees, irregular migrants and health workers engaged in intensive work. Another point that should be taken into account in surveys on the effects of COVID-19 is the lack of internet access on the part of the most affected groups due to various reasons such as the closure of service providers during the pandemic and the inabi-
lity to pay phone and internet bills following the loss of jobs and income. Those limitations of under representativeness could be regarded as legitimate during an emergency situation like a pandemic, however they must be reported to avoid misleading and overgeneralized research findings.

In conclusion, while determining the optimal survey mode to be used in during the COVID-19 pandemic there are many factors to be taken into consideration such as the socio-demographic characteristics of the target population, availability of a representative sample framework, available resources, urgency in reaching conclusions and the presence of sensitive questions. Comparing sample selection strategies, it is recommended to use, in both phone and internet surveys, a sample framework obtained through probability-based sampling prior to the COVID-19 pandemic. It should be noted that, this framework should include a list of persons with their current communication information and be representative of the society.

References


